Systematic approaches for nominating combination therapies in cancer

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

Systematic Approaches for Nominating Combination Therapies in Cancer

presented by Ciyue Shen

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: _____

Typed Name: Dr. Kornelia Polyak

Signature: Markus Basan _____

Typed Name: Dr. Markus Basan

Signature: _____

Typed Name: Dr. Walter Fontana

Signature: Forest White
Forest White (May 4, 2022 11:51 EDT) _____

Typed Name: Dr. Forest White

Date: May 02, 2022

Systematic Approaches for Nominating Combination Therapies in Cancer

A DISSERTATION PRESENTED

BY

CIYUE SHEN

TO

THE DIVISION OF MEDICAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

BIOLOGICAL AND BIOMEDICAL SCIENCES

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

MAY 2022

Systematic Approaches for Nominating Combination Therapies in Cancer

Abstract

Combination therapy in cancer can provide enhanced anticancer efficacy and reduce the risk of drug resistance. Focused molecular experiments to discover combinations in a large search space are inefficient while high-throughput screening approaches do not provide sufficient mechanistic insights. We present two systematic approaches for nominating combination therapies, i) computational prediction of cellular response to unseen combination perturbations based on network models of cell biology, and ii) unbiased proteomic profiling of cellular response upon drug treatment to identify resistance mechanisms.

Systematic perturbation of cells followed by comprehensive measurements of molecular and phenotypic responses provides information for constructing computational models of cell biology. Using a perturbation-response dataset of a melanoma cell line after drug treatments as a testbed, we developed a hybrid approach that combines explicit mathematical models of cell dynamics with a machine-learning framework to quantitatively predict cell behavior in response to perturbation of molecular targets. When sufficient perturbation-response data is available for model training, this model can predict cellular response to a vast number of unseen combination perturbations and therefore efficiently narrow down the search space and nominate promising sets of experimentally testable combination candidates.

Investigation of resistance mechanisms can be used for the rational design of combination therapy. In an alternative approach, we used unbiased quantitative protein mass spectrometry to assess the cellular response profile to a small number of anti-cancer drug perturbations in ovarian cancer cells. Data-driven protein network analysis revealed known and novel markers of resistance, which we used to propose combination drug candidates. In a first round of validation experiments, synergistic and additive effects were observed for some combination candidates across multiple ovarian cancer cell lines, suggesting potential therapeutic value for future pre-clinical and clinical studies.

Both systematic approaches can be used to effectively nominate combination therapies, provided the availability of sufficiently informative perturbation-response data. With further validation, the proposed combination candidates may contribute to the development of novel effective cancer therapeutics. We believe the approaches can be generalized to other cancer types and potentially be applied to other areas of cell biology.

Contents

Listing of figures

Listing of tables

To my father, we both have been waiting for this.

Acknowledgments

I would like to express my deepest appreciation to my thesis advisor, Chris Sander, for providing me the opportunity to pursue my PhD degree. I have learnt so much from you both as a researcher and as a person. I believe with all our efforts, we will cure cancer one day. I would like to extend my deepest gratitude to my dissertation advisory committee members, Kornelia Polyak, for always being critical, Nir Hacohen, for providing me the opportunity to write my first R script and being supportive, Michael Springer, for supporting my career plan and providing useful feedback. I would like to thank my defense committee members, Markus Basan, Walter Fontana, and Forest White, for your time and support.

I am extremely grateful to all the current and past Sander lab members, and collaborators, for your valuable contribution and insightful discussion. Bo Yuan, thank you for all the informative educational sessions and genuine support throughout the tough days. Augustin Luna, thank you for teaching me so much about coding, communication, everything and for being a great friend. Alex Franz, thank you for teaching and sharing your expertise on experimentation. Nicholas Gauthier, thank you for introducing and driving the projects. Frank Poelwijk, thank you for sharing insightful scientific opinions and for being an invaluable friend. Michael Stiffler, thank you for the helpful suggestions and those delightful teatime. Many thanks to other people without whom this work would not have been completed, Debora Marks, John Ingraham, Anil Korkut, Stefan Peidli, Nils Blüthgen, Tessa Green, Matthias Mann, Fabian Coscia, Lea Charaoui, Nathanael Fillmore, Nhan Van Do, Soumya Raychaudhuri, and those who have helped me during my graduate studies.

To the Biological and Biomedical Sciences program, Anne O'shea, Daniel Gonzalez, Kate Hodgins, thank you for making a family for us here and walking us through this process. I appreciate the Department of Data Sciences in Dana-Farber Cancer Institute and the Department of Cell Biology at Harvard Medical School for proving all the resources for my research.

Deep gratitude to all my friends who have helped and supported me. Special thanks to my BBS friends, Soun Lee, for all the deep, science or non-science, cheerful or depressing discussions, Yuchen Xiao, for being there and always shining as a role model, Samantha Wong, for all the mentally relaxing but physically stressful outdoor activities, and many others for being each other's support. Special thanks to my friends from other stages and parts of my life, my college friends Ruoxi Wang, Michelle Liu, Tianci Zhong, Chaoqun Wang, Huiyao Ouyang, for always being there for each other, Yanding Zhao, for constantly reminding me to do good science, Han Yang, for the good times about food and music, Guanlan Dong, for being my company for the tough quarantine days and a

constant driving force for good restaurants, and many others.

I am deeply indebted to my parents, Mrs. Wei Shen and Mr. Jianping Shen, for your constant support throughout my graduate school. Thank you for teaching me about love, kindness, and integrity, and for making me who I am today.

Lastly, I would like to thank my partner Anthony Tian, for always enlightening me when I'm lost, for always pulling me out of self-doubt, and for always bringing even more delight to bright days. Thank you for being by my side as I complete this work.

0

Introduction

0.1 Cancer and combination therapy

Cancer represents a collection of hundreds of diseases sharing one common feature, which is the uncontrollable and rapid growth of abnormal cells that can invade other parts of the body. Cancer is a leading cause of death worldwide in the 21st Century with nearly 10 million deaths in 2020[1], but our understanding of cancer has come a long way since the first discovery of cancer recorded

around 3000 BC in the "Edwin Smith Papyrus"[2]. Hundreds of years of scientific research on cancer makes it no longer a disease with "no treatment", but rather a disease with clear mechanisms of origin, multiple acquired capabilities for development, and effective treatment strategies to prolong patients' lives. The foundation of cancer origin has been revealed by R.A.Weinberg[3] as mutations in two different types of genes that lead to abnormal behaviors of cells, which are activating mutations in proto-oncogenes, and inactivating mutations in tumor suppressor genes. Proto-oncogenes are usually responsible for growth-stimulating signals, while tumor suppressor genes prevent inappropriate cell growth. Therefore, the gain-of-function and loss-of-function mutations in proto-oncogenes and tumor suppressor genes, respectively, both contribute to excessive multiplication of cells, and eventually the development of cancer. As rapid advance in cancer research has been achieved in the past decades, our knowledge about the acquired capabilities for cancer growth and progression is getting more complete and has been summarized as the well-known "hallmarks of cancer" (Figure 1)[4,5,6]. Beyond the originally-proposed mechanisms of sustaining proliferative signaling and evading growth suppressors, more complicated mechanisms, such as inducing replicative immortality, angiogenesis, genome instability, and etc., have been confirmed by recent research. One of the most prominent fruits coming from these mechanisms of cancer pathogenesis is the development of targeted therapies in cancer treatment.

The first treatment of cancer with detailed notes can be dated back to the 1st-2nd century AD, when a Roman medical author and practitioner Archigenes used surgical operation to remove a cancerous growth[7]. After centuries of development and improvement of technical expertise, surgeons can now perform much more precise surgical removal of tumors while minimizing the amounts of normal tissue removed during operations. While less invasive treatment approaches are used and actively developed, surgery still remains first-line therapy for solid and localized cancer[8]. The application of radiation in cancer diagnosis and treatment was made possible after the first discovery of x-rays in 1896 by a German physicist, Wilhelm Conrad Röntgen[9]. Radiation therapy has then

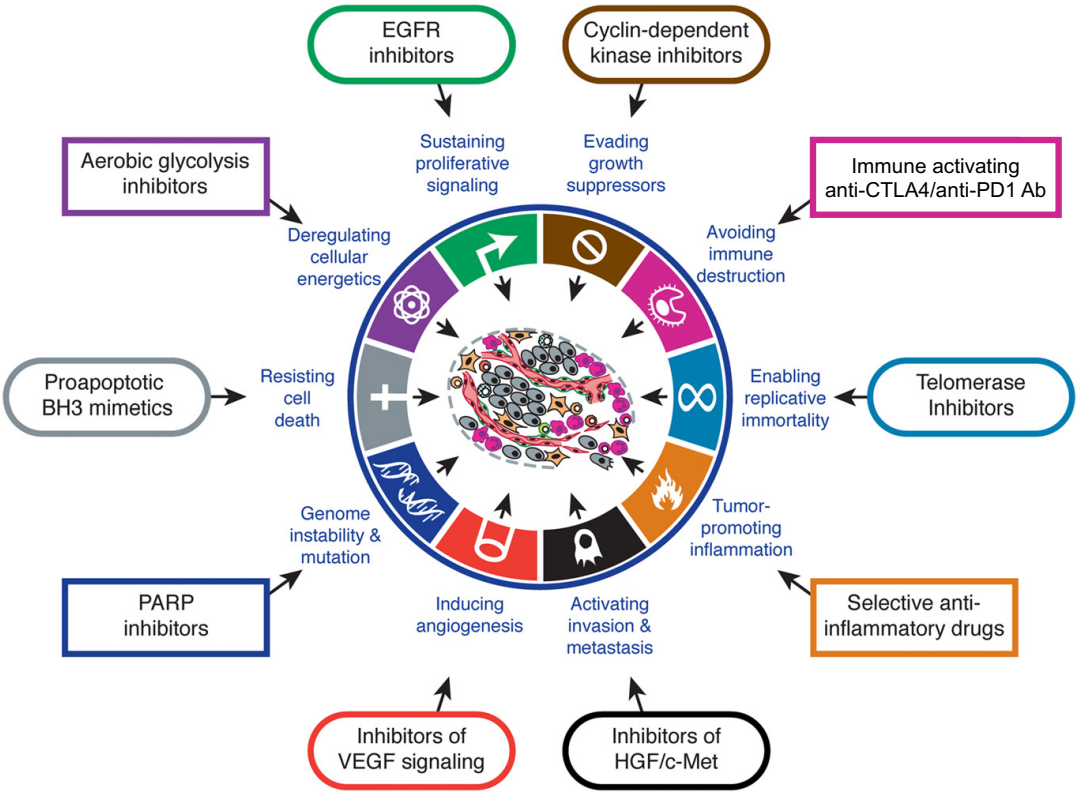


Figure 1: Hallmarks of cancer and therapeutic targeting approaches, adapted from D.Hanahan and R.A.Weinberg[5]. Ten hallmark capabilities of cancer have been proposed and research in the past decades have revealed the mechanisms behind most of the hallmarks. Many of the therapies in clinical use and under investigation align with these hallmarks with some examples illustrated. More candidate drugs are being developed targeting each of these hallmarks.

been under rapid development and different technologies have been applied to improve the precision of the aim of the radiation. Chemotherapy is probably the best-known cancer treatment in the past decades. The discovery of chemotherapy arose from the study of mustard gas as a military weapon during World War II[10]. Nitrogen mustard, a compound that worked against lymphoma, and aminopterin, a compound that worked against acute leukemia, led to the discovery of a series of drugs that inhibit cell replication. Chemotherapy, due to its unspecific targeting of dividing cells, usually causes damage to normal tissue in tumor-bearing host and therefore side effects in patients and their immune systems. This is where our knowledge of the molecular biology of cancer comes

into play and guides the development of targeted therapies.

Most targeted therapies under clinical use or development fall into the categories that generally correspond to the hallmarks of cancer, as shown in Figure 1. Monoclonal antibodies or small-molecule inhibitors are most commonly used as targeted therapies to disrupt the function of their targeted proteins. An example of monoclonal antibodies, which are usually designed to target receptors on the cell surface, is Trastuzumab (sold under the brand name Herceptin)[11]. Receptor tyrosine-protein kinase erbB-2 (HER2), a member of the human epidermal growth factor receptor (EGFR), is responsible for promoting cell growth and division. Its coding gene ERBB2 is amplified in different types of cancer, especially in a significant portion of breast cancer cases. Trastuzumab binds to the HER2 receptor, blocks the downstream signaling pathways that promote cell proliferation, and therefore effectively controls tumor progression in HER2 over-expressed breast cancer. A well-known example of small-molecule inhibitors, which usually target intracellular proteins, is Imatinib (sold under the brand name Gleevec)[12]. Most of the patients with chronic myelogenous leukemia (CML) have a chromosomal abnormality that results in a constitutively active tyrosine kinase Bcr-Abl that promotes cell proliferation. Imatinib selectively and competitively binds to the active site of the tyrosine kinase and therefore leads to its decreased activity. As extensive information has been collected on cancer genomics through the efforts of nationwide and worldwide consortia, including the Cancer Genome Atlas (TCGA) (https://www.cancer.gov/tcga) and the International Cancer Genome Consortium (ICGC)[13], we are presented with an unprecedented opportunity to continously develop targeted therapies for cancer. Another emerging therapeutic approach, originated from Wilhelm Busch and Friedrich Fehleisen in the 19th century, deploys the human's own immune systems to fight against cancer[14]. Cancer immunotherapies today generally fall into three different categories, including immune checkpoiont blockade, adoptive cellular therapies, and cancer vaccines[15]. Partially due to the long-lasting effects, immunotherapies have gained much popularity in the last two decades.

Combination therapy, a treatment approach that combines two or more therapeutic agents, has achieved some success in clinical oncology [16,17]. Combination therapy was first demonstrated effective in 1965, when Emil Frei, James F. Holland and Emil J. Freireich treated children with acute lymphocytic leukemia with the combination of four drugs, known as POMP (6-mercaptopurine, vincristine, methotrexate, and prednisone) [18]. Investigating combinatorial therapeutic approaches for synergistic or additive effects has become a focused area of research in cancer therapy. Such combination therapies include but are not limited to the combinations of chemotherapies and targeted therapies, the combinations of multiple targeted therapies [19], and the combinations of immunotherapy and targeted therapies [20].

Combination therapy is of particular clinical interest for multiple reasons. At an individual patient level, synergistic combinations provide better anticancer effects, and reduced toxicity when individual drugs are used at reduced dosages, which still holds true for combinations with additive anticancer effects [21]. Simultaneously targeting different molecular pathways likely blocks the mechanisms associated with adaptive resistance and therefore reduces the emergence of drug resistance to single anticancer agents [22,23,24,25]. In addition, tumor clonal heterogeneity are more likely overcome by combination therapy and improved patient outcome can be observed more obviously at a population level [26,27]. Practically speaking, if one or more drugs used in the combination therapy has already been approved by the FDA (U.S. Food and Drug Administration), the overall cost and time-span of the therapeutic development will be much reduced, significantly increasing the cost efficiency of cancer therapy.

0.2 High grade serous ovarian cancer

Ovarian cancer is one of the leading cause of cancer-related deaths in women. The current 5-year survival rate is about 49% in the United States and there has been minimal improvement in mor-

tality in the past decades[1]. Ovarian cancer can be divided into different subtypes based on the histology and cells of origin, and each histological subtype has different clinical features and treatments[28]. High-grade serous carcinoma (HGSOC) is the most common subtype (70%) and accounts for the majority of ovarian cancer deaths. Studies have shown that HGSOC probably originate in the fallopian tube. Large-scale genomic and molecular profiling on HGSOC patient samples have characterized HGSOC as having a high degree of copy number alterations and a relatively low mutation frequency[29]. More than 95% of tumors have mutations in the TP53 gene, a well-known tumor suppressor gene responsible for controlling cell proliferation. Germline BRCA1 and BRCA2 mutations have been observed in a total of 17% of the patients with additional 3% with somatic mutations in a cohort of 489 HGSOC patients reported in 2011[13]. Mutations in BRCA genes are known risk factors also in other types of cancers, especially breast cancer. Mutations and epigenetic changes are observed in other genes of the homologous recombination (HR) pathway in HGSOC, accounting for 51% of the cases, suggesting HR as a promising target for therapy (Figure 2). Other altered pathways identified include retinoblastoma tumor suppressor (RB) signaling, phosphatidylinositol3-kinase (PI3K)/RAS signaling, and NOTCH signaling. However, most of the alterations are copy number variation or changes in gene expression, and predominant recurring point mutations that can serve as drug targets are lacking.

Efforts have been made to extensively study the molecular profiles of HGSOC patient samples in order to better understand the mechanisms and develop therapeutics. mRNA expression has been used to further categorize HGSOC cases into four subgroups, proliferative, immunoreactive, differentiated, and mesenchymal, each with distinct molecular profiles[30,29]. However, survival analysis did not show significant differences among the subgroups, suggesting unclear clinical relevance of the subgroups. Many later studies on the "subtypes" of HGSOC have contradicting findings: subtypes are not mutually exclusive as multiple signatures can be observed in individual tumor samples[31]; survival difference was observed among the subtypes[32]; no more than 3 subtypes should exist

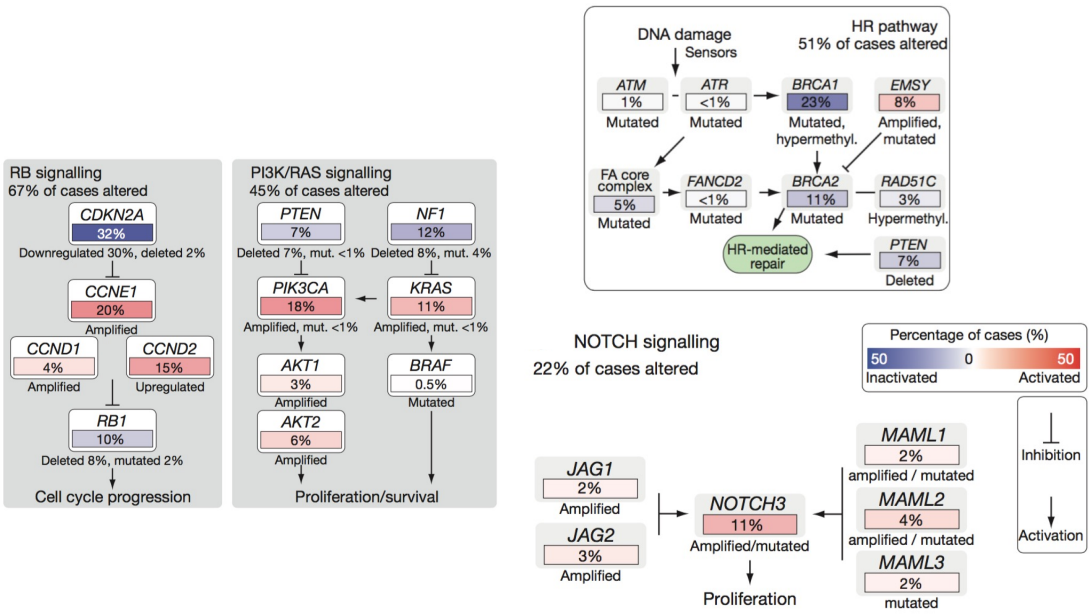


Figure 2: Altered pathways in HGSOC patients, adapted from TCGA Research Network[13]. Altered pathways were identified by different types of analysis on HGSOC patient samples. Alterations are defined by somatic mutations, DNA copy number changes or, significant change of mRNA expression in diploid tumours. The homologous recombination (HR) pathway is altered in up to 51% of the cases.

based on cross-population analysis[33]. A comprehensive clustering analysis suggests that the subtypes might largely depend on the types of molecular profiles and the clustering algorithms used[34]. Another proteomics analysis partially confirmed the mRNA subtypes but no survival differences were observed[35]. Although inconclusive, the results suggest that HGSOC is a genetically complicated disease and developing therapeutics is challenging.

The primary treatment of HGSCO is surgical cytoreduction followed by platinum-based chemotherapy, with either carboplatin and paclitaxel or carboplatin alone[28]. For the patients who are not suitable for initial surgery, neoadjuvant chemotherapy (NACT) has been used as an alternative treatment, which consists of carboplatin and paclitaxel for three cycles followed by interval surgical cytoreduction and additional chemotherapy. Despite the surgical removal and extensive chemotherapy, approximately 80% of the advanced-stage tumors relapse, resulting in a 5-year survival rate

of less than 30%[36]. Recurrent ovarian cancer in most cases cannot be cured. Current treatment options for platinum-sensitive cases include secondary surgical cytoreduction, re-use of platinum-based chemotherapy (with the risk of potentially life-threatening platinum drug allergies) or platinum-based combination chemotherapy, and poly(ADP-ribose) polymerase (PARP) inhibitors. PARPi have been approved by the FDA as single-agent, second-line/maintenance treatment for patients with BRCA mutations[37]. This class of therapeutics targets the DNA repair mechanism on which the tumors with defective HR repair capabilities are more dependent. Other therapies under development for HGSOC include drugs targeting angiogenesis, PI3K signaling, RAS signaling, and NOTCH signaling[38,39,40].

Recent pre-clinical studies and clinical trials suggest that combination therapies can be effective in HGSOC. PARPi, tested in combination with other targeted therapies, such as angiogenesis inhibitor[41], CDK (cyclin-dependent kinase) inhibitor[42], PI3K inhibitor[43], and anti-PD1 immunotherapy[44] has produced ovarian cancer remissions much more efficient than the single agent. Some studies have used combinations of PI3K inhibitors with chemotherapy, RAS/ERK inhibitors, inhibitors of BCL-2-family proteins, and other combinations of targeted therapies[45,46,47,48], suggesting combination therapy might be a solution to achieve better efficacy and overcome resistance in HGSOC.

0.3 Experimental approaches for discovery of combination therapies

Historically, the initial discovery of combination therapies was largely dependent on empirical experimentation of multiple effective agents in clinical settings, while the mechanisms behind were insufficiently investigated[49]. As the development of increasing numbers of anticancer drugs and the advancement of experimental techniques in molecular biology, resistance to single agent has been observed more frequently and the resistance mechanisms have been elucidated[24,25]. Combination

therapies can therefore be designed based on the molecular mechanisms of the original treatment. This approach is commonly used to combine targeted therapies such as kinase inhibitors. The rationale of combination design generally fall into three categories, as summarized by Al-Lazikani et al. [50].

i) Combining drugs targeting the same protein. One example is the combination of Trastuzumab with other Her2 inhibitors. The resistance mechanisms of Trastuzumab include inaccessiblity of drug recognition site, and continued activation of HER2 signaling through other mechanisms such as HER2/HER3 dimerization [51]. Trastuzumab was then used in combination with other HER2 inhibitors, Pertuzumab [52] or Lapatinib [53]. Pertuzumab binds at a different site on HER2 to inhibit dimerization, and Lapatinib works intracellularly to prevent phosphorylation of HER2.

ii) Combining drugs "vertically" targeting proteins in one signaling pathway. There are usually multiple kinases responsible for signal transduction in one signaling pathway and simultaneously targeting them is more likely to completely inhibit the signaling. One example is combination therapy with BRAF and MEK inhibitors in melanoma [54]. About half of patients with melanoma carries a mutation in the serine-threonine kinase BRAF (V600), resulting in activated mitogen-activated protein kinase (MAPK) pathway, which is responsible for communicating signals for cellular proliferation. It involves a series of kinases and therefore is also known as Ras-Raf-MEK-ERK pathway. Concurrent inhibition of the upstream BRAF and the downstream MEK can result in stronger and longer duration of anti-tumor effects and decrease MAPK-driven acquired resistance. This combination has become standard of care for patients with advanced-stage melanoma harboring BRAF V600 mutations [55].

iii) Combining drugs "horizontally" targeting parallel pathways. One example is the combinatorial inhibition of the well-studied MAPK pathway and PI3K/AKT pathway. Studies have suggested that PI3K/AKT activation is a resistance mechanism upon treatment of MEK inhibitors [56,57]. Combinations of MEK inhibitors and PI3K/AKT inhibitors, or the downstream mTOR inhibitors, have

shown synergistic effects and produced durable remission[58,59,60].

Beyond these rationales for combination design based on resistance mechanisms, hallmarks of cancer provide many more directions for combining therapeutics targeting different acquired capabilities for cancer growth. Statistics have shown that more than 330 anticancer drugs have been approved by FDA from 2009 to 2020, with 57 new approvals in 2020 alone[61]. This provides us with more than 50,000 possible two-drug combinations. The number of possible combinations goes to 845,000 if we include more than 1,300 anticancer drugs that are currently under development. It is laborious and practically unrealistic to use focused molecular experiments to explore all the possible combinations or study the drug resistance mechanisms for the rational design of combinations. Systematic high-throughput screening approaches have been developed and used as a complementary, unbiased approach.

Large phenotypic screens using drugs[62], or RNAi and CRISPR(Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9[63] have revealed cancer vulnerabilities and dependencies in different cancer types. These results contribute to the foundation of a cancer dependency map and can facilitate the prioritization of therapeutic targets. The development of an orthologous CRISPR-Cas9 system allows combinatorial genetic screens and therefore enables the exploration of cancer dependencies in a combinatorial fashion[64]. In addition to phenotypic screens, gene expression profiles upon perturbations can be informative in identifying small molecules with desired targets and biomarkers for dependencies[65]. Efforts have been made to extensively profile molecular response upon drug treatment in a collection of cancer cell lines[66,67,68]. The molecular response measurements mainly focus on gene expression, and measurements of proteomic and epigenomic changes have also been collected recently[69]. While the perturbation-response data directly provides information on dependencies and resistance mechanisms, baseline profiling of the cancer cell lines prevalently used by researches as models for different types of cancers also provides essential information, especially when integrated with response data[70]. Platforms have been made for researchers to easily

access the data, including the Cancer Cell Line Encyclopedia (CCLE[70]), Cancer Therapeutics Response Portal (CTRP[71]), and Genomics of Drug Sensitivity in Cancer (GDSC[72]). Tools have been built for data integration and analysis across databases and platforms[73].

0.4 Computational approaches to predict cell response and identify combinations

Beyond the hypothesis-driven empiric approaches and unbiased high-throughput technologies to experimentally identify effective drug combinations, computational methods are powerful in extracting information from experimental data and exploring the high-dimensional space of drug combinations that cannot be fully covered by experimental approaches. The first attempt to mathematically predict the maximum effects of combined drugs can be traced back to Loewe[74]. The Loewe Additivity model is still being widely used today to evaluate the combinatorial effects of drug combinations. Other mathematical methods to quantify the synergy, additivity or antagonism of drug combinations have also contributed to the foundation of determining drug combinations of interest, including Bliss Independence[75], Chou and Talalay's combination index[76], and the ZIP model[77]. Under the situation where cellular response to combinatorial perturbations or even individual perturbations was not available, regression models can be applied to directly predict cellular responses based on molecular perturbation-response measurements and linear interaction models[78,79]. Other methods find dissimilarities among the response profiles of drug treatment and propose drug combinations targeting independent vulnerabilities[80]. Although shown to be effective, these approaches usually focus on the pair-wise relations of drug treatment and lack mechanistic insights on a system level. A different flavor of computational approaches attempt to construct network models to better understand the mechanisms and predict cell response.

Before diving into the details of different types of computational methods for network con-

struction, we should be aware that years of molecular and biochemical experiments have helped researchers collect rich information on biological pathways and protein-protein interactions. Efforts have been made to curate such information to open-access database to facilitate basic and clinical research. One famous example is Reactome, which is an open source, manually curated and peer-reviewed pathway database with "signaling and metabolic molecules and their relations organized into biological pathways and processes"[81]. Other widely-recognized efforts include BioGRID[82], which focus on biological pathways with disease relevance, and KEGG[83], which collects information on various levels of molecular interactions. Some databases also attempts to collect computationally predicted interactions beyond experimentally validated ones, such as STRING[84]. Additional efforts to further integrate information from various databases provide a user-friendly platform, Pathway Commons, to access, analyse, and integrate pathways and interactions from multiple sources[85].

Various computational methods have been developed to build network models of cell biology. Static models, such as co-expression models[86,87,88], maximum entropy networks[89,90], and mutual information related methods[91,92], can use either baseline or perturbation-response data to construct network models of molecular interactions[93,94]. Dynamic models can provide information on signal propagation throughout the network and therefore provide more mechanisms rationalizing the design of combination choices. These models can be generally categorized into three different types, as summarized by Al-Lazikani et al.[50] (Figure 3).

i) Logic-based models, models that provide qualitative descriptions of the system behaviors. Each node in the network represents a molecular entity that can be in discrete states, e.g. ON and OFF. The nodes in the system are connected through a set of logic gates, such as AND, and OR. Discrete two-state logic models are known as Boolean models, which requires little prior knowledge on mechanistic details. Despite its simplicity, Boolean model can still predict accurate, qualitative system behaviors in some systems[95]. More complex logic-based methods, such as fuzzy-logic models[96],

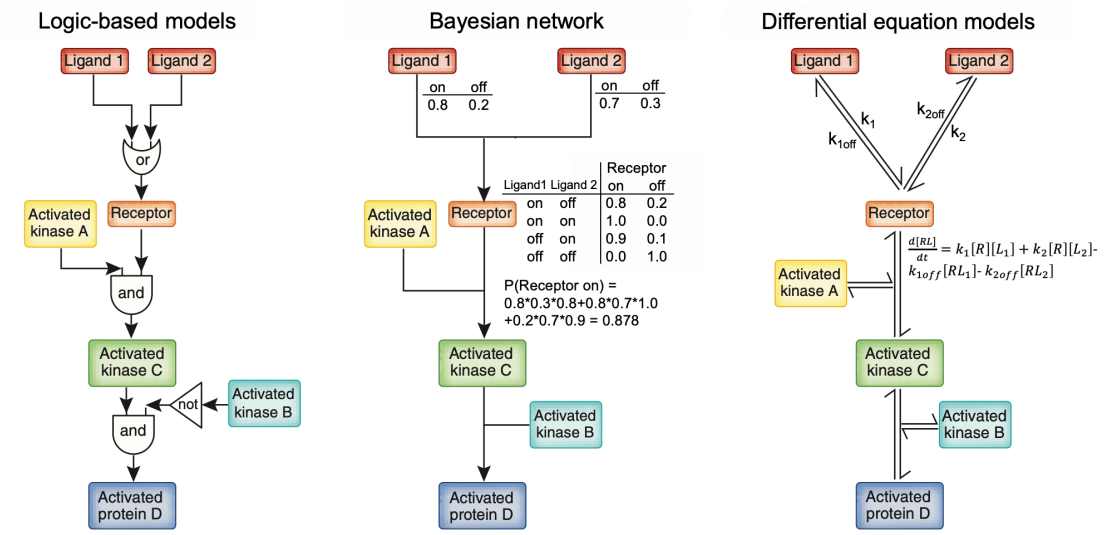


Figure 3: Different types of network-based computational models, adapted from Al-Lazikani et al.[50]. A hypothetical pathway is provided to facilitate the interpretation of the three main categories of network-based computational models. Either of the ligands can bind the receptor for its activation. Kinase A is then recruited for downstream activation of kinase C, followed by protein D, provided kinase B is inactivated. Logic models use a set of logic gates to represent such connections. Bayesian networks use a set of probabilities conditioned on previous events. A set of hypothetical conditional probabilities is given for ligand-receptor interactions. Differential equation models use a set of differential equations to connect the molecular entities. In particular, a mass action model has interactions represented as reaction equilibrium with underlying kinetics.

allow nodes to be in multiple discrete or even continuous states. Although these more complicated approaches can potentially provide more precise description of the system, they usually require parameters that were previously unknown or biochemically irrelevant[97].

ii) Bayesian network models, probabilistic graphical models that connects a set of variables, i.e. different molecular entities such as genes and proteins in systems biology, based on their conditional dependencies (probabilities) via a directed acyclic graph (DAG)[98]. Bayesian networks can be used to predict the likelihood of outcomes given the possible causes and therefore are useful for predicting drug response. The structure of the graph, i.e. how the nodes in the network are connected, is usually defined based on prior knowledge of the interactions among the variables in the system. The parameter inference is then performed to learn the joint probability distribution for each of

the variables. Dynamic Bayesian network (DBN), a derivative of classic Bayesian network, connects variable over adjacent time steps, and therefore allows prediction of dynamic cellular response over time [99,100]. While providing rich information on the system, inference of DBN models depend on accurate time-series response data upon perturbations.

iii) Differential equation models, which are mostly ordinary differential equation (ODE) models involving a set of differential equations connecting the (time) derivatives of the independent variables to other variables in the system. The variables are usually quantitative measurements of the molecular entities such as proteins, mRNA, and metabolites. One of the most commonly used ODEs is the mass action model, which essentially captures the process of chemical reactions and uses kinetic parameters [101,102]. A recent effort demonstrated that efficient parameter estimation allows the construction of large-scale mechanistic ODE models capable of predicting response to drug combinations [103]. While computation is used to infer most of the kinetic parameters in the large system, the model still depends on the prior knowledge of the equilibrium equations connecting the molecular components of the network, which is usually aggregated from disparate experimental sources and can suffer from the lack of uniform context. Other data-driven ODE models independent of prior knowledge on chemical reactions use slightly more abstract representations to capture the relations between the molecular components in the system [104,105]. Such an approach usually has a large number of parameters to infer due to very few constraints on interactions and therefore has to rely on accurate and efficient parameter inference algorithms. Dynamic optimization approaches, such as Monte Carlo (MC) methods and belief-propagation (BP) algorithms, have been used for parameter inference in data-driven network models [104,106,107,108,109]. However, they might not efficiently scale to larger systems (e.g., MC) or might require excessive approximations for the chosen mathematical model to facilitate efficient exploration of solution space (e.g., independent row approximation in BP). Therefore, to achieve good accuracy of parameter inference for larger systems and to gain the ability to generalize to more sophisticated kinetic models, a more general and poten-

tially more powerful data-driven modeling framework is needed.

Recently, deep learning has become an effective data-driven framework capable of generating predictions for large and complex systems. Gradient descent implemented with automatic differentiation, which has been broadly used in training graphical models, allows efficient parameter optimization in complex network systems. This framework has been successfully applied to many domains of biomedical research, from pathology image classification [110,111] to sequence motif detection [112]. Although the predictive power of deep-learning models is often impressive, their interpretation, which is crucial for providing understandable and, therefore, more trustable predictions, remains challenging. The complex multilayer network architecture of most deep-learning models lacks explicit representations and consequent direct interpretation. This difficulty is sometimes called the "black-box" problem [113]. Efforts have been made to increase interpretability of deep learning models by incorporating existing knowledge of biological pathways. Such interpretable models have shown some success in cancer discovery and predictive medicine [114,115]

0.5 Challenges in designing combination therapies in cancer and our solutions

Combination therapy in cancer can provide enhanced anticancer efficacy and reduce the risk of drug resistance. Many clinical trials and pre-clinical studies have demonstrated the success and great potentials of combination therapy in different types of cancer. We have a particular interest in developing combination therapies in ovarian cancer, especially HGSOC, which is the most and malignant subtype. Due to frequent relapse after first treatment and minimal improvement in survival rate, combination therapy might be a promising solution. Focused molecular experiments to discover combinations in a large search space are inefficient while high-throughput screening approaches do not provide sufficient mechanistic insights. We aim to develop high-throughput systematic ap-

proaches for nominating combination therapy while providing mechanistic rationales. We present here two systematic approaches for nominating combination therapies, i) computational prediction of cellular response to unseen combination perturbations based on network models of cell biology, and ii) unbiased proteomic profiling of cellular response upon drug treatment to identify resistance mechanisms.

