Faculty of Biological and Environmental Sciences
University of Helsinki
Finland

Computational tools for high-throughput drug combination screening, synergy scoring and predictive modelling in cancer

Liye He

Institute for Molecular Medicine Finland (FIMM)
University of Helsinki
and
Doctoral Program in Integrative Life Science (ILS)
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public discussion with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki on the 22nd of October, 2020 at 13 o'clock. The defence is open for audience through remote access.

Helsinki 2020

Supervisors
Prof. Tero Aittokallio, PhD
Institute for Molecular
Medicine Finland (FIMM),
University of Helsinki,
Helsinki, Finland

Asst. Prof. Jing Tang, PhD
Faculty of Medicine,
University of Helsinki,
Helsinki, Finland

Thesis Advisory Committee
Prof. Satu Mustjoki, MD, PhD
Department of Clinical Chemistry and
Hematology,
Faculty of Medicine,
University of Helsinki,
Helsinki, Finland

Henri Xhaard, PhD
Division of Pharmaceutical Chemistry and
Technology,
Faculty of Pharmacy,
University of Helsinki,
Helsinki, Finland

Thesis Reviewers
Adj. Prof. Jorrit Enserink, PhD
Department of Molecular Cell Biology,
Institute for Cancer Research,
The Norwegian Radium Hospital,
Oslo, Norway

Prof. Oliver Taboureau, PhD
Unité de Biologie Fonctionnelle &
Adaptative (BFA) - CNRS UMR 8251,
Université Paris Diderot,
Paris, France

Opponent
Pedro Ballester, PhD
Cancer Research Center of Marseille (CRCM),
Marseille, France

Custos
Prof. Kari Keinänen, PhD
Faculty of Biological and Environmental Sciences,
University of Helsinki,
Helsinki, Finland

ISBN 978-951-51-6566-4 (paperback)
ISBN 978-951-51-6567-1 (PDF)
https://ethesis.helsinki.fi
Unigrafia Oy
Helsinki 2020

”Patience is bitter, but its fruit is sweet.”
- Jean-Jacques Rousseau

Table of Contents

List of original publications

This thesis is based on the following publications, which are referred to in the text by their Roman numerals:

Publication I

He, L., Wennerberg, K., Aittokallio, T., and Tang, J. (2015) TIMMA-R: an R package for predicting synergistic multi-targeted drug combinations in cancer cell lines or patient-derived samples. Bioinformatics, 31, 1866-1868.

Publication II

He, L.*, Tang, J.*, Andersson, E.*, Timonen, S., Wennerberg, K., Mustjoki, S., and Aittokallio, T. (2018). Patient-customized Drug Combination Prediction and Testing for T-cell Prolymphocytic Leukemia Patients. Cancer Research, 78, 2407-2418.

Publication III

He, L., Kulesskiy, E., Saarela, J., Turunen, L., Wennerberg, K., Aittokallio, T., and Tang, J. (2018). Methods for High-Throughput Drug Combination Screening and Synergy Scoring. In: von Stechow L. (eds) Cancer Systems Biology. Methods in Molecular Biology, vol 1711. Humana Press, New York, NY.

* Equal contribution

The articles are reproduced with permission from the copyright holders

Publications related to the study but not included in the thesis:

Aleksandr, I., **He, L**., Aittokallio, T., and Tang, J. (2017). SynergyFinder: a web application for analyzing drug combination dose-response matrix data. Bioinformatics, 33, 2413-2415.

Author's contributions

Publication I

The author implemented the open-source R-package TIMMA. The author performed data analysis on the two case studies shown in the publication. The author also drafted the manuscript together with other contributing authors, as well as wrote reference manual and user instructions documentation for the TIMMA R-package.

Publication II

The author designed the study and implemented the drug combination prediction and testing (DCPT) platform. In the three T-PLL patient case studies, the author preprocessed the data, performed computational analysis, made drug combination predictions, as well as investigated the mechanisms of action of the personalized drug combinations. The author prepared the figures and tables for the manuscript and co-wrote the manuscript.

Publication III

The author developed the in-house tool FIMMCherry for designing the drug combination experiments to facilitate the drug screening process. The author also co-developed the open-source R-package SynergyFinder to prepare, analyze, and visualize the high-throughput drug combination screening data. The author participated in the manuscript writing and prepared the figures for the manuscript.

Abbreviations

ACVR1B	activin receptor type-1B
AL	acute leukemia
ALL	acute lymphocytic leukemia
AML	acute myeloid leukemia
AR	androgen receptor
AUC	area under the curve
BCL2	B-cell lymphoma 2
BM	bone marrow
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BWA	burrows wheeler aligner
BzCl	benzethonium chloride
CD52	cluster of differentiation 52
CCLE	cancer cell line encyclopedia
cDNA	complementary DNA
CDK	cyclin-dependent kinase
CI	combination index
CL	chronic leukemia
CLL	chronic lymphocytic leukemia
CM	conditioned medium
CMap	connectivity map
CML	chronic myeloid leukemia
CNV	copy number variation
CTNNB1	catenin beta 1
dbSNP	Single Nucleotide Polymorphism Database
DCPT	drug combination prediction and testing
DDLS	dedifferentiated liposarcoma
DIGRE	drug-induced genomic residual effect
DLBCL	diffuse large B-cell lymphoma
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DREAM	dialogue for reverse engineering assessments and methods
DSRT	drug sensitivity and resistance testing
DSS	drug sensitivity score
EC_{50}	half maximal effective concentration
EGFR	epidermal growth factor receptor
ERK	extracellular signal–regulated kinase

ES	exome sequencing
FC	fold-change
FDA	food and drug administration
FLT	fms related tyrosine kinase
FKBP	FK506-binding protein
FNR	false negative rate
FPKM	fragments per kilobase of exon model per million reads mapped
GDSC	genomics of drug sensitivity in cancer
GI_{50}	half maximal growth inhibition
GRC	genome reference consortium
HCL	hairy cell leukemia
HDAC	histone deacetylase
HAS	highest single agent
IC_{50}	half maximal inhibitory concentration
IGF1R	insulin like growth factor 1 receptor
IGV	integrative genomic viewer
JAK	Janus kinase
K_d	equilibrium dissociation constant
K_i	inhibitor constant
IL2RG	interleukin 2 receptor subunit gamma
IL5	interleukin 5
LINCS	library of integrated network-based cellular signatures
LL	lymphoblastic leukemia
LOO-CV	leave-one-out cross validation
MACS	medicinal algorithmic combinatorial screen
MAE	mean absolute error
MAF	minor allele frequency
MAPK	mitogen-activated protein kinase
MCM	mononuclear cell medium
MDM2	murine double minute 2
MDS	myelodysplastic syndromes
MET	mesenchymal epithelial transition
ML	myelogenous leukemia
MNC	mononuclear cell
MoAbs	monoclonal antibodies
mRNA	messenger RNA
mTOR	mammalian target of rapamycin
ncRNA	non-coding RNA
NGS	next-generation sequencing

NR3C1	nuclear receptor subfamily 3, group C, member 1
PB	peripheral blood
PDGFR	platelet-derived growth factor receptor
PEA	probability ensemble approach
PI	proteasome inhibitor
PI3K	phosphatidylinositol 3 kinase
PGR	progesterone receptor
pre-mRNA	precursor messenger RNA
QCA	qualitative comparative analysis
RACS	ranking-system of anti-cancer synergy
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA sequencing
rRNA	ribosomal RNA
STAT	signal transducer and activator of transcription
TACS	therapeutic algorithmic combinatorial screen
TIMMA	Target Inhibition Interaction using Maximization and Minimization Averaging
T-PLL	T-cell prolymphocytic leukemia
TPR	true positive rate
TTD	therapeutic target database
UCSC	university of santa cruz
WHO	world health organization
ZIP	zero interaction potency

Abstract

Cancer is a dynamic disease driven by complex molecular and environmental interactions. Therefore, the traditional "one gene, one drug, one disease" strategy is insufficient to treat most cancer patients. Drug combination therapy targeting various molecular mechanisms has become increasingly popular in treating cancer and other complex diseases. In comparison to monotherapy, combination therapy has the following advantages: possibility to reduce the dose of each single drug to minimize toxic side effects; achieving at least additive, multi-targeting effects, or even "greater-than-additive" effects, so called synergy; and reducing the likelihood of treatment resistance. However, even with the advanced high-throughput technologies currently used in drug combination screening, it remains infeasible to test systematically all the possible drug combinations across different cancer types, as the number of combination experiments grows exponentially. In addition, it remains difficult to understand the synergy mechanisms of many drug combinations, which poses further challenges for their clinical approval. Therefore, there is a timely need for computational tools that can help in identifying synergistic drug combinations for each individual patient, revealing the mechanisms of action of the drug combinations in the specific cellular context, as well as discovering potential predictive biomarkers for the synergistic responses in a systematic and reproducible way.

In this thesis, I implemented a systematic computational framework for identification and validation of synergistic drug combinations for each individual patient. Firstly, I developed machine learning models to predict drug combination effects by utilizing drug-target interaction networks, drug sensitivity profiles as well as genomic profiles of each patient. The models further enable one to identify both synergistic and cancer-selective drug combinations specific for each patient, therefore avoiding broadly toxic combinations. Secondly, to experimentally validate the predicted synergistic drug combinations, I developed software tools to help designing multi-dose drug combination experiments, to facilitate the semi-automated drug screening process, as well as to analyze the high-throughput drug combination screening data. Thirdly, I demonstrated the potential of constructing patient-specific cancer vulnerability networks to investigate the mechanisms of action of the personalized drug combinations, which is of importance to accelerate anticancer therapy discovery for precision oncology. All of the computational tools developed in this thesis are distributed as open-source tools, making it possible for others to reproduce the results and

apply to their data, or to extend the tools to new applications, as well as to integrate the tools as part of in-house drug combination analysis pipelines.

Introduction

Cancer is a collection of related diseases (over 277 different types of cancers) in which malignant cancer cells grow and divide at a rapid speed without control. Collectively, cancer is the second most lethal disease worldwide (Hassanpour and Dehghani, 2017; Fang et al., 2015). The number of individuals diagnosed with cancer as the primary disease is increasing during the last decades, which poses a serious threat to the health care system globally (IARC, 2014).

Cancer is a complex and dynamic disease. During the course of the disease, cancers evolve gradually to gain fitness benefits compared to other cell types. Most cancers also demonstrate wide heterogeneity, both intertumoral heterogeneity (patient by patient) and intratumoral heterogeneity (within a single tumor or one patient). Intertumoral heterogeneity refers to variability between patients with the same tumor type, which originates from patient-specific factors such as germline variation, somatic mutations, epigenetic differences or diverse environmental factors (Dagogo-Jack and Shaw, 2017; Liu et al., 2018). In contrast, intratumoral heterogeneity refers to the heterogeneity within one individual patient. Such heterogeneity can be present either as diverse tumor subpopulations located in different disease sites, or even within a single disease site (spatial heterogeneity), or as individual tumor subpopulations dynamically changing over time (temporal heterogeneity).

The heterogenous nature of cancers often results in drug resistance, which is one of the major reasons causing the failure of monotherapies. Drug resistance is a commonly seen phenomenon in pharmaceutical treatments for different diseases, and in cancer it occurs when tumors become tolerant to the pharmaceutical treatments (Housman et al., 2014). Certain types of drug resistance phenomena are disease or patient-specific, whereas other types developed over time while the disease evolve. For instance, many human drug-resistant cancers are initially responding to chemotherapy and then develop resistance over time through new mutations, metabolic changes, and other mechanisms (acquired resistance). With improved understanding of molecular biology of cancer and advances in high-throughput technology, the focus of cancer therapy development has shifted during the last few decades from identification of traditional chemotoxic treatments to discoveries of targeted therapy which targets certain pathways or cancer-specific proteins involved in cancer development (Bayat Mokhtari et al., 2017). For example, vemurafenib (BRAF inhibitor) has led to improved overall survival in melanoma patients with BRAF-V600 mutation (Fisher and Larkin,

2012). However, such targeted therapy still suffers from drug resistance since constant monotherapy with a single agent often causes cancer cells to recruit to alternative survival pathways (Madani Tonekaboni et al., 2016).

To reduce or even overcome monotherapy resistance, drug combination therapy is a promising therapeutic strategy, which has advantages over conventional cancer therapies. Combination therapy is targeting multiple cancer-inducing or resistance-driving pathways simultaneously. In this way, the chance of inducing cancer cells to recruit alternative pathways can be reduced, eventually resulting in reducing the likelihood of incidence of drug resistance. Furthermore, the adverse side effects of traditional chemotherapy and cytotoxic monotherapy, such as high toxicity and non-selective killing of both healthy and cancerous cells, make them less beneficial as modern treatment strategies for cancer. In comparison, drug combination therapies can often be used at lower doses of the single agents that may lead to reduced toxic effects. Furthermore, combination therapies offer also possibilities for increased therapeutic efficacy by synergistic effects, that is, the combined effect having greater than additive effect compared to the single agents.

Although there are many advantages of drug combination therapy for precision oncology, it is still challenging to identify synergistic drug combinations and get them approved for clinical use in an effective way. The enormous possibilities of drug combinations across different cancer types make it not feasible to screen them exhaustively, even with advanced high-throughput technology currently used in drug screening. Furthermore, many of the combinations are identified after intensive trials and errors, which makes it difficult to understand the synergy mechanisms, posing further challenges for clinical approval. Therefore, there is a timely need for computational tools that can help both in predicting and prioritizing of most synergistic drug combinations in a systematic way for combinatory therapy discovery.

The goal of this thesis was to develop computational models and software tools for predicting synergistic drug combinations and analyzing high-throughput drug combination screening data. The original publications of the thesis collectively demonstrate how to implement a systematic computational framework, including machine learning models to predict combination responses and computational approaches to identify synergistic combinations, along with experimental design and validation of the predictions, and all the way to construct networks to explore the potential mechanisms behind synergistic combinations

in individual patients. Such approaches are of importance to accelerate cancer therapy discovery for precision oncology.