Systematic perturbation of cells followed by comprehensive measurements of molecular and phenotypic responses provides information for constructing computational models of cell biology. Current approaches for network modeling have the following limitations: i) modeling of large system requires prior knowledge that is usually aggregated from different biological systems and therefore lacks consistency. ii) data-driven models are difficult to scale due to the lack of efficient and accurate parameter inference methods. We aim to develop a computational modeling approach of cell biology that is context-specific and can be scaled to large systems. Using a perturbation-response dataset of a melanoma cell line after drug treatments as a testbed, we developed a hybrid approach that combines explicit mathematical models of cell dynamics with a machine-learning framework to quantitatively predict cell behavior in response to perturbation of molecular targets. When sufficient perturbation-response data is available for model training, this model can predict cellular response to a vast number of unseen combination perturbations and therefore efficiently narrow down the search space and nominate promising sets of experimentally testable combination candidates.

Investigation of resistance mechanisms can be used for rational design of combination therapy. Current experimental approaches to identify resistance mechanisms have the following limitations: i) targeted approaches focus on a few key proteins or signaling pathways and cannot provide a comprehensive picture of resistance landscape. ii) large molecular profiling of drug response usually focus on the changes in gene expression (transcripts) instead of proteins, which are the direct functional targets of most anticancer drugs. We aimed to perform proteomic profiling for comprehen-

sive discovery of resistance mechanisms. We used unbiased quantitative protein mass spectrometry to assess the cellular response profile to a small number of anti-cancer drug perturbations in ovarian cancer cells. Data-driven protein network analysis revealed known and novel markers of resistance, which we used to propose combination drug candidates. In preliminary validation experiments, synergistic and additive effects were observed for some combination candidates across multiple ovarian cancer cell lines, suggesting potential therapeutic value for future pre-clinical and clinical studies.

eXplainable AI proposes creating a suite of ML techniques that enable humans to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

A.B.Arrieta et al. [116]

1

Computational modeling of cellular response to perturbations

Most of the work presented in this chapter except for the last section "Perturbation biology resource" has been published in the paper titled "CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy" by B. Yuan and C. Shen et al., *Cell Systems*, **12**, 128-140 (2021) [117].

Author list

Bo Yuan, Ciyue Shen, Augustin Luna, Anil Korkut, Debora S. Marks, John Ingraham, Chris Sander

Author contributions

Conceptualization, C. Sander and J.I.

Methodology and Software, B.Y., C. Shen, A.L., and J.I.

Writing - Original Draft, B.Y., C. Shen, and C. Sander

Writing - Review & Editing, B.Y., C. Shen, A.L., A.K., D.S.M., J.I., and C. Sander

Resources, A.K., and C. Sander

Visualization, B.Y. and C. Shen

Funding Acquisition, C. Sander

Supervision, C. Sander.

1.1 Rationale and model overview

Systematic perturbation of cells followed by comprehensive measurements of molecular and phenotypic responses provides informative data resources for constructing computational models of cell biology. Models that generalize well beyond training data can be used to identify combinatorial perturbations of potential therapeutic interest. Major challenges for machine learning on large biological datasets are to find global optima in a complex multidimensional space and mechanistically interpret the solutions. To address these challenges, we introduce a hybrid approach that combines explicit mathematical models of cell dynamics with a machine-learning framework, implemented in TensorFlow, which we call it CellBox. CellBox is designed to be a framework for computational modeling of cellular responses to perturbations that (1) links perturbations to molecular and phenotypic changes in a unified computational model, (2) quantifies time-dependent (dy-

namic) cellular responses, (3) promises training efficiency and scalability for large-scale systems, and (4) is interpretable in terms of interactions that can be compared with established models of molecular biology, such as signaling pathways. Here, we construct a nonlinear ordinary differential equation(ODE)-based model representing a biological network of 99 components connecting perturbations, protein response, and phenotypes to simulate dynamic cellular behavior. The network connections are directly learned from post-perturbational data under 89 experimental conditions with the objective of accurately reproducing the molecular and cellular responses on training data.

In order to construct a data-driven model to predict the dynamics of molecular and cellular behavior under combinations of drug treatments, the perturbation data must have (1) paired measurements of changes in protein levels and cellular behavior for a set of perturbations and (2) training and withheld data to test model performance. Here, we used a perturbation dataset for the melanoma cell line SK-Mel-133 [104], which contains molecular and phenotypic response profiles of cells treated with 12 different drugs and their pairwise combinations (Figures 1.1A). For each of the 89 perturbation conditions, levels of 82 selected proteins and phosphoproteins were measured in cell lysates before and 24 h after perturbation on antibody-based reverse-phase protein arrays (RPPA). In parallel, cellular phenotypes were assayed, including cell-cycle progression and cell viability. With parallel measurements of proteomic and phenotypic responses to a systematic set of perturbations, this dataset provides sufficient information to construct network models that quantitatively link molecular changes to cellular responses.

We used a set of ODEs with a nonlinear envelope (Figure 1.1B) to model the dynamic responses of the system to drug perturbations (Method details). The parameters of the ODEs (w_{ij}, $\sim$10,000 in total) are the interaction strengths between the entities in the network model. The simplicity of the interaction dynamics (Figure 1.1B), the nonlinear envelope, as well as the restoration term $-\alpha_i x_i(t)$ are computational devices, roughly analogous to mean-field approaches, to account for the fact that the data are limited to a relatively small fraction of all cellular components and to avoid instabili-

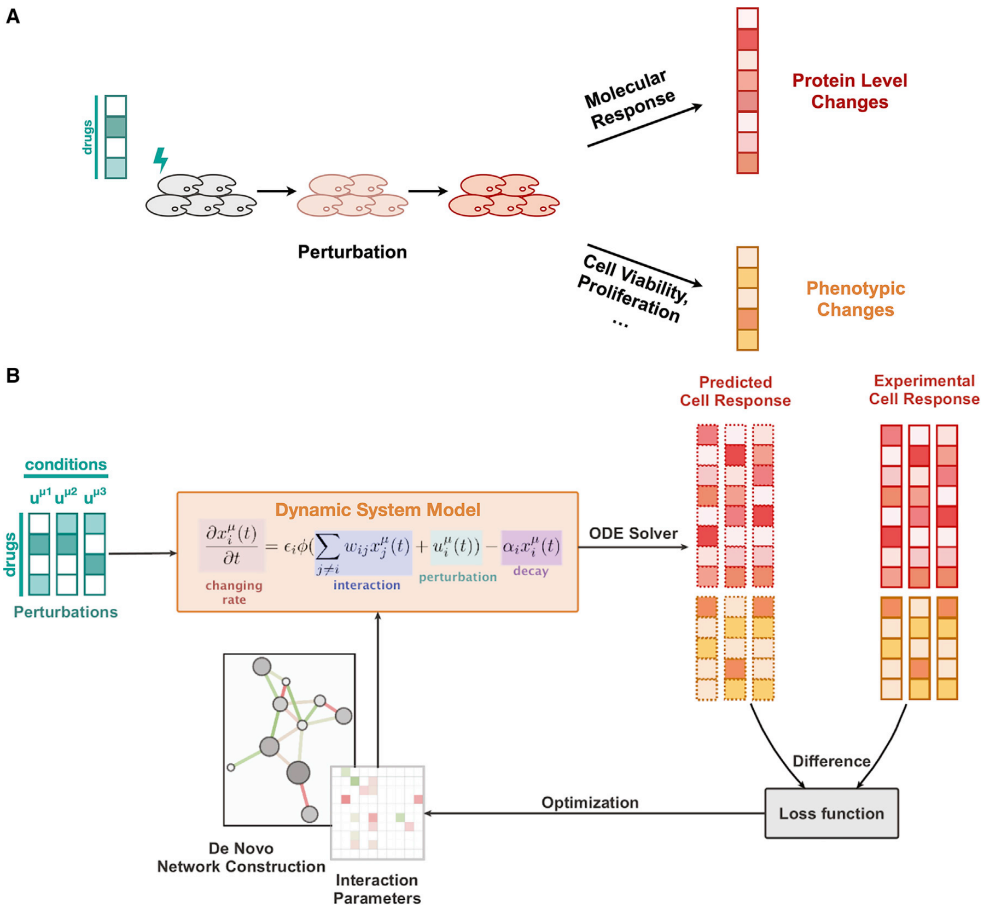


Figure 1.1: CellBox: dynamic modeling of cellular systems with perturbation data. (A) Perturbations such as drugs were used to disturb the cellular system. The cell responses, including protein- and phosphoprotein-level changes, and phenotypic changes, were measured to provide information for model construction. (B) Systematic responses of the cells under various drug perturbations were used to construct an interpretable machine-learning model. CellBox models system behavior in terms of interaction parameters connecting molecular (proteins and phosphoproteins) and phenotypic variables using a set of differential equations. CellBox was trained iteratively by optimizing interaction parameters to fit the numerically simulated system response to experimental observations. After training on pairwise data of input perturbation and output system behavior, the CellBox model can be used to predict the cellular response to arbitrary perturbation conditions.

ties [104,107,118]. The interaction parameters were randomly initialized and updated throughout the model training process, with the objective of minimizing a loss function. For the loss function, we chose the Euclidean distance between experimental data and the results of the numerical simula-

tion of the ODE model, plus an L1 regularization penalty on network density to avoid overfitting (Equation 1.3). We used Heun's ODE solver[119] to numerically simulate the ODE system and the Adam optimizer[120] with automatic differentiation to minimize the loss function. Taken together, we constructed an ODE model of a cell-biological system that was trained by using perturbation data, which we named CellBox.

1.2 Model performance in training and cross-validation

CellBox Can Be Trained on Perturbation Data to Predict Cell Response Accurately

In order to test the prediction performance of this training scheme, we randomly selected 70% of the perturbation data (n = 62 conditions) for training and withheld the rest 30% (n = 27 conditions) for testing. 20% of the training data were used as a validation set to stop model training when the performance on the validation set did not further improve. We manually fine-tuned the hyperparameters, including learning rate, regularization, and ODE simulation time, to increase the training efficiency. At the end of the training, the numerical solutions of the ODE model converged efficiently to experimental data (Figure 1.2A,B). We repeated the modeling scheme with 1,000 independent random data partitions to construct models for each partition. The average predictions on test sets across all models and all conditions correlate with experimental data with a Pearson's correlation coefficient of 0.93 (Figure 1.2C). A more refined analysis of individual perturbation conditions showed that the model trains equally well for all conditions (Figure 1.2D), and that model performance is independent of data scaling that is normally applied to protein-profiling data. The results illustrate that the CellBox models can be efficiently trained with perturbation data to predict cell response to experimentally applied perturbations accurately.

Even though ~70% of the models reached steady solutions of the ODEs (Figure 1.2B), some models converged to oscillatory solutions. In order to test whether the oscillation is an artifact of

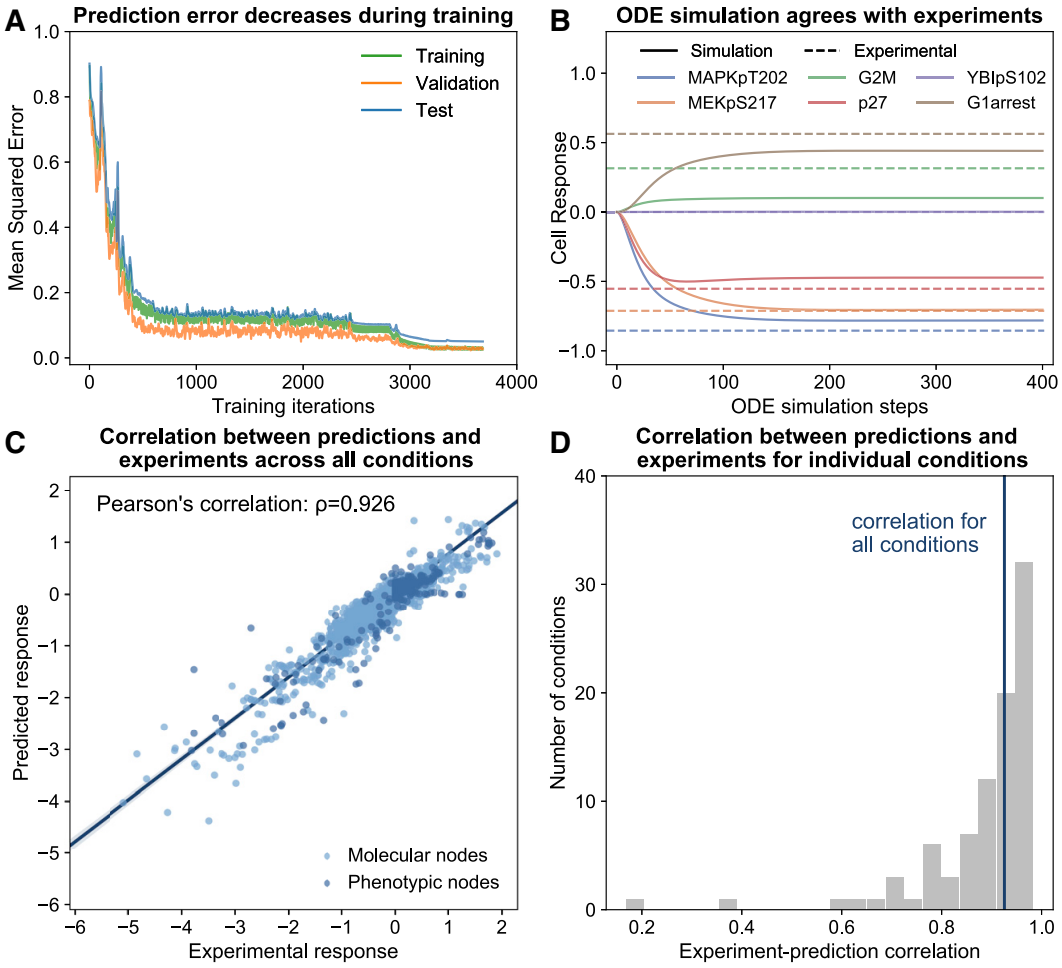


Figure 1.2: CellBox convergence and prediction accuracy on training-test datasets. (A) Over training iterations, the mean-squared error on the training set (56% of the entire dataset), validation set (14%), and test set (30%) decreased nearly monotonically, and the models converged at the end of the training. (B) The predicted molecular and phenotypic responses at the steady state of the ODE simulations agree with the experimental data on the test set. A subset of molecular measurements (MAPKpT202, YB1pS102, MEKpS217, and p27) and phenotypic measurements (G2M and G1arrest) are shown. Cell response is defined as the log2 ratio of post- and pre-perturbation measurements. (C) Across 1,000 models trained with different data partitions, the average predicted responses correlate with experimental observations (Pearson's correlation r = 0.926, regression line in dark blue with 95% confidence interval). Each point represents one measurement, either molecular or phenotypic, in one perturbation condition. (D) Nearly all predictions for individual conditions have high correlations with experimental measurements.

data partitioning during model training, we retrained the models with the same train-test data partitioning but with multiple different random seeds for the computational optimizer (Method details). Each partition of the training data can result in both steady and oscillatory solutions. The emergence of oscillatory solutions is independent of ODE solvers. On the basis of the assumption that the population average of cell response reaches a stable and non-oscillating steady state 24 h after drug treatment, we excluded the oscillatory models in the following analysis (Method details). Altogether, these results indicate that CellBox, a data-driven ODE-based cellular system model, can be trained to accurately predict dynamics of cell response without any requirement of prior knowledge about the relationship between particular protein levels or phenotypes.

CellBox Model Predicts Cell Response for Single-to-Combo and Leave-One-Drug-Out Cross-Validations

Even though the model makes accurate predictions with different training data, data partitioning, especially random partitioning, raises the concern of information sharing between training and test datasets. Combinatorial conditions in both datasets might share the same drugs such that the test set might not be truly independent of training and, therefore, is suboptimal for rigorous evaluation of the model performance. Moreover, the ability to predict the combinatorial effect of a drug, e.g., dominant, additive, synergistic, when none of its combinations has been seen by the model, is a nontrivial challenge in the context of making accurate predictions of experimentally untested drug combinations.

In order to address these points, rather than training the model with random data partitioning, we instead designed more rigorous tasks: single-to-combo (Figures 3A) and leave-one-drug-out cross-validation (Figures 1.3B,C) for each drug. In the single-to-combo analysis, all single-drug-treatment conditions were used for training, and predictions were made on all combinatorial drug conditions. In leave-one-drug-out cross-validation, all the combination conditions containing the

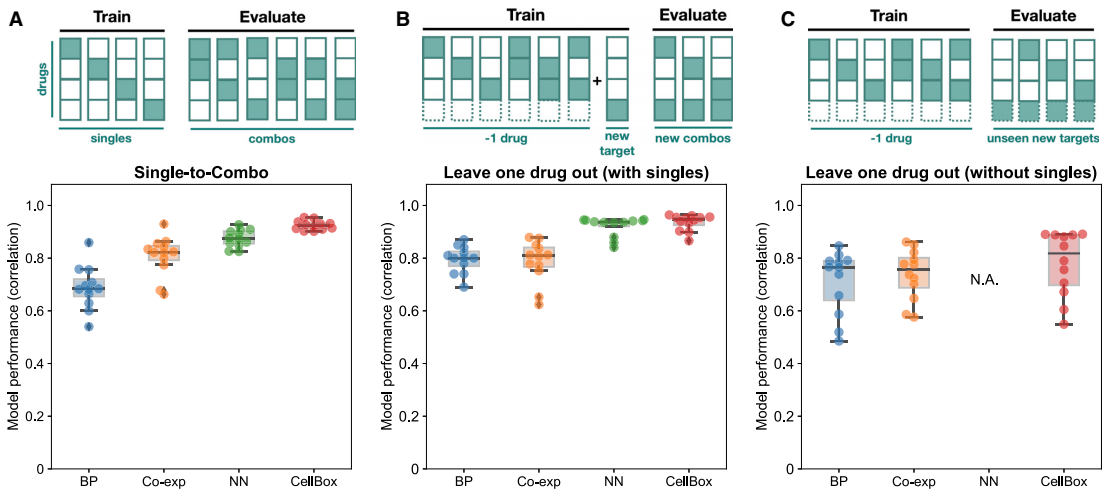


Figure 1.3: CellBox prediction accuracy for two types of cross-validations. (A) When only single conditions were used for training (single-to-combo), the CellBox models predict the effects of combinatorial conditions with high accuracy and outperform the dynamic network model inferred by using BP, the static co-expression network model (Co-exp), and a neural network regression model (NN) trained on the same data. (B) When combinatorial conditions associated with one drug were withheld from training, the CellBox models retain high accuracy for predicting the effects of unseen drug pairs. (C) When all conditions associated with one drug were withheld from training, the ODE network models predict the effects of the withheld drug with reduced accuracy, but direct-regression models such as NN cannot generalize to unseen targets at all. For each model type, the performance was evaluated by Pearson's correlation between predicted cell response and experimental cell response. The box charts indicate the group means and standard deviations.

treatment of a particular drug with or without the corresponding single-drug conditions were withheld and the rest of the conditions were used for training. In these more stringent tests, we found that the predicted values for withheld data were still highly correlated with the experimental observations (average Pearson's correlation: 0.93 for single-to-combo; 0.94 for leave-one-drug-out with single conditions, similar to that of the training with random partition; 0.79 for complete leave-one-drug-out). Under all three scenarios, on this dataset, CellBox outperforms the Belief Propagation (BP) dynamic model approach previously used in perturbation biology [104] in terms of predictive accuracy. These results indicate that the CellBox model can be trained with a relatively small set of perturbation data and that its predictions can be generalized to unseen combinatorial perturbations. In particular, CellBox models predict more accurately than linear models in the single-to-combo

scenario, suggesting that CellBox can capture the non-additive (synergistic or antagonistic) effects, which is particularly useful in nominating therapeutic drug combinations.

CellBox models are dynamic network models of a cell-biological system. To test whether such interpretable network models of molecular interactions help increase model predictive power, we compared the results to those of a static biological network model and a deep neural network model. The static network model was constructed by learning co-expression correlation for each pair of protein nodes (Co-exp) while the deep neural network model was trained to directly regress phenotypic changes against parameterized perturbations (NN) (Method details). In all three tasks, the static network models had lower accuracy relative to the dynamic CellBox. The NN had comparable performance to CellBox in the cross-validation for individual drugs, but its performance dropped significantly in the single-to-combo analysis (Figure 1.3A). Furthermore, the NN was unable to generalize to unseen targets whose information is completely excluded from training (Figure 1.3C). Altogether, because of the lack of mechanistic and dynamic information, static network or direct-regression models appear to be less suitable for facilitating the search for combinatorial targets.

1.3 Model robustness against data noise and data reduction

To examine model robustness of the CellBox models against a reduction in training data, we tested the stability of model performance when either the data quality or quantity is compromised. To test the former, we introduced different levels of multiplicative Gaussian noise (Method details) into the input molecular and cellular response data and trained models on the resultant noisy datasets. The assumption behind such multiplicative noise is that the uncertainty and noise in experimental measurements arise around the true values. When comparing the predicted response in test sets with the experimental data, we found that the predictions from training on the noisy data retain similarly high correlations to experimental data as those trained on the original data, even with the addition

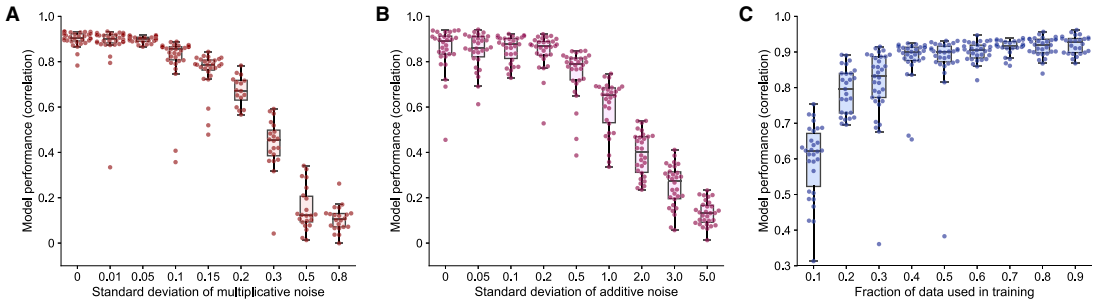


Figure 1.4: Model robustness against data noise and data reduction. (A and B) Correlation between predicted responses and experimental responses in the test set decreases as an increased level of multiplicative noise (A) or additive noise (B) is added to the training data (each dot represents one model). The CellBox models can tolerate up to σ_{mul} = 0.05 multiplicative noise or σ_{add} = 0.20 additive noise. (C) Correlation between predicted responses and experimental responses in the test set increases with an increasing quantity of data used for model training. For the current dataset, the correlation plateaus when 40% of the original dataset is used. The box charts indicate the group means and standard deviations.

of 5% multiplicative Gaussian noise (Figure 4A). As the magnitude of the noise increases, the model performance decreases gradually in terms of both convergence and predictive power (Figure 1.4A). We observed similar behavior when challenging the CellBox models with additive Gaussian noise (Figure 1.4B). The predictivity of CellBox models can tolerate an addition of up to σ_{add} = 0.20 additive Gaussian noise, i.e., about half the data standard deviation σ_{add} = 0.46. We conclude that model performance is stable in the presence of moderate experimental errors.

To test the dependency of model performance on data quantity, we trained the model on subsamples of the experimental dataset. We trained models with varying amounts of data (from 10% to 90% in steps of 10%) and found that the models could make accurate predictions of withheld data with as little as 40% of the complete dataset (Figure 1.4C). We found that increasing the size of the training set further has diminishing returns in terms of model performance. This implies, on the dataset used here with an interaction network of ~100 components, that a comparatively small number of perturbation conditions (40–100, rather than directly testing all ~3,000 possible combinations) is sufficient for constructing reasonably predictive models. This example might be a useful

guide for power calculations for systems with hundreds of measured components, which would be of considerable interest.

1.4 De Novo network recapitulates some known interactions

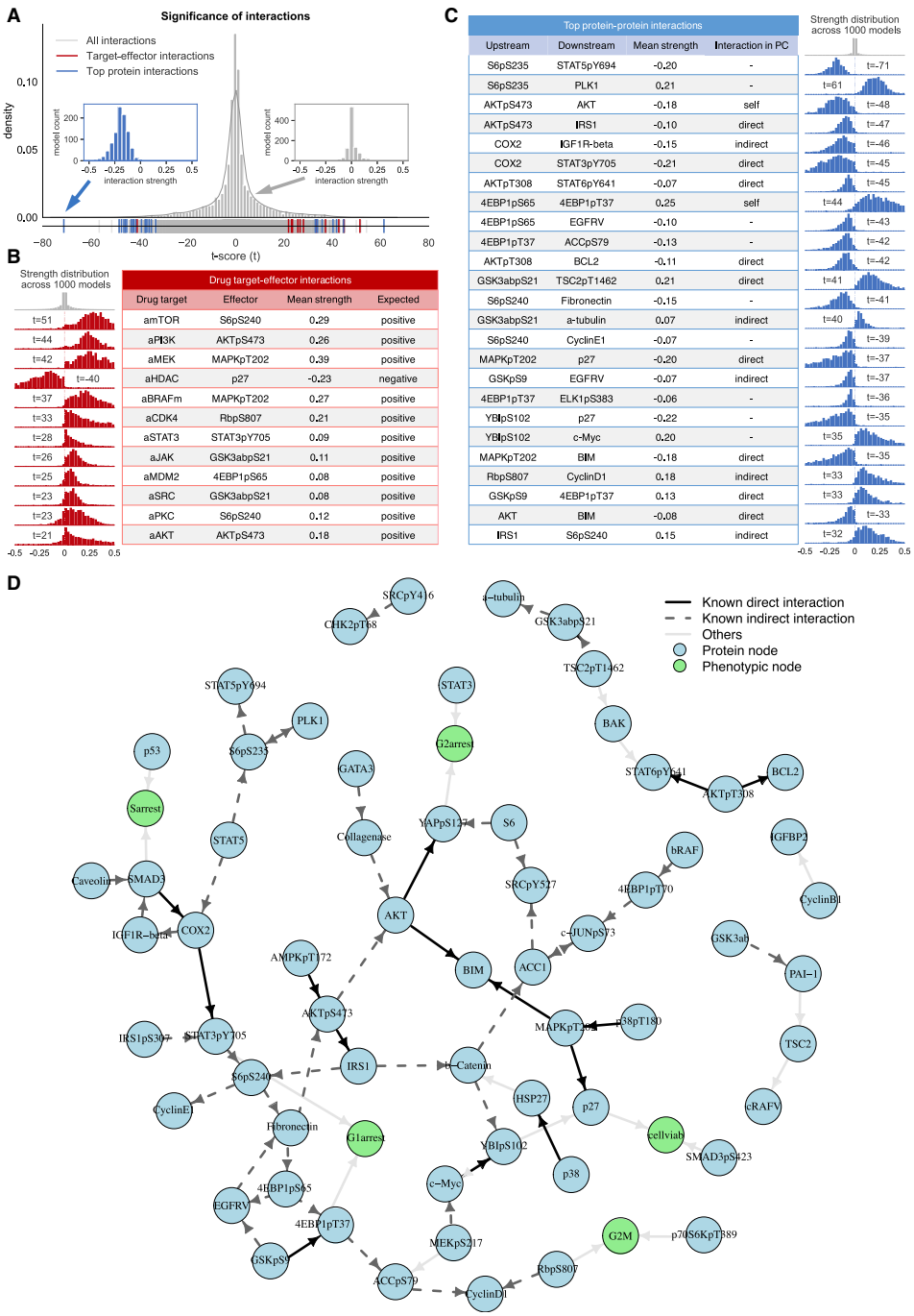
We used ODEs as the core framework of the current version of the CellBox mathematical model. Each parameter in the model represents the strength and direction of a biological interaction. In order to investigate whether the inferred interactions are consistent with current knowledge of biological pathways, we used the entire dataset as training data to generate 1,000 full models and examined the resulting de novo network edges learned from training. In order to measure both the strength and stability of edge inference, we used t scores (Method details) to assess the statistical significance of each interaction, where a higher absolute t score indicates higher interaction strength and lower variance across the models (Figure 1.5A). Using the primary targets of the drugs as the ground truth, we first examined the interactions between the drug-activity nodes and their down-stream effectors. We found that all 12 drug-activity nodes had significant edge connections to their known primary down-stream protein effectors with the interaction directions consistent with their expected effects (Figure 1.5B), suggesting the models are able to capture the literature-provided interactions between the drug target and their downstream effectors.

To further investigate to what extent the inferred networks represent known pathway interactions, we compared the interactions in the CellBox models to corresponding molecular interactions extracted from the Pathway Commons (PC) resource [121], which is an aggregation of ∼20 curated publicly available pathway interaction databases. In this comparison between interactions inferred by CellBox models and prior knowledge, model interactions can, in principle, be found in PC either as one-step (A-B), two-step (A-X-B), or logical (>2 steps) interactions, where the logical interactions can be usefully predictive or perhaps erroneous. Direct interactions (A-B) in PC represent one component (A) affecting the expression, phosphorylation, or state-of-change of the other component (B) (Figure 1.5C,D). For example, models and prior knowledge agree for the phosphorylation of the AKT protein kinase (AKTpS473), which negatively affects the insulin receptor substrate 1 (IRS1) [122], and for the activation of the mitogen-activated protein kinase (MAPKpT202) that affects p27 protein levels [123,124]. Some model interactions can be found as indirect two-step interactions (A-X-B) in PC, meaning at least one path with one intermediate component (X) can be found between the two components (A and B). For example, retinoblastoma protein (Rb1) affects cyclinD through p21 [125,126], and mitogen-activated protein kinase kinase (MEK1) indirectly interacts in two steps with the transcription factor c-Myc by a phosphorylation mechanism via ERK1/2 [127,128,129,130]. Additional evidence of functional links between proteins identified in the models includes protein-protein edges in the STRING database [131], such as links that represent expression (mRNA) correlation, and PubMed-derived links from text mining [132].

In order to confirm that such agreement between the inferred networks and prior knowledge is not an artifact, we examined solution stability as well as compared the inferred networks to random networks. We conducted stability-selection tests [133] on the interaction parameters, and the results indicate that the solutions are reproducible and robust. We compared the inferred networks to networks with randomly drawn interactions (Method details). In the inferred networks, we observed a statistically significant higher number of top interactions consistent with the knowledge in PC,

Figure 1.5 *(following page)*: Comparison of the network models with prior knowledge about pathways. (A) The t score distribution of all interactions across 1,000 full models suggests that a small fraction of interaction strengths is significantly different from zero. Insets are two examples of interaction strength distributions across models.(B) All 12 interactions between drug target (drug-activity nodes) and their downstream effectors (red bars in A) are significant, and the interaction directions are consistent with the literature. (C) Most of the top significant protein-protein interactions (blue bars in A) can be found as direct or indirect interactions in PC. The distributions of interaction strength across 1,000 models for each interaction in the two tables with corresponding colors are centered away from zero, in contrast to the background distributions of aggregated interactions across models (gray, all interactions with drug-activity nodes in A, all protein-protein interactions in C). (D) Network visualization of the top interactions (top two interactions acting from each node and onto each node) highlights the level of agreement between model-inferred interactions and those in PC.

Figure 1.5: (continued)



indicating that such agreement is not an artifact of the densely connected signaling networks in the pathway database. The remaining model interactions are between components that are more than two pathway steps away or cannot be connected by any path in PC. These interactions can be interpreted as either logical interactions important for predictive purposes or potential new physical interactions that are yet to be discovered in molecular experiments. Also, the PC database aggregates interaction information from multiple biological systems, and, therefore, complete consistency is not expected for any particular system, e.g., the melanoma cell line used here. Given that the CellBox network models are constructed in a completely data-driven way without any a priori intention to recapitulate known molecular interactions, the partial agreement between the model-inferred molecular interactions and the experimentally known ones from the literature is evidence of the validity of the modeling approach.

1.5 Prediction of unseen combination perturbations

Our results so far indicate that the CellBox models can be efficiently trained on a relatively small set of experimental data to parameterize the differential equations that model the behavior of the entire system of nodes and interactions at a reasonable level of predictive accuracy. This model can then predict cell responses to a full range of single and combinatorial unseen perturbations, which would be laborious and costly to test exhaustively by experiments. In order to nominate effective drug combinations for a much reduced number of focused experiments, we used simulations of the 1,000 full models to quantitatively predict the dynamic cell responses to ~110,000 in silico perturbations, including different dosages of single perturbations on each protein node as well as all pairwise combinations (Method details). For each perturbation condition, we averaged the predictions across all models and ranked the perturbations by predicted phenotypic changes (Figure 1.6A).

Previous models on the same dataset, using the same differential equations but parameterized by

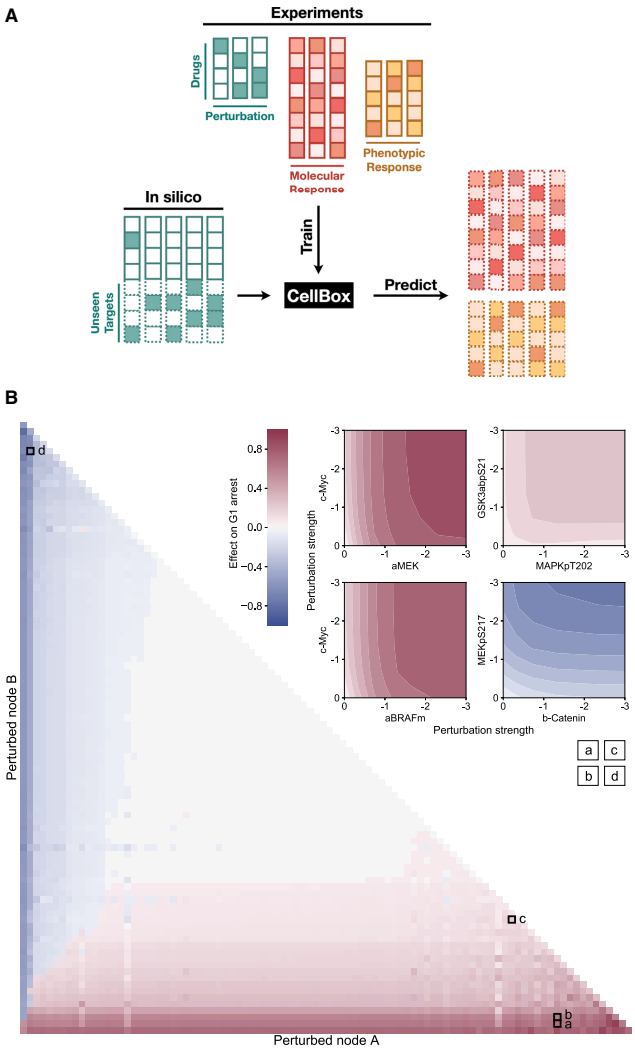


Figure 1.6: CellBox provides testable predictions of cell phenotype under synthetic perturbations. (A) For each (phospho)protein node in the network, CellBox was used to simulate all single and paired inhibitions and to predict the phenotypic changes. The phenotypic effects are the average prediction of 1,000 independent models trained on the full datasets. (B) The anti-proliferation effects of two perturbation pairs whose effects on cell-cycle arrest have been experimentally tested were closely examined (left two panels, c-Myc+MEKi [a], and c-Myc+RAFi [b]), as well as two other in silico conditions (right two panels, GSK3p+MAPKp [c] and MEKp+b-catenin [d]), by simulating with combinatorial perturbation strengths. The effects on cell-cycle arrest of pair-wise combinatorial perturbation of all (phospho) proteins in the network were simulated and can be used to nominate effective pharmaceutical candidates. These in silico inhibitory perturbations can result in anti-proliferation effects (red, bottom right) or pro-proliferation effects (blue, top left).

using BP, had predicted that two drug pairs, MEKi+c-Myc and RAFi+c-Myc, would increase G1 cell-cycle arrest, and this prediction was confirmed by experiments [104]. We found that the CellBox model predicts similar effects for these two drug pairs (Figure 1.6B, panel a, b). In order to identify additional therapeutic candidates, we examined the effects of all possible single and pairwise perturbations on cell-cycle arrest (Figure 1.6B). The top-ranked candidates included dominant anti-proliferative inhibition (uniform colors in rows or columns) of proteins in the Wnt, MAPK, and ERK/MEK pathways, which are known to be cancer related. Besides strong single candidates, synergistic drug pairs are of potential therapeutic interest (Figure 1.6B; departure from uniform colors). Inhibitory perturbations predicted to have pro-proliferation effects, which are undesirable as such, can also lead to effective anti-proliferative candidates via indirectly activating perturbations (top left corner). For example, protein nodes can, in principle, be activated by reducing upstream inhibition or degradation. As the CellBox models are completely data driven, the de novo predictions represent system-specific predictions independent of prior knowledge.

1.6 Significance and future application

Quantitative models that are predictive of dynamic cellular responses can be used to design combination therapies in cancer. To provide predictions with sufficient accuracy and potential mechanistic insight, we integrated machine-learning methods with dynamic modeling: we applied an optimization algorithm used in deep learning to a biologically interpretable differential equation (ODE) system. Our model CellBox can be trained efficiently and independently of prior knowledge to predict molecular and phenotypic responses to unseen perturbations with high accuracy. Although trained on a relatively small set of experiments, the model is capable of simulating cell responses to numerous arbitrary combinatorial perturbations applied to nodes repeatedly measured under different perturbation conditions. Ranking of cellular responses to the in silico combinatorial

perturbations by the desired phenotypic outcome, such as decreased proliferation, potentially leads to specific therapeutic hypotheses.

Interpretability of models that are to be used for practical decisions, such as the design of combination therapy, helps increase confidence and facilitates the design of focused validation experiments and is, therefore, as important as accuracy [134]. CellBox features interpretability in two aspects: transparency and traceability. Transparency: by using a well-defined mathematical model, CellBox is designed to be explicitly interpretable. In the current ODE model, each parameter represents a directed and quantitative interaction between cellular components or phenotypic quantities. Traceability: given a perturbation to the cellular system, the ODE simulation indicates how the effects of the perturbation propagate throughout the directed network in a time-dependent manner. The models can, therefore, provide mechanistic hypotheses of how the perturbations cause the observable cellular responses. Our model aims to meet the emerging demands of explainable artificial intelligence (XAI), which is essential for a socially acceptable application of machine-learning models in biomedicine [135,136].

In principle, CellBox is generalizable to other types of systems and larger systems. Other types of models will presumably benefit from automatic differentiation (AD) combined with stochastic gradient descent that performs optimization directly for any given mathematical ansatz and, therefore, can avoid oversimplified approximations [137]. The AD framework allows the flexible use of various mathematical forms of cellular dynamics. Whether alternative dynamics models lead to higher concordance between interactions inferred from the data-driven CellBox models with interactions in prior-knowledge mechanistic models or with biophysical interactions separately verified by direct experiments is to be investigated. We certainly expect increasingly detailed model-inferred interactions and closer agreement with biophysical interactions as larger datasets become available, i.e., with perturbation-response data covering many hundreds if not thousands of changes in molecular levels or covalent modifications.

The ability to model larger systems depends both on the availability of larger datasets and scalable modeling methods. Larger datasets can be obtained by measuring diverse types of molecular data, for example, transcriptomic, epigenomic, and metabolomic changes [138,139]. A major opportunity for larger datasets might arise from recent cell barcoding techniques that significantly increase perturbation throughput relative to arrayed experiments [78,140] by measuring levels of transcripts by sequencing, levels of proteins detected by antibodies labeled by oligonucleotides or isotopes at the single-cell level, or levels of both by multiplexing [141,142,143,144]. A key advantage of single-cell perturbation approaches for data-driven inference of dynamic networks is scale, as barcoded perturbation experiments can be pooled and individual cells identified on the basis of sequence tags.

As CellBox is implemented in the Google TensorFlow framework, it can make use of various advanced machine-learning techniques, such as dropout, mini-batching, and GPU boosting [145,146], to improve training efficiency, which partially addresses the issue of scalability [147]. As most state-of-the-art single-cell sequencing technologies still suffer from limited sequencing depth and inevitable noise [148], a new set of computational challenges arises when applying CellBox models to datasets with single-cell readouts. Future efforts are needed to resolve issues of sparsity and stochasticity, e.g., using robust techniques of dimensionality reduction [149,150,151]. Recent work has highlighted mathematical analogies between recurrent neural networks (RNN) and ODEs [152], suggesting a potential merge of the two fields for scientific machine-learning (SciML) models [153].

A particular advantage of our modeling approach is its complete independence from prior knowledge. This is in contrast, e.g., to other frameworks for interpretable models [103] that incorporate prior knowledge by predefining the connections in the network model based on large-scale curation of molecular and biochemical pathways. Such models have been further developed and incorporated with proteome/phosphoproteome data [154]. Although our current modeling framework is completely data driven, prior knowledge of cellular interactions could also be included in the optimization function by adding a term, which rewards the agreement between the inferred and

prior-knowledge values for each interaction parameter. However, considering the incompleteness of current pathway knowledge bases and the relatively high prediction accuracy of CellBox, we speculate that rewarding agreement with prior knowledge might give limited improvement in model performance. Besides, it would also be of interest to explore whether one can combine the efficiency advantage described in Fröhlich et al. [103] via adjoint analysis [155] with the advantages of the current implementation of CellBox.

For broad translational applicability, a tantalizing but challenging prospect is to incorporate individual tumor background via genetic perturbations and to propose optimal, personalized combinations of targeted therapeutics. We envision the systems biology approach described here to be broadly applicable to other areas of biology, such as developmental biology or synthetic biology, provided that suitable perturbation-response data become available. Key future challenges are, therefore, the design of experiments for each biological context of interest and the further development of transferable and scalable machine-learning methods.

1.7 Method details and data availability

Data Availability

The RPPA data used during this study is deposited and publicly available at https://github.com/dfci/CellBox under the data folder. All original code is publicly available at https://github.com/dfci/CellBox as release v0.3.1. A shareable, online, interactive environment using Binder is provided with all necessary dependencies pre-installed. The Quick Start section provides instructions to quickly try an example script that runs a shortened training process. The scripts used to generate the figures reported are available at https://github.com/dfci/CellBox under the manuscript folder.

Perturbation Dataset Overview

The CellBox models were trained using a perturbation-response dataset of the SK-Mel-133

melanoma cell line [104]. The cells were treated with 12 different single drugs, each at two different concentrations and 66 pairwise combinations of these drugs at IC40 concentrations. 24 hrs after drug treatment, Reverse Phase Protein Arrays (RPPA) was used to measure the level of 45 proteins and 37 phosphoproteins of interest. Cell-cycle progression, including G1 arrest, G2 arrest, G2/M transition, and S arrest, was measured by flow-cytometry. Cell viability was measured 72 hrs after drug treatment by the resazurin assay. Our experimental data were normalized using median normalization, which is a standard approach to processing RPPA data [156]. We used the log2 ratio of measurements in perturbed conditions over unperturbed conditions as the system variables x_i in the models. The dataset was initialized with 12 drug activity nodes representing the inhibition strengths of different drugs to their targets [118]. The resulting dataset has 89 perturbation conditions and 99 observed nodes (82 protein and phosphoproteins, 5 phenotypes, and 12 drug activity). A more detailed description of the experimental dataset is available in Korkut et al. [104].

Model Configuration

The models were constructed using Python 3.6 and Google TensorFlow [157] (version = 1.15.0). The molecular and phenotypic changes are linked in a unified biological network model using a system of ordinary differential equations

$$\frac{\partial x_i^{\mu}(t)}{\partial t} = \varepsilon_i \varphi\left(\sum_{j \neq i} w_{ij} x_j^{\mu}(t) + u_i^{\mu}(t)\right) - \alpha_i x_i^{\mu}(t) \tag{1.1}$$

where $x_i^{\mu}(t)$ represents the log_2-normalized relative change of each (phospho)protein or phenotype levels relative to control levels under condition μ. $u_i^{\mu}(t)$ quantifies the strength of the perturbation on target i. Here the drug effect is assumed to be constant and, therefore, $u(t) = u$ for $t > t_0$ is the perturbation strength determined using the endpoint level change of the primary target of the particular drug. α_i characterizes the effect of decay, meaning the tendency of protein i to return to the original level before perturbation. The interaction parameters w_{ij} indicate interactions between

network node j on network node i, assumed to be a constant property of the pair of molecules in this given cellular setting. We constrain the interaction parameters w_{ij} by disallowing three classes of interactions:

i) ingoing connections for drug nodes (drugs cannot be acted upon by any other node)

ii) outgoing connections for phenotypic nodes (phenotypes cannot act on any other nodes)

iii) self-interaction (nodes cannot act on themselves)

We use a sigmoid function $\varphi(x) = tanh(x)$, to introduce a saturation effect of the interaction term so that it is bounded by the constant value of ε_i. Imposing bounds on the absolute value of the contribution of the interaction term to the derivatives is in part motivated by the fact that we are only modeling a small fraction of cellular components and by the experimental observation of system stability in these experiments. We also tested different alternatives of envelope forms, including a clipped linear function (hard tanh) $\varphi(x) = max(-1, min(1, x))$, symmetric polynomial function $\varphi(x) = \frac{\tilde{x}^n - 1}{\tilde{x}^n + 1}, \tilde{x} = max(0, x)$ (adapted from Hill's equation), and no envelope function $\varphi(x) = x$ (linear).

The biological network interactions were constructed de novo without any prior knowledge input, meaning the network was fully connected with interaction parameters. The interaction parameters w_{ij} were randomly initialized following a normal distribution $\sim N(0.01, 1)$. The other two coefficients α_i and ε_i were initialized as 1.0.

Taken together, with this formulation of system dynamics, the effect of each new input perturbation is quantified by the response of downstream protein and phenotypic nodes, and the effect of the perturbation is simulated by propagating the input across the inferred interaction network according to the differential equations. After learning from perturbation-response data, the model can therefore make predictions for responses of perturbations on any experimentally probed node, including nodes not perturbed in the experimental dataset from which the models are derived. The ODE system was numerically solved using Heun's method [119] (Equation 1.2, time steps N_t = 400),

which is an improved variant of Euler's method. Model performance was evaluated by disagreement between the experimental cell responses and the numerical steady-state levels.

$$\tilde{x}_i(t+h) = x_i(t) + hf(t, x_i(t))$$
$$x_i(t+h) = x_i(t) + \frac{h}{2}[f(t, x_i(t)) + f(t+h, \tilde{x}_i(t+h)] \tag{1.2}$$

where h is the step size, $f(t, x_i(t)) = x'(t), y(t_0) = log(1) = 0$.

The loss function $L(w)$ is defined as a weighted sum of prediction error and complexity penalty in order to avoid overfitting. Here a mean squared error (MSE) and an L1-loss regularization term are used, as defined in Equation 1.3. We have tested different regularizations, including L1, L2, and both combined (elastic net) with various strengths (regularization strengths for L1, L2 term λ_1, λ_2= 0, 0.01, 0.0001, pairwise combination), and found no significant difference in model performance. The interaction parameters were optimized end-to-end using the Adam optimizer [120], with the objective of minimizing the loss function.

$$L(w) = \sum_{\mu} \sum_{i} \|\hat{x}_i^{\mu}(w_{ij}, t) - x_i^{\mu *}\|_2 + \lambda_1 \|w_{ij}\|_1 \tag{1.3}$$

$\hat{x}_i^{\mu}(w_{ij}, t)$ is calculated as the converged value of the numerical simulation of the ODE with the interaction parameters w_{ij} and defined i simulation timestep t. $x_i^{\mu *}$ indicates experimental measurement, which is used as a gold standard in training. The dataset was divided into training, validation, and test sets, in order to optimize parameters, provide an indication for stopping training, and test model performance, respectively. Optimization was conducted with an initial learning rate for the Adam optimizer (lr=0.1) and regularization strength (λ=0.01). It has been shown that a gradually decreasing learning rate is helpful for model convergence [145]. The model training was stopped when

the loss function of the validation set does not further decrease for a continuous of 20 iterations (stopping patience).

The model was trained with mini-batching: a random 80% portion of the training set was used to optimize parameters for each iteration, and a fixed batch size of 4 was used in each epoch. Models that failed to converge (MSE for training set > 0.05) were excluded as unsuitable. For a larger dataset, we recommend using a fixed batch size of 8 or 16, as documented in the latest version of the software on GitHub.

Model Training

For initial model training and analysis of model performance, the cell line perturbation-response dataset was randomly partitioned into training, validation, and test set in the proportion of 56% (n = 50 conditions), 14% (n = 12 conditions), and 30% (n = 27 conditions). 1,500 models were generated on 1,500 independently random-partitioned datasets.

The models were examined and categorized into non-oscillating and oscillating solutions based on time derivatives at the final time step of the ODE simulation. The non-oscillating solutions are defined as those with the average absolute value of time derivatives of all nodes and conditions in the training set smaller than δ, i.e. $\frac{1}{m}\sum_{i=1}^{m}|\frac{\partial x_i^{\mu}(t)}{\partial t}| < \delta; \delta = 1e-03$. In each category, twenty models were randomly selected, and each re-trained with the original data partitioning but forty different random seeds, covering all the random processes in training, including parameter initialization and mini-batching sampling. Oscillating solutions comprise about 30 percent of all models. In the following analysis, models that converged to oscillating solutions were excluded.

Under the training scheme using random data partition, we examined the sensitivity to data scaling and to the choice of ODE solver, by applying the following modifications to the data or the model and evaluating the changes of model performance. i) In addition to the log2 scale of the ratio of measurements in perturbed conditions over unperturbed conditions, we applied other scalings to

the raw data to generate model input, including a non-log (linear) scale. ii) It has been argued that the selection of numerical methods could have a significant influence on model training[158]. To test such limitations, we tested other numerical ODE solvers, including Euler methods, Midpoint methods, and Runge-Kutta methods, in addition to Heun's ODE solver. On the current dataset, these different methods perform similarly.

Cross-Validations

In the single-to-combo task, all single-drug conditions were allocated to the training set (n = 23 conditions), and the combination perturbation conditions were randomly distributed among the validation and test set (n = 53 conditions) in a 20/80 ratio. To evaluate model performance by cross-validation for each drug, the data were partitioned into training (n = 78 conditions) and test (n = 11 conditions) sets, where each test set contains all the drug combination conditions with the particular drug. 20% of training conditions (n = 15) are used as a validation set. The predictions on the test set were averaged over 100 models. In the more stringent leave-one-drug-out task where all perturbation conditions involving one drug were withheld as the test set, the evaluation of model performance was conducted by applying perturbation of the same strength directly on the downstream effector node of each withheld drug activity node.

The Belief Propagation (BP) models for both single-to-combo and cross-validation prediction were performed as in our earlier publication (https://github.com/korkutlab/pertbio). The predictions on the test set were averaged over 100 models. The deep neural network model (NN) network had 2 densely connected hidden layers (hidden layer H1: 20 neurons, H2: 100 neurons), connecting the parameterized perturbation tensor with the cell response tensor. We used the tanh activation function for the hidden layers and the leaky rectified linear unit function (ReLU) for the output layer. The NN models were constructed in the TensorFlow framework in Python and optimized using the same optimization methods (Adam optimizer). The co-expression static model (Co-exp)

was constructed in the Python environment using the sklearn (version = 0.21.3, https://scikit-learn.org/stable/) MultiTaskElasticNetCV module, which is a linear regression model with L1 and L2 regularizations whose strengths are automatically determined using built-in cross-validation. The model was trained to use the changes in levels of each pair of protein nodes (input) to predict levels for the rest of the proteins and phenotypes (output). To compare the four different predictive models, CellBox, BP, NN, and Co- exp, the paired t-test on two related samples was used to analyze the difference between model performance (test set correlation) and to assign p-values.

Sensitivity Analysis

We conducted a sensitivity analysis of our model to evaluate the robustness of its prediction in response to noise. To model experimental noise, we first applied varying levels of multiplicative Gaussian noise to the input molecular and phenotypic data (Equation 1.4). The assumption is that the uncertainty and noise in experimental measurements arise around the true values with a scaling factor depending on the experimental approach.

$$x_i(t) = x_i^*(t) * N(1, \delta_{mul}) \tag{1.4}$$

The scaling factor for each node and each condition was independently drawn from a Gaussian distribution $N(1, \delta_{mul})$ with a mean of 1 and a standard deviation of δ_{mul}. $x_i^*(t)$ represents the experimental measurements, and $x_i(t)$ represents the values with noise added. For each noise level, we evaluated 15 different training/validation/test partitioning, and each with 5 independent random noise patterns. Model training was performed on noisy training and validation sets, while the model performance evaluation was performed on the original, noise-free test data. For each noise level, the percentage of successful models, defined as those that converged in terms of both MSE and oscillation filters, was recorded.

To further evaluate model performance from a technical perspective, we applied additive Gaus-

sian noise to the input data (Equation 1.5) and evaluated model performance as above.

$$x_i(t) = x_i^*(t) + N(0, \delta_{add}) \tag{1.5}$$

We examined model sensitivity to training and validation set size. We reduced the combined size of the training and validation set, from 90% to 10% in steps of 10%, while keeping their relative size constant, 4:1. The remaining data were allocated to the test set. For each training set size, the percentage of successful models, defined as those that converged in terms of both MSE and oscillation filters, was reported.

Network Interpretation

The entire dataset was used to generate 1,000 full successful models, each with an independent data partitioning of training (n = 71 conditions) and validation (n = 18 conditions). For each inferred interaction (w_{ij}) between two nodes from the 1,000 full models, a t-score ($\frac{w_{ij}}{s\sqrt{m}}$) (which is effectively a one-sample Student's t test with the null hypothesis that the population mean is zero) was calculated as an indication of the confidence level of obtaining a value different from zero, where wij is the average interaction strength across models, s is the standard deviation, m is the number of models (m = 1,000).

In order to compare the model inferred interactions to those present in prior-knowledge pathway databases, all the proteins and phosphoproteins nodes were identified by their corresponding gene names. The interactions were compared against the Pathway Commons (PC) database (current version at https://www.pathwaycommons.org/archives/PC2/v11) using the paxtoolsr [159] software. The database was filtered down to direct interactions, which include "controls expression of", "controls phosphorylation of", "controls state change of", "controls production of", "controls transport of", and "controls transport of chemical". All the direct interactions were converted to a directed PC graph using igraph (version = 1.2.4.1, https://igraph.org/r/). A distance was calculated as the

length of the shortest path(s) between two components in the PC graph.

To provide an additional comparison with a well-known alternative method for network inference, the partial information decomposition and context (PIDC) algorithm was implemented in the Python environment based on the detailed mathematical description of the algorithm in Chan et al.[92]. The same number of PIDC networks (m = 1,000, same as for CellBox) were generated based on 1,000 independent random partitions of the entire dataset (80% of the conditions were used to infer PIDC networks). The t test was performed to compare the difference in means between the number of interactions inferred by CellBox network models and PIDC networks consistent with the PC database.

Model Predictions

We used the models trained with the full (non-partitioned) dataset to simulate responses of novel, experimentally unobserved, in silico perturbation conditions. These conditions included different doses of single perturbations (five different levels of perturbation strengths $u \in [0, 3]$ on all individual (phospho)protein and drug activity nodes within the network and as well as all pairwise combinations ($n_{all} = 94 * 5 + C_9 4^2 * 5^2 = 109,745$ conditions). The cell responses were dynamically simulated with u as the input perturbation with the same number of steps as in training ($N_t = 400$). For each perturbation condition, predictions for cell responses were averaged across 1,000 different models. Perturbations were nominated as therapeutic candidates by ranking the predicted magnitude of the phenotypic change in terms of cell cycle arrest.

1.8 Perturbation biology resource

This is an ongoing project with the following contributing authors: Stefan Peidli, Tessa Green, Joseph Min, Jake Taylor-King, Augustin Luna, Torsten Gross, Nils Blüthgen, Chris Sander.

With the latest functional genomics techniques, we can now perform large scale CRISPR screens

with single cell readouts providing a remarkable map between genotype, transcriptome [160], and in some cases phenotype [79]. Such large screens designed by researchers to study particular systems of their interests result in a heterogeneous collection of single-cell perturbation-response datasets with a range of different cell types, such as immortalized cell lines and iPSC-derived models, and different emerging perturbation technologies, including knockouts, activation, interference, base editing, and prime editing [161]. While computational approaches with various purposes have been developed on individual datasets [78,79], methods that are generalizable across datasets are in urgent need for extensive information extraction and potential data integration. We aim to provide a resource to standardize a number of datasets combining perturbations with single cell readouts and therefore facilitate the development and benchmarking of computational approaches in system biology.

Dataset ID	Modality	Tissue	Perturbation
AdamsonWeissman2016 [140]	RNA	leukemia cell line	CRISPR
AissaBenevolenskaya2021 [162]	RNA	cancer cell lines, patient tissue, mouse xenografts	drugs
DatlingerBock2017 [163]	RNA	leukemia cell line	CRISPR, TCR stimulation
DixitRegev2016 [78]	RNA	BMDC, leukemia cell line	CRISPR
FrangiehIzar2021 [164]	RNA+protein	melanoma and immune cells	CRISPR
GasperiniShendure2019 [165]	RNA	leukemia cell line	CRISPR
GehringPachter2019 [166]	RNA	neural stem cells	drugs
JaitinAmit2016 [167]	RNA	BMDC	CRISPR
McFarlandTsherniak2020 [168]	RNA	cancer cell lines	drugs
NormanWeissman2019 [79]	RNA	leukemia cell line	CRISPR
OrenRegev2021 [169]	RNA	cancer cell lines	drugs
PapalexiSatija2021 [170]	RNA+protein	leukemia cell line	CRISPR
PierceGreenleaf2021 [171]	ATAC	leukemia and breast cancer cell line	CRISPR
RubinKhavari2019 [172]	ATAC	lymphocytes and primary human epidermal cells	CRISPR
SchiebingerLander2019 [173]	RNA	iPSCs	cytokines
SchraivogelSteinmetz2020 [144]	RNA	BMDC, leukemia and lung cancer cell lines	CRISPR
ShifrutMarson2018 [174]	RNA	Primary human T cells	CRISPR, TCR stimulation
SrivatsanTrapnell2020 [175]	RNA	cancer cell lines	drugs
TianKampmann2019 [176]	RNA	iPSC-derived neurons	CRISPR
TianKampmann2021 [177]	RNA	iPSC-derived neurons	CRISPR
WeinrebKlein2020 [178]	RNA	mouse bone marrow cells	cytokines
XieHon2017 [179]	RNA	K562 and 293T cell lines	CRISPR
ZhaoSims2021 [180]	RNA	glioblastoma patient tissue	drugs

Table 1.1: Overview of datasets for perturbation biology resource. A collection of 23 datasets are included in the study. Datasets are named by the last name of the first author, last author and year of publication (Dataset ID) of the paper with which the experiments are reported in. Modality represents the type(s) of molecular readouts. The model organisms used for the experiments (Tissue) and the type(s) of perturbations are included.

To provide a resource for the community of perturbation biology, we collected a set of 23 publicly-available perturbation-response datasets (Table 1.1). Each dataset contains at least three perturbation conditions and one or more of the molecular readouts, including transcriptomics, proteomics,

and epigenomics. We processed the data into uniform h5 format and unified the annotations for the shared features of the observations (cells) and variables (genes, and proteins) to facilitate direct comparison across datasets and data integration (Table 1.7). We then performed quality control metrics on all the datasets to assess the data quality and quantified properties specific to perturbation datasets, e.g. number of cells per perturbation condition. A user-friendly web interface will be created for data access and analysis visualization. Users will be able to filter and download datasets of their interests, as well as interactively select datasets for visualizing any analysis results from this study.

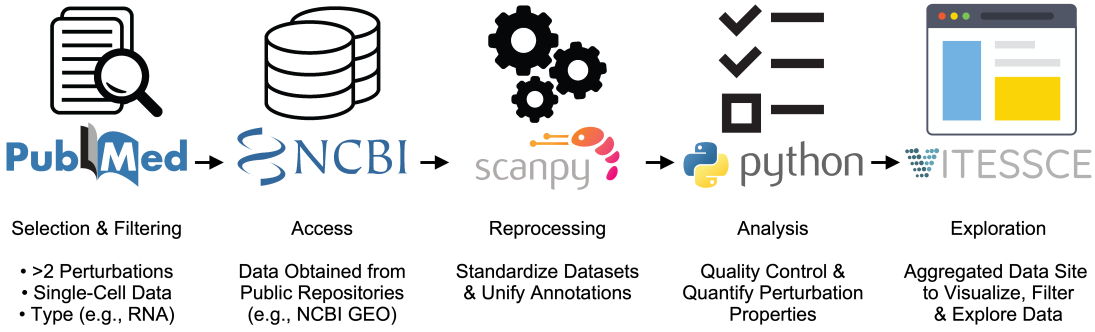


Figure 1.7: Workflow of collecting perturbation biology resource. Perturbation-response datasets were collected from literature with specified filtering criteria: i) at least three perturbation conditions of one or more types, including small-molecule drugs, CRISPR genome editing, and cytokine stimulation; ii) one or more of the molecular phenotypes, including transcriptomics, proteomics, and epigenomics. Datasets were obtained from public repositories, mostly from NCBI Gene Expression Omnibus (GEO) and a few from private repositories. Datasets were then processed into standardized h5 format using scanpy package [181] in python. Annotations for observations and variables were unified using customized python scripts. Additional quality control metrics were applied to assess data sparsity, quality, and properties specific to perturbations, e.g. the number of perturbations conditions, the number of cells captured per condition, and etc. A public user interface will be created for free data download and visualization of analysis results.

There are as many underlying mechanisms of resistance as there are patients with cancer.

N.Vasan, J.Baselga, & D.M.Hyman [25]

2

Proteomic profiling of drug response to identify resistance mechanism

This is an ongoing project with the following contributing authors: Alexandra Franz, Fabian Coscia, Lea Charaoui, Kenneth Munroe, Augustin Luna, Nicholas P. Gauthier, Matthias Mann, Chris Sander.

Author contributions

Conceptualization, N.P.G., and C. Sander

Investigation, A.F., C. Shen, L.C., F.C., and K.M.

Formal analysis, C. Shen, A.F., F.C., and A.L.

Visualization, C. Shen, and A.F.

Writing - Original Draft, A.F., C. Shen, and C. Sander

Resources, M.M, and C. Sander

Supervision, M.M., and C. Sander.

Part of the work presented in this chapter has been published in "Molecular response to PARP1 inhibition in ovarian cancer cells as determined by mass spectrometry based proteomics" by A. Franz et al., *Journal of ovarian research*, **14**, 1-14 (2021)[182].

A.F. designed the study, collected the samples, analyzed and interpreted the proteomics data, performed combination dose-response experiments, analyzed live-cell imaging data and wrote the manuscript. F.C. provided advice for the design of the study, acquired and analyzed the mass spectrometry data and critically revised the paper. L.C. carried out the dose-response experiments, analyzed live-cell imaging data, prepared the samples for MS measurements and edited the manuscript. C. Shen participated in the design of the study, analyzed the WES data, performed combination dose-response experiments, analyzed live-cell imaging data and edited the paper. M.M. and C. Sander designed and supervised the study and critically revised the paper. All authors have approved the final version.

2.1 Background and rationale

Investigation of resistance mechanisms can be used for rational design of combination therapy. In addition to the targeted molecular experiments that are usually used to examine resistance mechanisms to individual drug treatment, deep characterization of cellular response profiles upon drug

treatments has been applied to study the adaptive response. Such efforts have been pioneered by the large-scale collections of gene expression response profiles to drug perturbations across many cancer cell lines known as the Connectivity Map[66], followed by the NIH LINCS (the Library of Integrated Network-Based Cellular Signatures) program that collects changes in other cellular processes than gene expression[67,68]. Although these datasets provide comprehensive molecular changes that are informative of drug response, proteins, rather than mRNAs (gene expression), are the direct targets of most small molecule drugs and the direct regulators of the phenotypic response of the cells, e.g. cell proliferation. Protein expression profiles can therefore provide more direct information reflecting cellular changes upon drug treatment. Efforts have been made to collect proteomic response profiles to drug perturbations[69,183,184,185] but have been limited in studying ovarian cancer. We aim to use unbiased deep mass spectrometry (MS)-based proteomics to comprehensively assess the proteomic changes induced by drug perturbations in order to identify markers of resistance and provide information for drug combination strategies in ovarian cancer.

The resistance mechanisms we aim to identify here for the design of combination therapies refer to the short-term adaptive response of cells that can be reflected on proteomic changes. The adaptive response of cells likely counteracts the cytotoxic effects of drugs to restore proliferation. Therefore, the resistance mechanisms can be identified by searching for pro-proliferative proteins or biological processes whose activities are increased upon drug treatment. By simultaneously targeting such adaptive response, it is more likely that we completely block the proliferation of cancer cells.

2.2 Experimental design of mass-spectrometry profiling

For our study, we selected six anti-cancer drugs that are currently under development for the treatment of ovarian cancer patients. The selected drugs, MK-2206, Venetoclax, CHIR-99021, PD-0325901, Bisindolylmaleimide VIII (BIM VIII), and Bosutinib, each respectively target a frequently

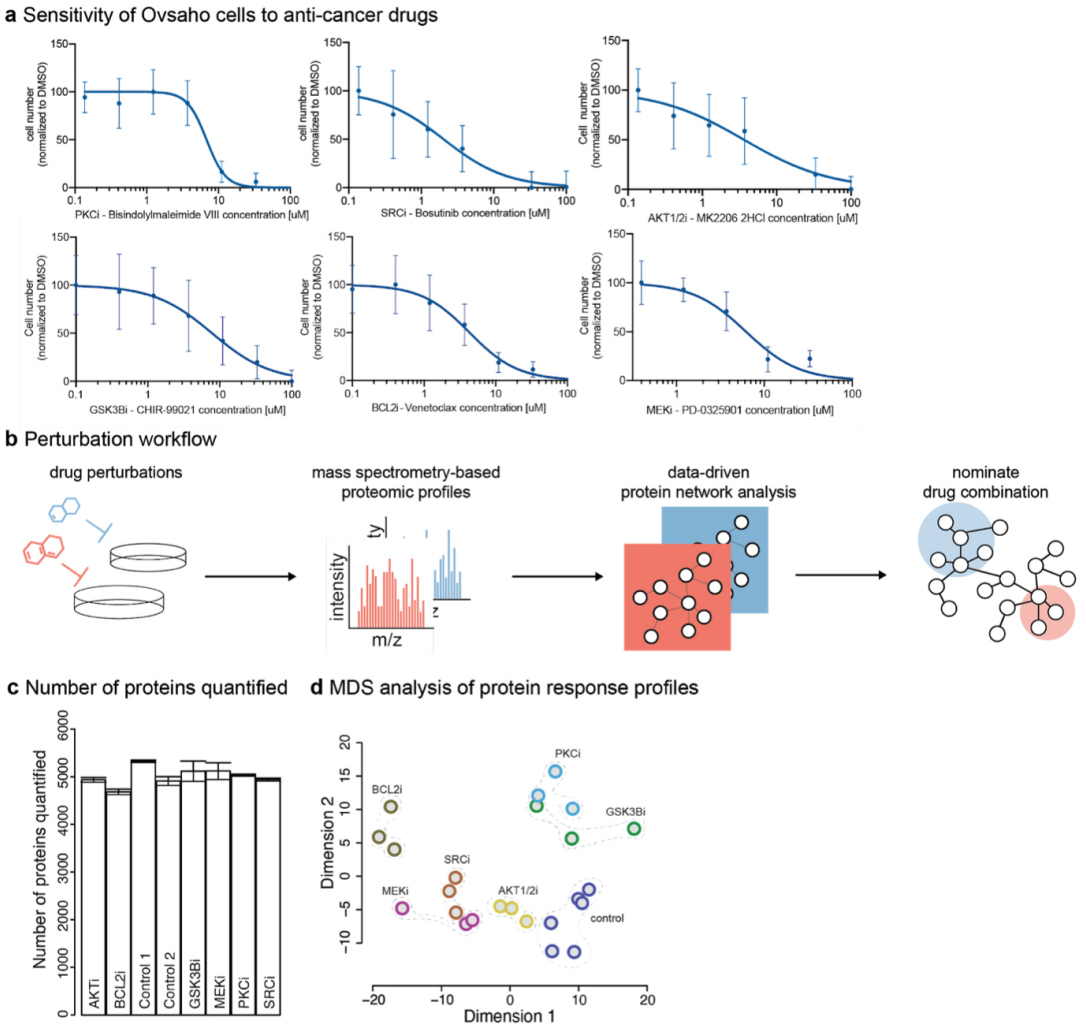


Figure 2.1: Systematic proteomic profiling of molecularly targeted drug response. (a) Dose-response curves of individual small-molecule inhibitors in OVSAHO cells, including AKT inhibitor MK-2206 (AKTi), BCL2 inhibitor Venetoclax (BCL2i), GSK3β inhibitor CHIR-99021 (GSK3βi), MEK inhibitor PD-0325901 (MEKi), PKC inhibitor Bisindolylmaleimide VIII (BIM VIII, PKCi), and SRC inhibitor Bosutinib (SRCi). Data is aggregated from three biological replicates each with three technical replicates and error bars represent standard deviations of the data. (b) Label-free mass spectrometry (MS)-based proteomics workflow used to measure protein changes in response to drug treatment at IC50. Data-driven network analysis was later applied to the acquired MS data to identify resistance mechanisms for the nomination of combination candidates. (c) Number of proteins quantified by MS upon each drug treatment are reported. Error bars are the standard deviation of three biological replicates for each condition. (d) Multi-dimensional scaling (MDS) on Euclidean distance of protein levels measured after treatment. Colors indicate different drug perturbation conditions with dotted lines wrapped around the biological replicates to indicate the proximity in distance of the replicates.

deregulated oncogenic signaling pathway or biological process, including Akt (protein kinase B) pathway, apoptosis regulated by Bcl-2 (B-cell lymphoma 2), GSK3 (glycogen synthase kinase 3) pathway, MAPK/ERK pathway (mitogen-activated protein kinase cascade), PKC (protein kinase C) signaling, and SRC (protein tyrosine kinase) signaling. We used OVSAHO cells as a preclinical HGSOC model system since they reflect well the genomic and proteomic features of HGSOC [186,187]. To explore the effects of the drugs on the proteome, we perturbed the cells with selected inhibitors at their IC50 concentration (half-maximal inhibitory concentration) and measured the proteomic changes using MS-based proteomics. The IC50 was determined based on the dose-response curves derived from the cell viability measured 72 hrs after the drug treatment. The treatment time of 72 hrs is sufficient to capture the adaptive proteomic response of cells upon drug treatment. Cell viability was measured by cell counting using live-cell imaging (Material and Methods). Treatment with individual inhibitors resulted in a dose-dependent reduction in cell viability with IC50 values ranging from 4 μM to 12 μM (Figure 2.1a). Cells were then treated at identified drug IC50 concentrations for 72 hrs in three biological replicates, and proteomic changes were profiled by MS (Figure 2.1b). To quantify protein expression levels across drug perturbation conditions, we applied data-independent acquisition (DIA) combined with label-free based quantification (Material and Methods). We quantified the proteome at a median depth of 5,000 protein groups per perturbation measurement (Figure 2.1c). Unsupervised multi-dimensional scaling (MDS) on Euclidean distance of biological replicates confirmed the reproducibility of proteomic responses upon different drug treatments (Figure 2.1d).