Review of the literature

1 Leukemia types

Leukemia is one of the hematological malignances, referring to a group of cancers in blood or bone marrow (Kasteng et al., 2007; Polychronakis et al., 2013). Depending on which types of blood cells are affected, leukemia can be classified as lymphocytic leukemia (also called lymphoid or lymphoblastic leukemia, LL) or myeloid leukemia (also called myelogenous leukemia, ML). LL develops from cells which eventually become lymphoid cells. ML develops from early myeloid cells which eventually become white blood cells (other than lymphocytes), red blood cells, or megakaryocytes. Depending on how quickly the disease develops, leukemia can be also classified as acute leukemia (AL) or chronic leukemia (CL). AL starts abruptly and progresses rapidly, which results in an accumulation of immature and functionless cells called blasts. Patients with AL often have poor survival. In contrast, CL progresses more slowly, and results in an accumulation of relatively more mature cells that can perform some of their functions. Patients with CL can usually live much longer. Although CL patients often live longer than AL patients, it is in general more difficult to cure CL than AL (Goldsmith, 2011). Based on above classification of leukemia, there are four major types of leukemia: acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), chronic myeloid leukemia (CML), and chronic lymphocytic leukemia (CLL). There are also several subtypes of these four major types and other types of leukemia, such as hairy cell leukemia (HCL) and myelodysplastic syndromes (MDS).

1.1 T-cell prolymphocytic leukemia

T-cell prolymphocytic leukemia (T-PLL) is a rare mature T-cell leukemia with distinct features and an aggressive clinical course from presentation (Dearden, 2012). As a rare disease, clinicians do not see a case of T-PLL very often (once every 5-10 years) (Vivekanandarajah et al., 2013; Figueroa-Jiménez et al., 2016), which makes the diagnosis of T-PLL very challenging. Correct diagnosis requires appropriate diagnostic tests and manual interpretation of the test results.

In general, T-PLL patients are resistant to conventional chemotherapy, resulting in very poor prognosis (a median survival of only 4 months for non-responders). Currently, the best treatment option for T-PLL is alemtuzumab, a monoclonal antibody targeting CD52 mature lymphocytes, followed by consolidation with a

hematologic stem cell transplant where possible. Such treatment has extended the median survival to more than 4 years, but most patients eventually relapse, hence calling for novel targeted combination therapies.

2 Cancer therapy

Many different types of cancer therapies have been developed during the past decades. Traditional cancer therapies such as surgery, radiation therapy, and chemotherapy, are still widely used in cancer treatment though there is the shift from traditional cancer therapies to more advanced therapies such as immune therapy and targeted combination therapy. Each cancer therapy has its own advantages and disadvantages. The selection of cancer therapy depends on many factors such as the types of cancer, the stage of cancer, age of the patient, and the locality of cancer (Abbas and Rehman, 2018).

2.1 Surgery

Surgery, resection, or operation is one of the preferred treatments for firm malignant tumors as it removes the cancerous tumor and it often causes less damage to the healthy tissues surrounding it, compared with chemotherapy and radiation therapy (Abbas and Rehman, 2018). Surgery can also help doctors to find out if the tumor has spread or not. The outcome of surgery depends on many factors such as the location, the type of cancer, the condition of the patient, and the stage of cancer. Surgery usually yields good results for early stage cancer and it can also be combined with other types of cancer therapies.

2.2 Chemotherapy

Chemotherapy is affecting the abnormally fast cell dividing capacity of cancer cells and enforcing them to programmed cell death (apoptosis). The therapeutic effect is based on the general fact that cancer cells are dividing faster than normal cells. There are two different types of chemotherapy agents: cytostatic agents and cytotoxic agents. Cytostatic agents such as molecularly-targeted drugs inhibiting specific targets induce growth-inhibitory effects with lower toxic effects in normal tissues (Kummar et al., 2006). Hence, cytostatic agents may stop the growth of tumors without directly killing cancer cells (Rixe and Fojo, 2007). In contrast, cytotoxic agents lead to cell killing directly, which eventually shrink the tumor. As chemotherapeutic agents are also targeting healthy cells, their usage leads to various types of side effects such as hair loss, vomiting, and nausea, and

eventually some patients become immunocompromised (Abbas and Rehman, 2018).

2.3 Radiation therapy

Radiation therapy utilizes ionizing radiation to transfer high energy from the rays to cancer cells. This high energy destroys cancer cells directly, or alters the DNA of cancer cells, resulting in the halt of cell division and eventually the death of cancer cells. In general, radiation therapy is targeting directly at the main tumorous mass to avoid damaging healthy cells. Radiation therapy can be given from outside the body by a machine or inside the body by introducing a radioactive source into the body. Possible side effects such as hair loss, loss of appetite, and tiredness could occur, depending on which areas of the body are receiving the radiation therapy.

2.4 Hormone therapy

Hormone therapy is in general used as treatment for hormone sensitive cancer such as breast, prostate, and ovarian cancer (Holle et al., 2015). In such cancers, hormones play important roles in the growth and regulation of malignant cells. Hence, hormone therapy can be as effective as cytotoxic therapy. Unlike cytotoxic chemotherapy, however, hormone therapy does not lead to so high degrees of cytotoxicity since hormones are natural substances in human bodies (Carlson et al., 2014; EBCTCG, 2005). In general, hormone therapy focuses on either blocking/lowering the production of hormones or changing the behavior of hormones in cancer cell growth. Unwanted side effects such as diarrhea, hot flashes, and weakened bones could happen during receiving hormone therapy.

2.5 Targeted therapy

Targeted therapy is one of the advanced cancer therapies and prompted by the major advances in understanding of cancer biology especially the molecular mechanisms that leads to cancer (Yan et al., 2011). During the past few decades, many oncogenes and tumor suppressor genes have been identified, and major signaling pathways involved in the development of cancer have been studied. Such studies not only deepen our understanding of how cancer occurs and evolves, but also accelerates the development of targeted cancer therapies which are designed to interfere with certain proteins regulating the signaling pathways related to tumor growth or progression (Wu et al., 2006). Currently, targeted

cancer therapies encompass various types of agents, of which some agents such as monoclonal antibodies (MoAbs) are designed to interact with cell surface antigens, growth factor receptors, or ligands, to alter signaling pathways, whereas some are inhibiting the activities of the intracellular cascade pathways, and there are also agents that are developed to affect tumor micro-environment via preventing vessel to grow or destroying existing vessel (Lheureux et al., 2017; Holle et al., 2015; Herbst, 2005). In contrast with traditional cancer therapies, targeted cancer therapies are more selective, meaning that they only modulate a specific target, such as cancer-driving oncogene, to avoid damaging normal cells.

2.6 Immunotherapy

During the past few decades, our understanding of the complex immune system has progressed rapidly, which has shed light on how to utilize our own immune system to fight cancer (Ventola, 2017). Our immune system has the so-called "immune surveillance" mechanism, which is able to recognize and eliminate abnormal cells (Pennock and Chow, 2015). In the process of immune surveillance, the adaptive and innate compartments of the immune system participate in tumor recognition and elimination (Finn, 2012). Such duel protective system is effectively detecting tumor cells and can destroy them in the elimination phase (Vesely et al., 2011). However, when the immune system is not able to eliminate the tumor cells completely, but the survived tumor cells are still under the immune surveillance, the immunological equilibrium phase is reached. In this phase, the tumor growth has been suppressed over a period of time and it remains under control. In such a case, tumor cells can eventually break the immunological equilibrium and escape the immune surveillance (escape phase), resulting in more progressive tumor cell growth and suppressed antitumor immune response. To prevent tumor cells escaping from the immune surveillance, immunotherapy focuses on enhancing the host immune response and changing the tumor-promoting immune response to tumor-rejecting immune response, in order to achieve better tumor recognition and elimination. As one of the most advanced cancer treatment modalities, immunotherapy holds promise of a life-long cure. However, the current challenge is to identify patients who respond to different types of immunotherapies, and combination of targeted chemotherapies with immunotherapies has been suggested as more personalized treatment regimen.

2.7 Stem cell therapy

Stem cells are undifferentiated cells which can differentiate into various cell types. Stem cells have many special features, such as regenerating original body

tissues and immune modulation (Fleifel et al., 2018). Such special abilities provide various strategies in cancer treatment. For example, engineering stem cells to stimulate immune response against cancer cells. Stem cell therapy has a great potential in curing cancer, but there are also many challenges to overcome such as how to isolate and identify stem cells correctly from patients' tissues.

2.8 Combination therapy

For most malignant diseases, such as cancer, driven by complex molecular and environmental interactions, treatment with traditional one-gene-one-drug approach is insufficient to disrupt the complex mechanisms (Wu et al., 2015). Therefore, combining drugs working against different molecular mechanisms has become increasingly popular in cancer treatment. The primary goals of combination therapy are: 1) reducing the dose of each single drug to sufficient response level to minimize the toxic side effects; 2) achieving at least additive effects or even "greater-than-additive" effects, so called synergy, in the biochemical activities of two drugs; 3) modulating selectively on activity of targets found in tumor cells but not in healthy cells; and 4) reducing the likelihood of drug resistance (Fitzgerald et al., 2006; Al-Lazikani et al., 2012).

3 Personalized drug combination therapy in cancer

3.1 Personalized medicine approach

Personalized medicine is a rapidly evolving field of health care. In cancer research, recent advances in high-throughput technology has allowed us to gain the knowledge of each individual patient's genetic and molecular information. With such unique genomic information, along with the clinical information, personalized medicine aims to provide more accurate treatment and prevention strategies for particular subgroups of patients (Jovanović et al., 2018). Such personalized treatment and prevention strategies suit for complex diseases like cancer due to its heterogeneity.

3.2 Identification of synergistic effects in combination experiments

The concept of synergistic effects, or synergy generated by the interactions between multiple chemical agents, has been studied in the biomedical world over a century (Roell et al., 2017). For complex diseases, such as cancer, which is

often treated with drug cocktails, this concept is often used to understand the interactions among the drugs in order to maximize the therapeutic effects while minimizing the side effects.

In general, the interactions between multiple agents can be classified as additive, synergistic, and antagonistic. An additive effect is considered as the baseline effect which is the theoretically expected effect when the agents are not interacting with each other. Any significant deviation from the additive effect is considered as either synergistic effect or antagonistic effect. Synergistic effect is usually defined as the effect that is greater than the expected theoretical effect, namely, the additive effect. In opposite of synergistic effect, antagonistic effect is usually defined as the effect that is lower than the additive effect. However, how to mathematically define the additive effect has been under debate over the last century because simply summing up two or more effects does not accurately reflect what happens in the cells. There are many models proposed to quantify the expected effect that are based on different empirical or biological assumptions (Tang et al., 2015; Lehár et al., 2009). These so-called synergy models inevitably lead to inconsistent results, due to the inherent differences among these assumptions, but no single universally best model exists (Vlot et al., 2019), as was already admitted by an agreement from a conference held in Saariselkä, Finland in 1992 (Greco et al., 1992).

Highest Single Agent (HSA) model (Berenbaum, 1989) is the simplest and most straight forward synergy scoring model. It takes the higher response of the two single agents at corresponding concentrations as the baseline effect. The formula for HSA model is $y_{HSA} = \max(y_A - y_B)$, where y_{HSA} is the additive response, y_A is the response of agent A, and y_B is the response of agent B. With HSA model, any additional effect over the higher response is classified as HSA synergy.

Bliss independence model (Bliss, 1939) is one of the most used baseline or "null hypothesis" models. It is based on the assumption that if two agents are behaving independently, namely, under non-interaction (Greco et al., 1995), the baseline effect can be simply calculated as the product of two individual responses: $y_{Bliss} = y_A * y_B$. In probability theory, the Bliss independence model can be interpreted as the probability of two independent events, namely, $P(AB) = P(A) * P(B)$.

As an alternative model to Bliss independence model, Loewe additivity model (Loewe, 1953; Loewe and Muischnek, 1926) is based on the idea of "sham experiment" in which a single agent is combined with itself. In the sham

experiment, no interaction is expected since a single agent cannot interact with itself. Similarly, a single agent is not expected to interact with an agent with similar structure and targets. In contrast to HSA model and Bliss independence model, the Loewe additivity model requires additional information about the dose-response relationships of each single agent (Tang et al., 2015), namely, $y = f_A(x)$ and $y = f_B(x)$, where x is the concentration of the agent. Loewe additivity model assumes that ratio of doses of agent A and agent B ($R = \frac{x_A}{x_B}$) is fixed when agent A at dose x_A yields the same response as agent B at dose x_B (Hennessey et al., 2010; Lederer et al., 2018). Under this assumption, the response of agent A at dose x_a is equivalent to the response of agent B at dose $x_a * \left(\frac{x_B}{x_A}\right)$, namely,

$$f_B\left(x_a * \left(\frac{x_B}{x_A}\right)\right) = f_A(x_a). \quad (1)$$

A dose of combination (x_a, x_b) which yields the same response y as each individual agent should satisfy:

$$f_A(x_a) + f_B(x_b) = f_A(x_A) = f_B(x_B). \quad (2)$$

By combining equation (1) and equation (2), the dose combination (x_a, x_b) must therefore satisfy:

$$\frac{x_a}{x_A} + \frac{x_b}{x_B} = 1. \quad (3)$$

Under the sham scenario, $x_a = x_b = \frac{1}{2}x_A = \frac{1}{2}x_B$, equation (3) is trivially satisfied.