2.3 Comparison of proteomic response across drug perturbations

We first examined the measured global proteomic response to directly compare the drug-induced cellular changes across drug perturbations. We identified a total of 5912 proteins of which 4480

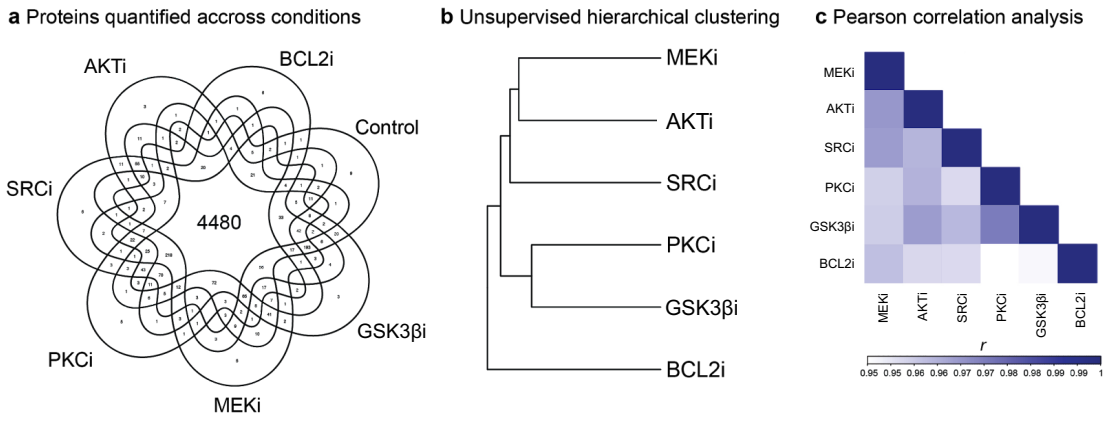


Figure 2.2: Comparison of drug-induced responses based on global proteomic changes. (a) Venn diagram illustrates the overlap of proteins captured and quantified across treatment conditions with a total of 5912 proteins identified, and 4480 proteins found in all perturbation conditions. (b) The similarity of protein responses to different drug perturbations was assessed by unsupervised hierarchical clustering on z-normalized mean protein expression values of replicate measurements using Euclidean distances. Drugs that are known to affect parallel signaling pathways, e.g. MEKi and AKTi, or kinase and its downstream effector, e.g. PKCi and GSK3βi, are clustered together, suggesting proteomic profiles can reflect known drug targets and pathway relationships. (c) The similarity of protein responses upon drug perturbations was assessed by Pearson correlation and pair-wise correlation coefficients are reported.

(75.7%) proteins were captured in all perturbation conditions (Figure 2.2a). We next applied unsupervised hierarchical clustering on z-normalized mean protein expression values of replicate measurements using Euclidean distances to investigate the similarities between drug effects (Material and Methods). Our analysis revealed three major clusters, clearly separating BCL2i from the remaining kinase-inhibitors (Figure 2.2b). The first cluster contains closely related MEKi and AKTi, as well as SRCi. It is known that ERK and AKT signaling pathways can work in a parallel manner to induce downstream signaling [188], and AKT activation is a resistance mechanism to the treatment of MEK inhibitors [57,189]. Src kinases regulate a broad spectrum of signaling pathways, including ERK and AKT signaling [190]. The second cluster groups together PKCi and GSK3βi, which can be partially explained by the function of PKC to phosphorylate GSK3β [191]. In agreement with these observations, we found that the Pearson correlation between proteome response profiles was overall high except for BCL2i, which has a lower correlation with the remaining inhibitors (Figure 2.2c).

We believe the overall high correlations between response profiles are driven by a large number of "housekeeping" proteins whose expression stays steady upon drug treatment. In general, these observations suggest that cell responses upon treatment of inhibitors targeting closely connected pathways are similar, indicating that the proteomic response profiles strongly reflect the functions of drug targets and our existing knowledge of biology.

We next assessed which proteins and cellular processes were strongly affected by the drug treatment. We define strongly responsive proteins as proteins with an expression level change upon drug treatment compared to the control treatment of at least 0.5 or -0.5 log2 units (log2ratio) with p-value < 0.05 at a false-discovery rate (FDR) cutoff at < 0.2 as determined by t-test statistics with Benjamini-Hochberg correction (Materials and Methods). We found that the vast majority of proteins (>80%) did not respond strongly across treatment conditions (Figure 2.3a). Drug treatments resulted in different proportions of strongly responsive proteins, which might reflect different scopes of downstream signaling for different drugs. For example, while AKTi induced expression level changes in a small number of proteins (8% of the measured proteins), inhibition of SRC affected 18%. SRC is known to have key roles in regulating multiple pathways [192] and therefore its inhibition is likely to affect a broader range of protein expression. By comparing strongly responsive proteins across treatment conditions, we found that the expression levels of a total of 2714 proteins changed significantly in at least one drug condition (Figure 2.3b), in which 127 proteins changed significantly in at least four perturbation conditions. Frequently responsive proteins generally had expression level changes in the same direction across all drug perturbations (Figure 2.3c), indicating a potential general resistance response. The remaining majority of strongly responsive proteins are relatively specific to each drug treatment.

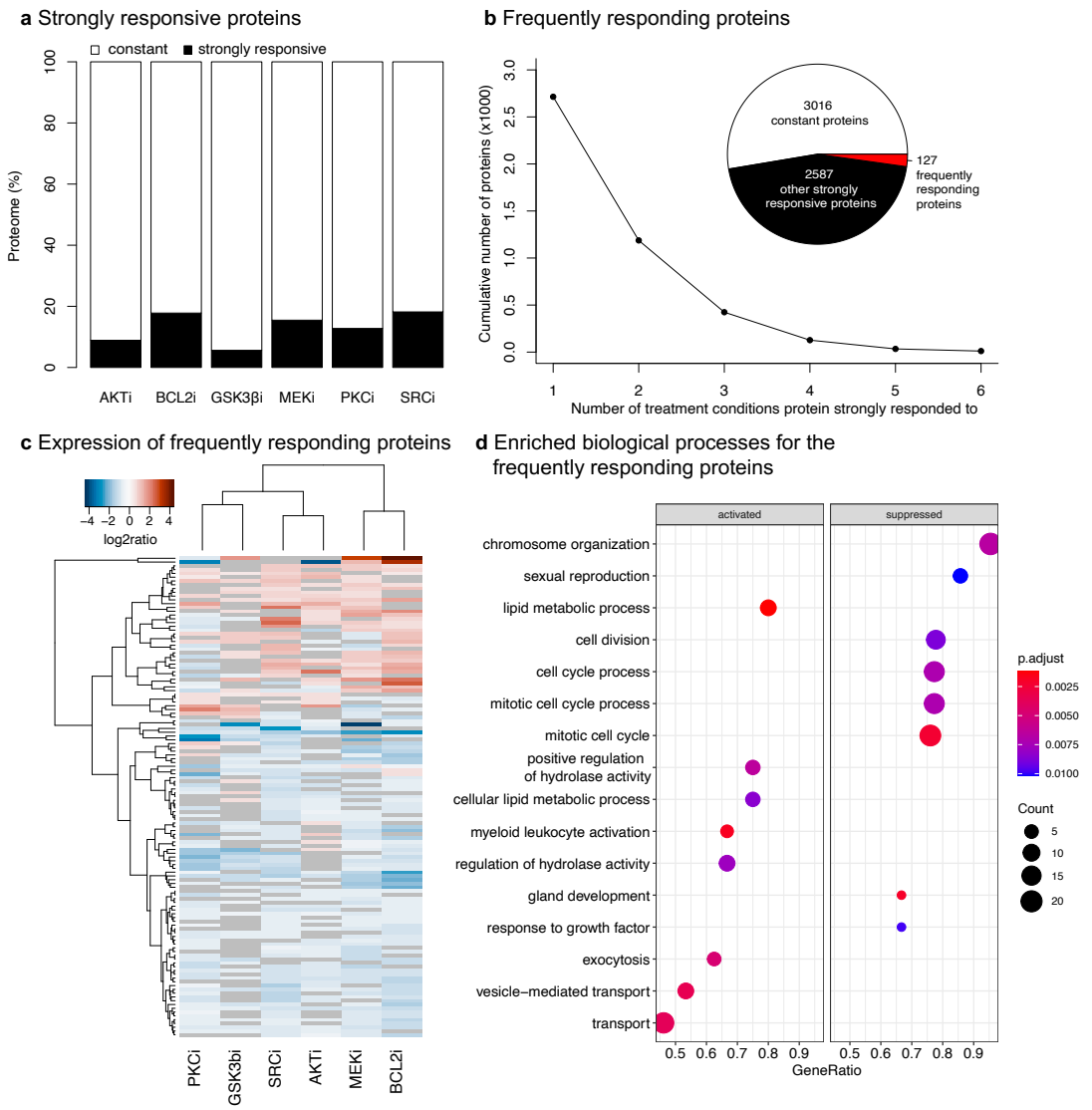


Figure 2.3: Identification of frequently responsive proteins to drug treatment. (a) Bar chart shows the proportion of proteomes per treatment condition that are changed compared to the control treatment. Strongly responsive proteins are defined as proteins whose absolute log2ratio is at least 0.5 (p-value < 0.05 with BH-based FDR < 0.2 in t-test). Constant proteins are defined as the remaining proteins. (b) Cumulative numbers of proteins that strongly respond to a certain number of perturbation conditions are reported. Proteins that strongly respond in 4 or more treatment conditions are defined as frequently responsive proteins. Pie chart shows the proportions of frequently responsive proteins, other strongly responsive proteins (respond to at most 3 drug treatments), and constant proteins. (c) Heatmap shows the protein expression change of identified frequently responsive proteins. Most proteins respond with consistent trends, either increased or decreased expression, across perturbation conditions, indicating a general stress response. (d) Gene set enrichment analysis (GSEA) identified both activated (increased protein expression) and suppressed (decreased protein expression) biological processes. Suppressed processes likely reflect the cytotoxic effects of drugs while activated processes might suggest a general resistance mechanism.

2.4 Identification of general resistance markers

Several well-known resistance mechanisms have been observed to develop in cancer cells against different pharmacological compounds that are structurally and functionally unrelated [193]. Therefore, we next assessed the functional properties of the 127 frequently responsive proteins. A gene set enrichment analysis (GSEA) [194] suggested that proteins involved in cell survival and proliferation were significantly decreased, including biological processes such as chromosome organization (GO:0051276, p-value = 6.63E-03), sexual reproduction (GO:0019953, p-value = 1.01E-02), and cell cycle process (GO:0022402, p-value = 7.18E-03) (Figure 2.3d). This observation could plausibly be associated with the cytotoxic effects of the drugs. In contrast, several biological processes are enriched for proteins with significantly increased expression, including lipid metabolic process (GO:0006629, p-value = 1.19E-03), vesicle-mediated transport (GO:0016192, p-value = 3.54E-03), and positive regulation of hydrolase activity (GO:0051345, p-value = 6.43E-03). Consistent with some previous research on the response of cancer cells under stressed environments [195,196], these up-regulated processes are likely general adaptive responses of ovarian cancer cells to pharmacological inhibition and might be of therapeutic interest.

2.5 Nomination of combination candidates based on drug-specific resistance markers

We next focused our analysis on strongly responsive proteins for individual drug treatment conditions to gain further insight into drug-specific cellular changes. We have applied three independent and complementary computational approaches to discover responsive proteins and pathways, which are module detection using Netbox [197,198], GSEA, and individual protein expression analysis. Netbox algorithm combines prior knowledge of protein interaction networks with a clustering algorithm to identify functional protein modules across the boundaries of curated and pre-defined lists

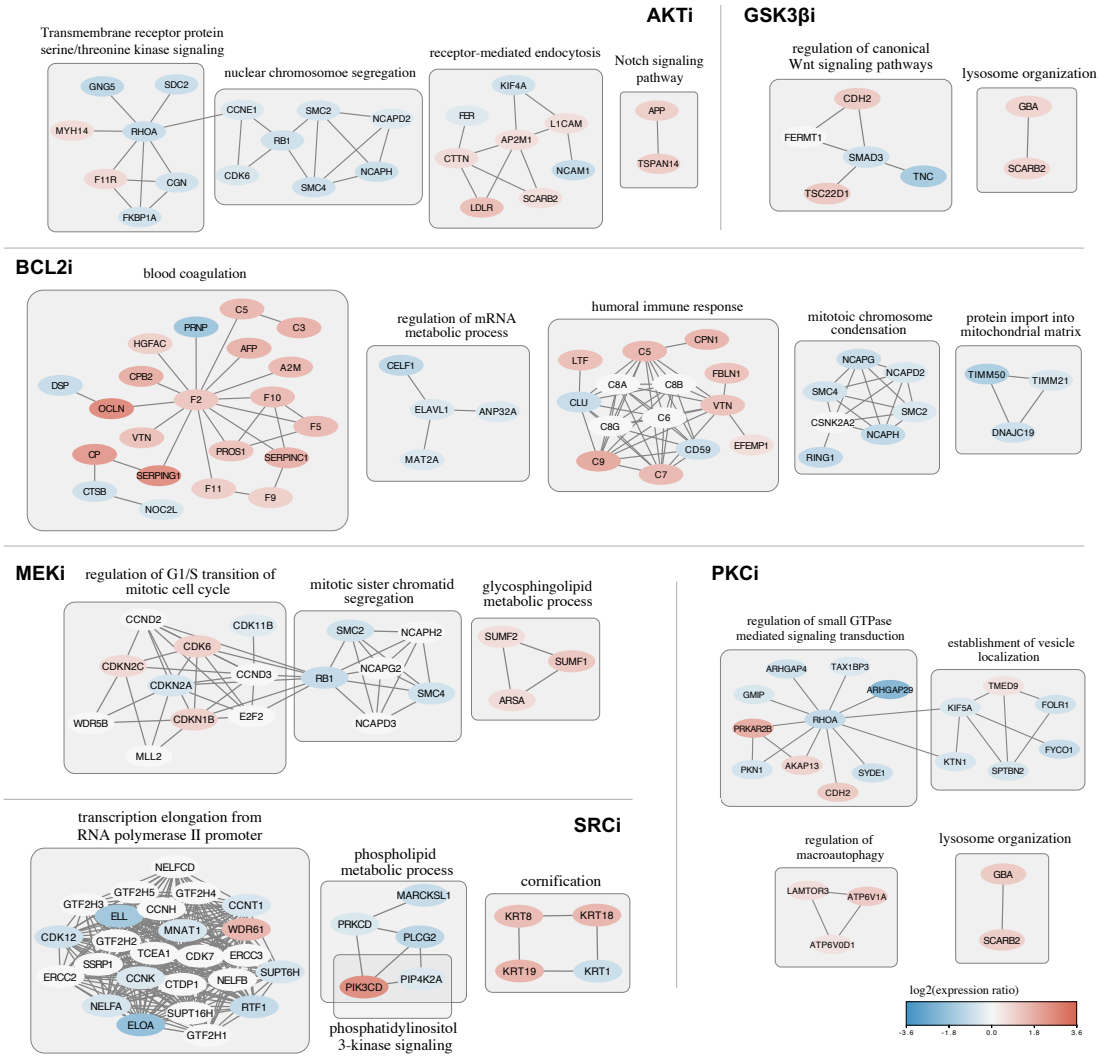


Figure 2.4: Drug-specific protein response networks identified using Netbox. Drug-specific protein responses were grouped into functional modules using the Netbox algorithm and the modules were characterized by functional labels based on enrichment analysis (Material and Methods). Subsets of protein modules for each drug perturbation were chosen based on their relevance to the known functions of the inhibitors and their implications for potential resistance mechanisms. Nodes are proteins colored by protein expression change (log2ratio). Edges are undirected protein interactions from the background network used by the Netbox algorithm. Relevant modules detected using either Reactome FI Network or INDRA network as the background network are reported.

of proteins. We first used the Netbox algorithm to identify functional protein modules and then applied pathway enrichment to label the identified modules (Material and Methods). We then applied GSEA to independently identify enriched biological processes upon each drug inhibition. In addition, we examined individual responsive proteins that are of interest in known cancer signalings. For each of the small-molecule drugs, the identified protein modules with decreased activities, indicated by the decrease in protein expressions, generally agree with the known functions of the drugs (Figure 2.4). We have also identified several protein modules and individual proteins with increased expression, which we will discuss in detail later in the section. These modules and proteins are of particular interest as they are likely indicative of potential resistance mechanisms to the small-molecule inhibitors.

Tested drug (A)	Combination partner (B)	Target of B	Mechanism of B	Clinical stage of B	Analysis evidence
AKTi (MK-2206)	**AL101 (BMS-906024)**	NOTCH signaling	γ-secretase inhibitor	FDA fast track designation for adenoid cystic carcinoma	Netbox analysis
BCL2i (Venetoclax)	**Infigratinib**	FGFR signaling	FGFR1-3 inhibitor	FDA-approved for cholangiocarcinoma	individual protein
BCL2i (Venetoclax)	**Everolimus**	MTOR signaling	mTORC1 inhibitor	FDA-approved for breast cancer, renal cell carcinoma, etc.	Individual protein
GSK3βi (CHIR-99021)	**Ivosidenib**	IDH1	Isocitrate dehydrogenase-1 inhibitor	FDA-approved for acute myeloid leukemia and cholangiocarcinoma	Individual protein
MEKi (PD-0325901)	**Erlotinib**	EGFR signaling	EGFR tyrosine kinase inhibitor	FDA-approved for non-small cell lung cancer (NSCLC) and pancreatic cancer	Individual protein
MEKi (PD-0325901)	**Palbociclib**	CDK4/6	CDK4/6 inhibitor	FDA-approved for breast cancer	Individual protein
PKCi (BIM VIII)	**Ivosidenib**	IDH1	Isocitrate dehydrogenase-1 inhibitor	FDA-approved for acute myeloid leukemia and cholangiocarcinoma	Individual protein
PKCi (BIM VIII)	**TVB-2640**	FASN	Fatty acid synthetase inhibitor	FDA fast track designation for nonalcoholic steatohepatitis (NASH)	GSEA analysis
SRCi (Bosutinib)	**Idelalisib**	PI3K signaling	PI3K delta isoform inhibitor	FDA-approved for chronic lymphocytic leukemia and certain types of lymphoma	Netbox analysis
SRCi (Bosutinib)	**Ivosidenib**	IDH1	Isocitrate dehydrogenase-1 inhibitor	FDA-approved for acute myeloid leukemia and cholangiocarcinoma	Individual protein
SRCi (Bosutinib)	**Auranofin**	Redox enzymes	Inhibition of redox enzymes	FDA-approved for rheumatoid arthritis, clinical trials in cancer	Individual protein & Netbox

Table 2.1: Proposed combination drug candidates based on potential resistance mechanisms. Potential resistance mechanisms were identified based on the analysis of proteomic profiling after drug treatment using three complementary approaches, i.e. Netbox analysis, GSEA analysis, and individual protein expression analysis. Small molecule inhibitors were selected to target the corresponding pathways indicative of resistance as combination partners of the originally profiled drugs. Drugs are prioritized if they are more relevant in clinical settings, i.e. FDA-approved or in late-stage clinical trials.

We identify resistance mechanisms as the proteins or biological processes that are both pro-

proliferative and up-regulated in activities, indicated by an increase in protein expression. Whether a protein or a biological process is pro-proliferative or anti-proliferative is defined as the individual functional score of the gene or the average functional score of the genes in the process (Materials and Methods). To nominate the combination therapies, we therefore simultaneously target the resistance mechanisms identified by our three computational approaches while applying the original small-molecule inhibition, with the rationale to completely block the opportunity of cancer cells to restore proliferation. The candidate drugs are chosen as specific inhibitors of the selected pathways or proteins (Table 2.1). We then ranked the candidates based on their clinical stages, prioritizing those that have been approved by the Food and Drug Administration (FDA). To further filter the candidates, we also reviewed previous literature on how widely these candidates have been used in pre-clinical studies and additional supporting information on their resistance mechanisms and potential partners in combination therapies.

AKTi

In agreement with AKT's central function in controlling key kinase signaling pathways, cell proliferation, and glucose metabolism[199,200], we found that the responsive protein network modules upon AKT inhibition are enriched for several biological pathways, including transmembrane receptor protein serine/threonine kinase signaling (GO:0007178, adj. p-value = 0.0024), muscle contraction (GO:0006936, adj. p-value = 2.10E-05), nuclear chromosome segregation (GO:0098813, adj. p-value = 0), and regulation of fibroblast proliferation (GO:0048145, adj. p-value = 0.044). The reduced activities in these biological processes, indicated by the decrease in protein expression, likely led to the decrease in cell viability.

We observed that upon AKT inhibition, a Netbox-identified protein module enriched for Notch signaling pathway (GO:0007219, adj. p-value = 2.37E-02) is upregulated with an average log2ratio of 0.99 over two genes, APP and TSPAN14 (Figure 2.4). Activation of Notch signaling is found in

different types of cancer and inhibition of Notch signaling has been shown as a potential therapeutic approach for cancers[201]. We believe the upregulation of proteins in Notch signaling pathways might be an adaptive mechanism for ovarian cancer cells to resist AKT inhibition. Therefore, we propose to use AL101 (BMS-906024), a γ-secretase inhibitor, to target Notch signalings, in combination with the AKT inhibitor (Table 2.1).

BCL2i

By analyzing the protein network modules derived upon BCL2 inhibition, we found that BCL2i negatively affected proteins involved in mitochondrial translation (GO:0032543, adj. p-value = 0), and protein import into the mitochondrial matrix (GO:0030150, adj. p-value = 0). This is in agreement with the role of BCL2 in mitochondrial apoptosis and metabolism[202].

We identified two pro-proliferative genes whose protein expression increased upon BCL2 inhibition, FGFR3 (log2ratio = 0.874, p-value = 2.00E-03) and MTOR (log2ratio = 3.49E-02), which are essential players of well-known cancer signalings, FGFR signaling, and mTOR signaling. With additional evidence from previous literature reporting enhanced anti-tumor effects of dual inhibition (Bcl-2 with FGFR, and Bcl-2 with mTOR)[203,204], we believe both signaling pathways may be resistance mechanisms to BCL2 inhibition and can be chosen as combination targets with BCL2i. We chose Infigratinib, a selective FGFR1-3 inhibitor to target FGFR signaling, and Everolimus, which inhibits the formation of the mTOR complex, in combination with BCL2i (Table 2.1).

In addition, our analysis from Netbox identified significant enrichment of proteins with increased expression in two biological processes, blood coagulation (GO:0007596, adj. p-value = 0), and humoral immune response (GO:0006959, adj p-value = 0) (Figure 2.4). Although the mechanisms linking these two processes and BCL2 are unclear, the strong enrichment and up-regulation of protein expressions support the hypothesis that they might be emerging resistance mechanisms to BCL2 inhibition.

GSK3βi

Inhibition of GSK3β led to the disrupted activities of several biological pathways, including Golgi vesicle transport (GO:0048193, adj. p-value = 0), and regulation of canonical Wnt signaling pathway (GO:0060828, adj. p-value = 0.0060), which are consistent with the known functions of GSK3β to regulate Golgi membrane trafficking[205] and affect Wnt/β-catenin signaling[206].

We found that the protein expression of isocitrate dehydrogenase 1 (IDH1) increased upon GSK3β inhibition (log2ratio = 0.767, p-value = 1.40E-03). It has been previously reported that pharmacological inhibition or knockdown of IDH1 decreased proliferation by inducing senescence in multiple HGSOC cell lines[207]. Therefore, we believe IDH1 can be a resistance marker for GSK3β inhibition. We select Ivosidenib, an FDA-approved inhibitor for IDH1, as the combination partner for GSK3βi (Table 2.1).

We also found that the Netbox-identified protein module enriched for lysosome organization (GO:0007040, adj. p-value = 1.18E-03) is activated (Figure 2.4). This response has been previously reported by Albrecht et. al[208] who observed that macropinocytosis, triggered by GSK3-inhibition, induced catabolic activity of lysosomes. We believe increased lysosome activity might be a resistance mechanism to GSK3β inhibition. The lysosome is linked to many hallmarks of cancer and a wide range of agents have shown to affect multiple aspects of lysosome activities in clinical trials[209]. We believe our analysis does not provide sufficient evidence to support our choice of drug(s) inhibiting a particular lysosomal target. Therefore, experimental validation of co-targeting GSK3β and lysosome activities, which potentially need to be done in a drug screen, might be beyond the scope of this study.

MEKi

We found that the responsive protein modules upon MEK inhibition are enriched for biological processes involved in cell cycle and proliferation, including mitotic sister chromatid segregation

(GO:0000070, adj. p-value = 0), mitotic spindle assembly (GO:00090307, adj. p-value = 2.17E-05), and ribosome biogenesis (GO:0006364, adj. p-value = 0). These observations are consistent with the major function of the MAPK/ERK signaling pathway to regulate cell cycle entry, transcription, and translation [210].

The protein expression of EGFR (log2ratio = 1.19, p-value = 1.01E-02) increased upon MEK inhibition. Previous literature demonstrated that combined inhibition of MEK and EGFR prevents the emergence of resistance in EGFR-mutant lung cancer. We therefore selected the most widely used EGFR inhibitor Erlotinib as the combination partner with our MEKi (Table 2.1). We also observed that the expression of cyclin-dependent kinase 6 (CDK6) increased (log2ratio = 0.931, p-value = 4.55E-03), which positively regulate cell cycle. Since CDK6 is a potential resistance marker of MEKi, as well as a therapeutic target for ovarian cancer patients [211], we selected palbociclib as the specific inhibitor of CDK4/6 to use in combination with MEKi (Table 2.1).

We found an increase in the expression of proteins enriched for the glycosphingolipid metabolic process (GO:0006687, adj. p-value = 3.51E-07) using Netbox (Figure 2.4). Interestingly, metabolic rewiring in response to MEK inhibition has been previously described as a possible mechanism of resistance in melanoma cells [212,213], and targeting of lipid metabolism has been previously reported in ovarian cancer cells [214]. Although targeting highly expressed metabolic enzymes in the glycosphingolipid metabolic process might provide a potential strategy to overcome MEKi resistance in ovarian cancer, this hypothesis has yet to be tested due to the lack of specific inhibitors of the pathway.

PKCi

Enriched biological processes upon PKC inhibition include extracellular matrix organization (GO:0030198, adj. p-value = 0), establishment of vesicle localization (GO:0051650, adj. p-value = 3.17E-05), regulation of small GTPase mediated signal transduction (GO:0051056, adj. p-value = 1.54E-06), and etc. Altered activities in these processes agree with known regulatory functions of

PKC, including but not limited to membrane structure, vesicle permeability, and phosphorylation of downstream proteins[215].

The results of GSEA suggested that proteins involved in lipid metabolic process (GO:0006629, p-value = 1.13E-03) are enriched upon PKC inhibition. Lipid metabolic processes are also identified as the general stress response and resistance mechanism for small-molecule inhibition. Therefore, we selected TVB-2640, a fatty acid synthase inhibitor to target this resistance mechanism. Similar to our observations with GSK3β inhibition, IDH1 (log2ratio = 0.694, p-value = 1.81E-03), as well as proteins involved in the process of lysosome organization (GO:0007040, adj. p-value = 1.24E-05) (Figure 2.4) have increased expression upon PKCi. We propose Ivosidenib with PKCi as a potential combination candidate (Table 2.1).

SRCi

Protein modules identified upon SRC inhibition are enriched mainly for biological processes that are essential for cell survival and proliferation, including covalent chromatin modification (GO:0016569, adj. p-value = 0), transcription elongation from RNA polymerase II promoter (GO:0006368, adj. p-value = 0), and etc. We found that one responsive protein module identified by Netbox is enriched for proteins involved in the phospholipid metabolic process (GO:0006644, adj. p-value = 4.86E-04) (Figure 2.4). An interesting observation is that two proteins, PI3K delta isoform (PIK3CD) and phosphatidylinositol-5-phosphate 4-kinase type-2 alpha (PIP4K2A) in this module are also involved in PI3K signaling pathway, which is critical for cell survival and frequently altered in ovarian cancer. Although SRC inhibition has been previously reported to suppress the PI3K pathway[216], our observation suggested that the protein expression of the PI3K delta isoform significantly increased upon SRC inhibition (log2ratio = 2.48, p-value = 1.77E-03). Therefore, we propose to use Idelalisib, a specific PI3K delta isoform inhibitor, in combination with SRCi (Table 2.1). We also observed an increase in protein expression of IDH1 upon SRC inhibition (log2ratio =

0.644, p-value = 2.76E-03) and therefore propose Ivosidenib and SRCi as a combination candidate.

In addition, we observed that the protein expressions of peroxiredoxins II (log2ratio = 0.754, p-value = 2.87E-03) and V (log2ratio = 0.821, p-value = 8.95E-04) (PRDX) are significantly increased. PRDXs are antioxidant enzymes that play key roles in regulating peroxide levels within cells controlling various physiological functions[217]. Several studies have implicated an increase of reactive oxygen species (ROS) in carcinogenesis due to a loss of proper redox control[218]. Most notably, in ovarian cancer cells, PRDX expression was found to be associated with platinum drug resistance[219] since increased levels of antioxidants inhibit apoptosis[220]. Our Netbox analysis with SRCi-specific responsive proteins (strongly responsive in at most two other perturbation conditions) identified a module labeled detoxification of ROS (data not shown). Our observations together with the reported results of others suggest that PRDXs are promising resistance markers to be targeted together with SRCi. Therefore, to target peroxiredoxins and the antioxidation process, we select auranofin, which inhibits the redox enzymes (Table 2.1). Although not specific, auranofin is the closest to clinical use in cancer compared to all other inhibitors of antioxidation[221].

2.6 Experimental validation of the combination candidates

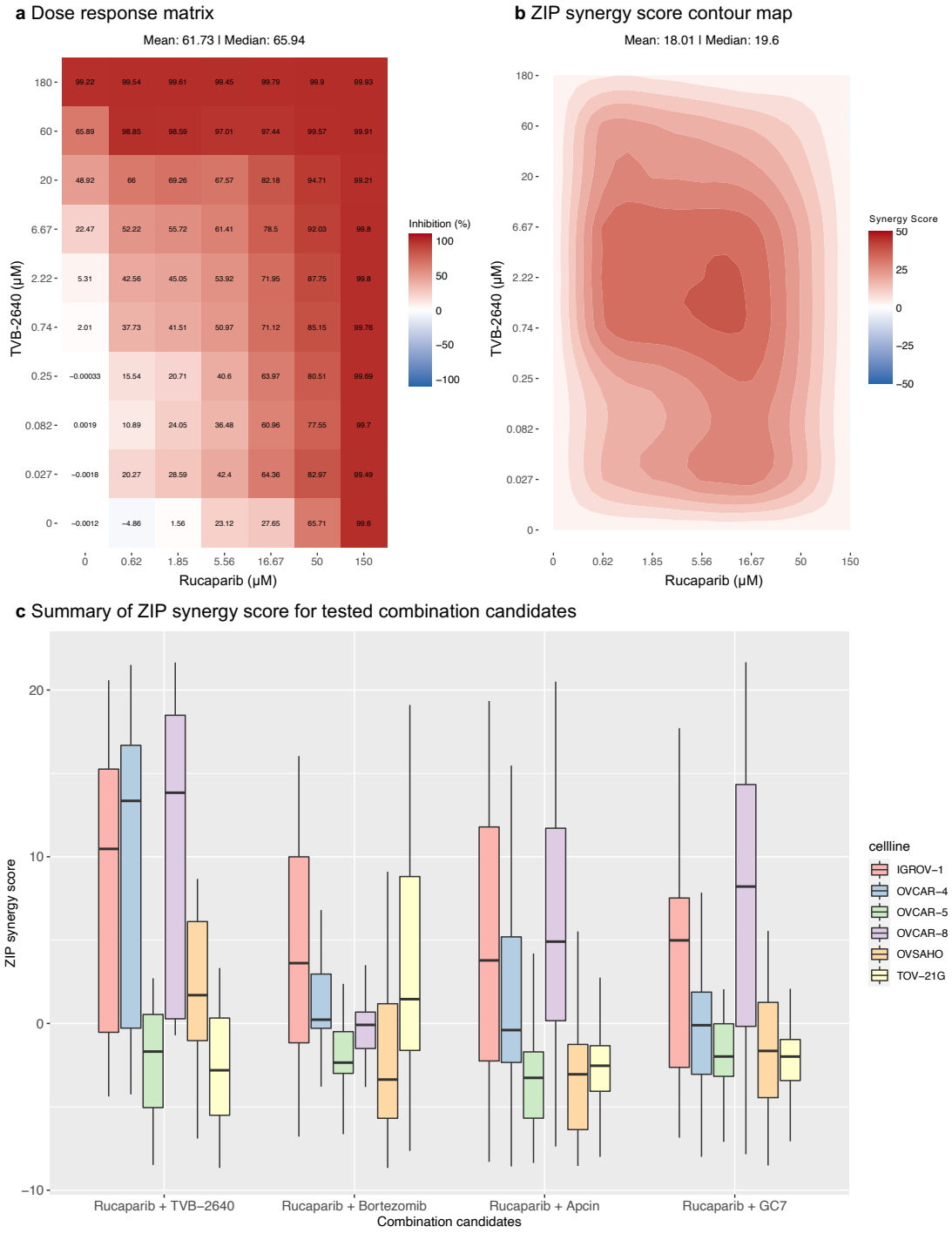
To validate the proposed combination candidates, we designed experiments to test their combinatorial effects. To demonstrate the generalizability of these combinations in HGSOC, in addition to the OVSAHO cell line, we chose another 5 HGSOC cell lines with different genetic backgrounds. In addition to the proposed combination candidates in this study, we validated another 4 drug combinations proposed in our previous work using a similar approach [182], which are PARP inhibitor in combination with fatty acid synthase inhibitor TVB-2640, proteasome inhibitor Bortezomib, APC inhibitor Apcin, and eIF5A-2 inhibitor GC7. For the PARP inhibitor, instead of using the small molecule AG-14361 used in the original discovery, we used an FDA-approved PARP inhibitor, Rucaparib, which is structurally very similar to AG-14361, to further facilitate the process of pushing the combinations into pre-clinical and clinical studies, provided the validation results are promising.

We performed checkerboard assays for each combination candidate in each cell line (Material and Methods), resulting in individual dose response matrices (Figure 2.5a). ZIP synergy scores [77] were calculated for each dose combination using SynergyFinder R package [222] (Figure 2.5b). We summarized the results for the combination candidates that have already been tested, which are the four combinations involving PARPi Rucaparib (Figure 2.5c). The combination candidates proposed in this study are in the process of being tested.