As the name, zero interaction potency (ZIP) suggests, the ZIP model assumes that there should be zero potency shift if there is no interaction between two drugs, meaning the potency of each drug's dose-response curve should remain under non-interaction (Yadav et al., 2015). When using the four-parameter log-logistic functions to model the dose-response relationship, and assuming complete inhibition of the cell growth rates (i.e., $E_{min} = 0$ and $E_{max} = 1$), then the dose-response curve for drug 1 after adding drug 2 becomes under no interaction:

$$y_{1\leftarrow 2} = \frac{y_2+(\frac{x}{m_1})^{\lambda_1}}{1+(\frac{x}{m_1})^{\lambda_1}}, \tag{4}$$

where $y_{1\leftarrow 2}$ is the response after adding drug 2 to drug1, λ_1 is the slope of the dose-response curve of drug 1, and m_1 is the half maximal inhibitory concentration (IC_{50}) (Yung-Chi and Prusoff, 1973). In such case, adding drug 2 should only increase the baseline response of drug 1 from E_{min}=0 to E_{min}=y_2 with $m_1$and λ_1 remaining unchanged as no potency shift occurs in the dose-response curve. At dose pair (x_1, x_2), the expected response of adding drug 2 to drug 1 becomes

$$y_{ZIP}^{1\leftarrow 2} = \frac{y_2+(\frac{x_1}{m_1})^{\lambda_1}}{1+(\frac{x_1}{m_1})^{\lambda_1}} = \frac{\frac{1}{1+(\frac{m_2}{x_2})^{\lambda_2}}+(\frac{x_1}{m_1})^{\lambda_1}}{1+(\frac{x_1}{m_1})^{\lambda_1}}. \tag{5}$$

In the same way, the expected response of adding drug 1 to drug 2 becomes

$$y_{ZIP}^{2\leftarrow 1} = \frac{y_1+(\frac{x_2}{m_2})^{\lambda_2}}{1+(\frac{x_2}{m_2})^{\lambda_2}} = \frac{\frac{1}{1+(\frac{m_1}{x_1})^{\lambda_1}}+(\frac{x_2}{m_2})^{\lambda_2}}{1+(\frac{x_2}{m_2})^{\lambda_2}}. \tag{6}$$

It was proved that $y_{1\leftarrow 2}$ and $y_{2\leftarrow 1}$are equal (Yadav et al., 2015), which means that the order of adding each individual drug does not matter in ZIP model. The expected response can be factorized as a multiplicative of single dose–response curves:

$$y_{ZIP} = \frac{(\frac{x_1}{m_1})^{\lambda_1}}{1+(\frac{x_1}{m_1})^{\lambda_1}} + \frac{(\frac{x_2}{m_2})^{\lambda_2}}{1+(\frac{x_2}{m_2})^{\lambda_2}} - \frac{(\frac{x_1}{m_1})^{\lambda_1}}{1+(\frac{x_1}{m_1})^{\lambda_1}}\frac{(\frac{x_2}{m_2})^{\lambda_2}}{1+(\frac{x_2}{m_2})^{\lambda_2}}, \tag{7}$$

which can be interpreted as the probabilistic independence between the two drugs.

As illustrated in Figure 1, the observed combination response is modelled in ZIP in a two-way manner as well:

$$y_c^{1\leftarrow 2} = \frac{y_2+(\frac{x_1}{m_{1\leftarrow 2}})^{\lambda_{1\leftarrow 2}}}{1+(\frac{x_1}{m_{1\leftarrow 2}})^{\lambda_{1\leftarrow 2}}} = \frac{\frac{1}{1+(\frac{m_2}{x_2})^{\lambda_2}}+(\frac{x_1}{m_{1\leftarrow 2}})^{\lambda_{1\leftarrow 2}}}{1+(\frac{x_1}{m_{1\leftarrow 2}})^{\lambda_{1\leftarrow 2}}}, \tag{8}$$

$$y_c^{2\leftarrow 1} = \frac{y_1+(\frac{x_2}{m_{2\leftarrow 1}})^{\lambda_{2\leftarrow 1}}}{1+(\frac{x_2}{m_{2\leftarrow 1}})^{\lambda_{2\leftarrow 1}}} = \frac{\frac{1}{1+(\frac{m_1}{x_1})^{\lambda_1}}+(\frac{x_2}{m_{2\leftarrow 1}})^{\lambda_{2\leftarrow 1}}}{1+(\frac{x_2}{m_{2\leftarrow 1}})^{\lambda_{2\leftarrow 1}}}. \tag{9}$$

In ZIP model, a delta score is defined to quantify the overall potency shift:

$$\delta(\theta) = \frac{y_c^{1\leftarrow 2} - y_{ZIP}^{1\leftarrow 2}}{2} + \frac{y_c^{2\leftarrow 1} - y_{ZIP}^{2\leftarrow 1}}{2}. \quad (10)$$

When $\delta = 0$, there is zero interaction between two drugs under ZIP model. It is also shown (Yadav et al., 2015) that $\delta = 0$ also holds when combining two identical drugs in a sham experiment. When $\delta > 0$, it indicates synergistic effect between two drugs. When $\delta < 0$, it indicates antagonistic effect between two drugs. As ZIP model makes use of the advantages of both the Loewe and the Bliss models by taking the probabilistic independence between the two drugs and sham experiment into consideration, ZIP model may be considered as an integration of these two synergy models.

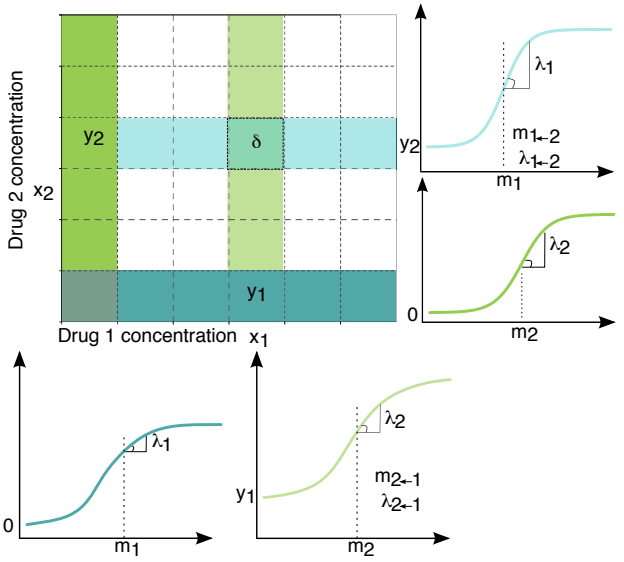

Figure 1: A schematic illustration of modelling the observed combination response in a two-way manner in ZIP model. Reproduced with permission from Elsevier (Yadav et al., 2015).

3.3 Prediction of synergistic drug combinations in silico

Despite the many benefits of drug combination therapy in curing cancer, compared to monotherapy, identifying novel synergistic drug combinations is still challenging due to the enormous search space of all possible drug combinations. In addition, getting novel synergistic drug combinations approved for clinical usage is challenging, because most of drug combinations in clinical

use are based on intensive trials and error testing, rather than rational understanding of their mechanism of action (Madani Tonekaboni et al., 2016).

Even though high-throughput screening technologies have been implemented successfully for testing the phenotypic responses of single drug or drug combinations, it is not feasible to test all possible drug combinations via *in vitro* and *in vivo* screening across different cancer types as the number of experiments grows rapidly (Madani Tonekaboni et al., 2016; Fang et al., 2015). Therefore, there is a need to develop computational methods to prioritize drug combinations for *in vitro* and *in vivo* testing. Identifying synergistic drug combinations via computational methods is less time-consuming compared with direct experiment approaches. Moreover, computational approaches, such as machine learning models, accompanied with new technologies such as large-scale genomic and systems or network biology, provide powerful approaches to making predictions of the responses of drug combinations.

There are generally three types of computational approaches to predict synergistic drug combinations: explicit mathematical methods, stochastic search algorithms, and machine learning methods. Each type has its own advantages and disadvantages.

Explicit mathematical methods refer to methods which utilizes one or several direct mathematical models or statistical tests to predict the responses of drug combinations. These can be based on assumptions of the relationship between drug combination effects and transcriptomic changes. For example, drug-induced genomic residual effect (DIGRE) (Yang et al., 2015), the best performing method in one of the community challenges on drug combinations launched by the Dialogue for Reverse Engineering Assessments and Methods (DREAM) consortium, hypothesized that in a setting that cancer cells are treated with two drugs sequentially, cancer cells treated after the first drug would contribute to the transcriptomic changes induced by the second drug. There are three main steps in DIGRE. In the first step, a drug-pair similarity score is obtained by comparing the genomic or transcriptome changes induced by two drugs. In the second step, a mathematical model integrates the similarity score and the dose-response curve information to estimate the induced cell death. Finally, the average of the induced cell death values estimated from two possible sequential orders of treatment is used as the predicted synergy score.

Stochastic search algorithms or genetic algorithms are search methods that can be used for solving many types of search problems, such as optimization of a

function or determination of the proper order of a sequence (Forrest, 1993). Such search method can be applied to search for synergistic drug combinations. For example, a genetic algorithm called Medicinal Algorithmic Combinatorial Screen (MACS) was developed to walk through a stepwise process of enriching successive pools of drug cocktails for increased fitness, which indicates a more optimal solution in the genetic algorithm (Zinner et al., 2009). This kind of algorithms show the feasibility of predicting drug combinations of arbitrary size. However, it is likely that such algorithms converge to different drug cocktails in different runs, due to a complex search landscape, suggesting a local optima rather than a global optima.

Machine learning methods utilize various types of omics data to predict the combinatory effects of drug combinations. The data commonly used in drug combination prediction are sample-specific data and drug-specific data. Sample -specific data are collected from the tested samples (either cell lines or patient-derived samples), such as gene expression, point mutation, and copy number variation data. Drug-specific data contain information about the drugs, such as similarity of structures or biochemical properties between drugs, responses of the cancer cells to drugs, and target information of the drugs (Madani Tonekaboni et al., 2016). With such data, there are generally three types of machine learnings methods, namely, supervised, unsupervised, and semi-supervised that can be used for drug combination prediction.

Supervised machine learning methods learn the structure or the pattern of labeled data. The labels of the training data could be either the synergy scores of the drug combinations for a regression problem or classes, such as “synergistic” and “non-synergistic”, for a classification problem. Supervised machine learning methods try to learn the relationship between the data and the labels. There are some supervised machine learning methods developed to predict the synergy classes or the quantification of drug combination responses. For example, a Bayesian network approach called, termed probability ensemble approach (PEA) (Li et al., 2015), was developed to integrate data from different sources such as the chemical and pharmacological features of drugs to predict the efficacy of drug combinations. Supervised machine learning methods can often make good predictions with good-quality labeled data. However, there are many obstacles as well. For example, one of biggest obstacles is lack of labeled drug combination data, which requires a large amount of high-quality drug screenings experiments.

In contrast to supervised machine learning methods, unsupervised machine learning methods aim at discovering the hidden structure of unlabeled data. For

example, a novel tool called DrugComboRanker (Huang et al., 2014) was developed to prioritize drug combinations and to uncover the underlying mechanisms of action. DrugComboRanker constructs a drugs' functional network based on their transcriptomic response profiles from Connectivity Map (CMap) (Lamb et al., 2006), and then divides the network into drug network communities using a Bayesian non-negative matrix factorization approach. Under the assumption that drugs within overlapping community share common mechanisms of action, DrugComboRanker next predicts the unknown targets of drugs by applying a recommendation system on the drug communities. With a disease-specific signaling network built with genomic and interactome data and using the predicted targets, DrugComboRanker searched for such drug combinations that have their targets enriched in the complementary signaling modules of the disease signaling network, which are predicted to be synergistic.

As its name suggests, semi-supervised machine learning methods utilize unlabeled data, along with typically small amounts of labeled data. Semi-supervised machine learning methods can address issues like lack of labeled data and data imbalance (insufficient number of positive or negative labeled data) in the drug combination prediction field. Ranking-system of Anti-Cancer Synergy (RACS) (Sun et al., 2015) is one of the semi-supervised machine learning methods developed to predict synergistic drug combinations, which is based on a manifold ranking method. RACS was developed with 26 effective combinations (labeled data) and 502 pairs of drug combinations (unlabeled data). The performance of RACS was demonstrated with drug combination data from the 2014 DREAM challenge. It outperformed the best performing method of the challenge with an AUC value of 0.85.

4 Functional and molecular profiling for personalized therapy

4.1 Next-generation sequencing

Next-generation sequencing (NGS), also known as high-throughput sequencing, is a technology to perform sequencing of millions of small fragments in parallel (Behjati and Tarpey, 2013). NGS offers the possibility to produce a large amount of sequencing data much quicker and cheaper than previous gold standard sequencing technology, namely, Sanger sequencing. NGS enables researchers to perform a wide variety of applications including whole genome sequencing, whole exome sequencing, RNA sequencing, and targeted amplicon sequencing

for more in-depth genomic knowledge. Such applications have revolutionized research in different areas of genomics, epigenomics, transcriptomics, and metagenomics (Herzyk, 2014).

4.1.1 Exome sequencing

The exome refers to the sequence encompassing all exons of protein coding genes in the genome and it covers around 1%-2% of the human genome (Warr et al., 2015). Exome sequencing (ES) is used for identification of rare and common genetic variants in the protein coding regions (Belkadi et al., 2015). As roughly 85% of diseases-related variants are falling into the protein coding regions, ES is a highly efficient approach to detect diseases related variants (Normand et al., 2018). In addition, ES is more cost-effective than whole-genome sequencing as exome sequencing only targets at a small subset of the genome.

ES targets all the genes in the protein coding regions, regardless whether they are associated with the disease or not. Such a systematic way enables us to collect unbiased information to decide if a new gene is associated with the disease or not, even when it has not been previously identified or described in the literature or public databases (Normand et al., 2018). Since we are still in the process of enriching our knowledge of gene-disease relationships, it is possible to re-analyze the ES data without performing ES on a new sample, once some additional phenotypic information becomes available (Normand et al., 2018).

Even though performing ES does not require a clear diagnosis or target region ahead of the time, it requires accurate identification of the variants that are correlated with the patient's phenotype, as ES results often contain many rare variants that are not phenotype-related (Normand et al., 2018). Therefore, each variant has to be verified based on the relationship between patient's phenotype and gene's known function or disease association.

Although ES is increasingly used in both research and clinic context, there are some limitations to be noted. In contrast to multigene sequencing panels that simultaneously sequence a few to hundreds of genes, ES is targeting at the protein coding region, which is about 30 million base pairs and covers 20,000-25,000 genes (Bertier et al., 2016). Such a wide genomic range results in a lower coverage depth than multigene sequencing panels (Normand et al., 2018). ES has difficulties in targeting certain regions, such as CG rich regions and highly repetitive sequences. The lower coverage depth and difficulties in those regions may lead to false positives. Furthermore, ES is not reliable at detecting large

structural changes such as inversions, translocations, and nucleotide repeat expansions.