We observed that Rucaparib with TVB-2640 shows synergistic effects (ZIP scores > 10) in half of the HGSOC cell lines (3 out of 6). More importantly, the combination has strong synergistic effect in OVCAR-4 cells, which are relatively resistant to various PARP inhibitors based on data from GDSC [72] and are genomically very similar to HGSOC patient samples [186]. These results suggest that fatty acid synthase (FASN) inhibitor TVB-2640 can be a promising combination partner with PARP inhibitors to treat HGSOC and potentially sensitize the PARPi-resistance tumors. The other three combinations have mostly additive effects (ZIP score < 10 and > -10) in different cell

Figure 2.5 *(following page)*: Experimental validation of the proposed combination candidates. (a) The combinatorial effects of the proposed candidate pairs were experimentally tested in 6 HGSOC cell lines, including IGROV-1, OVCAR-4, OVCAR-5, OVCAR-8, OVSAHO, and TOV-21G. For each combination candidate and each cell line, a checkerboard assay was performed (Material and Methods) and the resulting dose response matrix was used for synergy calculation. An example dose response matrix for the combination Rucaparib and TVB-2640 in OVCAR-8 is shown. The inhibition strength is relative cell viability to DMSO control. (b) ZIP synergy scores were calculated using the R package SynergyFinder[222] for each combination concentration and plotted as contour map. Rucaparib and TVB-2640 exhibits synergistic effect (ZIP score > 10) in OVCAR-8 cell line. (c) ZIP synergy scores for all tested combination candidates in all cell lines are reported as boxplots. The combination of Rucaparib and TVB-2640 shows synergistic effects in three cell lines, while other combinations exhibit mostly additive effects (ZIP score < 10 and > -10). No combinations exhibit any antagonistic effects (ZIP score < -10).

Figure 2.5: (continued)



lines. However, they are still of therapeutic interest since combining them with lower concentrations to achieve similar anti-cancer effect can potentially reduce toxicity. We also observed that most combination candidates have mild or strong synergistic effects on the two cell lines IGROV-1 and OVCAR-8. IGROV-1 has a truncating mutation, which is a putative driver, in BRCA1 gene and OVCAR-8 has a missense mutation with unknown significance in ATR gene. We suspect that these two cell lines have some degrees of alteration in homologous recombination (HR) pathways, and are therefore more sensitive when PARP inhibitor is administrated with another anti-cancer agent.

2.7 Significance and future directions

We used mass-spec proteomics profiling to study resistance mechanisms for the design of combinatorial targeted therapies in ovarian cancer. Using data-driven network analysis, as well as enrichment analysis and examination of individual protein expression on the proteomic profiles after treatment of small-molecule drugs that are currently under development for ovarian cancer, we identified biological processes and proteins indicative of resistance, which are those whose functions are pro-proliferative and whose expression increased upon drug treatment. We proposed combination candidates targeting these resistance mechanisms together with the originally profiled drugs and performed experimental validation.

Our experimental results suggest that PARPi (Rucaparib) in combination with FASNi (TVB-2640) can be a promising therapy for ovarian cancer. We observed strong synergistic effects of the combination in 3 out of the 6 HGSOC cell lines tested while additive effects were observed for the remaining three cell lines. As previously discussed, there is evidence that lipid metabolic reprogramming can mediate resistance to anti-cancer agents including chemotherapy and targeted therapies in breast, lung and prostate cancer cell lines and xenograft models[223]. Indeed, targeting altered metabolic lipid pathways with pharmacologic interventions has become an emerging anti-

tumor strategy in several cancer types with currently 10 inhibitors being tested in clinical trials including TVB-2640[224]. While in ovarian cancer overexpressed lipogenic enzymes are increasingly seen as potential therapeutic opportunity to inhibit cancer growth and/or overcome resistance to chemotherapy or anti-angiogenic therapy[223,225,214], our study demonstrates that the combined inhibition of PARP1 and FSAN represents an synergistic anti-proliferative combination strategy. In this study we also observed that increased activity in lipid metabolism can potentially be a general resistance mechanism to pharmalogical inhibitors. Therefore, additional studies are needed to determine whether such synergistic effects arise from specific interactions between the HR pathways and lipid metabolism to gain further mechanistic insight. Other combinations with PARPi exhibited additive effects in most HGSOC cell lines, but they might still be useful clinically. Such combinations can reduce the likelihood of the emergence of resistance, minimize drug toxicity by using lower doses of individual drugs, or benefit patients on a population level by targeting genetically heterogeneous and differentially sensitive tumors.

We observed that all four combinations involving PARPi have mild to strong synergistic effects on IGROV-1 cell line, which has a loss-of-function truncating mutation in BRCA1 gene. Such trend also exists but is milder for OVCAR-8 cell line, which has a missense mutation with unknown function in ATR gene, a gene that encodes for a serine/threonine-specific protein kinase that plays an important role in DNA damage detection and repair. IGROV-1 and OVCAR-8 are slightly more sensitive to PARP inhibitors compared to the other lines as reported in GDSC database[72], although such differences were not pronounced in our experiments. We suspect that the alterations in HR pathways (BRCAness) might contribute to the sensitivity of the cells and therefore amplify the combinatorial effects. On the other hand, OVCAR-5 is relatively resistant to PARP inhibitors and we have not observed synergistic effects of any combinations in OVCAR-5. It is possible that BRCAness plays a role in combination targeted therapies with PARPi as in synergistic effects are more likely to be observed in cells with a defect HR pathway. However, no definitive conclusion can be

drawn from our results due to limited sample size. More combination candidates and more HG-SOC cell lines with different BRCAness are needed to test such hypothesis.

All the proposed combination candidates in this study are still under experimental validation. We expect that a few combination candidates might show synergistic effects and are worth pursing further in more clinically relevant model organisms, such as patient-derived organoids (PDO), or xenografts (PDX). Beyond desired effects of combination candidates on phenotypic response of ovarian cancer cells, molecular and biochemical experiments are indispensable to decipher the synergistic or anti-tumor effects on a molecular level. Understanding the molecular mechanisms of synergistic combinations further expands our knowledge on drug interactions and facilitates the design of combination therapies.

2.8 Material and methods

Drug perturbation experiments for proteomic profiling

OVSAHO cells were obtained from JCRB (Japanese Collection of Research Bioresources Cell Bank) cell bank (NIBIOHN, Cell No. JCRB 1046) and grown in MCDB105/199 medium supplemented with 10% FBS (Fisher, #10438026) and 1% Penicillin-Streptomycin. All cells were free of Mycoplasma and their identity was verified by whole-exome sequencing at the sequencing platform of the Broad Institute of MIT and Harvard. The following drugs used in the experiments were all commercially available: AKT inhibitor MK-2206 2HCl (Selleckchem, #S1078), BCL2 inhibitor Venetoclax (ABT-199, Selleckchem, #S8048), GSK3β inhibitor CHIR-99021 (Selleckchem, #S1263), MEK inhibitor PD-0325901 (Selleckchem, #S1036), PKC inhibitor Bisindolylmaleimide VIII (Caymanchem, #13333), and SRC inhibitor Bosutinib (SKI-606, Selleckchem, #S1014). To determine the IC50 for each of the drugs, cells were treated with DMSO control or varying concentrations (from 0.01 μM to 100 μM) of each drug 24 hours after seeding. The Incucyte NucLight

Rapid Red (NRR) Reagent (Essen Bioscience, #4717, 1:4000) was added to the medium to label nuclei of live cells. Cell viability was determined as the number of live cells 72 hr after drug treatment, counted using the live imaging of Incucyte and normalized to those of the DMSO control. Cell viability data from three biological replicates each consisting of three technical replicates were merged and fitted using a non-linear log(inhibitor) versus normalized response dose-response curves with a variable slope using GraphPad Prism. IC50 was determined as the drug concentration that gives 50% of the maximal inhibitory response on cell viability for each drug. To harvest cells for mass spectrometry proteomic profiling, cells were treated with DMSO control or the IC50 concentration of each drug 24 hrs after seeding in 10 cm dishes. Cells were harvested 72 hrs after drug treatment by gently scraping off from the plates into PBS and transferred into deep 2 mL 96 well plates. Cells were stored at -80°C and sent for mass spectrometry measurements.

Liquid chromatography-mass spectrometry (LC-MS) measurements and MS data analysis

Harvested cells were digested and the proteomic measurements were taken as previously described [182], section 'Liquid chromatograph-mass spectrometry (LC-MS) measurements' of Methods. Tryptic digestion of proteins from samples were followed by a nanoflow LC/MS-MS analysis of peptides using the quadrupole Orbitrap mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific, Bremen, Germany) [226] coupled to an EASYnLC 1200 ultra-high-pressure system (Thermo Fisher Scientific) through a nano-electrospray ion source. Briefly, about 300 ng peptides were loaded on a 50-cm HPLC-column with a 75-µm inner diameter (New Objective, USA; in-house packed with ReproSil-Pur C18-AQ 1.9-µm silica beads; Dr. Maisch GmbH, Germany). Peptides were separated using a linear gradient from 5 to 60% solvent B (80% ACN; 0.1% formic acid) in solvent A (0.1% formic acid). Total gradient length was 100 min and the column temperature was maintained at 60 °C with an in-house made column oven. The. The MS was operated in a data-independent acquisition (DIA) mode. The DIA method consisted of one MS1 scan (m/z

350 to 1650, resolution 60,000, maximum ion injection time 60 ms, AGC target 3E6) and 32 DIA segments with varying isolation widths ranging from 39.7 Th to 562.8 Th inversely proportional to the m/z distribution of tryptic peptides and with an overlap of 1 Th. MS2 scan resolution was set to 30,000, maximum injection time to 54 ms and the automated gain control (AGC) target to 3E6. A stepped collision energy of 25, 27.5 and 30 was used and the default MS2 charge state was 2. Data was acquired in profile mode.

The raw data obtained from mass spectrometry measurements was processed as previously described [182], section 'MS data analysis' of Methods. DIA raw files were analyzed with Spectronaut Pulsar X software (Biognosys, version 12.0.20491.20) using the default settings for targeted data extraction. We used a project specific OVSAHO cell line proteome dataset as spectral library comprising 7720 protein groups (103,554 precursors). The false discovery rates at the precursor and protein levels were < 1% with the 'mutated' decoy model. Prior to data analysis, data was filtered by 'No Decoy' and 'Quantification Data Filtering'. Protein expression matrix with rows as identified proteins and columns as samples was obtained after data processing.

Proteomics data analysis

The protein expression matrix contained protein measurements from three biological replicates of each drug treatment and two sets of three biological replicates of DMSO control (due to two batches of samples). Data analysis was performed using customized scripts in R (Version 3.6). Counting the number of measured proteins, multi-dimensional scaling (MDS) on Euclidean distance of protein measurements, and pair-wise Pearson correlation of individual samples were performed on the raw protein expression matrix. To identify differentially expressed proteins of each perturbation condition from negative controls, an unpaired t-test was used to compare protein expressions from the drug-treated samples with samples treated with DMSO. log2(expression ratio) (log2ratio) and corrected p-values for an FDR < 0.2 by the Benjamini-Hochberg (BH) method

were obtained for each drug perturbation. Strongly responsive proteins were defined as those whose log2ratio > 0.5 or < -0.5 with corrected p-value < 0.05. The protein measurements from the biological replicates were then pooled to obtain average protein expression values for each perturbation condition. The averaged protein expression across replicates was then used for unsupervised hierarchical clustering, pair-wise correlation of perturbation conditions, and all the following analyses.

Protein module detection and enrichment analysis

Netbox[197,198] was used to perform the detection of responsive protein modules. For each of the drug-treatment conditions, the input gene list contains strongly responsive proteins upon the drug treatment (absolute value of log2ratio > 0.5) whose p-value of the t-test of differential expression is smaller than 0.01. To ensure the most compact and confident modules findings, the adjusted p-value of the connected linker node was set to 0.005 with all the other parameters of Netbox kept at default. Two protein-protein interaction networks were used independently as the background network of Netbox analysis: 1) Reactome Functional Interaction (FI) Network (version 2020 from Reactome.org)[81] with predicted interactions filtered out, 2) Interactions collected by Integrated Network and Dynamical Reasoning Assembler (INDRA) with at least a belief score of 0.95[227], which typically indicate the interaction is supported by multiple independent reading systems or pathway databases. Enrichment analysis was performed on the identified modules using the clusterProfiler Bioconductor package[228] using gene annotations from the Gene Ontology with default parameters. Gene Set Enrichment Analysis (GSEA)[194], was performed on the responsive proteins using clusterProfiler with default parameters. The input gene lists for each drug treatment are consistent with those for Netbox analysis, and the ranking of the genes is based on the log2ratio of the corresponding proteins in decreasing order.

Determination of functional score

The list of cancer census genes was obtained from the Catalogue of Somatic Mutations In Can-

cer (COSMIC, cancer.sanger.ac.uk)[229]. The functional score of a gene is defined as positive (1, pro-proliferative) if the role of the gene in cancer is an oncogene, and negative (-1, anti-proliferative) if the role of the gene is a tumor suppressor gene (TSG). If a gene has both roles of oncogene and TSG, the functional score is not defined.

Combination perturbation experiments for candidate validation

Experimental validation of the proposed combination candidates was performed by Charles River Laboratories. Six cell lines were used in the experiments, including IGROV-1 (NCI #0507369), OVCAR-4 (NCI #0502527), OVCAR-5 (NCI #0507336), OVCAR-8 (NCI #0507407), OVSAHO, and TOV-21G (ATCC CRL-11730 Lot#58690706). OVSAHO cells were grown in MCDB105/199 medium supplemented with 10% FBS and 1% Penicillin-Streptomycin. TOV-21G cells were grown in RPMI medium supplemented with 15% FBS and 1% Penicillin-Streptomycin. All other four cell lines were grown in RPMI-1640 medium supplemented with 10% FBS and 1% Penicillin-Streptomycin. All the small-molecule drugs are commercially available: Rucaparib (Selleckchem, #S1098), TVB-2640 (Selleckchem, #S9714), Bortezomib (Selleckchem, #S1013), Apcin (Selleckchem, #S9605), GC7 (Sigma-Aldrich, #259545), AL101/BMS-906024 (MedChemExpress, #HY-15670), Infigratinib (Selleckchem, #S2183), Everolimus (Selleckchem, #S1120), Ivosidenib (Selleckchem, #S8206), Erlotinib (Selleckchem, #S7786), Palbociclib (Selleckchem, #S1116), Bosutinib (Selleckchem, #S1014), Idelalisib (Selleckchem, #S2226), and Auranofin (Selleckchem, #S4307).

Cells were seeded into 96-well plates for checkerboard assays, in which a 3-fold serial dilution of one compound in the combination candidate was performed vertically down the plate for 6 dilution points, and a 3-fold serial dilution of the other compound was performed horizontally across the plate for 9 dilution points. One column of DMSO controls were included for each plate. A 3-fold serial dilution of 6 dilution points was also performed for each compound in single-drug treatment

experiments. Different concentration ranges were tested for different compounds for the best outcome: Rucaparib (6 3-fold dilution from 150 μM to 0.62 μM), TVB-2640 (9 3-fold dilution from 180 μM to 0.027 μM), Bortezomib (9 3-fold dilution from 210 nM to 0.032 nM), Apcin and GC7 (9 3-fold dilution from 360 μM to 0.055 μM). Cell viability was measured using CellTiter-Glo assay 72 hr after drug treatment. Synergy scores were calculated using the SynergyFinder R package[222].

3

Conclusion

Combination therapy in cancer has demonstrated promising results in providing enhanced anti-cancer efficacy and reducing the risk of drug resistance. Existing methods of nominating combination therapies are usually either low-throughput or do not provide sufficient mechanistic insights. In this work we presented two systematic approaches for nominating combination therapies, i) computational prediction of cellular response to unseen combination perturbations based on network models of cell biology, and ii) unbiased proteomic profiling of cellular response upon drug treat-

ment to identify resistance mechanisms.

For the first approach, we aimed to use comprehensive measurements of molecular and phenotypic responses upon systematic perturbation of cells to construct computational models of cell biology. We developed a hybrid approach, which we called CellBox, that combines explicit mathematical models of cell dynamics with a machine-learning framework to predict cell behaviors. Using a perturbation-response dataset of a melanoma cell line after drug treatments as a testbed, we demonstrated that the model can be efficiently trained to accurately predict cellular behavior accurately. Even though completely data driven and independent of prior knowledge, the resulting de novo network models recapitulate some known interactions from a pathway knowledgebase. We showed that the trained models are capable of simulating cell response to numerous arbitrary combinatorial perturbations applied to the molecular entities in the network model that were repeatedly measured under different perturbation conditions. Ranking of cellular response to the in silico combinatorial perturbations by the desired phenotypic outcome, such as decreased proliferation, can efficiently narrow down the search space and nominate promising sets of experimentally testable combination candidates.

From the perspective of cancer research, the development of the CellBox model provides a novel computational approach to facilitate the search for combination therapy in cancer. By constructing data-driven models on a relatively small-scale experimental dataset, we can efficiently narrow the search space of the predicted cellular response to experimentally untested perturbations. From the perspective of cell biology, the inferred network models can provide insight into the molecular interactions in the systems of interest. The inferred interactions not only can be confirmative of known interactions but can also be novel interactions to be experimentally validated. From the perspective of machine learning, CellBox brings in interpretability, which has been lacking in traditional black-box neural network models but is essential for making more transparent and therefore more trustable predictions in biomedical research. From the perspective of basic biological research,

we envision that CellBox models can be generalized to various types of molecular measurements, perturbations, and systems of interest and therefore applicable to study different fields of biology provided the suitable perturbation datasets are available.

An essential requirement of a CellBox model is a sufficiently informative perturbation-response dataset. The number of in silico combinatorial perturbations that can be predicted by the model is determined by the number of molecular entities measured and therefore included in the network model. Based on our sensitivity analysis of data reduction and our knowledge of parameter inference[230,231], the number of perturbation conditions should be about the same order of magnitude as the number of molecular entities measured in order to avoid underdetermined systems in model training. If we aim to construct a network of thousands of nodes, at least hundreds of perturbation conditions are needed for accurate model training, which can be laborious if performed in an arrayed fashion. We believe the recently developed sequencing and barcoding techniques which allow perturbation and molecular measurement to be performed on single-cell levels could be a potential solution to perform a large number of perturbations in a pooled fashion[78,163], hence one of our motivations is to collect single-cell perturbation-response datasets as a resource. Most of these datasets have gene expression as molecular measurements. This leads to an mRNA-mRNA interaction network model if CellBox is applied, which can be predictive as shown by some preliminary results (not shown) but is difficult to interpret and compare to existing knowledge[232]. Until the innovation of reliable single=cell proteomic profiling techniques, caution needs to be applied when developing CellBox on single-cell RNA-seq data: mRNA gene expression is not always a good proxy for the expression of protein, which are primary drug target.

Another layer of complexity in acquiring sufficiently informative perturbation-response datasets comes from the design of perturbation targets. The chosen perturbation conditions should ideally serve as probes to extensively investigate the entire system of interest. Our preliminary analysis (not shown) suggests that the information provided by single-agent perturbations can be covered by a

sufficient number of dual-agent perturbations. In our attempt to collect a large-scale perturbation-response dataset, we designed pilot experiments to include perturbations that target different signaling pathways with the hope to maximize the portion of the system being perturbed, which led to our second approach for nominating combination therapies.

For the second approach, we aimed to use inference of resistance mechanisms of cancer cells for the rational design of combination therapy. The resistance mechanisms refer to the short-term adaptive response of cells to pharmacological inhibitors to restore cell proliferation. We used unbiased quantitative protein mass spectrometry to assess the cellular response profiles to small-molecule drug perturbations in ovarian cancer cells. Data-driven protein network analysis, enrichment analysis, and examination of individual protein expression identified pro-proliferative proteins and biological processes that are up-regulated, revealing general stress response, as well as known and novel markers of resistance specific to drug perturbations. We then designed combination therapies by simultaneously targeting these resistance mechanisms and the small-molecule drugs originally profiled with the expectation to completely block the proliferation of cancer cells. In the validation experiments, where a subset of combination candidates was tested in six different ovarian cancer cell lines, synergistic and additive effects were observed, suggesting potential therapeutic values in future pre-clinical and clinical studies. The remaining proposed combination candidates are under experimental validation.

From the perspective of cancer research, we nominated promising combination candidates in ovarian cancer, a cancer type with a high relapse rate and slow improvement in mortality. We have shown sufficient evidence for potential efficacy of the combination candidate(s) in vitro so that they can be further validated in model systems that are more clinically relevant to patients, such as patient-derived organoids (PDO) and xenografts (PDX). We selected these combinations to contain at least one FDA-approved drug in order to lower the barrier and minimize the efforts for clinical studies. From the perspective of cell biology, we have provided comprehensive proteomic profiling

of cell response upon treatment of inhibitors targeting well-known cancer signaling pathways. We applied relatively straightforward pathway analysis approaches to identify potential resistance mechanisms. In future, this type of dataset can further serve as a rich resource for the development and application of more sophisticated computational methods, as well as a hypothesis-generating source for molecular and biochemical experiments to decipher the detailed molecular mechanisms of drug resistance.

A limitation of this approach is the lack of phosphoprotein profiles as molecular response to drug perturbations. Phosphorylation states of proteins are critical for the functions of proteins especially in cancer signaling pathways[233]. Many small-molecule kinase inhibitors act through blocking the phosphorylation of targeted proteins or their downstream effectors to achieve cytotoxic effects. Therefore, capturing phosphoproteins provides important information for determining the activation status of many proteins. Due to limited resources in this study, we were not able to profile the phosphoproteomic response upon drug treatment, which we believe resulted in less powerful detection of resistance mechanisms. Increased activity of a biological process could be achieved by shifting the equilibrium between phosphorylated (usually activated) and unphosphorylated (usually inactivated) proteins rather than increasing the expression of total proteins. Detection of phosphoproteins should be used to complement the findings based on whoe-protein level data. This leads to another critical question of experimental design, the choice of timing for the detection of molecular responses. Changes in phosphorylation events usually occur first in response to drug treatment, as early as 5 min after the treatment[100,234]. The choice of 72 hrs in this study for steady-state adaptive response is largely based on empirical evidence from the literature[69] and our observation that phenotypic responses (cell viability) are usually stable by the end of three days after drug treatment. Cellular response reflected by proteomics and phosphoproteomics are therefore difficult to capture at a single time point. A time-course experiment can be beneficial to capture both early phosphorylation changes and late protein expression changes, and to elucidate the detailed resistance mechanisms of

cellular rewiring of signaling pathways to restore proliferation.

We have demonstrated two systematic approaches that can be used to nominate combination therapies in cancer, but they both have their limitations. The former requires a rich dataset for model construction, while the latter is focused only on combinations concerning a few drugs of our choice. We ultimately aim to combine the two approaches to collect rich proteomic profiling data and construct a large-scale model of cell biology for the nomination of combination therapy. In fact, the second approach presented here was derived from our effort to do so. We plan to apply hundreds of perturbations to ovarian cancer cells followed by mass-spectrometry proteomic profiling and phenotypic measurements. The collected data can then be used by CellBox to construct network models of thousands of nodes, resulting in millions of possible in silico pair-wise combinatorial perturbations, and even more higher-order combinations. Chemical compounds/drugs will be used as perturbations for the following reasons: i) they act on proteins, which are the direct functional regulators of cellular phenotypes, and therefore provide more direct perturbation-target relationship; ii) the alternative perturbation approach CRISPR-cut in our hands did not provide strong proteomic changes due to incomplete genome editing on a population level, and suffered from heterogenous genome patterns as a result of DNA repair (data not shown). CRISPRi on the other hand may be a useful alternative. Off-target effects of perturbation by drugs can be of concern, but our recent results [105] suggest that the model can perform data-driven derivation of perturbation targets. An improvement of the experimental approach is to collect multiple types of cellular phenotypes in addition to cell viability/proliferation, such as apoptosis, senescence, and quiescence. These phenotypes can also serve as desired treatment outcomes, as suggested by recent studies [235,236], for the ranking of promising combination candidates by the models.

We acknowledge that this is an ambitious plan since hundreds of arrayed experiments for mass-spectrometry profiling can be costly and labor intensive. As discussed previously, the number of perturbation conditions usually scales with the size of the network model for accurate parameter

inference, a potential approach to reduce the number of required perturbation conditions is to perform dimensionality reduction on the high-dimensional molecular measurements. While many data-driven machine learning approaches can be used for dimensionality reduction, such as principle components analysis (PCA), non-negative matrix factorization (NMF), autoencoder, an interpretable approach can be to collapse proteins into modules with similar functions or known signaling pathways. Additional analysis is needed to assess the information loss due to the compromise of resolution. While the inferred network models are context-specific and need to be constructed on experimental data collected in the biological background of interest, e.g. cell lines of a particular cancer type, recent computational methods [114,115] suggest that different genomic backgrounds can be taken into account for model prediction.

Combination therapies can be effective in treating cancer and many researchers are working on actively searching for promising combination candidates. We, among many of them, have presented here our attempts to nominate combination therapies using systematic approaches. We believe this work has provided testable candidates as well as powerful methods in the field of cancer combination therapy. We hope it can also be inspiring for future work and eventually contribute to the effective cure of cancer.

A

Additional research projects

During my graduate studies, I have conducted several additional research projects to understand the mechanisms in disease biology with the hope to contribute to potential therapeutics. I would like to include them here as useful resources for researchers who are interested and for the communities of relevance.

In the study presented in the first section, we investigated the potential functional impact of non-coding somatic mutations in cancer[237]. The discovery of drivers of cancer has traditionally focused

on protein-coding genes. The analyses of point mutations and structural variants in non-coding regions across 2,658 genomes from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have identified novel candidates of driver mutations in non-coding regions. We demonstrated that the mutations in two long non-coding RNA candidates, RMRP and RPPH1, largely fall within regions of RNA secondary and tertiary structures, and RNA-protein interaction sites. These results provide additional evidence supporting potential cancer driver discovery in these non-coding regions.

The projects reported in the other two sections focus on COVID-19 (Coronavirus disease 2019), a pandemic that started in December 2019 caused by a virus SARS-CoV-2 (the severe acute respiratory syndrome coronavirus 2). Effective, low-cost therapeutics are needed to prevent and treat COVID-19. Disulfiram is an approved oral drug used to treat alcohol use disorder that is a potent anti-inflammatory agent and an inhibitor of the viral proteases. We investigated the potential effects of disulfiram on SARS-CoV-2 infection and disease severity in an observational study using a large database of clinical records from the national US Veterans Affairs healthcare system [238]. A multivariable Cox regression adjusted for demographic information and diagnosis of alcohol use disorder revealed a reduced risk of SARS-CoV-2 infection with disulfiram use at a hazard ratio of 0.66 (34% lower risk, 95% confidence interval 24–43%). There were no COVID-19 related deaths among the 188 SARS-CoV-2 positive patients treated with disulfiram, in contrast to 5–6 statistically expected deaths based on the untreated population (P = 0.03). Our epidemiological results suggest that disulfiram may contribute to the reduced incidence and severity of COVID-19. These results support carefully planned clinical trials to assess the potential therapeutic effects of disulfiram in COVID-19.

In the third section of this chapter, we applied the approach of evolutionary couplings (ECs) to explore the RNA structures and RNA-RNA interactions in coronavirus to identify potential

therapeutic targets. Using ECs, we confirmed previously known secondary structures in coronavirus and provided a more precise characterization of the 3' pseudoknot region involved in a molecular switch mechanism of virus transcription, which might be of potential therapeutic interest. Long-range RNA-RNA interactions were not identified using our approach due to insufficient sequence diversity. However, we believe that the rapid growth of sequenced coronavirus genomes can provide much more power for the discovery of high-order RNA structures and interactions and therefore more information for the development of therapeutics.

A.1 Functional impact of non-coding somatic mutations in cancer

Part of the work presented in this section has been included in the paper titled "Analyses of non-coding somatic drivers in 2,658 cancer whole genomes" by E. Rheinbay et al., *Nature*, **578**, 102-111 (2020)[237].

The work is derived from the analyses of driver point mutations and structural variants in non-coding regions across 2,658 genomes from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA). Multiple driver-discovery algorithms followed by stringent filtering strategy have discovered 103 hits in non-coding elements as putative drivers for cancer. It has been found that both the exon and promoter of the non-coding RNA (ncRNA) RMRP were significantly mutated in multiple types of cancer (Figure A.1). RNase MRP (RMRP) is an enzymatically active ribonucleoprotein discovered in all eukaryotes that has the following known functions: generate RNA primers for mitochondrial DNA replication, process precursor rRNA in the nucleus, and in *S. cerevisiae* regulate cell cycle by cleaving cyclin B2 (CLB2) mRNA[239]. RMRP is not catalytically active without associated protein subunits (Pop1, Pop3/Rpp38, Pop4/Rpp29, Pop5, Pop6/Rpp25, Pop7/Rpp20, Pop8/Rpp14, Rpp1/Rpp30, and Rpp40)[239] and many of the previous studies have shown evidence

towards the complex structure. We then conducted analyses to further investigate the functional implications of the mutations in RMRP. The gene body mutations (7 SNVs in pan-cancer) show a significant bias towards high structural impact (rank-sum test, P = 0.011). Three of these are individually significant (each with P < 0.1, sample level permutation tests). Of the four gene-body indels, three are located in or near protein-binding sites (P = 0.08), including a deletion that is predicted to affect the secondary structure (Figure A.1).

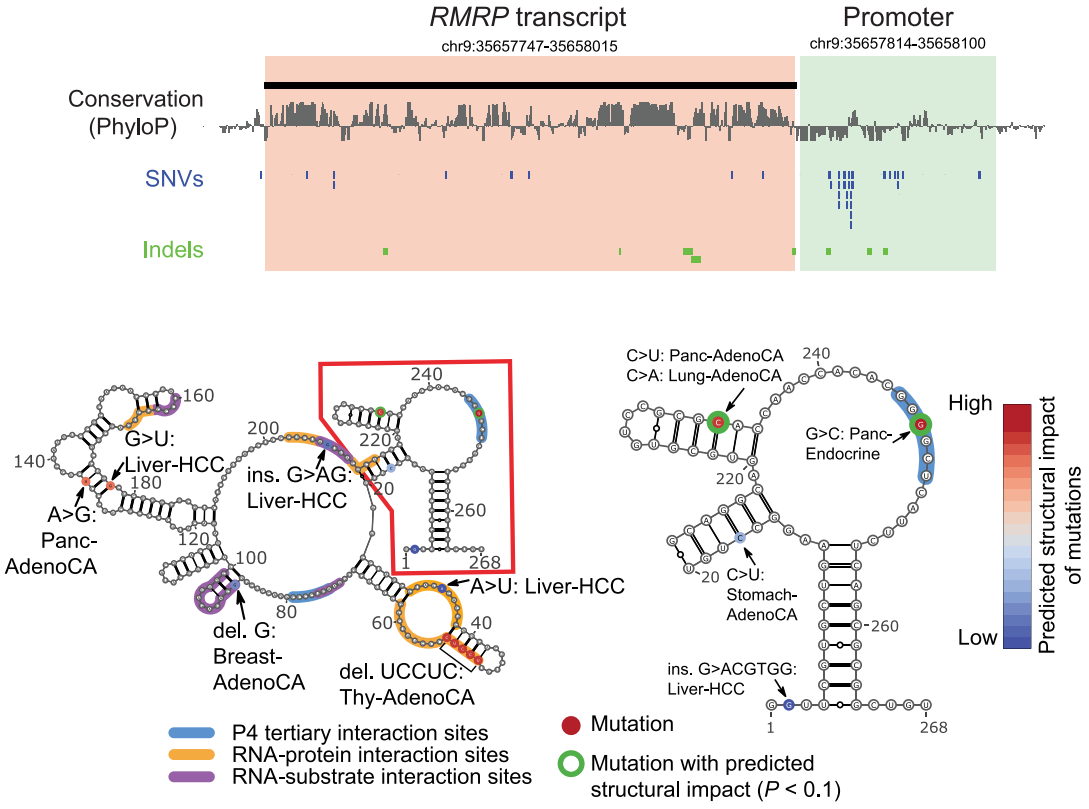


Figure A.1: Mutations in RMRP and their functional implications. Genomic locus of the RMRP transcript and promoter region (upper). RMRP is an RNA component of the endoribonuclease RNase MRP, the function of which depends on its RNA secondary and tertiary structure. The RNA secondary structure, tertiary structure interactions, protein and substrate interactions, and mutations with their predicted structural effect (bottom) of RMRP; lymphoma and melanoma mutations are excluded.

For each of the mutations in RMRP, we evaluated whether they appear within the locations of i) secondary structures of the RNA, ii) predicted tertiary structures of the RNA, or iii) putative RNA-protein contacts (Table A.1). The secondary structure pairs were determined based on the predicted secondary structures by Rfam (The RNA families database)[240] followed by confirmation using Evolutionary Couplings (EC) analysis[241,242] on the multiple sequence alignments. The tertiary structures were derived from the top 100 interactions predicted by ECs. The RNA-protein contacts were collected from results of previously reported experiments such as UV-crosslinking[243], footprinting[244,245], and a crystal structure of an RNA-protein complex of a yeast homolog[246,247]. Mutations that do not fall into any of the three categories were not included in the table.

RMRP	Secondary structure pair	Tertiary structure contact	RNA-protein contact
17 C>U:Stomach-AdenoCA	✓		
33 A>U:Liver-HCC			✓
49-54 UCCUCC>C:Thy-AdenoCA	✓		✓
53-58 CCCCUU>U:Thy-AdenoCA	✓		✓
89-90 GC>C:Breast-AdenoCA	✓		✓
135 A>G:Panc-AdenoCA	✓		
177 G>U:Liver-HCC	✓		
201 C>U:Skin-Melanoma		✓	
208 G>AG:Liver-HCC			✓
223 C>U:Panc-AdenoCA	✓		
223 C>A:Lung-AdenoCA	✓		
247 G>C:Panc-Endocrine		✓	
255 C>G:Skin-Melanoma	✓		

Table A.1: Functional implications of mutations in RMRP. Mutations identified in RMRP mostly localize within the regions of secondary structure, predicted tertiary structure, and RNA-protein interaction sites.

Fifteen ncRNAs hits have also passed the systematic post-filter, including RPPH1, the RNA component of the ribonuclease P (RNase P). RMRP is evolutionarily related to RNase P and share nine of the ten protein subunits as RNase P in human. We therefore performed the same analyses

on RPPH1, investigating the locations of the mutations relative to the secondary, tertiary structures and RNA-protein contacts (Table A.2).

RPPH1	Secondary structure pair	Tertiary structure contact	RNA-protein contact
36 C>G:Breast-AdenoCa			✓
80 C>G:Skin-Melanoma		✓	
118 C>A:Liver-HCC	✓		✓
133 G>A:Lymph-BNHL	✓		✓
235 G>C:Panc-AdenoCA	✓		
267 G>CG:Lymph-BNHL			✓
268 A>G:Prost-AdenoCA			✓
270 U>C:Liver-HCC			✓
285 G>UG:Biliary-AdenoCA	✓		

Table A.2: Functional implications of mutations in RPPH1. Mutations identified in RPPH1 mostly localize within the regions of secondary structure, predicted tertiary structure, and RNA-protein interaction sites.