4.1.2 RNA sequencing

RNA sequencing (RNA-seq) is a recently developed transcriptome profiling technology to measure RNA abundance from a sample at a particular time utilizing NGS technologies. There are three main steps in a typical RNA-seq experiment: RNA extraction, RNA-seq library construction, and sequencing on an NGS platform.

In the first step, RNA is extracted from tissue and mixed with deoxyribonuclease (DNase), which reduces the amount of genomic DNA. A successful RNA-seq experiment requires high-quality RNA in order to produce a library for sequencing (Kukurba and Montgomery, 2015). The quality of RNA is typically evaluated by an RNA Integrity Number (RIN), with a range from 1 (highly degraded) to 10 (highest integrity), which is produced by an Agilent Bioanalyzer (Herzog and Papenfort, 2018). RIN of ≥8 defines good-quality RNA (Hitzemann et al., 2014).

In the second step, an RNA-seq library is constructed by following different protocols depending on the selection of RNA species and NGS platforms (Kukurba and Montgomery, 2015). In general, construction of an RNA-seq library contains some common basic steps: selecting the desired RNA species/modules and converting RNA to complementary DNA (cDNA). There are four typical RNA modules: ribosomal RNA (rRNA), precursor messenger RNA (pre-mRNA), messenger RNA (mRNA), and different types of non-coding RNA (ncRNA). When selecting the RNA modules, the majority of the RNA modules for most cell types (up > 80%-90% of the total RNA), namely, rRNA is removed because rRNA is generally not of interest. In such a way, the experiment will have a better overall depth coverage and a better detection power for less-abundant RNAs. After selecting the desired RNA modules, a population of desired RNA modules is converted to a library of cDNA and sequencing adapters are added.

In the last step, each molecule is sequenced in an NGS platform to produce short reads from one end via single-end sequencing or both ends via pair-end sequencing (Wang et al., 2009). The length of the reads varies depending on the sequencing technology. After obtaining the reads, an RNA-seq analysis pipeline is needed to analyze the reads for different research purposes such as measuring

the gene expression level of the sequenced samples, transcript quantification, reference-based gene annotation, and de novo transcript assembly (Li and Li, 2018).

RNA-seq has many advantages over hybridization-based microarrays s that were previously developed for profiling gene expression. Firstly, RNA-seq can detect more novel transcripts that are differentially modulated, splice junctions, and non-coding transcripts because RNA-seq allows for full sequencing of the whole transcriptome, while microarrays probe only sequences of predefined transcripts/genes (Rai et al., 2018; Rao et al., 2019). Secondly, RNA-seq has very low background signal and it does not have an upper limit for quantification in contrast to microarrays. Such an advantage enables RNA-seq to detect genes with a large dynamic range of expression levels (Wang et al., 2009). Last but not least, as a hybridization-free approach, RNA-seq does not suffer from the errors caused during hybridization of microarrays as it does not depend on genome annotation for prior probe selection (Zhao et al., 2014).

Although there are many advantages in the RNA-seq technologies, RNA-seq still has its own limitations. For example, RNA-seq is subject to some systematic variations such as different library sizes across samples (Vafaee et al., 2019). Hence, RNA-seq results of different experiments or samples must be normalized before performing any comparison analysis across different experiments or samples. There are several normalization methods available for RNA-seq data (Abbas-Aghababazadeh et al., 2018).

4.1.3 Analyzing sequence data

NGS technologies generally generate significant amounts of output data, which requires bioinformatical tools to analyze the data for different applications. There are some common steps to analyze the data. Firstly, the quality of the raw sequencing data must be assessed, and quality-based trimmer is performed on the raw data to remove contaminating sequences and low-quality sequences. Then, the reads are aligned to the reference genome such as University of Santa Cruz (UCSC) (Casper et al., 2018) or Genome Reference Consortium (GRC) (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc) human reference genome using the aligner tools such as Burrows Wheeler aligner (BWA) (Li and Durbin, 2009) and Bowtie2 (Langmead and Salzberg, 2012). After the alignment, it is recommended to perform three steps of post-alignment processing, namely, read duplicate removal, indel realignment, and base quality score recalibration (BQSR) (Bao et al., 2014). Next, different types of variants including copy

number variants (CNV), germline variants, somatic variants, and structural variants are identified with different tools (Zhao et al., 2013; Kosugi et al., 2019). In cancer research, somatic mutations are often more interesting than germline mutations (Wadapurkar and Vyas, 2018). Somatic mutations are those that are present only in somatic cells while germline mutations are present in germ cells. After the variants are identified, annotation of the variants is needed as it provides more information about the variants such as the amino acid change, gene symbol, exonic function, and genomic features. In addition, some computational tools are providing additional information such as minor allele frequency (MAF) in normal populations and known cancer driver genes by integrating public databases such as dbSNP (Sherry et al., 2001). After annotating the variants, they can be visualized via different tools such as Integrative Genomic Viewer (IGV) (Thorvaldsdóttir et al., 2013) for manual curation.

4.2 Drug sensitivity testing

With advanced high-throughput technologies, a comprehensive functional strategy called drug sensitivity and resistance testing (DSRT) was developed to test a large panel of drugs including both clinically available chemotherapeutics and targeted agents on cancer cells in a systematic manner (Pemovska et al., 2013). Functional data from DSRT on cancer cells provide direct information about which drugs are effective on killing the cancer cells. Such information can help to select the most effective drugs for personalized treatment.

In DSRT data analysis, the dose-response of a drug across different concentrations is fitted with four-parameter log-logistic function:

$$y = d + \frac{a-d}{1+10^{b(c-x)}}, \tag{11}$$

where y is the observed response (e.g., cell inhibition or viability), x is the concentration, a is the maximum drug response, b is the slope of the dose-response curve, c is the half maximal inhibitory concentration (IC_{50}) (Yung-Chi and Prusoff, 1973) or half maximal effective concentration (EC_{50}) (Sebaugh, 2011), or half maximal growth inhibition (GI_{50}) (Marx, 2013), depending how the response is measured, and d is the minimal drug response.

It has become popular to use the point estimates of the dose-response curve such as IC_{50}, EC_{50}, and GI_{50} as summary drug response measure. However, the point estimation is not always sufficiently accurate to quantify the actual response. For

example, different dose-response curves could have exactly the same point estimation, even if the curve shapes are quite different from each other (see Figure 2). This problem can be avoided by using a model that takes the whole dose-response curve into consideration to quantify the drug response. This model can then be used to calculate the area under the curve (AUC) or so- called drug sensitivity score (DSS) (Yadav et al., 2014) to represent the efficacy of the drug.

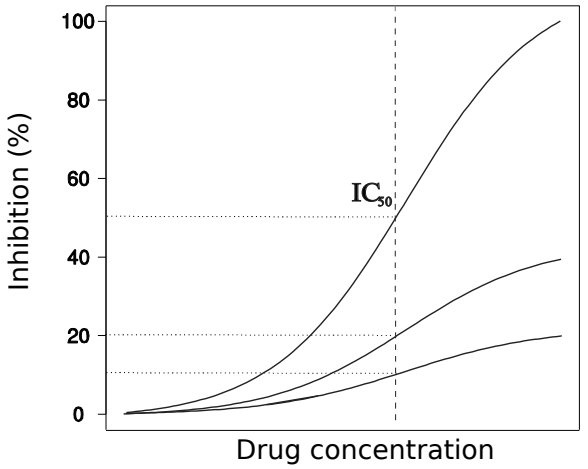

Figure 2: Schematic demonstration of the different dose-response curves with the same point estimate (IC50), but with different area under the dose-response curve (AUC).

Aims of the study

This thesis aimed to develop computational models and tools for accelerating the process of identifying personalized drug combination therapy for cancer patients in a systematic manner, covering all the phases of the discovery process from experiment design to predicting and validating synergistic drug combinations. To prioritize the most potential synergistic combinations for follow-up experiment validation and clinical translation, fast computational tools are needed. To make the results more reproducible, a systematic approach and open source-code implementation are needed.

The study consisted of four parts with specific objectives:

1. To develop computational models and open-code tools to predict patient-specific synergistic drug combinations.
2. To develop open-code tools for facilitating the design of high-throughput drug combination experiments.
3. To develop open-code tools to effectively analyze the data from high-throughput drug combination experiments.
4. To develop open-code tools to discover the underlying mechanisms of synergistic drug combinations for a particular patient.

Materials and methods

5 Compound datasets

In Publication I, drug-target interaction data from 65 kinase inhibitors and 322 kinase targets from a previous study (Tyner et al., 2013) were used. The drug-target interaction data were categorized into six classes that were labelled as [0, 1, 2, 3, 4, 5]. Class 0 represents the inactive group, meaning no interaction between the drug and the target based on bioactivity data from (Tyner et al., 2013). Class 1 means the target is the most potent target for the drug. The bioactivity values of target in class 1 were lower than 10-fold of the lowest bioactivity value of all targets for the drug. Classes 2 to 5 represent the intermediate interaction groups. The 65 kinase inhibitors were tested on 151 leukemia patient-derived samples *ex vivo*, where the drug responses were measured in a cell proliferation assay and quantified by point estimation IC_{50} from the dose-response curve (Tyner et al., 2013). For each sample, the IC_{50} values were normalized into [0, 1] interval in Publication I:

$$y = \frac{Max(IC_{50}) - IC_{50}}{Max(IC_{50}) - Min(IC_{50})}, \tag{12}$$

where y is the normalized drug response, $Max(IC_{50})$ is the maximum IC_{50} values for the sample, $Min(IC_{50})$ is the minimum IC_{50} values for the sample.

In addition to the data from Tyner study, Publication I also included data from another study (Miller et al., 2013), where only primary target information is available for each drug. The data from Miller study included a total 14 drugs that were tested in pairwise combinations in a dedifferentiated liposarcoma (DDLS)-derived cancer cell line (Table 1). Primary targets are import indicators about the drugs' mode of action, but not enough to understand their full polypharmacological interactions. Hence, additional drug-target interaction data for the 14 EMD Millipore drugs were obtained from a broad bioactivity screen (Gao et al., 2013). With further manual curation on the data, the drug-target data were enriched (Table 2). Each of the 13 drugs (Rapamycin was missing in the original publication's supplementary data) were tested on DDLS8817 cell line with 7 concentrations. The responses were measured as the relative cell metabolic activities and quantified by IC_{50}.

Drug	Abbreviation	Primary Targets	Company
AG538	AG	IGF1R	EMD Millipore
AKT 1/2 Inh	AK	AKT1/2	EMD Millipore
FR180204	ER	ERK1/2	EMD Millipore
Gefitinib	GF	EGFR	Tocris
HNHA	HN	HDAC	Cayman
PDGFR TKI III	PD	PDGFR	EMD Millipore
PI3Ka Inh IV	PI	PI3K (PIK3CA/B/C/D)	EMD Millipore
Rapamycin		mTOR	Selleckchem
Rottlerin	RT	PKC/CaMKIII	Tocris
Ryuvidine	RY	CDK4/6	Tocris
SL327	SL	MEK1/2	EMD Millipore
SRC Inh I	SR	SRC	EMD Millipore
Stattic	ST	STAT3	EMD Millipore
SU11247	SU	MET	EMD Millipore

Table 1: The drug information in the Miller study (Miller et al., 2013).

Drug	Abbreviation	Number of targets	PubMed reference
AG538	AG	4	12869569
AKT 1/2 Inh	AK	64	23398362
FR180204	ER	58	23398362
Gefitinib	GF	16	24065146
HNHA	HN	1	17353008
PDGFR TKI III	PD	126	23398362
PI3Ka Inh IV	PI	4	16837202
Rapamycin		9	24521231
Rottlerin	RT	3	24521231
Ryuvidine	RY	2	24902048
SL327	SL	26	17114005
SRC Inh I	SR	75	23398362
Stattic	ST	3	23398362
SU11247	SU	79	23398362

Table 2: The number of drug targets identified from the literature in the Miller study (Miller et al., 2013).

6 Patients samples

In Publication II, three T-PLL patient samples and three healthy donor samples were collected from Helsinki University Central Hospital (Finland), Hospital Universitario de la Princesa (Spain), and University Hospital Aachen (Germany). The study was approved by the Helsinki University Hospital (Finland) ethics committee and was conducted in accordance of the Helsinki Declaration principles (Andersson et al., 2017). Written informed consents were obtained from all the patients and healthy donors. Biomaterial samples from Aachen followed the regulations of the biomaterial bank and it was approved by the ethics committee of the RWTH Aachen Medical Faculty. Patients diagnosis fulfilled the old and current WHO diagnostic criteria of T-PLL (Swerdlow et al., 2016).

7 TIMMA-R workflow for drug combination prediction

In Publication I, a logic-based network algorithm, Target Inhibition Interaction using Maximization and Minimization Averaging (TIMMA), was implemented in R with major extensions from its original version (Tang et al., 2013), including modelling of multi-classes drug-target data and network visualization. TIMMA-R is much faster than the original MATLAB implementation and the TIMMA-R predictions are robust to the intrinsic noise in the experimental data (He et al., 2015).

7.1 Input data

TIMMA-R workflow takes two types of input data: drugs' polypharmacological profiles and drug sensitivity profiles (Figure 3). In the polypharmacological profiles of drugs, both strong and week drug-target interactions are taken into consideration in order to model the drug combination effect through their target interactions. To construct such profiles, publicly available proteome-wide quantitative drug-target interaction data from PubChem, ChEMBL (Gaulton et al., 2012) or CanSAR (Halling-Brown et al., 2012) databases were utilized. To integrate bioactivity data from different sources was advised to minimize the false positive or negative interactions (Tang et al., 2014). The drug sensitivity profiles can be obtained from either cancer cell lines such as the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2012)，and the Library of Integrated Network-based Cellular Signatures (LINCS) (Vempati et al., 2014) databases or patient-derived samples (Pemovska et al., 2013).

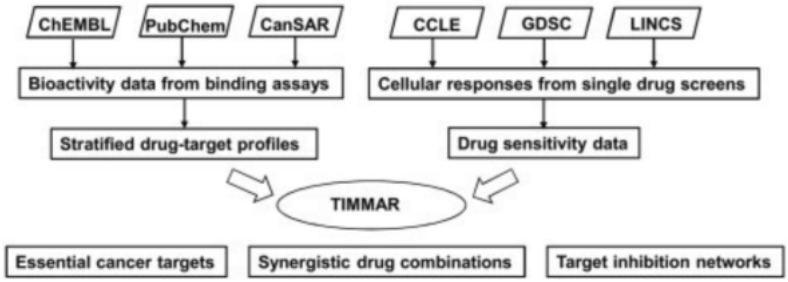

Figure 3: The TIMMA-R workflow for drug combination prediction. Reproduced with permission from Oxford University Press (He et al., 2015).

7.2 Modeling of multi-class drug-target data

Instead of utilizing conventional binary drug-target data for modelling, TIMMA-R models multi-class drug-target data by classifying the targets of given drugs into several classes based on their binding affinities (Tyner et al., 2013). In many databases such as DrugBank (Law et al., 2014) or TTD (Zhu et al., 2012), one can only find a few primary targets for many approved drugs without knowing the corresponding activity classification threshold. Such information is not sufficient to understand the drugs' mode of action, as the therapeutic efficacy of many approved drugs is not necessarily associated with higher binding affinity (Gleeson et al., 2011). Some databases such as ChEMBL database, on the other hand, provide comprehensive bioactivity profiles for many targeted drugs, revealing considerably promiscuous drug-target interactions at the proteome-level (Bento et al., 2014). However, for most compounds, the effect of many week or unintended drug-target interactions on signaling pathways in a cellular environment still remains mostly unknown. Therefore, using the conventional binary drug-target classification might not fully reflect the pharmacological significance of some drug-target interactions. In TIMMA-R, strong and week drug-target interactions are classified into different potency classes to make the target profile of a drug as comprehensive as possible.

With multi-class drug-target data, the set relationships such as target subsets and target supersets between two drugs still hold. In such way, modelling multi-class drug-target data is similar to binary class situation. However, for a drug combination with N targets and P interaction classes, the number of possible scenarios for a target combination increase from 2^n to P^N. The increased target combination space makes the prediction computationally more expensive, while

providing at least the same level of prediction accuracy, as using binary drug-target interaction data. In addition, the prediction performance also depends on how to classify the target classes, due to the fact that there are no standard protocols to classify active and inactive interactions from a typical drug-target binding assay. In practice, it is not recommended to have too many classes to over-interpret the drug-target interactions, which may easily bring noise in the predictions (He et al., 2015).

7.3 Network construction

To facilitate the interpretation of the drug combination predictions, TIMMA-R is utilizing Boolean expression to explore the minimal configurations in the drug-target profiles which can lead to effective drug sensitivities. By minimizing the Boolean expression with the enhanced Quine-McCluskey algorithm (Quine, 1952; Quine, 1955), TIMMA-R formulated the predictions as a complete truth table using the Qualitative Comparative Analysis (QCA) package (Duşa, 2008). The target combination scenarios that lead to the same maximal response are in the minimized Boolean expression. With the minimized Boolean expression, TIMMA-R constructs a two-terminal graph that includes series and parallel components to visualize the target inhibition network. Such a graph represents the predicted response of cancer survival signaling pathways which can help researchers to pinpoint target combinations that pass a certain sensitivity threshold. The network visualization originally constructed by TIMMA-R can be imported for downstream analysis in network analysis tools like Cytoscape (Shannon et al., 2003).

8 NGS methods and data analysis

8.1 Exome-sequencing analysis

In study III, DNA was isolated from sorted CD8+ and CD4+ T-cell tumor fractions from T-PLL patients' peripheral blood (PB) mononuclear cell (MNC) samples according to the Nucleospin® Tissue Kit instructions (Macherey-Nagel). The concentrations of DNA were measured with Qubit® 2.0 Flurometer (Invitrogen) and the quality was assessed by the Agilent 2100 Bioanalyzer (Agilent Technologies).

Exome sequencing was conducted with the Illumina GAII instrument to produce 82 bp paired end reads and somatic mutation calling were performed with the VarScan 2.2.3 somatic mutation caller using the CD8+ sample as the tumor and

the CD4+ sample as the normal control. More details about the exome sequencing experiment and data analysis pipeline can be found here (Koskela et al., 2012; Pemovska et al., 2013).

8.2 RNA-sequencing analysis

In study III, RNA was extracted from the CD4+ T-cell fractions, CD4+CD8+ T-cell fractions or CD3+ T-cell fractions (healthy controls), with the miRNeasy kit including DNAse I digestion (Qiagen) according to the manufacturer's instructions. The concentrations of RNA and the quality were measured with the same machines and the same methods as DNA extraction.

RNA sequencing was performed with Illumina HiSeq2000 platform to produce more than 40x106 paired end reads with 100bp read length. The reads from both patient and healthy control samples were aligned to hg19 human reference genome with TopHat (Trapnell et al., 2009) and TopHat2 (Kim et al., 2013). Then reads were mapped to Ensembl 64 gene models and counted using the HTSeq Python package (http://htseq.readthedocs.io/en/release_0.11.1). Differential gene expression analysis was performed with DeSeq2 R package (Love et al., 2014).

9 Compound library

In Publication II and III, FIMM oncology compound libraries were used. FIMM oncology compound library is a comprehensive and constantly evolving compound collection including both FDA-approved drugs as well as investigational compounds. In Publication II, only 218 targeted compounds from the library were used (Figure 4B). Approximately half of the 218 targeted compounds were kinase inhibitors and the rest were other types of compounds, such as differentiating/epigenetic or metabolic modifiers. In Publication III, the full collection was used, which contained 525 small-molecular oncology compounds (Figure 4A). Most of the 525 compounds were kinase inhibitors and conventional chemotherapeutics. Other signal transduction modulators were included as well.

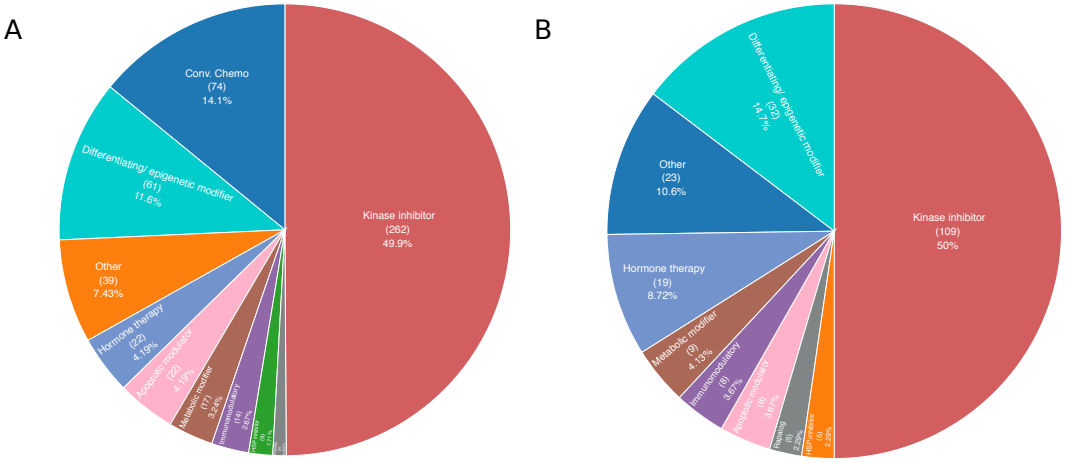

Figure 4: An overview of FIMM oncology compound library. (A) Compound classes for the full collection (525 compounds). (B) Compound classes for targeted compounds only (218 compounds). Reproduced with permission from Springer Nature (He et al., 2018a).

10 Drug sensitivity and resistance testing (DSRT)

10.1 DSRT platform

In study III, DSRT was performed on fresh peripheral blood (PB) mononuclear cell (MNC) samples, of which over 80% were leukemic cells, from T-PLL patients as well as healthy control samples. The compounds were dissolved in DMSO and preprinted on 384-well plates (Corning 3707) using an acoustic liquid handling device (Echo 550). The compounds were diluted into 10-fold dilution series of five different concentrations. The concentration ranges were chosen separately for each compound based on previous investigations of their dose-response relationships but most of compounds were ranged between 1 nM and 10000 nM. There were 16 positive controls (100 µM of benzethonium chloride) and 16 negative controls (DMSO) in each plate. The plates with compounds and controls then were kept in nitrogen pressurized StoragePods (Roylan Developments Ltd.). Before the cells were plated, the plates were suspended in 5 µl of Mononuclear cell media (MCM) and dissolved in a shaker for 10 minutes. Next, 20 microliters of single-cell suspension were plated into each well using a MultiDrop Combi (Thermo Scientific) dispenser for 72 hours' incubation at 37 °C and 5% CO_2. After the incubation, cell viability of each well was measured using the CellTiter-Glo luminescent assay (Promega) and a Pherastar FS (BMG

Labtech) plate reader. The percentage inhibition was computed by normalizing the cell viability to all the control wells.

10.2 Drug sensitivity scoring (DSS)

In study III, the compound responses in the samples *ex vivo* were quantified by drug sensitivity scoring (DSS) (Yadav et al., 2014) based on the dose-response curves. The dose-response curve is fitted with the four-parameter log-logistic function. DSS integrates the area under the fitted dose-response curve in a closed-form by summarizing all four fitted parameters: half maximal inhibitory concentration (IC_{50}), slope of the fitted dose-response curve, minimum and maximum responses of the fitted dose-response curve. The scale of DSS scores is between 0 and 50 and the calculation was done with the DSS R-package (https://bitbucket.org/BhagwanYadav/drug-sensitivity-score-dss-calculation).

11 Drug combination prediction and testing (DCPT) platform

11.1 Drug combination plate design using FIMMCherry

In Publication II, the drug combination testing was performed in 384-well plates, one of which could accommodate six pairs of drug combinations. Each drug combination pair was tested in an 8 × 8 dose matrix. In the matrix, the first row and the first column were cells treated with single drug with a series of one blank and seven half-log diluted concentrations, whereas the rest was cells treated with drug combination. DMSO as negative control is plated in the first well of the dose matrix while 100 µmol/L benzethonium chloride (BzCl) as positive control is in the last well.

To transfer the compounds into the plates according to the dose matrix defined above, a picking list containing information of where the source and destination plates are, transfer volumes for the compounds is needed. To generate such a complex list manually requires lots of time and prone to errors. To avoid errors and save time, an in-house tool called FIMMCherry (He et al., 2018a) was developed to generate the picking list automatically and effortlessly in the Publication III. FIMMCherry is a desktop application, which provides an easy-to-use interface with user guidelines (Figure 5). FIMMCherry is implemented in Python and Qt framework, which allows FIMMCherry to run in different operating systems such as Windows, MacOS, and Linux.

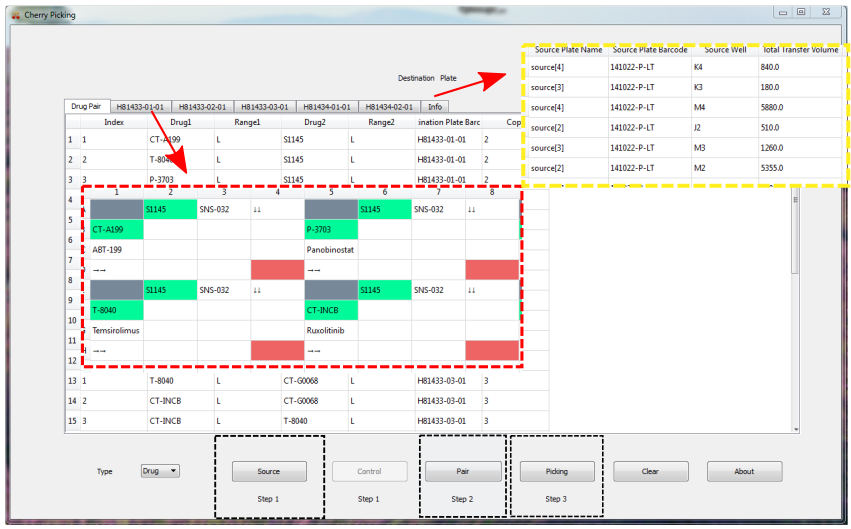

Figure 5: Drug combination plate design using FIMMCherry. Reproduced with permission from Springer Nature (He et al., 2018a).

There are three steps to generate the pick list with FIMMCherry. In the step one, FIMMCherry requires users to load two input files as tab-delimited text files: a source plate file providing information about the compound stocks (compound identification, available concentration ranges, source plate identification, and well identification) and a control file containing information where the controls are located on the plate. In the step two, FIMMCherry requires a drug combination file containing compound pairs to be tested.

After the required input files are loaded correctly, FIMMCherry displays the layout of the plates accordingly (the red inset box in Figure 5), as well as the input files (the yellow inset box in Figure 5). Then FIMMCherry generates a pick list for dispensing compounds in the step three. The generated pick list is compatible with Labcyte Echo dispenser, which can transfer the compounds from source wells to destination wells in a non-contact fashion in 2.5 nL droplets. The pick list is compatible with Echo Cherry Pick software to produce plates where compounds are pre-dispensed with no further modifications (Kulesskiy et al., 2015).

11.2 Drug combination testing

In Publication II, drug combination testing was performed on fresh peripheral blood (PB) mononuclear cell (MNC) samples from T-PLL patients as well as unsorted MNC cells from healthy donors. In a similar way as drug sensitivity testing, 20 µL of single-cell suspension were plated into the pre-dispensed plate using the plate design described above. After incubating the cells in the plates for 72 hours at 37 and 5% CO2, cell viability of each well was measured via reading luminescence in the plates using the CellTiter-Glo luminescent assay (Promega) and a Pherastar FS (BMG Labtech) plate reader. The percentage inhibition was calculated by normalizing the cell viability to the control wells.