A similar visualization of the mutations in RPPH1 in the context of the secondary structure, tertiary structure and RNA-protein contacts is shown here (Figure A.2). The structure and domain assignments of RPPH1 are very similar to those of RMRP. To better compare the two ncRNAs and examine the shared characteristics of the mutations, we aligned the mutations between RMRP and RPPH1 to their similar predicted secondary structures and mapped the protein and substrate interaction sites (Figure A.3). We found mutations in both RNAs in P3 and P4 regions with similar positions. Both the secondary structure P3 and the tertiary structure P4 are among the strongest ECs, suggesting these two structures are evolutionarily conserved and might be essential for the RNA functionality. In RMRP, one point mutation was found in position 247 (G>C), which exhibits strong coupling with position 77 with almost exclusive G-C or A-U pairs in reference genomes of all other organisms. The mutation is likely to disrupt P4 pairing and therefore the function of RMRP. Interestingly, the mutation in similar P4 position 80 of RPPH1 does not have strong evolutionary

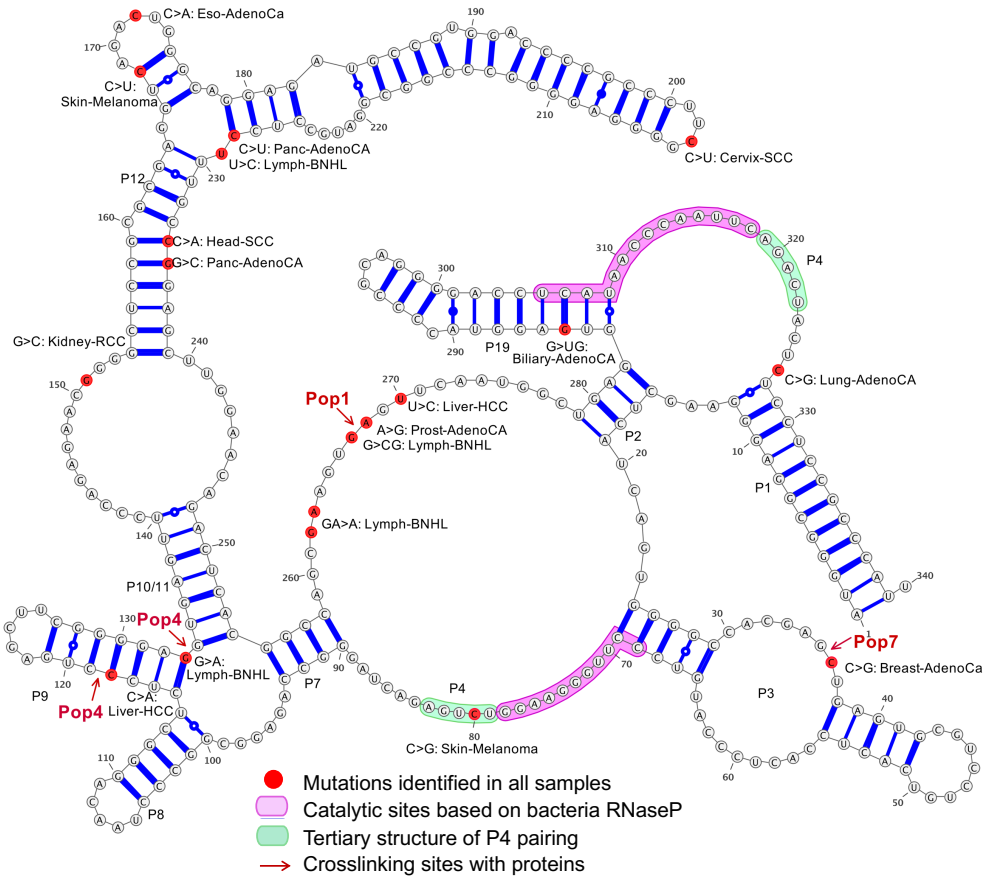


Figure A.2: Functional implications of mutations in RPPH1. RPPH1 is an RNA component of the ribonuclease P, the function of which depends on its RNA secondary and tertiary structure. The RNA secondary structure, tertiary structure interactions, and protein and substrate interactions are shown.

couplings with its corresponding paring position 320, suggesting this mutation might not be disruptive to the functionality of RPPH1.

It has been known that telomerase complex shares protein subunits Pop1, Rpp25/Pop6, and Rpp20/Pop7 with RNase MRP and RNase P due to a functional P3-like domain in telomerase RNA[248]. In addition, human telomerase reverse transcriptase (TERT) interacts with RNA component of RMRP to form a ribonucleoprotein complex with RNA-dependent RNA polymerase ac-

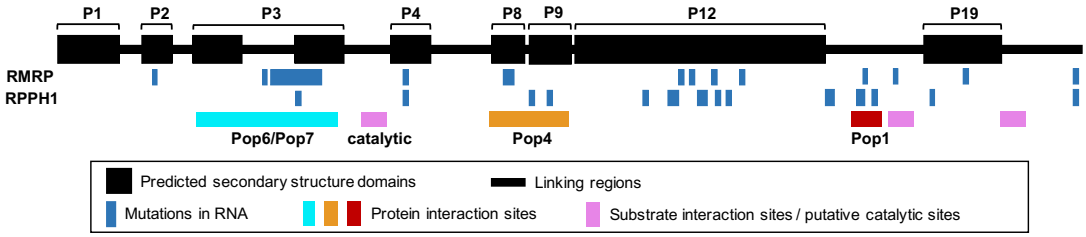


Figure A.3: Mapping of mutations to secondary structures and protein interaction sites. RMRP and RPPH1 share highly similar secondary structures and protein-RNA/substrate interaction sites.

tivity and produce double-stranded RNAs that are further processed by Dicer into small interfering RNA[249]. Given the increased telomerase activity in cancer, we propose a possible mechanism that the mutations in the RNA component of RMRP (or RNase P) increase the amount of telomerase, through loosening the interactions between RNA and proteins in RMRP, RNase P, and TERT-RMRP complexes and therefore sparing the three protein subunits to form telomerase complexes. This hypothesis may be worth pursuing but is beyond the scope of this project.

In summary, we performed analyses on exploring the functional impact of the non-coding somatic mutations in cancer. Evidence from secondary structures, tertiary structures from EC, and protein-RNA interactions suggest that some mutations in RMRP and RPPH1 are potential drivers in cancer. However, due to the existence of technical artifacts, potential mapping issues, and high germline mutation rates in the discovery, experiments further supporting the mutational mechanisms are needed to draw definitive conclusions.

A.2 Reduced risk of COVID-19 with the use of disulfiram

The work presented in this section has been published in the paper titled "Disulfiram use is associated with lower risk of COVID-19: A retrospective cohort study" by N. Fillmore, S. Bell, C. Shen et al., *Plos One*, 16(10), 2021 [238].

Author contributions

Conceptualization: Nathanael Fillmore, Steven Bell, Hao Wu, Judy Lieberman, Nhan Do, Chris Sander.

Data curation: Nathanael Fillmore, Steven Bell, Ciyue Shen, Vinh Nguyen, Jennifer La.

Formal analysis: Nathanael Fillmore, Steven Bell, Ciyue Shen, Vinh Nguyen, Chris Sander.

Funding acquisition: Nhan Do.

Investigation: Maureen Dubreuil, Judith Strymish, Mary Brophy, Gautam Mehta, Hao Wu, Judy Lieberman, Nhan Do.

Methodology: Nathanael Fillmore, Steven Bell, Vinh Nguyen. Resources: Nhan Do.

Supervision: Nhan Do, Chris Sander.

Validation: Nathanael Fillmore, Steven Bell, Ciyue Shen, Chris Sander.

Visualization: Nathanael Fillmore, Steven Bell, Ciyue Shen.

Writing – original draft: Nathanael Fillmore, Steven Bell, Ciyue Shen, Hao Wu, Judy Lieberman, Nhan Do, Chris Sander.

Writing – review & editing: Nathanael Fillmore, Steven Bell, Ciyue Shen, Vinh Nguyen, Jennifer La, Maureen Dubreuil, Judith Strymish, Mary Brophy, Gautam Mehta, Hao Wu, Judy Lieberman, Nhan Do, Chris Sander.

As of September 2021, there have been more than 210 million cases of coronavirus disease 2019 (COVID-19) worldwide with over 4.5 million deaths. Although several vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been approved and are being administered

internationally, there will still be a significant number of infections in people who are not vaccinated or for whom vaccination fails to fully protect, especially in regions with inadequate access or acceptance of vaccination. Moreover, there is a growing need for effective treatment of COVID-19, especially in light of reduced vaccine effectiveness in preventing infection with novel mutant viruses. Thus, identifying effective, inexpensive, and widely accessible treatment strategies is a high priority.

Repurposing of medications that are already approved for other indications is an attractive strategy, given the length of time and costs involved in developing new drugs[250]. With a growing number of candidates being identified, conducting a clinical trial for each remains a challenge due to the substantial time and financial resources required. While not equivalent to a randomized controlled study, analysis of real world clinical records can be used to investigate the incidence or severity of disease in those prescribed a medication compared to those not on the drug and provide observational evidence of a drug's potential effectiveness against COVID-19[250]. This study focuses on disulfiram (Antabuse), which has been used for over 60 years to deter alcohol use as it inhibits aldehyde dehydrogenase and causes alcohol flush reaction[251]. Disulfiram potentially targets multiple enzymes because it is a Cys-reactive drug. Despite its lack of specificity, it has an exceedingly favorable safety profile in the absence of concomitant alcohol use. The study was motivated both by promising observations about the effects of disulfiram in biochemical and cell-based screens and in mouse experiments[252,253,254,255,256,257,258,259,260,261].

There are several indications regarding the potential effects of disulfiram in reducing infections with SARS-CoV-2 and severity of COVID-19, as follows: (i) Inhibition of the main protease: Jin et al.[252] conducted high throughput screening of potential inhibitors of the SARS-CoV-2 main protease (Mpro), an enzyme needed for viral replication and identified disulfiram ($IC50 = 9.35\ \mu M$) among the candidates. (ii) Inhibition of the papain-like protease: earlier Lin et al.[253] demonstrated the inhibition of the papain-like proteases (PLpro) of MERS-CoV and SARS-CoV-1 by disulfiram through multiple enzymatic assays. The inhibition of PLpro of SARS-CoV-2 has been confirmed

by several groups, at IC50 values in biochemical assays ranging from 2–7 μM[254,255,256]. (iii) Inhibition of other viral proteins: disulfiram has been reported to inhibit the ATPase activity of Nsp13 and the exoribonuclease activity of Nsp14 of SARS-CoV-2 via its Zn^{2+}-ejector function[260]. (iv) In addition, disulfiram has general anti-inflammatory effects in sepsis. Based on the results of a drug screen of 3,752 compounds, Hu et al.[261] discovered a pronounced anti-inflammatory effect of disulfiram that reduced mortality in a mouse model of sepsis. Disulfiram inhibits inflammatory cytokine IL-1β release and pyroptosis (inflammatory cell death) by blocking the assembly of the gasdermin D pore by covalent modification of Cys191. These mechanistic studies support the hypothesis that disulfiram may be a polypharmacological therapeutic compound preventing both virus replication and life-threatening host response in coronavirus disease.

In this study, we looked for epidemiological evidence to be considered when prioritizing existing medications for clinical trials in both the early and late stages of COVID-19. We investigated the incidence and outcome of COVID-19 disease among persons taking disulfiram compared to the general population. We conducted a retrospective cohort study comparing Veterans who were treated with disulfiram to those who were not treated. The data is from the US Department of Veterans Affairs (VA) COVID-19 Shared Data Resource and the VA Corporate Data Warehouse (CDW), which consolidates electronic health record data from VA facilities nationwide. We identified a cohort of Veterans with at least one SARS-CoV-2 laboratory test result between Feb. 20, 2020 and Feb. 1, 2021. In order to ensure patients were receiving routine care at the VA, we required that each Veteran had visited a VA primary care provider in the 18 months before their first SARS-CoV-2 test.

Among 944,127 patients meeting the inclusion criteria, 2,233 had at least one pharmacy record for disulfiram on or after Feb. 20, 2019 and 100,873 had been diagnosed with AUD (Figure A.4). The patient population is racially and regionally heterogeneous, reflecting the VA's nation-wide patient population (Table S1). Patients treated with disulfiram differed in their basic demographic and clinical history from those who were not treated, e.g., they were younger, more likely to be white,

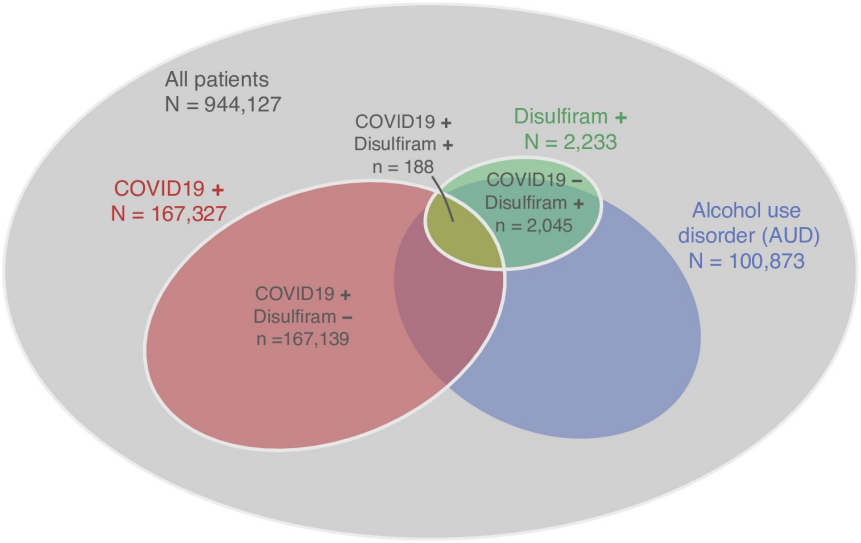


Figure A.4: Population distribution of the patient cohort used for the study. 944,127 Veterans have at least one SARS-CoV-2 laboratory test result between Feb. 20, 2020 and Feb. 1, 2021, and had visited a VA primary care provider in the 18 months before their first SARS-CoV-2 test. Disulfiram + (-) indicates at least one (no) pharmacy record for disulfiram on or after Feb. 20, 2019. COVID19 + (-) indicates a positive (no positive) SARS-CoV-2 laboratory test.

and having a lower burden of comorbidity ($P < 0.001$ for all tests). Therefore these potentially confounding variables were taken into account in the regression analysis.

The incidence of SARS-CoV-2 infection in patients treated with disulfiram was 188 patients infected during 1,674 person years of follow-up. In contrast, the incidence of infection in untreated patients was 167,139 patients infected during 838,279 person years of follow-up. Reflecting this difference, the incidence rate ratio of infection in disulfiram-treated vs untreated patients was 0.56 (95% confidence interval (CI): 0.49–0.65, $P < 0.001$), i.e., a 44% lower rate of infection in those treated with disulfiram.

In a simple univariable Cox regression model, we observed a hazard ratio (HR) of 0.50 (95% CI 0.44–0.58, $P < 0.001$), similar to the incidence rate ratio, as expected. In the more refined multivariable model (Table A.3, unrestricted analysis), after adjusting for age, gender, race/ethnicity, region, Charlson score (a comorbidity index), and AUD diagnosis, the association was still highly signifi-

cant. Patients treated with disulfiram had a 34% lower risk of SARS-CoV-2 infection compared to those not treated with disulfiram, i.e., an HR of 0.66 (95% CI 0.57–0.76, P < 0.001). The multivariable model also revealed significant associations of risk of infection with age, gender, race/ethnicity, region, and comorbidity (via the Charlson score), which are consistent with earlier results [262].

		Unrestricted analysis Hazard Ratio (95% CIs[a])	P-value	Restricted analysis Hazard Ratio (95% CIs)	P-value
Disulfiram	**Treated**	**0.66 (0.57, 0.76)**	**<0.001**	**0.68 (0.58, 0.80)**	**<0.001**
AUD[b]	Diagnosed	0.69 (0.68, 0.70)	<0.001		
Gender	Female	1.00 (ref.[c])			
	Male	1.18 (1.16, 1.20)	<0.001	1.12 (1.05, 1.20)	<0.001
Age	Continuous increase	1.00 (1.00, 1.00)	<0.001	0.99 (0.99, 1.00)	<0.001
Ethnicity	Non-Hispanic White	1.00 (ref.)			
	Non-Hispanic Black	1.10 (1.09, 1.11)	<0.001	1.17 (1.12, 1.21)	<0.001
	Hispanic	1.33 (1.31, 1.35)	<0.001	1.70 (1.61, 1.79)	<0.001
	Other/Unknown	1.12 (1.10, 1.14)	<0.001	1.19 (1.12, 1.27)	<0.001
Region	Continental	1.00 (ref.)			
	Midwest	1.02 (1.00, 1.03)	0.038	1.01 (0.96, 1.06)	0.736
	North Atlantic	0.71 (0.70, 0.72)	<0.001	0.77 (0.73, 0.81)	<0.001
	Pacific	0.65 (0.64, 0.66)	<0.001	0.79 (0.75, 0.84)	<0.001
	Southeast	0.78 (0.77, 0.79)	<0.001	0.78 (0.74, 0.83)	<0.001
Charlson score[d]	0	1.00 (ref.)			
	1–2	0.96 (0.95, 0.97)	<0.001	0.91 (0.87, 0.94)	<0.001
	3–4	0.87 (0.85, 0.88)	<0.001	0.89 (0.83, 0.94)	<0.001
	≥5	0.83 (0.81, 0.85)	<0.001	0.92 (0.85, 1.00)	0.046
	Unknown	1.06 (1.03, 1.09)	<0.001	0.95 (0.55, 1.63)	0.844

Unrestricted analysis stands for multivariable Cox regression on all 944,127 patients meeting the inclusion criteria; Restricted analysis was performed on the 100,873 patients diagnosed with AUD.
[a]Confidence intervals
[b]Alcohol Use Disorder
[c]Reference category
[d]Charlson comorbidity index.

Table A.3: Multivariable Cox regression results for a positive SARS-CoV-2 test against disulfiram treatment and other factors. Multivariable Cox regression from both unrestricted and restricted analysis showed that disulfiram use, together with other demographic factors, is associated with reduced risk of COVID-19.

To further separate the effect of AUD (HR = 0.69 in the unrestricted analysis) and disulfiram, we performed the multivariable Cox regression analysis restricted to those with diagnosis of AUD (100,873 out of 944,127 patients). We observed an HR of 0.68 (95% CI 0.58–0.80, P < 0.001) (Table A.3, restricted analysis). Within confidence limits, this result is consistent with the HR of 0.66 of the unrestricted analysis.

In the main analysis, motivated by potential incompleteness of data capture and the relatively

short length of followup (<1 year maximum) in our study, we assume that each patient continues disulfiram treatment starting from their first disulfiram pharmacy record through the end of their followup. As an additional sensitivity analysis, we fit a multivariable model with the same covariates but assumed that patients remained on disulfiram for only one month after each record of dispensation, and observed a similar association as in the primary analysis (HR = 0.53; 95% CI 0.44–0.64, P < 0.001).

Finally, it is possible that our results could be affected by SARS-CoV-2 vaccination, which began during the study period. As a third sensitivity analysis, we fit a multivariable model with the same covariates but with followup time censored on December 11, 2020, which is the date of the first dispensation of a SARS-CoV-2 vaccine at the VA, and we again observed a similar association (HR = 0.64; 95% CI 0.55–0.74, P < 0.001). Of note, patients treated with disulfiram were less likely to be vaccinated against SARS-CoV-2 during our study period (11.6%) compared to those untreated (15.7%; P < 0.001), which if anything would result in an underestimation of the HR for those on disulfiram.

We also analyzed whether disulfiram prescription was linked to severe clinical outcomes and a composite measure of severe outcomes (Table A.4). Inferences from these data are limited due to the low frequency of events. There were no statistically significant differences in ICU admissions and mechanical ventilation between the two cohorts. However, we did observe a statistically significant difference in occurrence of COVID-19 related death. In the untreated cohort, 3% of patients experienced COVID-19 related death. In contrast, among the 188 patients treated with disulfiram, no COVID-19 related deaths were reported at all (P = 0.03, Table A.4), contrasting with 5–6 that would have been expected under a 3% incidence proportion. Nevertheless, overall, there was no statistically significant difference between the proportion of the composite severe endpoints between the two cohorts. Given the low counts of severe clinical outcomes in this study and uncertainty regarding drug adherence and potential discontinuation on admission, epidemiological observations

in a larger dataset or in an inpatient setting are needed to probe whether disulfiram has an impact on the severity of COVID-19.

	No. of treated patients (N = 188)	No. of untreated patients (N = 167,139)	Incidence proportion in treated	Incidence proportion in untreated	P-value
ICU[a]	11	7,403	0.059	0.044	0.44
Mechanical ventilation	1	959	0.005	0.006	1.00
Death	**0**	**5,009**	**0.000**	**0.030**	**0.03**
All severe outcomes	12	13,371	0.063	0.080	0.45

[a]Intensive care unit.

Table A.4: Clinical outcomes in patients infected with SARS-CoV-2 by disulfiram treatment status. Disulfiram use is associated with lower death rate in COVID-19 patients, but statistical power is insufficient to draw conclusions on all other severe outcomes.

Our study is the largest using real world data to demonstrate a lower risk of SARS-CoV-2 infection in those taking disulfiram. Recently, in a preliminary report, Tamburin et al.[263] showed that patients taking disulfiram had a significantly lower incidence of symptoms compatible with COVID-19, such as fever and dyspnea, compared to a control group. However, their study was underpowered to discern a difference in actual risk of COVID-19 disease. Our study significantly extends these initial observations in two ways: (1) the VA patient population with relevant data is considerably larger, which enabled us to infer a statistically significant association between disulfiram use and risk of SARS-CoV-2 infection, and (2) we were able to document the clinical outcomes of those infected with COVID-19 taking disulfiram and estimate the risk of death. Other commonly used drugs have also been reported to reduce the risk of infection and disease severity[264]. Access to larger clinical datasets and meta-analysis across diverse health-care systems would be useful in future investigations on this topic.

The principal limitation of our study, as is true even for studies with controlled cohorts, is that association does not prove causation and that even very likely causation does not guarantee the success of the corresponding therapeutic intervention. There are several additional limitations. Since Veterans can receive care outside of the VA, diagnoses and outcomes may be incompletely recorded

and we do not know if patients actually took the prescribed disulfiram drug, or whether they were still taking the drug near the time of infection. We also cannot rule out the possibility that those prescribed disulfiram had other factors, such as behavioral differences or other confounding variables, that contributed to their protection against COVID-19. Additionally, we are unable to distinguish symptomatic from asymptomatic infection, as documentation of the relevant symptoms is inconsistent in the electronic health records used in our study. For inpatient settings, patients are routinely tested on admission. For outpatient settings, testing can occur for symptoms or concerns of exposure.

Our epidemiological results suggest that disulfiram may be efficacious in combating COVID-19. Additional information is expected from two small ongoing Phase II clinical trials of disulfiram involving early mild-to-moderate symptomatic (NCT04485130, 60 participants) or hospitalized (NCT04594343, 200 participants) COVID-19 patients. Reduction in disease progression in the first trial, if observed, may be primarily due to inhibition of viral replication. This hypothesis is based on clear evidence from biochemical experiments that disulfiram inhibits several enzymes involved in viral replication (Mpro, PLpro, Nsp13 and Nsp14) via covalent modification of cysteine residues in the active site of the Mpro and PLpro proteases and via weakening of Cys-coordinated Zn2+ ion binding sites in the finger domain of PLpro and in the RNA-modifying enzymes Nsp13 and Nsp14[252,253,254,255,256,257,258,259,260]. Better outcomes in the second trial, if observed, may be more influenced by the anti-inflammatory effects of disulfiram, as demonstrated in a mouse model of sepsis, via inhibition of the formation of the host gasdermin D pore and reduction of pyroptosis and inflammatory cytokine secretion[261].

Given the urgent medical need, larger clinical trials stratified for early and late COVID-19 disease as well as in uninfected populations are highly desirable to test the impact of disulfiram across the entire COVID-19 continuum, from infection to disease progression and severe disease. A positive drug effect would most likely also extend to the constantly evolving SARS-CoV-2 variants and re-

lated coronaviruses, as the known viral target sites of disulfiram are highly conserved Cys residues. As a repurposing candidate, disulfiram has particular advantages given its low cost, both in production and distribution, and favorable safety profile[251,265]. We therefore propose that clinical trials should be pursued proactively, given the current limited access to vaccination and the lack of generally available, low-cost proven therapeutic agents against this pandemic disease. Our view is that such trials require governmental or philanthropic funding and that time is of the essence. If successful in clinical trials, disulfiram would be a good candidate as a COVID-19 therapeutic for worldwide distribution, including to low-income populations.

A.3 Exploration of RNA structure and RNA-RNA interactions in coronavirus

Coronaviruses are a group of positive-sense single-stranded RNA viruses that can cause respiratory illness in humans and animals. Three recent coronaviruses, SARS-CoV, MERS-CoV, and SARS-CoV-2, which cause severe symptoms and even deaths, have raised the concern of the public. SHAPE analysis and thermodynamic predictions suggest that RNA secondary structures are well conserved in the coronavirus genome, including the 5' and 3' UTR regions, and the frameshifting stimulation element (FSE) in between the ORF1a and ORF1b [266,267,268,269,270,271,272]. Reverse genetic studies show that these structures play an important role in the replication and transcription (discontinuous synthesis of subgenomic RNAs) of the viruses. We are interested in using our evolutionary covariation approach [241,242] to study the RNA structures and their functional implication in the virus life cycle, in particular to i) recapitulate the known interactions, ii) discover new RNA-RNA interactions that might be important for structures and functions, iii) provide insight for potential therapy of recent pandemic from the RNA perspectives. Here, we report our results on our exploration of RNA structure and RNA-RNA interactions in coronavirus.

The well-accepted model of the RNA secondary structures in the coronavirus genome contains the multiple stem-looped 5' UTR, the 3' UTR with the pseudoknot, and the bulged stem-loop (BSL), and the FSE region with the pseudoknot (Figure A.5). Coronavirus genome sequences were collected from the National Center for Biotechnology Information (NCBI) database [273] and the GISAID database (March 30th, 2020). About five thousand complete genome sequences for coronavirus were collected and about one thousand unique sequences were found for the 5' UTR and the 3' UTR region. Multiple sequence alignments (MSA) were made for three sub-regions, including the SL2 stem-loop, the FSE, and the 3' pseudoknot (PK). The data from Rfam [240] was used as seed alignments and MUSCLE [274] was used to build the final MSA with the sequences manually extracted from the corresponding regions of the collected genome sequences.

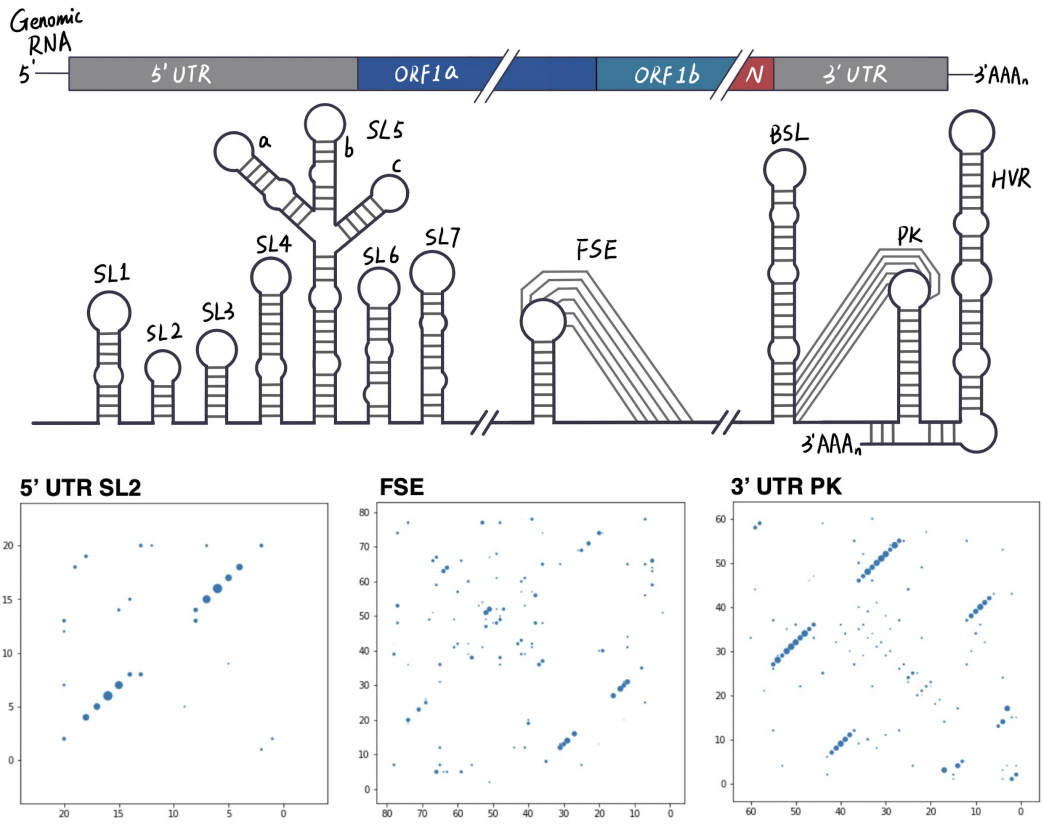


Figure A.5: RNA structures in the coronavirus genome and their evolutionary couplings. Known secondary structures mainly locate at the 5' UTR, FSE region, and the 3' UTR (adapted from Sola et al.[266]) (top). Contact maps of evolutionary couplings, including those for the 5' UTR SL2 (20 nucleotides, 133 sequences), FSE (79 nucleotides, 146 sequences), and 3' UTR PK (60 nucleotides, 94 sequences), reflect the RNA structures of the corresponding regions (bottom). The alignments and all following analysis are centered on the SARS-CoV-2 sequence.

Evolutionary couplings (ECs) between pairs of nucleotides were calculated using the standard plmc pipeline (https://github.com/debbiemarkslab/plmc) and plotted based on the coupling strengths (Figure A.5). The strongest ECs correspond to the base-pairing in the stem-loops and the interactions in the pseudoknots. Beyond the thermodynamic models, this covariation analysis indicates that these structures are evolutionarily conserved and therefore functionally important.

The 3' PK is a hypothetical molecular switch that regulates negative-strand RNA synthesis[275]:

The the bulged stem-loop (BSL) is fully closed until the beginning of transcription by the primase (nsp8); then the formation of the pseudoknot enables the loading of the RNA-dependent RNA polymerase (nsp12) (Figure A.5 upper). However, the evolutionary couplings within the left part of the pseudoknot suggest that the RNA should form a short hairpin structure, and the pseudoknot should instead be a kissing stem-loop interaction (Figure A.5bottom). Although the ECs did not provide any supporting or opposing evidence for the molecular switch model, they give a more precise characterization of the 3' PK region and a better basis for future studies of the hypothesis.

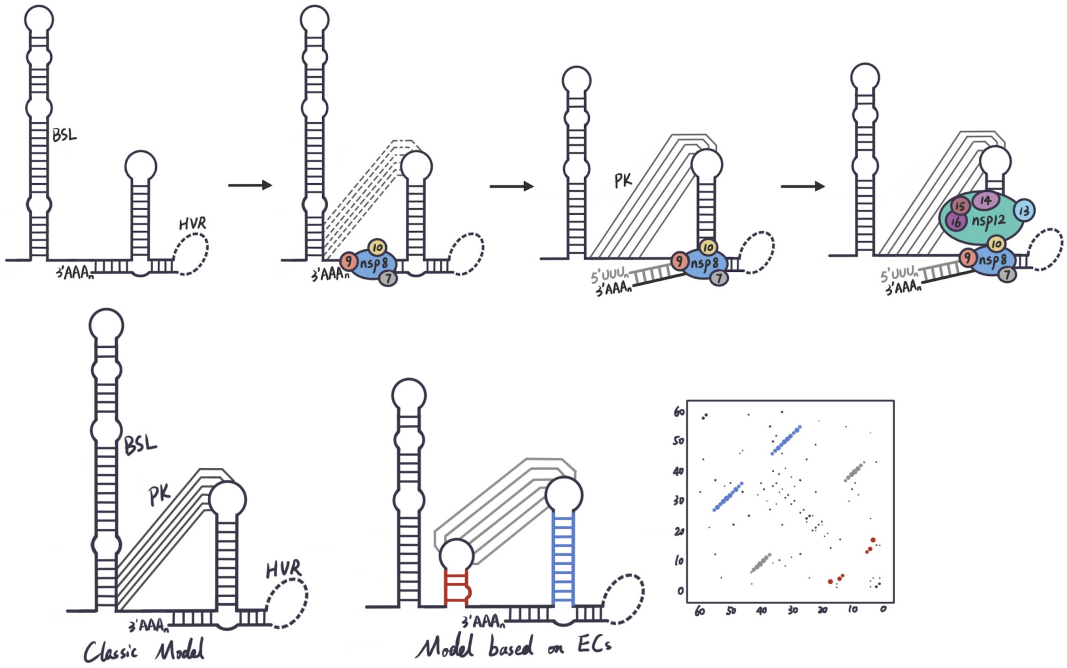


Figure A.6: Hypothetical molecular switch in the 3' UTR of coronavirus genome. The hypothetical molecular switch plays a role in the process of RNA synthesis (adapted from Zust et al.[275]). The primase (nsp8) and other transcription cofactors are loaded onto the 3' UTR while the bulged stem-loop (BSL) is fully closed. After the transcription of the first tens of nucleotides, the formation of pseudoknot induces the conformational change, which allows the loading of the RNA-dependent RNA polymerase (nsp12) and other cofactors (top). The evolutionary covariations suggest that the pseudoknot interaction should be more precisely characterized as a kissing stem-loop interaction (bottom).

In addition to local secondary structures, high-order RNA structures can be discovered by evolu-

tionary covariations. In the case of the 5' UTR and 3' UTR regions, we, however, were not able to find confidently RNA structures or interactions spanning couple hundreds of nucleotides mainly due to insufficient sequence diversity of coronaviruses in the current database.

Mateos-Gomez et al.[276] identified a long-distance RNA-RNA interaction of a 9-nt base-pairing between the 5' end and 3' end of the coronavirus genome (Figure A.7 top). This interaction is conserved in Alphacoronavirus I species and is important for the cyclization of the genome during sub-genomic RNA synthesis. To explore whether such interaction is conserved among other species of coronaviruses, we employed both the forward approach and reverse approach with the assumption that both the location of and the secondary structures around the base-pairing interaction are conserved. In the forward approach, we screened the regions with the proper secondary structures for such base-pairing interaction, while in the reverse approach, we examined all possible 9-nt base-pairing for proper secondary structures. Such base-pairing interactions were only discovered in a few coronavirus species in addition to the three Alphacoronavirus I species, and we were not able to confidently align the interactions among the species (Figure A.7 bottom). Without experimental validation, we cannot conclude that such long-distance RNA-RNA interaction is functionally conserved or contributes significantly to the sub-genomic RNA synthesis in other coronavirus species.