11.3 Feature integration

For a polypharmacology-based machine learning method to predict patient/control sample-specific drug combination responses, a patient/control sample-specific feature vector combining a target feature vector with a genomic feature vector was formulated for each drug combination. A target feature vector contains binary values which indicate potency of the targets for each single drug or their combinations. One in the vector indicates the target is a potent target while zero means the target is not potent. For a single drug, all the potent targets are marked as 1, while the non-potent targets are marked as 0. The target feature for drug combinations is constructed in a similar way by marking potent targets of both drugs as 1 while the rest as 0 in the combined target space of both compounds. To construct the target feature, we collected the bioactivity data across the compounds and their cancer-specific targets from our crowdsourcing bioactivity data platform, DrugTargetCommons (https://drugtargetcommons.fimm.fi) (Tang et al., 2018; Tanoli et al., 2018). For each compound-target pair, we obtained the normalized bioactivity value by comparing the median value of the log-transformed bioactivity values such as K_d and K_i again the smallest bioactivity value among all the targets of this compound. After binarization with a compound-specific cutoff (log-fold-change less than 2), we obtained the target feature of 366 targets for each single drug or their combinations.

A genomic feature vector contains the patient/control gene expression data extracted from RNA-seq and mutation profiles of the same sample extracted from exome sequencing. To construct the patient/control gene expression feature, a set of 634 genes were selected due to their functionalities associated with cancer progress, tumor suppressors, or drug response or resistance. In the vector, the raw counts and normalized fragments per kilobase of exon model per million reads

mapped (FPKM) data for these genes are utilized. The mutation vector is a binary vector of 68 entries as in total 68 mutated genes were identified in the three T-PLL patients using VarScan2 algorithm with significance cut-off ($P \leq 0.01$). In the mutation vector, one indicates a mutation is identified in the particular patient.

11.4 Predictive modeling

In DCPT, a supervised machine learning method called random forest was applied to predict the quantitative responses of two combined drugs. Random forest is an ensemble machine learning method which averages multiple deep decision trees (Breiman, 2001). The integrated feature vectors defined above were used as input features to train a model for predicting single-compound DSS values in each sample separately. To evaluate the model performance and to avoid model overfitting, one of the commonly used cross validation methods, called leave-one-out cross validation (LOO-CV), was utilized. When using LOO-CV, the model is applied once to each of data as test data while using the rest for model training. In such way, LOO-CV could estimate the general model performance.

With the trained, tuned, and sample-specific models, the compound combination responses were predicted by treating a compound combination as a new single "compound", whose targets are combined targets of each single compound. For each sample, the combination responses of all the possible pairwise combinations among 218 compounds (23,653 combinations) were predicted. The predicted combination responses were then compared against the predicted single compound responses to classify each the combination as a synergistic or non-synergistic combination for each sample individually, using the highest single agent (HSA) model. More specifically, if the combination response of a compound combination A+B was higher than the maximum response of compound A and compound B, the combination A+B was classified as synergistic. Otherwise, combination A+B was classified as non-synergistic.

Those combinations that were found to be synergistic in the patient samples using HSA model were required to be non-synergistic in the healthy control sample. To further avoid toxic combination effects caused by any toxic single compound, the potential combinations were further filtered with single-compound responses (DSS < 20 in both patient and control samples with some exceptions, where DSS < 20 was achieved only when excluding the highest concentration point). After the filtering, the remaining combinations were considered as patient-specific as well as cancer selective combinations. Such combinations are likely to have less

severe side effects, which makes the translation into clinical practice more easily. The R scripts for the predictive modelling are freely available at github (https://github.com/hly89/ComboPred).

11.5 Quantification of drug combination synergy using SynergyFinder

After selecting the model-predicted synergistic combinations which are patient-specific and cancer-selective, the next step is to validate these combinations *ex vivo* using combination plate designed using tools from Publication III. To quantify the degree of combination effects, we calculated an overall synergy score across all the tested concentration combinations using the open-source R-package SynergyFinder available at bioconductor from Publication III (https://bioconductor.org/packages/release/bioc/html/synergyfinder.html) as well as its web application (https://synergyfinder.fimm.fi) (Ianevski et al., 2017; Ianevski et al., 2020).

In Publication III, we demonstrated how to use R package SynergyFinder to compute synergy scores for drug combinations (ibrutinib combined with ispinesib and ibrutinib combined with canertinib) tested in 6 × 6 matrix design for treatment of diffuse large B-cell lymphoma (DLBCL) (Mathews Griner et al., 2014). The input data in csv format which contains the following information about the dose-response matrix for a drug combination: BlockID (an identifier for a drug combination), Row and Col (indexes of the wells in the drug combination plate), DrugRow (the row-wise drug name), DrugCol (the column-wise drug name), and Response (normalized drug responses). SynergyFinder provides functions to fit a four-parameter log-logistic model to summarize the single drug effects and to visualize the effects via a dose-response curve (Figure 6). In addition, the input dose-response matrix can be visualized as a heatmap (Figure 6).

To quantify the degree of combination effects, a reference model is required. There are four reference models available from SynergyFinder, namely, highest single agent (HSA) model, Loewe model, Bliss model, and zero interaction potency (ZIP) model. Once a reference model is selected, SynergyFinder provides functions to calculate the synergy scores across all the tested concentration combinations and to visualize the results as 2D and 3D synergy plots (Figure 7). Such plots are very informative when identifying the dose regions of interest such as the most synergistic dose regions. In addition, SynergyFinder provides an option to correct the negative response values by using the average of the minimum responses of the two single drugs.

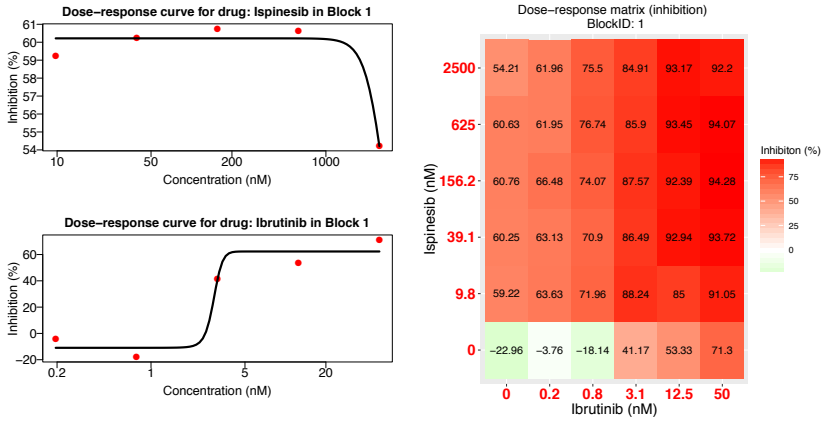

Figure 6: Plots for single-drug dose-response curves and drug combination dose-response matrices (ibrutinib and ispinesib). Reproduced with permission from Springer Nature (He et al., 2018a).

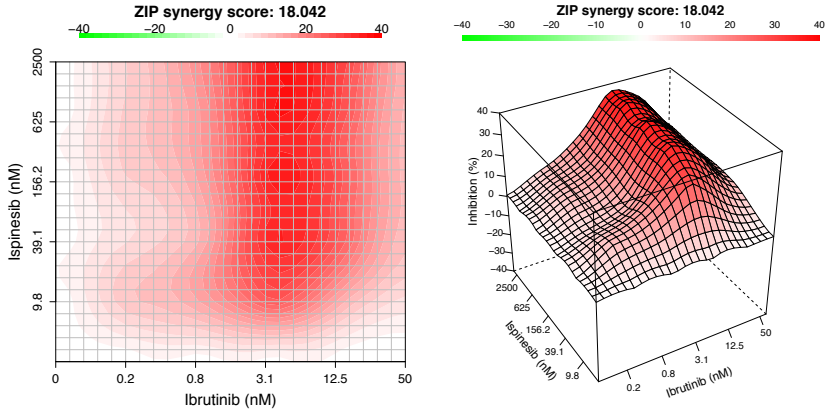

Figure 7: 2D and 3D synergy plots for ibrutinib and ispinesib based on ZIP model. Reproduced with permission from Springer Nature (He et al., 2018a).

In Publication II, ZIP model was used to compute synergy scores for all the drug combinations tested in DCPT patient samples as well as healthy controls. ZIP model assumes the effect of non-interaction, which is used as reference to compute the synergy score, to be achieved when two drugs do not potentiate each other. In such case, both the assumptions of the Loewe model and the Bliss model could be met, making ZIP model a synthesis of both Loewe and Bliss models.

11.6 Patient-specific network visualization

For each patient-specific drug combination prediction, the last step of the DCPT pipeline was to construct a cancer vulnerability network, which was based on comprehensive cancer signaling networks including 6,296 proteins and 79,083 interactions (Zaman et al., 2013; Kirouac et al., 2012). Firstly, the selective drug targets of the predicted drug combinations and patient-specific genetic aberrations and molecular changes including mutated or dysregulated genes were mapped to the cancer signaling networks. Then, the shortest paths to connect the drug targets with patient-specific genetic aberrations and molecular changes were computed to construct the pathway cross-talk network. Such patient-customized networks provide a systematic way to investigate the mechanisms of action of the drug combinations in the patient's cellular context, as well as to discover the potential predictive biomarkers for the synergistic responses.

An easy-to-use R/Shiny web-application with user detailed user instructions (https://patientnet.fimm.fi) was developed in Publication II to construct the cancer vulnerability networks automatically. The required input data are patient-specific exome/RNA-seq data such as gene names and their expression levels, single compound responses such as DSS, and target annotations containing drug names and their target names. The output of the web-application is the patient-customized network diagram which can be visualized in the web browsers and downloaded in an html format. The output is also compatible with Cytoscape (versions 2.7 and 3.4).

Results

12 Predicting synergistic multi-targeted drug combinations using TIMMA-R

12.1 Computational speed

TIMMA-R implementation was designed to be efficient and scalable for large-scale analyses by using vectorization and multi-dimensional matrix computation. It takes less than 50 seconds on average to run predictions for a typical dataset with 50 drugs and 500 targets on a normal desktop computer, which is much faster than the original MATLAB implementation that took over 250 seconds (He et al., 2015). The current implementation aggregates the information from each drug sequentially, which could be further speeded-up by aggregating parallelly using multiple processors.

12.2 Sensitivity analysis

TIMMA-R uses the polypharmacological information to predict the combination responses, according to the recent proteome-wide bioactivity studies that have shown that there are many drugs that interact with more than one or two target (Reddy and Zhang, 2013; Davis et al., 2011). The experimental uncertainties in the interactions between a drug and its multiple targets could affect the robustness. Therefore, we performed a sensitivity analysis to test the robustness of TIMMA-R.

In the sensitivity analysis, we used 50 drugs and 100 targets with simulated binary drug-target interaction, where 1 indicates a target and 0 indicates a non-target. To introduce experimental uncertainties in the drug-target interaction data, we flipped a target to non-target (1 to 0, false negative) or vice versa (0 to 1, false positive) for up to 30% of the data. The number of positive interactions in the data was set to 30% of total interactions. We collected the prediction results after repeating the flipping process for 100 times. To compare the averaged predictions from before and after the flipping, we used the RV-coefficient (Escoufier, 1973; Robert and Escoufier, 1976). As expected, the RV-coefficient decreased when the false positive rate or the false negative rate increased (Figure 8), but being still significantly higher than the random predictions ($P < 0.001$, paired t-test). Such result shows that TIMMA-R can be applied to real case studies, where the

drug-target interactions have been either experimentally validated or manually curated.

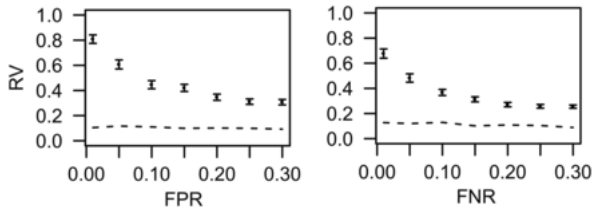

Figure 8: RV-coefficients for simulated drug target interaction with different false positive rates or false negative rates. Reproduced with permission from Oxford University Press (He et al., 2015).

12.3 Leukemia case study

We tested the performance of TIMMA-R using data from a study on the kinase pathway dependence in leukemia patients using 65 small-molecule kinase inhibitors as chemical probes, tested on 151 leukemia patient-derived samples *ex vivo* (Tyner et al., 2013). In this case study, the effective drug or drug combinations for these patient-derived samples could be readily utilized for clinical applications.

From the Tyner study, we collected the drug-target interaction data including 65 small molecule kinase inhibitors and 322 targets. Each drug-target interaction has been classified into one of the six classes labelled as [0, 1, 2, 3, 4, 5]. We scaled the drug sensitivities (IC_{50}) into [0, 1] interval, where a higher value indicates a stronger effect, using equation (12).

We evaluated if multi-class drug-target interaction can lead to any improvement by comparing with using binary drug-target interaction data. The binary drug-target interaction data were obtained by treating all the non-zero classes as one group, which includes both the strong and weak drug-target interactions. We used mean absolute error (MAE) in the leave-one-out cross-validation (LOO-CV) to compare the performance.

As shown in Figure 9, no significant difference was observed when using either binary drug-target interaction data or multi-class drug-target interaction data in terms of the prediction accuracy. The average differences in MAE across the 151 samples is 0.0091, which is quite minimal compared to the actual MAEs. The results indicate that introducing more potency classes may not always lead to a better prediction accuracy. When introducing more classes, the model search

space increases from 2^N to P^N, where P is the number of classes and N is the number of targets, which in general requires more training data to acquire better performance. In addition, if the number of drugs or the number of targets does not increase significantly, more classes will lead to very sparse training data, which introduces higher uncertainty in the model predictions.

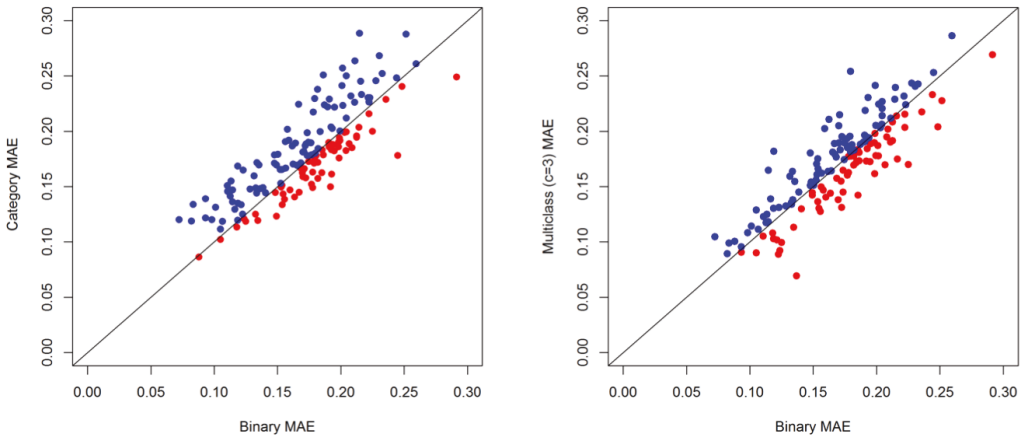

Figure 9: Scatterplots of MAE (Mean absolute error) in the LOO-CV using either binary or multiclass drug-target interaction data. Reproduced with permission from Oxford University Press (He et al., 2015).

To further explore the impact of classification on the prediction accuracy, we tested a simpler scheme by considering only 3 classes: class 0 is still the inactive group; class 4 and 5 are merged into one group as the strong interaction group; the rest are grouped together as the weak interaction group. Such a simplified classification indeed improved the average MAE from 0.1792 to 0.1754, with increased number of cases (from n = 61 to n = 74), where it outperformed the binary data (Figure 9). However, the overall difference remained minimal, which concluded that if the drug-target interaction data is of high-quality (e.g. manually curated from experimental results), the classification schemes seem to have little impact on the prediction accuracy of TIMMA-R.

12.4 Dedifferentiated liposarcoma study

To further evaluate the performance of TIMMA-R, we tested TIMMA-R on data from another study (Miller et al., 2013). Unlike the Tyner study which contains large-scale drug-target interaction data, only primary target information for 14 drugs (Table 1) is available in the original Miller study, which makes the TIMMA analysis less straightforward. Such primary target information is useful for

explaining the intended therapeutical actions of the compounds, but provide limited information on the polypharmacological effects of a drug combination. In such case, we still need to obtain the quantitative drug-target interaction profiles at the proteome level. We found that 9 of these 14 drugs from the EMD Millipore company are also included in another study which has profiled 158 drugs using a broad activity screen (Gao et al., 2013). With data obtained from a global drug-target profiling (Gao et al., 2013; Tyner et al., 2013), as well as data from the integration of multiple studies (Tang et al., 2014), we obtained enriched drug-target data shown in Table 2.

In the original Miller study, the dose-response values of single and paired drug combinations on the tumor-derived cell line DDLS8817 are available. However, only 13 of 14 drugs were found, whereas Rapamycin sensitivities were missing. Each of the 13 drugs was tested with 7 doses, where the responses were measured as the relative cell metabolic activities. The cell viability IC_{50} was derived using the logistic curve-fitting with the drc package in R (Ritz et al., 2015). The drug sensitivity values were calculated in the same way as in the Tyner study.

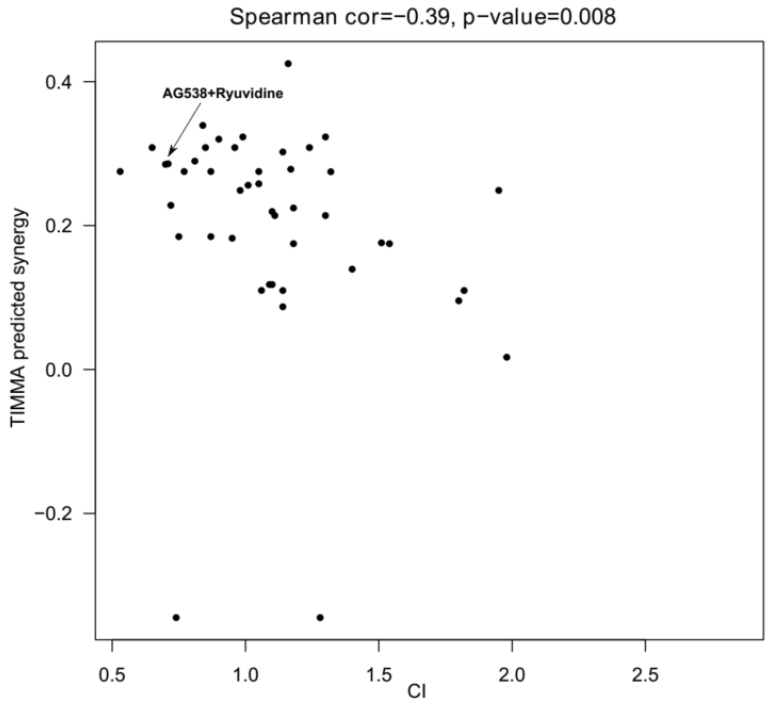

Figure 10: Comparison of the predicted synergy scores and the combination index (CI) over 12 compounds from the Miller study (Miller et al., 2013). Reproduced with permission from Oxford University Press (He et al., 2015).

We performed the TIMMA analysis on the 13 drugs (Rapamycin sensitivities were missing). The spearman correlation between the TIMMA- predicted synergy scores and the combination index (CI) scores from the Miller study was -0.39 with a p-value of 0.008 (Figure 10). As low CI indicates stronger synergistic effect, negative correlation was expected. In the TIMMA prediction, AG538 is synergistic with ryuvidine with a synergy score of 0.285. Their targets IGF1R and CDK4 in combination was also predicted a sensitivity of 0.997, which is in line with the major finding in the Miller study (Miller et al., 2013). Furthermore, in the derived target inhibition network alternative target combinations, such as inhibitions of ABL1 and FLT3, or ABL1 and ACVR1B, might be also synergistically blocking the cancer survival pathway (Figure 11).

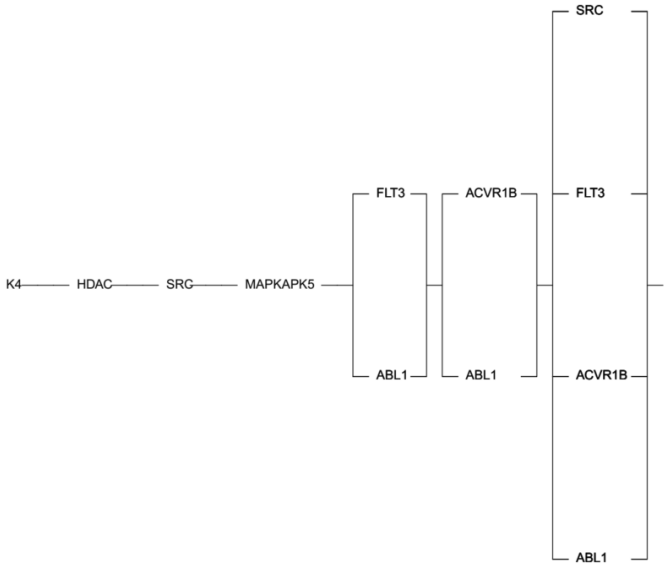

Figure 11: The target inhibition network for the DDLS-derived cancer cells. The network connections indicate inhibiting FLT3 and ABL1 or ACVR1B and ABL1 might block the cancer survival pathway. Reproduced with permission from Oxford University Press (He et al., 2015).

13 Predicting patient-specific and cancer-selective drug combinations using DCPT

The molecular pathways driving disease progression and treatment resistance are highly redundant. Even within the same type of cancer, those pathways often vary among each individual patient. To tackle this complex rewiring of pathway cross-

talk, we implemented a computational and experimental drug combination prediction and testing (DCPT) platform as an extension of TIMMA model (He et al., 2015; He et al., 2018b; Tang et al., 2013). DCPT was designed to prioritize potential drug combinations *in silico* and test them *ex vivo* in patient-derived samples in order to identify patient-specific and cancer-selective drug combinations for individual cancer patients.

Using DCPT platform, we carried out successful drug combination predictions for three T-PLL patients with varying phenotypes and genotypes (Figure 12). For each patient, a patient-customized network was constructed by connecting the targets of the most effective drug combinations with the patient-specific genomic and molecular alterations to understand the potential mechanisms of these drug combinations in the particular T-PLL case.

	1508	1263	1409	Control
Phenotype	CD4+	CD4+CD8+	CD4+	CD3+
ATM	Deletion	Deletion	Deletion	Wild type
TCL1A	Translocation	Translocation	Wild type	Wild type
IL2RG	K315E, 55%	Wild type	Wild type	Wild type
STAT5B	P702S, 18%	Wild type	Wild type	Wild type

Figure 12: T-cell phenotypic and genetic characteristics of the three T-PLL patient cases and the healthy control. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

13.1 Case 1263

For T-PLL patient case 1263 with a generally sensitive *ex vivo* single drug response profile based on her treatment-naïve diagnostic sample (Figure 13), the DCPT platform was still able to identify selective and synergistic combinations. One of such combinations was bryostatin 1 (investigational protein kinase C activator) with nutlin-3 (apoptotic modulator) (Figure 14A). This combination

inhibits the interaction between MDM2 and p53, activating p53 at higher concentrations. When testing bryostatin 1 combined with three other MDM2 inhibitors (Figure 15), selective synergy was also shown in at lower concentrations of SAR405838, which effectively stabilizes p53 and activates the p53 pathway (Bill et al., 2016). The other two MDM2 inhibitors (idasanutlin and AMG-232), however, did not show synergy in this patient.

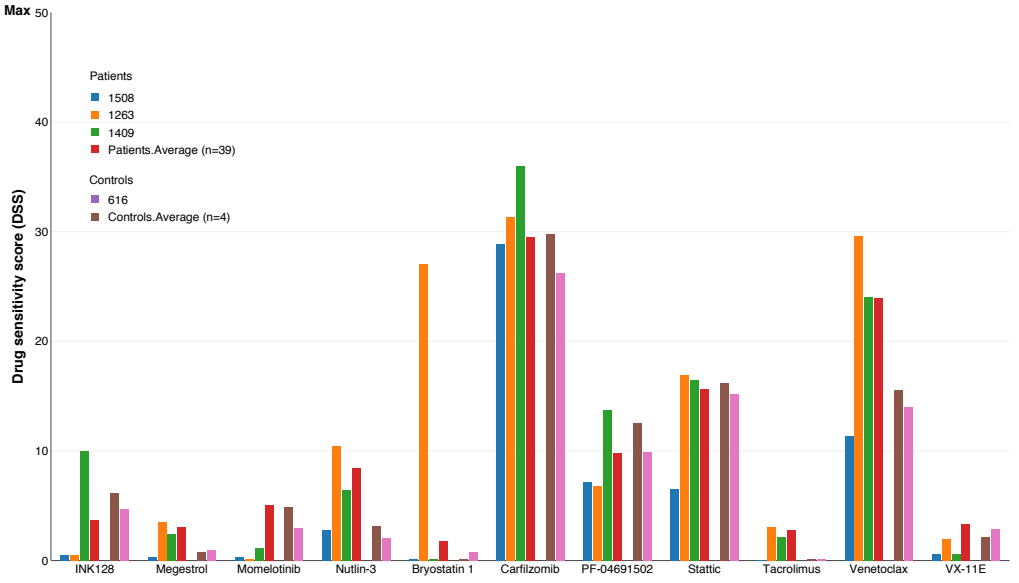

Figure 13: The *ex vivo* drug sensitivity profiles of the single compounds tested in combinations in the patient cases as well the control sample. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

Due to the dose and exposure-dependent activity of bryostatin 1, it may be difficult to treat patients with bryostatin 1 clinically. Therefore, we predicted and experimentally validated more clinically actionable combinations for this patient. Selective synergy was confirmed between ventoclax (BCL2 inhibitor) and tacrolimus binding to intracellular FKBP12 (Figure 16A). Tacrolimus inhibits calcineurin, which has been shown to mainly activate overexpressed NFAT transcription factors in T-PLL tumor cells (Dutta et al., 2017), including patient 1263.

To further investigate the role of FKBP12 in this patient, we combined ventoclax with another approved drug, temsirolimus, which binds to FKBP12 as well to inhibit the activity of mTORC1 (Figure 17). Synergy was shown at lower concentrations of these two drugs.