In summary, we applied the approach of evolutionary couplings to study the RNA structures in coronavirus that are of potential functional implications. We were able to confirm known secondary structures using ECs. In addition, compared to the existing model, we provided a more precise characterization of the 3' pseudoknot region involved in a molecular switch mechanism of virus transcription, which might be of potential therapeutic interest. Due to limited number of coronavirus sequences and insufficient sequence diversity, we were not able to identify new RNA structures or long-range RNA-RNA interactions using ECs. However, this work was done in the beginning of the pandemic of SARS-CoV-2, when only about thousands of coronavirus genome sequences were available. By April 2022, almost 10 million genome sequences of SARS-CoV-2 have become avail-

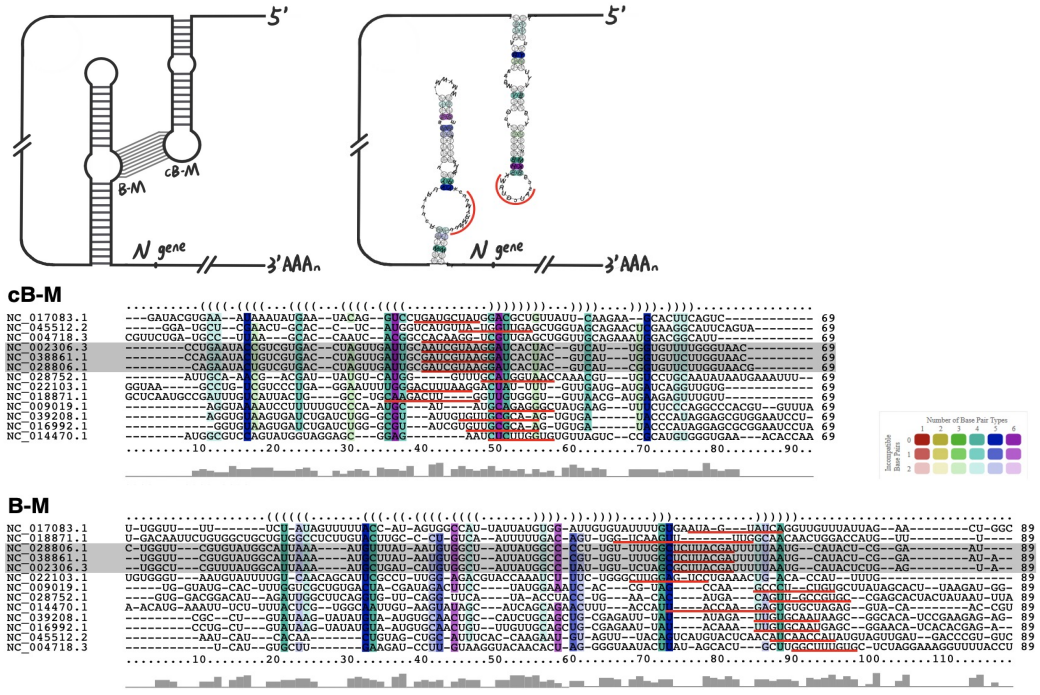


Figure A.7: A long-distance RNA-RNA interaction in the coronavirus genome. Visualization and secondary structure representation of an RNA-RNA interaction of 9-nucleotide base-pairing (cB-M, and B-M) that is more than 25,000 nt apart in Alphacoronavirus I species discovered by Mateos-Gomez et al.[276] (adapted from Mateos-Gomez et al.[276], top). The multiple sequence alignments by LocARNA[277] of the regions around the hypothetical cB-M and B-M (underlined in red) in 13 coronavirus species, including the three Alphacoronavirus I species (shaded in gray) and others, such as SARS-CoV and SARS-CoV-2 (bottom).

able[278], which will likely provide much more power for the discovery of high-order RNA structures in coronavirus and therefore provide useful information for the development of treatment.

References

[1] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2021. *CA: a cancer journal for clinicians*, 71(1):7–33, 2021.

[2] Steven I Hajdu. A note from history: landmarks in history of cancer, part 1. *Cancer*, 117(5):1097–1102, 2011.

[3] Robert A Weinberg. How cancer arises. *Scientific American*, 275(3):62–70, 1996.

[4] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[5] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.

[6] Douglas Hanahan. Hallmarks of cancer: New dimensions. *Cancer Discovery*, 12(1):31–46, 2022.

[7] Niki Papavramidou, Theodossis Papavramidis, and Thespis Demetriou. Ancient greek and greco–roman methods in modern surgical treatment of cancer. *Annals of surgical oncology*, 17(3):665–667, 2010.

[8] Lynda Wyld, Riccardo A Audisio, and Graeme J Poston. The evolution of cancer surgery and future perspectives. *Nature Reviews Clinical Oncology*, 12(2):115–124, 2015.

[9] Rajamanickam Baskar, Kuo Ann Lee, Richard Yeo, and Kheng-Wei Yeoh. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, 9(3):193, 2012.

[10] Vincent T DeVita and Edward Chu. A history of cancer chemotherapy. *Cancer research*, 68(21):8643–8653, 2008.

[11] Greg L Plosker and Susan J Keam. Trastuzumab. *Drugs*, 66(4):449–475, 2006.

[12] Nida Iqbal and Naveed Iqbal. Imatinib: a breakthrough of targeted therapy in cancer. *Chemotherapy research and practice*, 2014, 2014.

[13] International Cancer Genome Consortium et al. International network of cancer genome projects. *Nature*, 464(7291):993, 2010.

[14] F Fehleisen. Ueber die züchtung der erysipelkokken auf künstlichem nährboden und ihre übertragbarkeit auf den menschen. *Dtsch Med Wochenschr*, 8(31):553–4, 1882.

[15] Alex D Waldman, Jill M Fritz, and Michael J Lenardo. A guide to cancer immunotherapy: from t cell basic science to clinical practice. *Nature Reviews Immunology*, 20(11):651–668, 2020.

[16] Reza Bayat Mokhtari, Tina S Homayouni, Narges Baluch, Evgeniya Morgatskaya, Sushil Kumar, Bikul Das, and Herman Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23):38022, 2017.

[17] Nehad M Ayoub. Novel combination therapies for the treatment of solid cancers. *Frontiers in Oncology*, 11, 2021.

[18] EMIL FREI III, Myron Karon, Robert H Levin, EMIL J FREIREICH, Robert J Taylor, Juliet Hananian, Oleg Selawry, James F Holland, Barth Hoogstraten, Irving J Wolman, et al. The effectiveness of combinations of antileukemic agents in inducing and maintaining remission in children with acute leukemia. *Blood*, 26(5):642–656, 1965.

[19] Timothy A Yap, Aurelius Omlin, and Johann S De Bono. Development of therapeutic combinations targeting major cancer signaling pathways. *Journal of Clinical Oncology*, 31(12):1592–1605, 2013.

[20] Aaron C Tan, Stephen J Bagley, Patrick Y Wen, Michael Lim, Michael Platten, Howard Colman, David M Ashley, Wolfgang Wick, Susan M Chang, Evanthia Galanis, et al. Systematic review of combinations of targeted or immunotherapy in advanced solid tumors. *Journal for immunotherapy of cancer*, 9(7), 2021.

[21] Barbara Bukowska, Arkadiusz Gajek, and Agnieszka Marczak. Two drugs are better than one. a short history of combined therapy of ovarian cancer. *Contemporary Oncology*, 19(5):350, 2015.

[22] Jonathan B Fitzgerald, Birgit Schoeberl, Ulrik B Nielsen, and Peter K Sorger. Systems biology and combination therapy in the quest for clinical efficacy. *Nature chemical biology*, 2(9):458–466, 2006.

[23] Levi A Garraway and Pasi A Jänne. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer discovery*, 2(3):214–226, 2012.

[24] Behzad Mansoori, Ali Mohammadi, Sadaf Davudian, Solmaz Shirjang, and Behzad Baradaran. The different mechanisms of cancer drug resistance: a brief review. *Advanced pharmaceutical bulletin*, 7(3):339, 2017.

[25] Neil Vasan, José Baselga, and David M Hyman. A view on drug resistance in cancer. *Nature*, 575(7782):299–309, 2019.

[26] James H Doroshow and Richard M Simon. On the design of combination cancer therapy. *Cell*, 171(7):1476–1478, 2017.

[27] Adam C Palmer and Peter K Sorger. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell*, 171(7):1678–1691, 2017.

[28] Ursula A Matulonis, Anil K Sood, Lesley Fallowfield, Brooke E Howitt, Jalid Sehouli, and Beth Y Karlan. Ovarian cancer. *Nature reviews Disease primers*, 2(1):1–22, 2016.

[29] Cancer Genome Atlas Research Network et al. Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353):609, 2011.

[30] Richard W Tothill, Anna V Tinker, Joshy George, Robert Brown, Stephen B Fox, Stephen Lade, Daryl S Johnson, Melanie K Trivett, Dariush Etemadmoghadam, Bianca Locandro, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research*, 14(16):5198–5208, 2008.

[31] Roel GW Verhaak, Pablo Tamayo, Ji-Yeon Yang, Diana Hubbard, Hailei Zhang, Chad J Creighton, Sian Fereday, Michael Lawrence, Scott L Carter, Craig H Mermel, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *The Journal of clinical investigation*, 123(1), 2012.

[32] Gottfried E Konecny, Chen Wang, Habib Hamidi, Boris Winterhoff, Kimberly R Kalli, Judy Dering, Charles Ginther, Hsiao-Wang Chen, Sean Dowdy, William Cliby, et al. Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *JNCI: Journal of the national cancer institute*, 106(10), 2014.

[33] Gregory P Way, James Rudd, Chen Wang, Habib Hamidi, Brooke L Fridley, Gottfried E Konecny, Ellen L Goode, Casey S Greene, and Jennifer A Doherty. Comprehensive cross-population analysis of high-grade serous ovarian cancer supports no more than three subtypes. *G3: Genes, Genomes, Genetics*, 6(12):4097–4103, 2016.

[34] Broad Institute TCGA Genome Data Analysis Center. Analysis overview for ovarian serous cystadenocarcinoma (primary solid tumor cohort) - 28 january 2016, 2016.

[35] Hui Zhang, Tao Liu, Zhen Zhang, Samuel H Payne, Bai Zhang, Jason E McDermott, Jian-Ying Zhou, Vladislav A Petyuk, Li Chen, Debjit Ray, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*, 166(3):755–765, 2016.

[36] Nicoletta Colombo, Domenica Lorusso, and Paolo Scollo. Impact of recurrence of ovarian cancer on quality of life and outlook for the future. *International Journal of Gynecologic Cancer*, 27(6), 2017.

[37] Panagiotis A Konstantinopoulos, Stephanie Lheureux, and Kathleen N Moore. Parp inhibitors for ovarian cancer: current indications, future combinations, and novel assets in development to target dna damage repair. *American Society of Clinical Oncology Educational Book*, 40:e116–e131, 2020.

[38] Kezia Gaitskell, Igor Martinek, Andrew Bryant, Sean Kehoe, Shibani Nicum, and Jo Morrison. Angiogenesis inhibitors for the treatment of ovarian cancer. *Cochrane Database of Systematic Reviews*, (9), 2011.

[39] Zachary C Dobbin and Charles N Landen. The importance of the pi3k/akt/mtor pathway in the progression of ovarian cancer. *International journal of molecular sciences*, 14(4):8213–8227, 2013.

[40] Hyun-Jin Choi, Guillermo N Armaiz Pena, Sunila Pradeep, Min Soon Cho, Robert L Coleman, and Anil K Sood. Anti-vascular therapies in ovarian cancer: moving beyond anti-vegf approaches. *Cancer and Metastasis Reviews*, 34(1):19–40, 2015.

[41] S Percy Ivy, Joyce F Liu, Jung-Min Lee, Ursula A Matulonis, and Elise C Kohn. Cediranib, a pan-vegfr inhibitor, and olaparib, a parp inhibitor, in combination therapy for high grade serous ovarian cancer. *Expert opinion on investigational drugs*, 25(5):597–611, 2016.

[42] Gottfried E Konecny. Combining parp and cdk4/6 inhibitors in myc driven ovarian cancer. *EBioMedicine*, 43:9, 2019.

[43] Panagiotis A Konstantinopoulos, William T Barry, Michael Birrer, Shannon N Westin, Karen A Cadoo, Geoffrey I Shapiro, Erica L Mayer, Roisin E O'Cearbhaill, Robert L Coleman, Bose Kochupurakkal, et al. Olaparib and α-specific pi3k inhibitor alpelisib for patients with epithelial ovarian cancer: a dose-escalation and dose-expansion phase 1b trial. *The Lancet Oncology*, 20(4):570–580, 2019.

[44] Rowan E Miller, Amy J Lewis, and Melanie E Powell. Parp inhibitors and immunotherapy in ovarian and endometrial cancers. *The British Journal of Radiology*, 94(1128):20210002, 2021.

[45] Karen E Sheppard, Carleen Cullinane, Katherine M Hannan, Meaghan Wall, Joanna Chan, Frances Barber, Jung Foo, Donald Cameron, Amelia Neilsen, Pui Ng, et al. Synergistic inhibition of ovarian cancer cell growth by combining selective pi3k/mtor and ras/erk pathway inhibitors. *European journal of cancer*, 49(18):3936–3944, 2013.

[46] Elias E Stratikopoulos, Meaghan Dendy, Matthias Szabolcs, Alan J Khaykin, Celine Lefebvre, Ming-Ming Zhou, and Ramon Parsons. Kinase and bet inhibitors together clamp inhibition of pi3k signaling and overcome resistance to therapy. *Cancer cell*, 27(6):837–851, 2015.

[47] Jeffrey J Wallin, Jane Guan, Wei Wei Prior, Kyle A Edgar, Robert Kassees, Deepak Sampath, Marcia Belvin, and Lori S Friedman. Nuclear phospho-akt increase predicts synergy of pi3k inhibition and doxorubicin in breast and ovarian cancer. *Science translational medicine*, 2(48):48ra66–48ra66, 2010.

[48] Abdelghani Jebahi, Marie Villedieu, Cécile Pétigny-Lechartier, Emilie Brotin, Marie-Hélène Louis, Edwige Abeilard, Florence Giffard, Marika Guercio, Mélanie Briand, Pascal Gauduchon, et al. Pi3k/mtor dual inhibitor nvp-bez235 decreases mcl-1 expression and sensitizes ovarian carcinoma cells to bcl-xl-targeting strategies, provided that bim expression is induced. *Cancer letters*, 348(1-2):38–49, 2014.

[49] Adam C Palmer, Christopher Chidley, and Peter K Sorger. A curative combination cancer therapy achieves high fractional cell killing through low cross-resistance and drug additivity. *Elife*, 8:e50036, 2019.

[50] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. Combinatorial drug therapy for cancer in the post-genomic era. *Nature biotechnology*, 30(7):679–692, 2012.

[51] Afshin Derakhshani, Zohreh Rezaei, Hossein Safarpour, Morteza Sabri, Atefeh Mir, Mohammad Amin Sanati, Fatemeh Vahidian, Ali Gholamiyan Moghadam, Ali Aghadoukht, Khalil Hajiasgharzadeh, et al. Overcoming trastuzumab resistance in her2-positive breast cancer using combination therapy. *Journal of Cellular Physiology*, 235(4):3142–3156, 2020.

[52] Rita Nahta, Mien-Chie Hung, and Francisco J Esteva. The her-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells. *Cancer research*, 64(7):2343–2346, 2004.

[53] Zhi-qiao Xu, Yan Zhang, Ning Li, Pei-jie Liu, Ling Gao, Xin Gao, and Xiao-jing Tie. Efficacy and safety of lapatinib and trastuzumab for her2-positive breast cancer: a systematic review and meta-analysis of randomised controlled trials. *BMJ open*, 7(3):e013053, 2017.

[54] Zeynep Eroglu and Antoni Ribas. Combination therapy with braf and mek inhibitors for melanoma: latest evidence and place in therapy. *Therapeutic advances in medical oncology*, 8(1):48–56, 2016.

[55] National Comprehensive Cancer Network et al. Clinical practice guidelines in oncology: melanoma, 2017.

[56] Chia-Hung Chen, Te-Chun Hsia, Ming-Hsin Yeh, Tsung-Wei Chen, Yun-Ju Chen, Jung-Tsu Chen, Ya-Ling Wei, Chih-Yen Tu, and Wei-Chien Huang. Mek inhibitors induce akt activation and drug resistance by suppressing negative feedback erk-mediated her 2 phosphorylation at thr701. *Molecular oncology*, 11(9):1273–1287, 2017.

[57] Pietro Paolo Vitiello, Claudia Cardone, Giulia Martini, Davide Ciardiello, Valentina Belli, Nunzia Matrone, Giusi Barra, Stefania Napolitano, Carmina Della Corte, Mimmo Turano, et al. Receptor tyrosine kinase-dependent pi3k activation is an escape mechanism to vertical suppression of the egfr/ras/mapk pathway in kras-mutated human colorectal cancer cell lines. *Journal of Experimental & Clinical Cancer Research*, 38(1):1–12, 2019.

[58] EJ Haagensen, S Kyle, GS Beale, RJ Maxwell, and DR Newell. The synergistic interaction of mek and pi3k inhibitors is modulated by mtor inhibition. *British journal of cancer*, 106(8):1386–1394, 2012.

[59] F Coussy, R El Botty, M Lavigne, C Gu, L Fuhrmann, A Briaux, L de Koning, A Dahmani, E Montaudon, L Morisset, et al. Combination of pi3k and mek inhibitors yields durable remission in pdx models of pik3ca-mutated metaplastic breast cancers. *Journal of hematology & oncology*, 13(1):1–10, 2020.

[60] E Kun, YTM Tsang, CW Ng, DM Gershenson, and KK Wong. Mek inhibitor resistance mechanisms and recent developments in combination trials. *Cancer treatment reviews*, 92:102137, 2021.

[61] Timothée Olivier, Alyson Haslam, and Vinay Prasad. Anticancer drugs approved by the us food and drug administration from 2009 to 2020 according to their mechanism of action. *JAMA network open*, 4(12):e2138793–e2138793, 2021.

[62] Mathew J Garnett, Elena J Edelman, Sonja J Heidorn, Chris D Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, I Richard Thompson, Xi Luo, Jorge Soares, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, 2012.

[63] Aviad Tsherniak, Francisca Vazquez, Phil G Montgomery, Barbara A Weir, Gregory Kryukov, Glenn S Cowley, Stanley Gill, William F Harrington, Sasha Pantel, John M Krill-Burger, et al. Defining a cancer dependency map. *Cell*, 170(3):564–576, 2017.

[64] Fadi J Najm, Christine Strand, Katherine F Donovan, Mudra Hegde, Kendall R Sanson, Emma W Vaimberg, Meagan E Sullender, Ella Hartenian, Zohra Kalani, Nicolo Fusi, et al. Orthologous crispr–cas9 enzymes for combinatorial genetic screens. *Nature biotechnology*, 36(2):179–189, 2018.

[65] Guo Wei, Adam A Margolin, Leila Haery, Emily Brown, Lisa Cucolo, Bina Julian, Shyemaa Shehata, Andrew L Kung, Rameen Beroukhim, and Todd R Golub. Chemical genomics identifies small-molecule mcl1 repressors and bcl-xl as a predictor of mcl1 dependency. *Cancer cell*, 21(4):547–562, 2012.

[66] Justin Lamb, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N Ross, et al.

The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *science*, 313(5795):1929–1935, 2006.

[67] Mario Niepel, Marc Hafner, Qiaonan Duan, Zichen Wang, Evan O Paull, Mirra Chung, Xiaodong Lu, Joshua M Stuart, Todd R Golub, Aravind Subramanian, et al. Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling. *Nature communications*, 8(1):1–11, 2017.

[68] Amar Koleti, Raymond Terryn, Vasileios Stathias, Caty Chung, Daniel J Cooper, John P Turner, Dušica Vidović, Michele Forlin, Tanya T Kelley, Alessandro D'Urso, et al. Data portal for the library of integrated network-based cellular signatures (lincs) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic acids research*, 46(D1):D558–D566, 2018.

[69] Wei Zhao, Jun Li, Mei-Ju M Chen, Yikai Luo, Zhenlin Ju, Nicole K Nesser, Katie Johnson-Camacho, Christopher T Boniface, Yancey Lawrence, Nupur T Pande, et al. Large-scale characterization of drug responses of clinically relevant proteins in cancer cell lines. *Cancer Cell*, 38(6):829–843, 2020.

[70] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, 2012.

[71] Amrita Basu, Nicole E Bodycombe, Jaime H Cheah, Edmund V Price, Ke Liu, Giannina I Schaefer, Richard Y Ebright, Michelle L Stewart, Daisuke Ito, Stephanie Wang, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, 154(5):1151–1161, 2013.

[72] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A Smith, I Richard Thompson, et al. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(D1):D955–D961, 2012.

[73] Augustin Luna, Fathi Elloumi, Sudhir Varma, Yanghsin Wang, Vinodh N Rajapakse, Mirit I Aladjem, Jacques Robert, Chris Sander, Yves Pommier, and William C Reinhold. Cellminer cross-database (cellminercdb) version 1.2: Exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Research*, 49(D1):D1083–D1093, 2021.

[74] S Loewe. The quantitation problem of pharmacology [in german]. *Ergebn Physiol*, 27:47–187, 1928.

[75] Chester I Bliss. The toxicity of poisons applied jointly 1. *Annals of applied biology*, 26(3):585–615, 1939.

[76] Ting-Chao Chou and Paul Talalay. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Advances in enzyme regulation*, 22:27–55, 1984.

[77] Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, and Jing Tang. Searching for drug synergy in complex dose–response landscapes using an interaction potency model. *Computational and structural biotechnology journal*, 13:504–513, 2015.

[78] Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *cell*, 167(7):1853–1866, 2016.

[79] Thomas M Norman, Max A Horlbeck, Joseph M Replogle, Alex Y Ge, Albert Xu, Marco Jost, Luke A Gilbert, and Jonathan S Weissman. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*, 365(6455):786–793, 2019.

[80] Ravi S Narayan, Piet Molenaar, Jian Teng, Fleur MG Cornelissen, Irene Roelofs, Renee Menezes, Rogier Dik, Tonny Lagerweij, Yoran Broersma, Naomi Petersen, et al. A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nature communications*, 11(1):1–14, 2020.

[81] Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, et al. The reactome pathway knowledgebase 2022. *Nucleic acids research*, 50(D1):D687–D692, 2022.

[82] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. Biogrid: a general repository for interaction datasets. *Nucleic acids research*, 34(suppl_1):D535–D539, 2006.

[83] Minoru Kanehisa, Miho Furumichi, Yoko Sato, Mari Ishiguro-Watanabe, and Mao Tanabe. Kegg: integrating viruses and cellular organisms. *Nucleic acids research*, 49(D1):D545–D551, 2021.

[84] Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, et al. The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, 49(D1):D605–D612, 2021.

[85] Igor Rodchenkov, Ozgun Babur, Augustin Luna, Bulent Arman Aksoy, Jeffrey V Wong, Dylan Fong, Max Franz, Metin Can Siper, Manfred Cheung, Michael Wrana, et al. Pathway commons 2019 update: integration, analysis and exploration of pathway data. *Nucleic acids research*, 48(D1):D489–D497, 2020.

[86] Scott L Carter, Christian M Brechbühler, Michael Griffin, and Andrew T Bond. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics*, 20(14):2242–2250, 2004.

[87] Kai Wang, Masumichi Saito, Brygida C Bisikirska, Mariano J Alvarez, Wei Keat Lim, Presha Rajbhandari, Qiong Shen, Ilya Nemenman, Katia Basso, Adam A Margolin, et al. Genome-wide identification of post-translational modulators of transcription factor activity in human b cells. *Nature biotechnology*, 27(9):829–837, 2009.

[88] Özgün Babur, Emek Demir, Mithat Gönen, Chris Sander, and Ugur Dogrusoz. Discovering modulators of gene expression. *Nucleic acids research*, 38(17):5648–5656, 2010.

[89] Timothy R Lezon, Jayanth R Banavar, Marek Cieplak, Amos Maritan, and Nina V Fedoroff. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proceedings of the National Academy of Sciences*, 103(50):19033–19038, 2006.

[90] Jason W Locasale and Alejandro Wolf-Yadlin. Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data. *PloS one*, 4(8):e6522, 2009.

[91] Patrick E Meyer, Frederic Lafitte, and Gianluca Bontempi. minet: Ar/bioconductor package for inferring large transcriptional networks using mutual information. *BMC bioinformatics*, 9(1):1–10, 2008.

[92] Thalia E Chan, Michael PH Stumpf, and Ann C Babtie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell systems*, 5(3):251–267, 2017.

[93] Yasin Şenbabaoğlu, Selçuk Onur Sümer, Francisco Sanchez-Vega, Debra Bemis, Giovanni Ciriello, Nikolaus Schultz, and Chris Sander. A multi-method approach for proteomic network inference in 11 human cancers. *PLoS computational biology*, 12(2):e1004765, 2016.

[94] Song Yi, Shengda Lin, Yongsheng Li, Wei Zhao, Gordon B Mills, and Nidhi Sahni. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nature Reviews Genetics*, 18(7):395–410, 2017.

[95] Ozgur E Akman, Steven Watterson, Andrew Parton, Nigel Binns, Andrew J Millar, and Peter Ghazal. Digital clocks: simple boolean models can quantitatively describe circadian systems. *Journal of The Royal Society Interface*, 9(74):2365–2382, 2012.

[96] Bree B Aldridge, Julio Saez-Rodriguez, Jeremy L Muhlich, Peter K Sorger, and Douglas A Lauffenburger. Fuzzy logic analysis of kinase pathway crosstalk in tnf/egf/insulin-induced signaling. *PLoS computational biology*, 5(4):e1000340, 2009.

[97] Michelle L Wynn, Nikita Consul, Sofia D Merajver, and Santiago Schnell. Logic-based models in systems biology: a predictive and parameter-free network analysis method. *Integrative biology*, 4(11):1323–1337, 2012.

[98] Chengwei Su, Angeline Andrew, Margaret R Karagas, and Mark E Borsuk. Using bayesian networks to discover relations between genes, environment, and disease. *BioData mining*, 6(1):1–21, 2013.

[99] Min Zou and Suzanne D Conzen. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71–79, 2005.

[100] Steven M Hill, Nicole K Nesser, Katie Johnson-Camacho, Mara Jeffress, Aimee Johnson, Chris Boniface, Simon EF Spencer, Yiling Lu, Laura M Heiser, Yancey Lawrence, et al. Context specificity in causal signaling networks revealed by phosphoprotein profiling. *Cell systems*, 4(1):73–83, 2017.

[101] Han Yan, Bo Zhang, Shao Li, and Qianchuan Zhao. A formal model for analyzing drug combination effects and its application in tnf-α-induced nfκb pathway. *BMC systems biology*, 4(1):1–12, 2010.

[102] Sergio Iadevaia, Yiling Lu, Fabiana C Morales, Gordon B Mills, and Prahlad T Ram. Identification of optimal drug combinations targeting cellular networks: integrating phosphoproteomics and computational network analysis. *Cancer research*, 70(17):6704–6714, 2010.

[103] Fabian Fröhlich, Thomas Kessler, Daniel Weindl, Alexey Shadrin, Leonard Schmiester, Hendrik Hache, Artur Muradyan, Moritz Schütte, Ji-Hyun Lim, Matthias Heinig, et al. Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. *Cell systems*, 7(6):567–579, 2018.

[104] Anil Korkut, Weiqing Wang, Emek Demir, Bülent Arman Aksoy, Xiaohong Jing, Evan J Molinelli, Özgün Babur, Debra L Bemis, Selcuk Onur Sumer, David B Solit, et al. Perturbation biology nominates upstream–downstream drug combinations in raf inhibitor resistant melanoma cells. *Elife*, 4:e04640, 2015.

[105] Elin Nyman, Richard R Stein, Xiaohong Jing, Weiqing Wang, Benjamin Marks, Ioannis K Zervantonakis, Anil Korkut, Nicholas P Gauthier, and Chris Sander. Perturbation biology links temporal protein changes to drug responses in a melanoma cell line. *PLoS computational biology*, 16(7):e1007909, 2020.

[106] Frank J Bruggeman, Hans V Westerhoff, Jan B Hoek, and Boris N Kholodenko. Modular response analysis of cellular regulatory networks. *Journal of theoretical biology*, 218(4):507–520, 2002.

[107] Sven Nelander, Weiqing Wang, Björn Nilsson, Qing-Bai She, Christine Pratilas, Neal Rosen, Peter Gennemark, and Chris Sander. Models from experiments: combinatorial drug perturbations of cancer cells. *Molecular systems biology*, 4(1):216, 2008.

[108] S Hug, A Raue, J Hasenauer, J Bachmann, U Klingmüller, Jens Timmer, and FJ Theis. High-dimensional bayesian parameter estimation: Case study for a model of jak2/stat5 signaling. *Mathematical biosciences*, 246(2):293–304, 2013.

[109] Bertram Klinger, Anja Sieber, Raphaela Fritsche-Guenther, Franziska Witzel, Leanne Berry, Dirk Schumacher, Yibing Yan, Pawel Durek, Mark Merchant, Reinhold Schäfer, et al. Network quantification of egfr signaling unveils potential for targeted combination therapy. *Molecular systems biology*, 9(1):673, 2013.

[110] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2424–2433, 2016.

[111] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[112] Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature methods*, 12(10):931–934, 2015.

[113] Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

[114] Haitham A Elmarakeby, Justin Hwang, Rand Arafeh, Jett Crowdis, Sydney Gang, David Liu, Saud H AlDubayan, Keyan Salari, Steven Kregel, Camden Richter, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880):348–352, 2021.

[115] Brent M Kuenzi, Jisoo Park, Samson H Fong, Kyle S Sanchez, John Lee, Jason F Kreisberg, Jianzhu Ma, and Trey Ideker. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer cell*, 38(5):672–684, 2020.

[116] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[117] Bo Yuan, Ciyue Shen, Augustin Luna, Anil Korkut, Debora S Marks, John Ingraham, and Chris Sander. Cellbox: interpretable machine learning for perturbation biology with application to the design of cancer combination therapy. *Cell systems*, 12(2):128–140, 2021.

[118] Evan J Molinelli, Anil Korkut, Weiqing Wang, Martin L Miller, Nicholas P Gauthier, Xiaohong Jing, Poorvi Kaushik, Qin He, Gordon Mills, David B Solit, et al. Perturbation biology: inferring signaling networks in cellular systems. *PLoS computational biology*, 9(12):e1003290, 2013.

[119] Endre Süli and David F Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003.

[120] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[121] Ethan G Cerami, Benjamin E Gross, Emek Demir, Igor Rodchenkov, Özgün Babur, Nadia Anwar, Nikolaus Schultz, Gary D Bader, and Chris Sander. Pathway commons, a web resource for biological pathway data. *Nucleic acids research*, 39(suppl_1):D685–D690, 2010.

[122] Sarat Chandarlapaty, Ayana Sawai, Maurizio Scaltriti, Vanessa Rodrik-Outmezguine, Olivera Grbovic-Huezo, Violeta Serra, Pradip K Majumder, Jose Baselga, and Neal Rosen. Akt inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer cell*, 19(1):58–71, 2011.

[123] Jeffrey CH Donovan, Andrea Milic, and Joyce M Slingerland. Constitutive mek/mapk activation leads to p27kip1deregulation and antiestrogen resistance in human breast cancer cells. *Journal of Biological Chemistry*, 276(44):40888–40895, 2001.

[124] Luciana H Osaki and Patricia Gama. Mapk signaling pathway regulates p27 phosphorylation at threonin 187 as part of the mechanism triggered by early-weaning to induce cell proliferation in rat gastric mucosa. *PloS one*, 8(6):e66651, 2013.

[125] Suzanne Carreira, Jane Goodall, Isil Aksan, S Anna La Rocca, Marie-Dominique Galibert, Laurence Denat, Lionel Larue, and Colin R Goding. Mitf cooperates with rb1 and activates p21cip1 expression to regulate cell cycle progression. *Nature*, 433(7027):764–769, 2005.

[126] Wanli Lei, Fan Liu, and Scott A Ness. Positive and negative regulation of c-myb by cyclin d1, cyclin-dependent kinases, and p27 kip1. *Blood*, 105(10):3855–3861, 2005.

[127] Shashi Gupta and Roger J Davis. Map kinase binds to the nh2-terminal activation domain of c-myc. *FEBS letters*, 353(3):281–285, 1994.

[128] Elizabeth R Butch and Kun-Liang Guan. Characterization of erk1 activation site mutants and the effect on recognition by mek1 and mek2 (□). *Journal of Biological Chemistry*, 271(8):4230–4235, 1996.

[129] Rosalie Sears, Faison Nuckolls, Eric Haura, Yoichi Taya, Katsuyuki Tamai, and Joseph R Nevins. Multiple ras-dependent phosphorylation pathways regulate myc protein stability. *Genes & development*, 14(19):2501–2514, 2000.

[130] Kazuhiro Aoki, Masashi Yamada, Katsuyuki Kunida, Shuhei Yasuda, and Michiyuki Matsuda. Processive phosphorylation of erk map kinase in mammalian cells. *Proceedings of the National Academy of Sciences*, 108(31):12675–12680, 2011.

[131] Christian Von Mering, Lars J Jensen, Berend Snel, Sean D Hooper, Markus Krupp, Mathilde Foglierini, Nelly Jouffre, Martijn A Huynen, and Peer Bork. String: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic acids research*, 33(suppl_1):D433–D437, 2005.

[132] Donna Maglott, Jim Ostell, Kim D Pruitt, and Tatiana Tatusova. Entrez gene: gene-centered information at ncbi. *Nucleic acids research*, 39(suppl_1):D52–D57, 2010.

[133] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[134] ZC Lipton. The mythos of model interpretability. arxiv 2016. *arXiv preprint arXiv:1606.03490*, 2019.

[135] Eva Thelisson. Towards trust, transparency and liability in ai/as systems. In *IJCAI*, pages 5215–5216, 2017.

[136] David Gunning and David Aha. Darpa's explainable artificial intelligence (xai) program. *AI magazine*, 40(2):44–58, 2019.

[137] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of Marchine Learning Research*, 18:1–43, 2018.

[138] Robert Brown, Edward Curry, Luca Magnani, Charlotte S Wilhelm-Benartzi, and Jane Borley. Poised epigenetic states and acquired drug resistance in cancer. *Nature Reviews Cancer*, 14(11):747–753, 2014.

[139] Esther A Zaal and Celia R Berkers. The influence of metabolism on drug response in cancer. *Frontiers in oncology*, page 500, 2018.

[140] Britt Adamson, Thomas M Norman, Marco Jost, Min Y Cho, James K Nuñez, Yuwen Chen, Jacqueline E Villalta, Luke A Gilbert, Max A Horlbeck, Marco Y Hein, et al. A multiplexed single-cell crispr screening platform enables systematic dissection of the unfolded protein response. *Cell*, 167(7):1867–1882, 2016.

[141] Andreas P Frei, Felice-Alessio Bava, Eli R Zunder, Elena WY Hsieh, Shih-Yu Chen, Garry P Nolan, and Pier Federico Gherardini. Highly multiplexed simultaneous detection of rnas and proteins in single cells. *Nature methods*, 13(3):269–275, 2016.

[142] Aleksandra Wroblewska, Maxime Dhainaut, Benjamin Ben-Zvi, Samuel A Rose, Eun Sook Park, El-Ad David Amir, Anela Bektesevic, Alessia Baccarini, Miriam Merad, Adeeb H Rahman, et al. Protein barcodes enable high-dimensional single-cell crispr screens. *Cell*, 175(4):1141–1155, 2018.

[143] Eleni P Mimitou, Anthony Cheng, Antonino Montalbano, Stephanie Hao, Marlon Stoeckius, Mateusz Legut, Timothy Roush, Alberto Herrera, Efthymia Papalexi, Zhengqing Ouyang, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and crispr perturbations in single cells. *Nature methods*, 16(5):409–412, 2019.

[144] Daniel Schraivogel, Andreas R Gschwind, Jennifer H Milbank, Daniel R Leonce, Petra Jakob, Lukas Mathur, Jan O Korbel, Christoph A Merten, Lars Velten, and Lars M Steinmetz. Targeted perturb-seq enables genome-scale genetic screens in single cells. *Nature methods*, 17(6):629–635, 2020.

[145] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pages 437–478. Springer, 2012.

[146] Jianglin Liang and Ruifang Liu. Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. In *2015 8th international congress on image and signal processing (CISP)*, pages 697–701. IEEE, 2015.

[147] Paul Stapor, Leonard Schmiester, Christoph Wierling, Simon Merkt, Dilan Pathirana, Bodo MH Lange, Daniel Weindl, and Jan Hasenauer. Mini-batch optimization enables training of ode models on large-scale datasets. *Nature communications*, 13(1):1–17, 2022.

[148] Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and TM Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, 17(2):147–154, 2020.

[149] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

[150] Gökcen Eraslan, Lukas M Simon, Maria Mircea, Nikola S Mueller, and Fabian J Theis. Single-cell rna-seq denoising using a deep count autoencoder. *Nature communications*, 10(1):1–14, 2019.

[151] Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scgen predicts single-cell perturbation responses. *Nature methods*, 16(8):715–721, 2019.

[152] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[153] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

[154] Leonard Schmiester, Yannik Schälte, Fabian Fröhlich, Jan Hasenauer, and Daniel Weindl. Efficient parameterization of large-scale dynamic models based on relative measurements. *Bioinformatics*, 36(2):594–602, 2020.

[155] Fabian Fröhlich, Barbara Kaltenbacher, Fabian J Theis, and Jan Hasenauer. Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS computational biology*, 13(1):e1005331, 2017.

[156] Raoul Tibes, YiHua Qiu, Yiling Lu, Bryan Hennessy, Michael Andreeff, Gordon B Mills, and Steven M Kornblau. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular cancer therapeutics*, 5(10):2512–2521, 2006.

[157] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[158] Philipp Städter, Yannik Schälte, Leonard Schmiester, Jan Hasenauer, and Paul L Stapor. Benchmarking of numerical integration methods for ode models of biological systems. *Scientific reports*, 11(1):1–11, 2021.

[159] Augustin Luna, Özgün Babur, Bülent Arman Aksoy, Emek Demir, and Chris Sander. Paxtoolsr: pathway analysis in r using pathway commons. *Bioinformatics*, 32(8):1262–1264, 2016.

[160] Joseph M Replogle, Reuben A Saunders, Angela N Pogson, Jeffrey A Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J Wagner, Karen Adelman, Jessica L Bonnar, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *bioRxiv*, 2021.

[161] Laralynne Przybyla and Luke A Gilbert. A new era in functional genomics screens. *Nature Reviews Genetics*, pages 1–15, 2021.

[162] Alexandre F Aissa, Abul BMMK Islam, Majd M Ariss, Cammille C Go, Alexandra E Rader, Ryan D Conrardy, Alexa M Gajda, Carlota Rubio-Perez, Klara Valyi-Nagy, Mary Pasquinelli, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nature communications*, 12(1):1–25, 2021.

[163] Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. Pooled crispr screening with single-cell transcriptome readout. *Nature methods*, 14(3):297–301, 2017.

[164] Chris J Frangieh, Johannes C Melms, Pratiksha I Thakore, Kathryn R Geiger-Schuller, Patricia Ho, Adrienne M Luoma, Brian Cleary, Livnat Jerby-Arnon, Shruti Malu, Michael S Cuoco, et al. Multimodal pooled perturb-cite-seq screens in patient models define mechanisms of cancer immune evasion. *Nature genetics*, 53(3):332–341, 2021.

[165] Molly Gasperini, Andrew J Hill, José L McFaline-Figueroa, Beth Martin, Seungsoo Kim, Melissa D Zhang, Dana Jackson, Anh Leith, Jacob Schreiber, William S Noble, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*, 176(1-2):377–390, 2019.

[166] Jase Gehring, Jong Hwee Park, Sisi Chen, Matthew Thomson, and Lior Pachter. Highly multiplexed single-cell rna-seq by dna oligonucleotide tagging of cellular proteins. *Nature Biotechnology*, 38(1):35–38, 2020.

[167] Diego Adhemar Jaitin, Assaf Weiner, Ido Yofe, David Lara-Astiaso, Hadas Keren-Shaul, Eyal David, Tomer Meir Salame, Amos Tanay, Alexander van Oudenaarden, and Ido Amit. Dissecting immune circuits by linking crispr-pooled screens with single-cell rna-seq. *Cell*, 167(7):1883–1896, 2016.

[168] James M McFarland, Brenton R Paolella, Allison Warren, Kathryn Geiger-Schuller, Tsukasa Shibue, Michael Rothberg, Olena Kuksenko, William N Colgan, Andrew Jones, Emily Chambers, et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nature communications*, 11(1):1–15, 2020.

[169] Yaara Oren, Michael Tsabar, Michael S Cuoco, Liat Amir-Zilberstein, Heidie F Cabanos, Jan-Christian Hütter, Bomiao Hu, Pratiksha I Thakore, Marcin Tabaka, Charles P Fulco, et al. Cycling cancer persister cells arise from lineages with distinct programs. *Nature*, 596(7873):576–582, 2021.

[170] Efthymia Papalexi, Eleni P Mimitou, Andrew W Butler, Samantha Foster, Bernadette Bracken, William M Mauck, Hans-Hermann Wessels, Yuhan Hao, Bertrand Z Yeung, Peter Smibert, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nature genetics*, 53(3):322–331, 2021.

[171] Sarah E Pierce, Jeffrey M Granja, and William J Greenleaf. High-throughput single-cell chromatin accessibility crispr screens enable unbiased identification of regulatory networks in cancer. *Nature communications*, 12(1):1–8, 2021.

[172] Adam J Rubin, Kevin R Parker, Ansuman T Satpathy, Yanyan Qi, Beijing Wu, Alvin J Ong, Maxwell R Mumbach, Andrew L Ji, Daniel S Kim, Seung Woo Cho, et al. Coupled single-cell crispr screening and epigenomic profiling reveals causal gene regulatory networks. *Cell*, 176(1-2):361–376, 2019.

[173] Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

[174] Eric Shifrut, Julia Carnevale, Victoria Tobin, Theodore L Roth, Jonathan M Woo, Christina T Bui, P Jonathan Li, Morgan E Diolaiti, Alan Ashworth, and Alexander Marson. Genome-wide crispr screens in primary human t cells reveal key regulators of immune function. *Cell*, 175(7):1958–1971, 2018.

[175] Sanjay R Srivatsan, José L McFaline-Figueroa, Vijay Ramani, Lauren Saunders, Junyue Cao, Jonathan Packer, Hannah A Pliner, Dana L Jackson, Riza M Daza, Lena Christiansen, et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science*, 367(6473):45–51, 2020.

[176] Ruilin Tian, Mariam A Gachechiladze, Connor H Ludwig, Matthew T Laurie, Jason Y Hong, Diane Nathaniel, Anika V Prabhu, Michael S Fernandopulle, Rajan Patel, Mehrnoosh Abshari, et al. Crispr interference-based platform for multimodal genetic screens in human ipsc-derived neurons. *Neuron*, 104(2):239–255, 2019.

[177] Ruilin Tian, Anthony Abarientos, Jason Hong, Sayed Hadi Hashemi, Rui Yan, Nina Dräger, Kun Leng, Mike A Nalls, Andrew B Singleton, Ke Xu, et al. Genome-wide crispri/a screens in human neurons link lysosomal failure to ferroptosis. *Nature Neuroscience*, 24(7):1020–1034, 2021.

[178] Caleb Weinreb, Alejo Rodriguez-Fraticelli, Fernando D Camargo, and Allon M Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479):eaaw3381, 2020.

[179] Shiqi Xie, Jialei Duan, Boxun Li, Pei Zhou, and Gary C Hon. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Molecular cell*, 66(2):285–299, 2017.

[180] Wenting Zhao, Athanassios Dovas, Eleonora Francesca Spinazzi, Hanna Mendes Levitin, Matei Alexandru Banu, Pavan Upadhyayula, Tejaswi Sudhakar, Tamara Marie, Marc L Otten, Michael B Sisti, et al. Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell rna-seq. *Genome medicine*, 13(1):1–15, 2021.

[181] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):1–5, 2018.

[182] Alexandra Franz, Fabian Coscia, Ciyue Shen, Lea Charaoui, Matthias Mann, and Chris Sander. Molecular response to parp1 inhibition in ovarian cancer cells as determined by mass spectrometry based proteomics. *Journal of ovarian research*, 14(1):1–14, 2021.

[183] Ronald FS Lee, Alexey Chernobrovkin, Dorothea Rutishauser, Claire S Allardyce, David Hacker, Kai Johnsson, Roman A Zubarev, and Paul J Dyson. Expression proteomics study to determine metallodrug targets and optimal drug combinations. *Scientific reports*, 7(1):1–11, 2017.

[184] Amir Ata Saei, Christian Michel Beusch, Alexey Chernobrovkin, Pierre Sabatier, Bo Zhang, Ülkü Güler Tokat, Eleni Stergiou, Massimiliano Gaetani, Ákos Végvári, and Roman A Zubarev. Protargetminer as a proteome signature library of anticancer molecules for functional discovery. *Nature communications*, 10(1):1–13, 2019.

[185] Benjamin Ruprecht, Julie Di Bernardo, Zhao Wang, Xuan Mo, Oleg Ursu, Matthew Christopher, Rafael B Fernandez, Li Zheng, Brian D Dill, Huijun Wang, et al. A mass spectrometry-based proteome map of drug action in lung cancer cell lines. *Nature chemical biology*, 16(10):1111–1119, 2020.

[186] Silvia Domcke, Rileen Sinha, Douglas A Levine, Chris Sander, and Nikolaus Schultz. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications*, 4(1):1–10, 2013.

[187] F Coscia, KM Watters, M Curtis, MA Eckert, CY Chiang, S Tyanova, A Montag, RR Lastra, E Lengyel, and M Mann. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nature communications*, 7(1):1–14, 2016.

[188] Jeremiah N Winter, Leonard S Jefferson, and Scot R Kimball. Erk and akt signaling pathways function through parallel mechanisms to promote mtorc1 signaling. *American Journal of Physiology-Cell Physiology*, 300(5):C1172–C1180, 2011.

[189] Alexa B Turke, Youngchul Song, Carlotta Costa, Rebecca Cook, Carlos L Arteaga, John M Asara, and Jeffrey A Engelman. Mek inhibition leads to pi3k/akt activation by relieving a negative feedback on erbb receptors. *Cancer research*, 72(13):3228–3237, 2012.

[190] Qinlian Jiao, Lei Bi, Yidan Ren, Shuliang Song, Qin Wang, and Yun-shan Wang. Advances in studies of tyrosine kinase inhibitors and their acquired resistance. *Molecular cancer*, 17(1):1–12, 2018.

[191] Nigel Goode, Kenneth Hughes, JR Woodgett, and PJ Parker. Differential regulation of glycogen synthase kinase-3 beta by protein kinase c isotypes. *Journal of Biological Chemistry*, 267(24):16878–16882, 1992.

[192] Erica L Mayer and Ian E Krop. Advances in targeting src in the treatment of breast cancer and other solid malignancies. *Clinical Cancer Research*, 16(14):3526–3532, 2010.

[193] Genevieve Housman, Shannon Byler, Sarah Heerboth, Karolina Lapinska, Mckenna Longacre, Nicole Snyder, and Sibaji Sarkar. Drug resistance in cancer: an overview. *Cancers*, 6(3):1769–1792, 2014.

[194] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005.

[195] Rimsha Munir, Jan Lisec, Johannes V Swinnen, and Nousheen Zaidi. Lipid metabolism in cancer cells under metabolic stress. *British journal of cancer*, 120(12):1090–1098, 2019.

[196] Qi Wu, Hanpu Zhang, Si Sun, Lijun Wang, and Shengrong Sun. Extracellular vesicles and immunogenic stress in cancer. *Cell Death & Disease*, 12(10):1–12, 2021.

[197] Ethan Cerami, Emek Demir, Nikolaus Schultz, Barry S Taylor, and Chris Sander. Automated network analysis identifies core pathways in glioblastoma. *PloS one*, 5(2):e8918, 2010.

[198] Eric Minwei Liu, Augustin Luna, Guanlan Dong, and Chris Sander. netboxr: Automated discovery of biological process modules by network analysis in r. *PloS one*, 15(11):e0234669, 2020.

[199] Brendan D Manning and Alex Toker. Akt/pkb signaling: navigating the network. *Cell*, 169(3):381–405, 2017.

[200] George Mihai Nitulescu, Maryna Van De Venter, Georgiana Nitulescu, Anca Ungurianu, Petras Juzenas, Qian Peng, Octavian Tudorel Olaru, Daniela Grădinaru, Aristides Tsatsakis, Dimitris Tsoukalas, et al. The akt pathway in oncology therapy and beyond. *International Journal of Oncology*, 53(6):2319–2331, 2018.

[201] Masuko Katoh and Masaru Katoh. Precision medicine for human cancers with notch signaling dysregulation. *International journal of molecular medicine*, 45(2):279–297, 2020.

[202] Alfredo Giménez-Cassina and Nika N Danial. Regulation of mitochondrial nutrient and energy metabolism by bcl-2 family proteins. *Trends in Endocrinology & Metabolism*, 26(4):165–175, 2015.

[203] Ayse Hande Nayman, Halime Siginc, Ebru Zemheri, Faruk Yencilek, Asif Yildirim, and Dilek Telci. Dual-inhibition of mtor and bcl-2 enhances the anti-tumor effect of everolimus against renal cell carcinoma in vitro and in vivo. *Journal of Cancer*, 10(6):1466, 2019.

[204] Leisl M Packer, Samantha J Stehbens, Vanessa F Bonazzi, Jennifer H Gunter, Robert J Ju, Micheal Ward, Michael G Gartside, Sara A Byron, and Pamela M Pollock. Bcl-2 inhibitors enhance fgfr inhibitor-induced mitochondrial-dependent cell death in fgfr2-mutant endometrial cancer. *Molecular oncology*, 13(4):738–756, 2019.

[205] Atsuhiro Adachi, Fumi Kano, Takashi Tsuboi, Morihisa Fujita, Yusuke Maeda, and Masayuki Murata. Golgi-associated gsk3β regulates the sorting process of post-golgi membrane trafficking. *Journal of cell science*, 123(19):3215–3225, 2010.

[206] Jieqiong Huang, Xinyue Guo, Weihong Li, and Haiyan Zhang. Activation of wnt/β-catenin signalling via gsk3 inhibitors direct differentiation of human adipose stem cells into functional hepatocytes. *Scientific reports*, 7(1):1–12, 2017.

[207] Erika S Dahl, Raquel Buj, Kelly E Leon, Jordan M Newell, Yuka Imamura, Benjamin G Bitler, Nathaniel W Snyder, and Katherine M Aird. Targeting idh1 as a prosenescent therapy in high-grade serous ovarian cancer. *Molecular Cancer Research*, 17(8):1710–1720, 2019.

[208] Lauren V Albrecht, Nydia Tejeda-Muñoz, Maggie H Bui, Andrew C Cicchetto, Daniele Di Biagio, Gabriele Colozza, Ernst Schmid, Stefano Piccolo, Heather R Christofk, and Edward M De Robertis. Gsk3 inhibits macropinocytosis and lysosomal activity through the wnt destruction complex machinery. *Cell reports*, 32(4):107973, 2020.

[209] Shawn M Davidson and Matthew G Vander Heiden. Critical functions of the lysosome in cancer biology. *Annual review of pharmacology and toxicology*, 57:481–507, 2017.

[210] Yan-Jun Guo, Wei-Wei Pan, Sheng-Bing Liu, Zhong-Fei Shen, Ying Xu, and Ling-Ling Hu. Erk/mapk signalling pathway and tumorigenesis. *Experimental and therapeutic medicine*, 19(3):1997–2007, 2020.

[211] Alessandra Dall'Acqua, Michele Bartoletti, Nastaran Masoudi-Khoram, Roberto Sorio, Fabio Puglisi, Barbara Belletti, and Gustavo Baldassarre. Inhibition of cdk4/6 as therapeutic approach for ovarian cancer patients: current evidences and future perspectives. *Cancers*, 13(12):3035, 2021.

[212] Maria Rosaria Ruocco, Angelica Avagliano, Giuseppina Granato, Elena Vigliar, Stefania Masone, Stefania Montagnani, and Alessandro Arcucci. Metabolic flexibility in melanoma: A potential therapeutic target. In *Seminars in Cancer Biology*, volume 59, pages 187–207. Elsevier, 2019.

[213] Mai Q Nguyen, Jessica LF Teh, Timothy J Purwin, Inna Chervoneva, Michael A Davies, Katherine L Nathanson, Phil F Cheng, Mitchell P Levesque, Reinhard Dummer, and Andrew E Aplin. Targeting phgdh upregulation reduces glutathione levels and resensitizes resistant nras-mutant melanoma to mapk kinase inhibition. *Journal of Investigative Dermatology*, 140(11):2242–2252, 2020.

[214] Rain R Chen, Mingo MH Yung, Yang Xuan, Shijie Zhan, Leanne L Leung, Rachel R Liang, Thomas HY Leung, Huijuan Yang, Dakang Xu, Rakesh Sharma, et al. Targeting of lipid metabolism with a metabolic inhibitor cocktail eradicates peritoneal metastases in ovarian cancer cells. *Communications biology*, 2(1):1–15, 2019.

[215] Adrian R Black and Jennifer D Black. Protein kinase c signaling and cell cycle regulation. *Frontiers in immunology*, 3:423, 2013.

[216] Thomas C Beadnell, Kelsey W Nassar, Madison M Rose, Erin G Clark, Brian P Danysh, Marie-Claude Hofmann, Nikita Pozdeyev, and Rebecca E Schweppe. Src-mediated regulation of the pi3k pathway in advanced papillary and anaplastic thyroid cancer. *Oncogenesis*, 7(2):1–14, 2018.

[217] Arden Perkins, Kimberly J Nelson, Derek Parsonage, Leslie B Poole, and P Andrew Karplus. Peroxiredoxins: guardians against oxidative stress and modulators of peroxide signaling. *Trends in biochemical sciences*, 40(8):435–445, 2015.

[218] Arianna Nicolussi, Sonia D'inzeo, Carlo Capalbo, Giuseppe Giannini, and Anna Coppa. The role of peroxiredoxins in cancer. *Molecular and clinical oncology*, 6(2):139–153, 2017.

[219] Xin-yan Wang, Hai-jiao Wang, and Xiu-qin Li. Peroxiredoxin iii protein expression is associated with platinum resistance in epithelial ovarian cancer. *Tumor Biology*, 34(4):2275–2281, 2013.

[220] E. V. Kalinina, T. T. Berezov, A. A. Shtil', N. N. Chernov, V. A. Glazunova, M. D. Novichkova, and N. K. Nurmuradov. Expression of peroxiredoxin 1, 2, 3, and 6 genes in cancer cells during drug resistance formation. *Bulletin of Experimental Biology and Medicine*, 153(6):879–882, October 2012.

[221] Malgorzata Bajor, Agnieszka Graczyk-Jarzynka, Katsiaryna Marhelava, Malgorzata Kurkowiak, Arman Rahman, Claudia Aura, Niamh Russell, Agata O. Zych, Malgorzata Firczuk, Magdalena Winiarska, William M. Gallagher, and Radoslaw Zagozdzon. Triple combination of ascorbate, menadione and the inhibition of peroxiredoxin-1 produces synergistic cytotoxic effects in triple-negative breast cancer cells. *Antioxidants*, 9(4):320, April 2020.

[222] Shuyu Zheng, Wenyu Wang, Jehad Aldahdooh, Alina Malyutina, Tolou Shadbahr, Ziaurrehman Tanoli, Alberto Pessia, and Jing Tang. SynergyFinder plus: Toward better interpretation and annotation of drug combination screening datasets. *Genomics, Proteomics &amp Bioinformatics*, January 2022.

[223] Nicolas Germain, Mélanie Dhayer, Marie Boileau, Quentin Fovez, Jerome Kluza, and Philippe Marchetti. Lipid metabolism and resistance to anticancer treatment. *Biology*, 9(12):474, December 2020.

[224] Lindsay A. Broadfield, Antonino Alejandro Pane, Ali Talebi, Johannes V. Swinnen, and Sarah-Maria Fendt. Lipid metabolism in cancer: New perspectives and emerging mechanisms. *Developmental Cell*, 56(10):1363–1393, May 2021.

[225] Matteo Curtarello, Martina Tognon, Carolina Venturoli, Micol Silic-Benussi, Angela Grassi, Martina Verza, Sonia Minuzzo, Marica Pinazza, Valentina Brillo, Giovanni Tosi, Ruggero Ferrazza, Graziano Guella, Egidio Iorio, Adrien Godfroid, Nor Eddine Sounni, Alberto Amadori, and Stefano Indraccolo. Rewiring of lipid metabolism and storage in ovarian cancer cells after anti-VEGF therapy. *Cells*, 8(12):1601, December 2019.

[226] Christian D Kelstrup, Dorte B Bekker-Jensen, Tabiwang N Arrey, Alexander Hogrebe, Alexander Harder, and Jesper V Olsen. Performance evaluation of the q exactive hf-x for shotgun proteomics. *Journal of proteome research*, 17(1):727–738, 2018.

[227] Benjamin M Gyori, John A Bachman, Kartik Subramanian, Jeremy L Muhlich, Lucian Galescu, and Peter K Sorger. From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology*, 13(11):954, November 2017.

[228] Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, May 2012.

[229] John G Tate, Sally Bamford, Harry C Jubb, Zbyslaw Sondka, David M Beare, Nidhi Bindal, Harry Boutselakis, Charlotte G Cole, Celestino Creatore, Elisabeth Dawson, Peter Fish, Bhavana Harsha, Charlie Hathaway, Steve C Jupe, Chai Yin Kok, Kate Noble, Laura Ponting, Christopher C Ramshaw, Claire E Rye, Helen E Speedy, Ray Stefancsik, Sam L Thompson, Shicai Wang, Sari Ward, Peter J Campbell, and Simon A Forbes. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Research*, 47(D1):D941–D947, October 2018.

[230] Biswa Nath Datta. *Numerical linear algebra and applications*, volume 116. Siam, 2010.

[231] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

[232] Sonia Guil and Manel Esteller. Rna–rna interactions in gene regulation: the coding and noncoding players. *Trends in biochemical sciences*, 40(5):248–256, 2015.

[233] Jacqueline S Gerritsen and Forest M White. Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells. *Expert Review of Proteomics*, 18(8):661–674, 2021.

[234] Lendert Gelens and Adrian T Saurin. Exploring the function of dynamic phosphorylation-dephosphorylation cycles. *Developmental cell*, 44(6):659–663, 2018.

[235] Lynda Wyld, Ilaria Bellantuono, Tamara Tchkonia, Jenna Morgan, Olivia Turner, Fiona Foss, Jayan George, Sarah Danson, and James L Kirkland. Senescence and cancer: a review of clinical implications of senescence and senotherapies. *Cancers*, 12(8):2134, 2020.

[236] Ariadna Recasens and Lenka Munoz. Targeting cancer cell dormancy. *Trends in pharmacological sciences*, 40(2):128–141, 2019.

[237] Esther Rheinbay, Morten Muhlig Nielsen, Federico Abascal, Jeremiah A Wala, Ofer Shapira, Grace Tiao, Henrik Hornshøj, Julian M Hess, Randi Istrup Juul, Ziao Lin, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*, 578(7793):102–111, 2020.

[238] Nathanael Fillmore, Steven Bell, Ciyue Shen, Vinh Nguyen, Jennifer La, Maureen Dubreuil, Judith Strymish, Mary Brophy, Gautam Mehta, Hao Wu, et al. Disulfiram use is associated with lower risk of covid-19: A retrospective cohort study. *PloS one*, 16(10):e0259061, 2021.

[239] Olga Esakova and Andrey S Krasilnikov. Of proteins and rna: the rnase p/mrp family. *Rna*, 16(9):1725–1747, 2010.

[240] Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, and Anton I Petrov. Rfam 13.0: shifting to a genome-centric resource for non-coding rna families. *Nucleic acids research*, 46(D1):D335–D342, 2018.

[241] Debora S Marks, Lucy J Colwell, Robert Sheridan, Thomas A Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Protein 3d structure computed from evolutionary sequence variation. *PloS one*, 6(12):e28766, 2011.

[242] Caleb Weinreb, Adam J Riesselman, John B Ingraham, Torsten Gross, Chris Sander, and Debora S Marks. 3d rna and functional interactions from evolutionary couplings. *Cell*, 165(4):963–975, 2016.

[243] Elena Khanova, Olga Esakova, Anna Perederina, Igor Berezin, and Andrey S Krasilnikov. Structural organizations of yeast rnase p and rnase mrp holoenzymes as revealed by uv-crosslinking studies of rna–protein interactions. *RNA*, 18(4):720–728, 2012.

[244] Olga Esakova, Anna Perederina, Igor Berezin, and Andrey S Krasilnikov. Conserved regions of ribonucleoprotein ribonuclease mrp are involved in interactions with its substrate. *Nucleic acids research*, 41(14):7084–7091, 2013.

[245] Robert D Fagerlund, Anna Perederina, Igor Berezin, and Andrey S Krasilnikov. Footprinting analysis of interactions between the largest eukaryotic rnase p/mrp protein pop1 and rnase p/mrp rna components. *RNA*, 21(9):1591–1605, 2015.

[246] Anna Perederina, Olga Esakova, Chao Quan, Elena Khanova, and Andrey S Krasilnikov. Eukaryotic ribonucleases p/mrp: the crystal structure of the p3 domain. *The EMBO journal*, 29(4):761–769, 2010.

[247] Katharina Hipp, Kyriaki Galani, Claire Batisse, Simone Prinz, and Bettina Böttcher. Modular architecture of eukaryotic rnase p and rnase mrp revealed by electron microscopy. *Nucleic acids research*, 40(7):3275–3288, 2012.

[248] Bruno Lemieux, Nancy Laterreur, Anna Perederina, Jean-François Noël, Marie-Line Dubois, Andrey S Krasilnikov, and Raymund J Wellinger. Active yeast telomerase shares subunits with ribonucleoproteins rnase p and rnase mrp. *Cell*, 165(5):1171–1181, 2016.

[249] Yoshiko Maida, Mami Yasukawa, Miho Furuuchi, Timo Lassmann, Richard Possemato, Naoko Okamoto, Vivi Kasim, Yoshihide Hayashizaki, William C Hahn, and Kenkichi Masutomi. An rna-dependent rna polymerase formed by tert and the rmrp rna. *Nature*, 461(7261):230–235, 2009.

[250] Sudeep Pushpakom, Francesco Iorio, Patrick A Eyers, K Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, et al. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18(1):41–58, 2019.

[251] Jesse J Suh, Helen M Pettinati, Kyle M Kampman, and Charles P O'Brien. The status of disulfiram: a half of a century later. *Journal of clinical psychopharmacology*, 26(3):290–302, 2006.

[252] Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, et al. Structure of mpro from sars-cov-2 and discovery of its inhibitors. *Nature*, 582(7811):289–293, 2020.

[253] Min-Han Lin, David C Moses, Chih-Hua Hsieh, Shu-Chun Cheng, Yau-Hung Chen, Chiao-Yin Sun, and Chi-Yuan Chou. Disulfiram can inhibit mers and sars coronavirus papain-like proteases via different modes. *Antiviral research*, 150:155–163, 2018.

[254] Chunlong Ma, Yanmei Hu, Julia Alma Townsend, Panagiotis I Lagarias, Michael Thomas Marty, Antonios Kolocouris, and Jun Wang. Ebselen, disulfiram, carmofur, px-12, tideglusib, and shikonin are nonspecific promiscuous sars-cov-2 main protease inhibitors. *ACS pharmacology & translational science*, 3(6):1265–1277, 2020.

[255] Emery Smith, Meredith E Davis-Gardner, Ruben D Garcia-Ordonez, Tu-Trinh Nguyen, Mitchell Hull, Emily Chen, Pierre Baillargeon, Louis Scampavia, Timothy Strutzenberg, Patrick R Griffin, et al. <? covid19?> high-throughput screening for drugs that inhibit papain-like protease in sars-cov-2. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, 25(10):1152–1161, 2020.

[256] Karen Sargsyan, Chien-Chu Lin, Ting Chen, Cédric Grauffel, Yi-Ping Chen, Wei-Zen Yang, Hanna S Yuan, and Carmay Lim. Multi-targeting of functional cysteines in multiple conserved sars-cov-2 domains by clinically safe zn-ejectors. *Chemical science*, 11(36):9904–9909, 2020.

[257] Chih-Jung Kuo, Tai-Ling Chao, Han-Chieh Kao, Ya-Min Tsai, Yi-Kai Liu, Lily Hui-Ching Wang, Ming-Chang Hsieh, Sui-Yuan Chang, and Po-Huang Liang. Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against sars-cov-2. *Antimicrobial Agents and chemotherapy*, 65(4):e02577–20, 2021.

[258] Bara Cihlova, Andrea Huskova, Jiri Böserle, Radim Nencka, Evzen Boura, and Jan Silhan. High-throughput fluorescent assay for inhibitor screening of proteases from rna viruses. *Molecules*, 26(13):3792, 2021.

[259] Ting Chen, Cheng-Yin Fei, Yi-Ping Chen, Karen Sargsyan, Chun-Ping Chang, Hanna S Yuan, and Carmay Lim. Synergistic inhibition of sars-cov-2 replication using disulfiram/ebselen and remdesivir. *ACS pharmacology & translational science*, 4(2):898–907, 2021.

[260] Naún Lobo-Galo, Manuel Terrazas-López, Alejandro Martínez-Martínez, and Ángel Gabriel Díaz-Sánchez. Fda-approved thiol-reacting drugs that potentially bind into the sars-cov-2 main protease, essential for viral replication. *Journal of Biomolecular Structure and Dynamics*, 39(9):3419–3427, 2021.

[261] Jun Jacob Hu, Xing Liu, Shiyu Xia, Zhibin Zhang, Ying Zhang, Jingxia Zhao, Jianbin Ruan, Xuemei Luo, Xiwen Lou, Yang Bai, et al. Fda-approved disulfiram inhibits pyroptosis by blocking gasdermin d pore formation. *Nature immunology*, 21(7):736–745, 2020.

[262] Nathanael R Fillmore, Jennifer La, Raphael E Szalat, David P Tuck, Vinh Nguyen, Cenk Yildirim, Nhan V Do, Mary T Brophy, and Nikhil C Munshi. Prevalence and outcome of covid-19 infection in cancer patients: a national veterans affairs study. *JNCI: Journal of the National Cancer Institute*, 113(6):691–698, 2021.

[263] Stefano Tamburin, Elisa Mantovani, Ernesto De Bernardis, Donato Zipeto, and Fabio Lugoboni. Covid-19 and related symptoms in patients under disulfiram for alcohol use disorder. *Internal and Emergency Medicine*, 16(6):1729–1731, 2021.

[264] Ariel Israel, Alejandro A Schäffer, Assi Cicurel, Kuoyuan Cheng, Sanju Sinha, Eyal Schiff, Ilan Feldhamer, Ameer Tal, Gil Lavie, and Eytan Ruppin. Identification of drugs associated with reduced severity of covid-19–a case-control study in a large population. *Elife*, 10:e68165, 2021.

[265] Pedro N Batalha, Luana SM Forezi, Carolina GS Lima, Fernanda P Pauli, Fernanda CS Boechat, Maria Cecilia BV de Souza, Anna C Cunha, Vitor F Ferreira, and Fernando de C da Silva. Drug repurposing for the treatment of covid-19: Pharmacological aspects and synthetic approaches. *Bioorganic chemistry*, 106:104488, 2021.

[266] Isabel Sola, Fernando Almazan, Sonia Zúñiga, and Luis Enjuanes. Continuous and discontinuous rna synthesis in coronaviruses. *Annual review of virology*, 2:265–288, 2015.

[267] PS Masters and PJM Rottier. Coronavirus reverse genetics by targeted rna recombination. *Coronavirus Replication and Reverse Genetics*, pages 133–159, 2005.

[268] Dong Yang and Julian L Leibowitz. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus research*, 206:120–133, 2015.

[269] Ramakanth Madhugiri, Nadja Karl, Daniel Petersen, Kevin Lamkiewicz, Markus Fricke, Ulrike Wend, Robina Scheuer, Manja Marz, and John Ziebuhr. Structural and functional conservation of cis-acting rna elements in coronavirus 5'-terminal genome regions. *Virology*, 517:44–55, 2018.

[270] Ewan P Plant, Gabriela C Pérez-Alvarado, Jonathan L Jacobs, Bani Mukhopadhyay, Mirko Hennig, and Jonathan D Dinman. A three-stemmed mrna pseudoknot in the sars coronavirus frameshift signal. *PLoS biology*, 3(6):e172, 2005.

[271] Michael P Robertson, Haller Igel, Robert Baertsch, David Haussler, Manuel Ares Jr, and William G Scott. The structure of a rigorously conserved rna element within the sars virus genome. *PLoS biology*, 3(1):e5, 2005.

[272] Torstein Tengs and Christine M Jonassen. Distribution and evolutionary history of the mobile genetic element s2m in coronaviruses. *Diseases*, 4(3):27, 2016.

[273] David L Wheeler, Tanya Barrett, Dennis A Benson, Stephen H Bryant, Kathi Canese, Vyacheslav Chetvernin, Deanna M Church, Michael DiCuccio, Ron Edgar, Scott Federhen, et al. Database resources of the national center for biotechnology information. *Nucleic acids research*, 36(suppl_1):D13–D21, 2007.

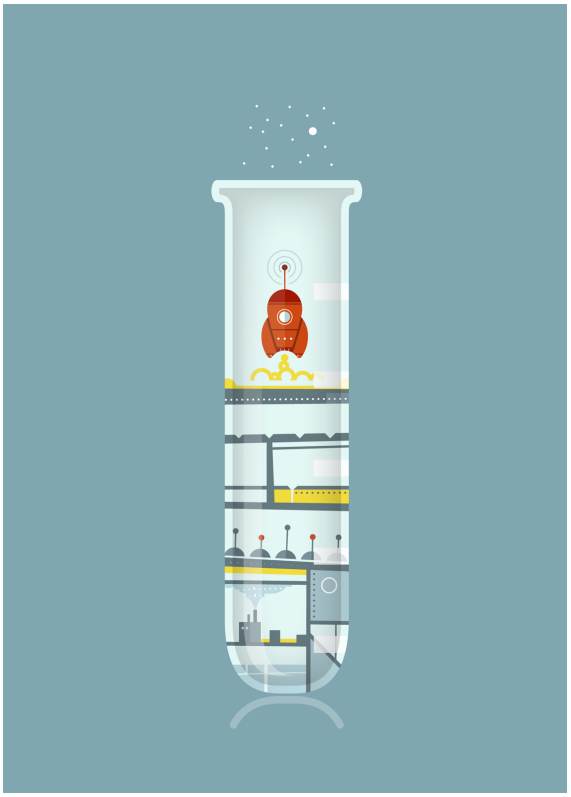
[274] Robert C Edgar. Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5(1):1–19, 2004.

[275] Roland Züst, Timothy B Miller, Scott J Goebel, Volker Thiel, and Paul S Masters. Genetic interactions between an essential 3′ cis-acting rna pseudoknot, replicase gene products, and the extreme 3′ end of the mouse coronavirus genome. *Journal of virology*, 82(3):1214–1228, 2008.

[276] Pedro A Mateos-Gomez, Lucia Morales, Sonia Zuñiga, Luis Enjuanes, and Isabel Sola. Long-distance rna-rna interactions in the coronavirus genome form high-order structures promoting discontinuous rna synthesis during transcription. *Journal of virology*, 87(1):177–186, 2013.

[277] Sebastian Will, Tejal Joshi, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Locarna-p: accurate boundary prediction and improved detection of structural rnas. *Rna*, 18(5):900–914, 2012.

[278] Shruti Khare, Céline Gurry, Lucas Freitas, Mark B Schultz, Gunter Bach, Amadou Diallo, Nancy Akite, Joses Ho, Raphael TC Lee, Winston Yeo, et al. Gisaid's role in pandemic response. *China CDC Weekly*, 3(49):1049, 2021.



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.