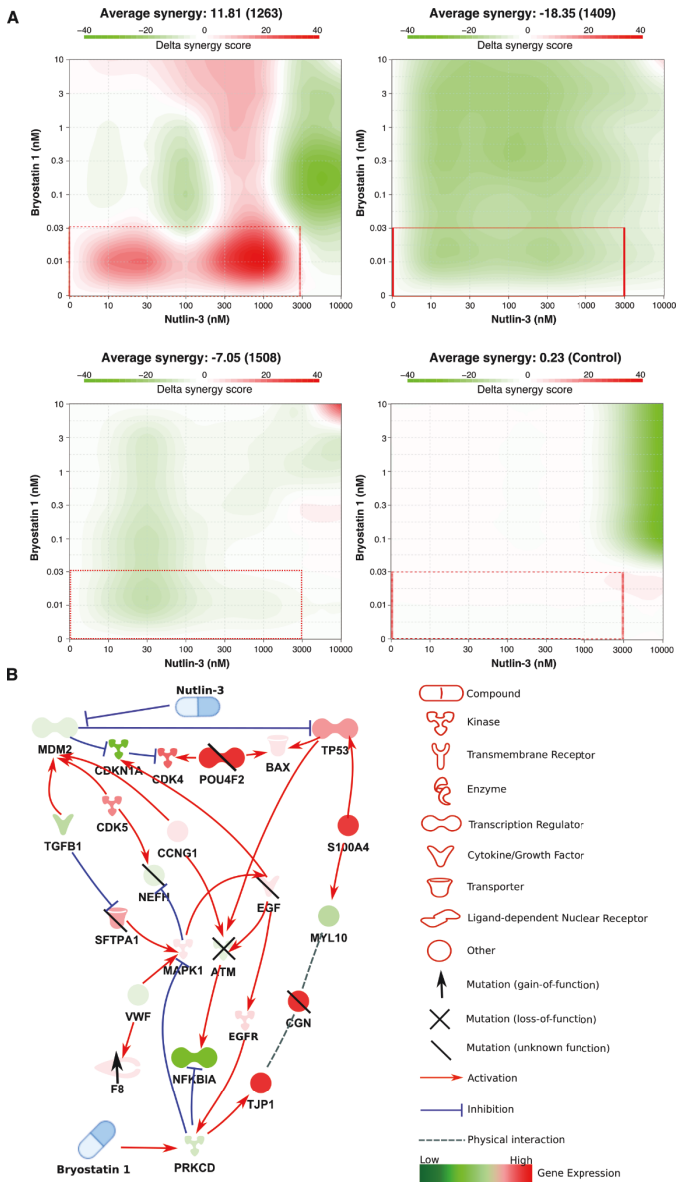

Figure 14: Experimental testing of the selective combination (bryostatin 1 and nutlin-3) for patient 1263. (A) the 2D synergy plots of the combination of bryostatin 1 and nutlin-3 tested in all patients and healthy control ex vivo. (B) patient 1263-specific signaling network constructed by connecting the targets of bryostatin 1 and nutlin-3/SAR405838 with the patient-specific somatic mutations. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

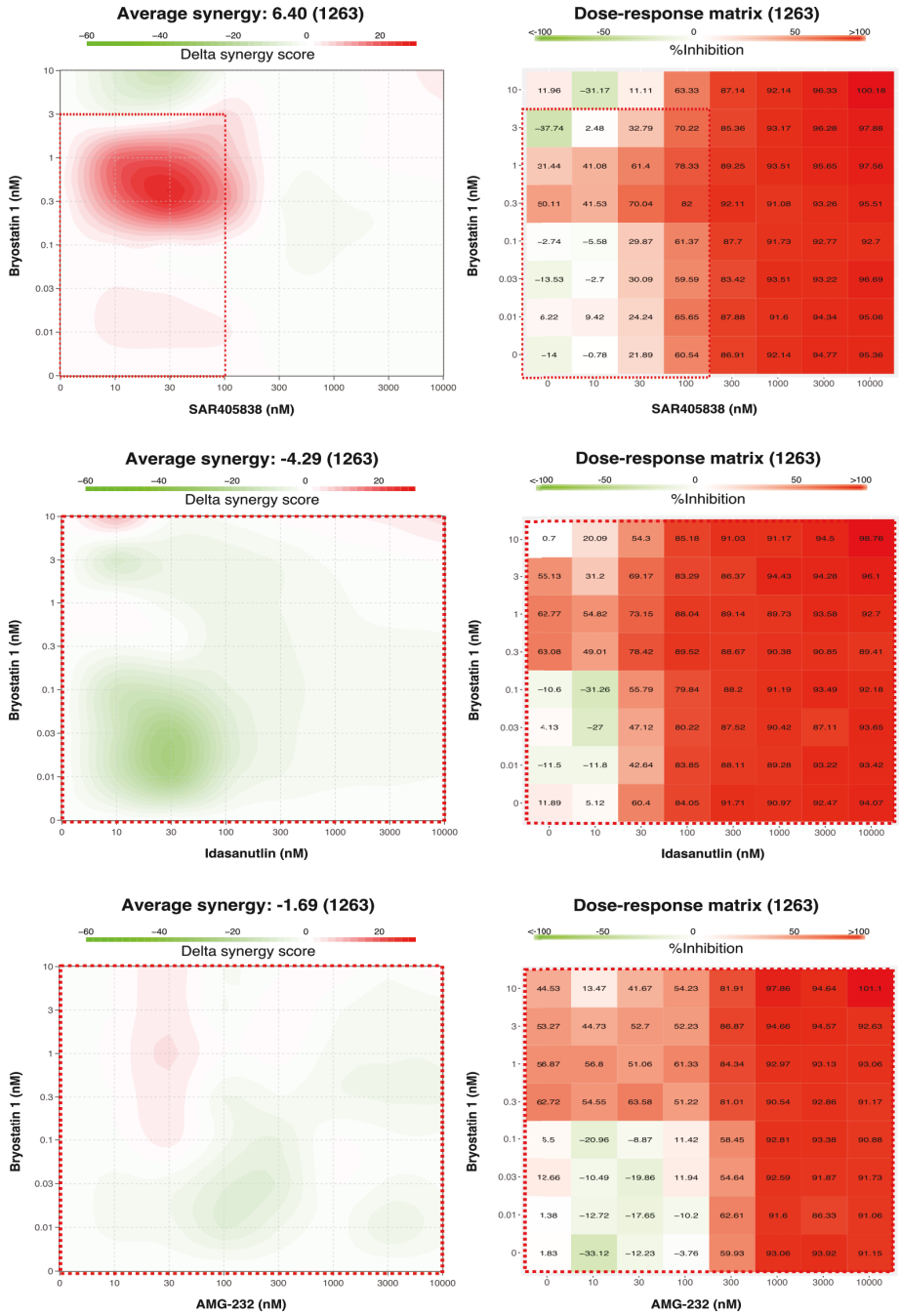

Figure 15: 2D synergy plots and dose-response heatmaps of combinations between PKC activator (bryostatin 1) and MDM2 inhibitors (SAR405838, idasanutlin, AMG-232) for patient 1263. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

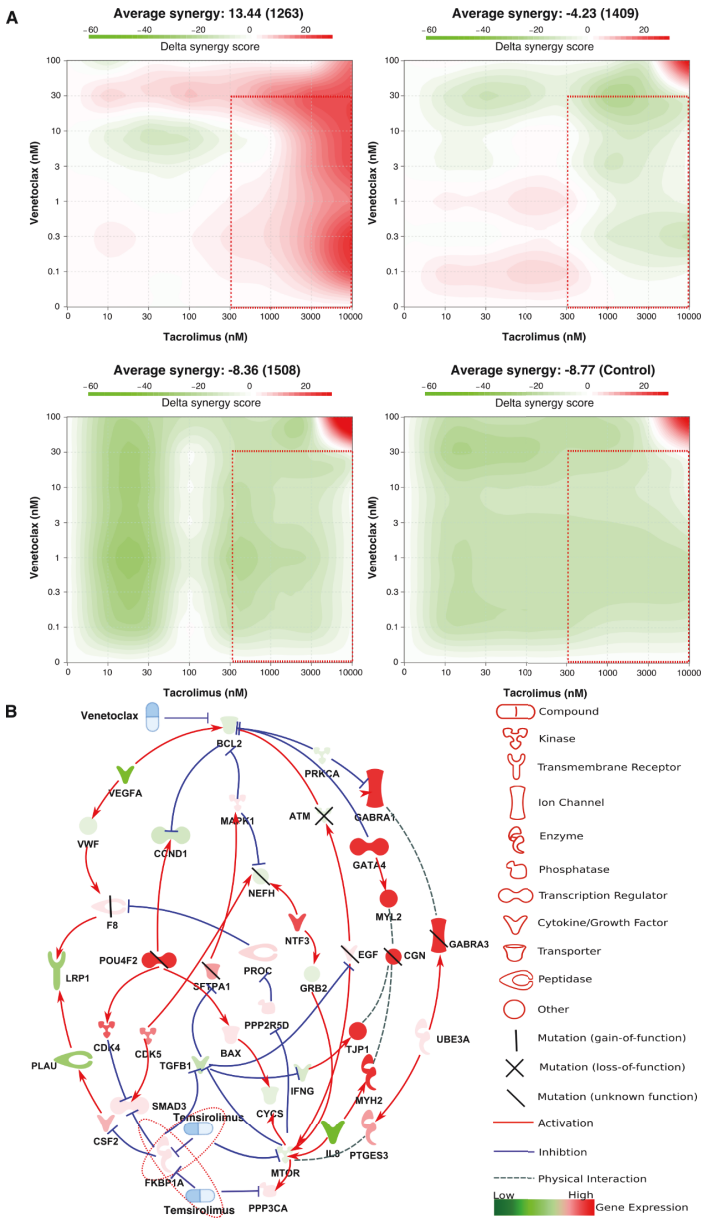

Figure 16: Experimental testing of the selective combination (venetoclax and tacrolimus) for patient 1263. (A) the 2D synergy plots of the combination of venetoclax and tacrolimus tested in all patients and healthy control ex vivo. (B) patient 1263-specific signaling network constructed by connecting the targets of venetoclax and tacrolimus/temsirolimus and FKBP1A with the patient-specific somatic mutations. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

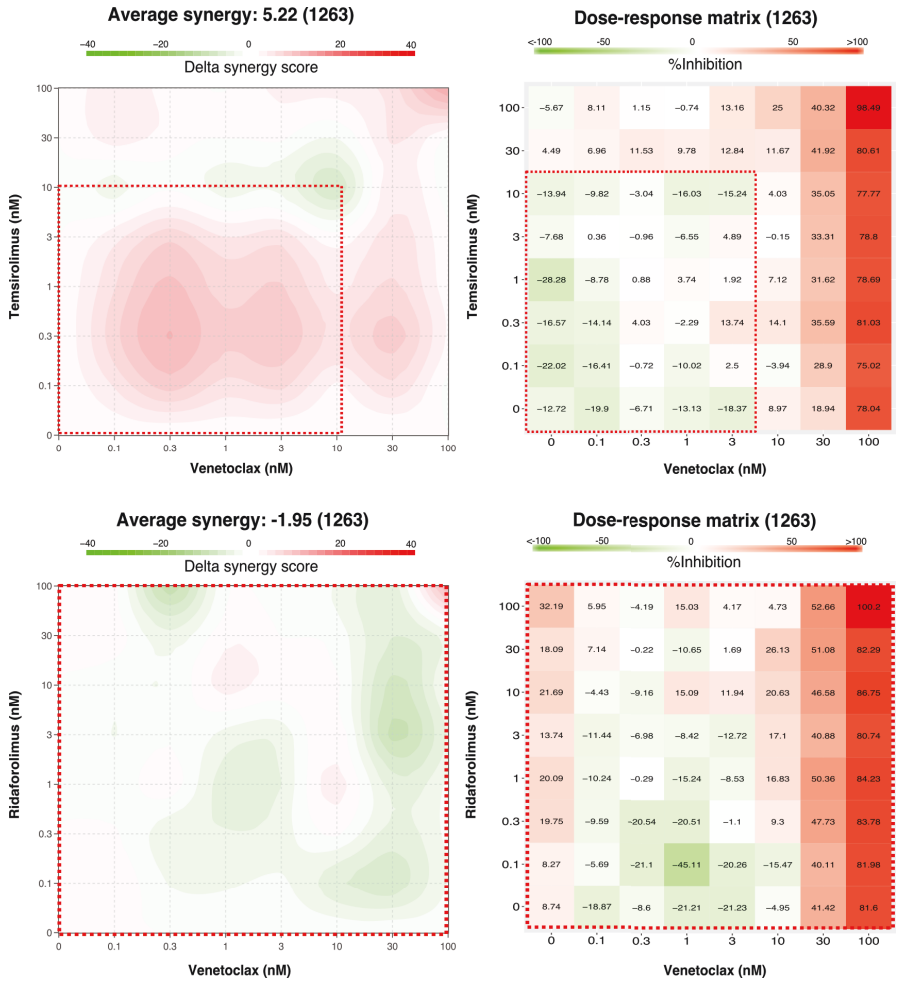

Figure 17: 2D synergy plots and dose-response heatmaps of combinations between BCL-2 inhibitor (venetoclax) and inhibitors binding to FKBP12 (temsirolimus, ridaforolimus) in patient case 1263. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

DCPT platform enabled us to construct customized cancer vulnerability network diagrams for all the selective combinations by connecting the targets of these combinations to patient-specific genetic and molecular aberrations via the shortest signaling paths (Figure 14B, 16B). The customized cancer vulnerability network for patient 1263 suggests that the selective synergies may be driven by the MAPK pathway. The selective synergy from another predicted 1263-specific combination, bryostatin 1 and the investigational JAK1/2 inhibitor momelotinib (a drug for myelofibrosis undergoing phase III clinical trials) further supported the functional importance of the JAK-MAPK pathway rewiring in patient 1263.

13.2 Case 1409

For T-PLL patient 1409, the *ex vivo* single-compound sensitivity profile was created based on patient's sample treated with alemtuzumab, a monoclonal antibody which binds to CD52 on the surface of mature T lymphocytes, allowing the immune system to pick out and kill the leukemic T cells. Using DCPT platform, combination of carfilzomib (proteasome inhibitor approved for the treatment of relapsed or refractory multiple myeloma) and static (STAT3 inhibitor) was identified to show selective synergistic combination effect in the post-treatment phase (Figure 18A). It was surprising to see that the selective synergistic effect was observed on the lower concentration end of carfilzomib (0-3 nmol/L) even though the effect was only moderate with delta score of 5.1. Since the sample from patient 1409 was not treatment-naïve, it may explain the degrees of synergies for patient 1409 were somewhat lower than other patient cases. However, it is often considered enough for clinical therapy benefits by having just additive effect that is just over the single-agent efficacies using HAS model for synergy scoring.

Based on the cancer vulnerability network diagram for patient 1409 (Figure 18B), the synergistic effect between carfilzomib and static may be driven by the rewired cross-talks between PI3K/mTOR pathways. To investigate this hypothesis, we further tested more combinations based on the prediction results from DCPT using two investigational kinase inhibitors, sapanisertib and PF-04691502, which are targeting at mTOR and/or PIK3CA pathways (Figure 18C). All the additional combinations showed selective synergy in patient 1409 samples comparing to the healthy control samples. With this case, we demonstrated how DCPT help researchers make hypotheses about the mechanisms of selective synergies for a single patient based on the customized cancer vulnerability networks. These hypotheses can be further validated or invalidated with additional target perturbations.

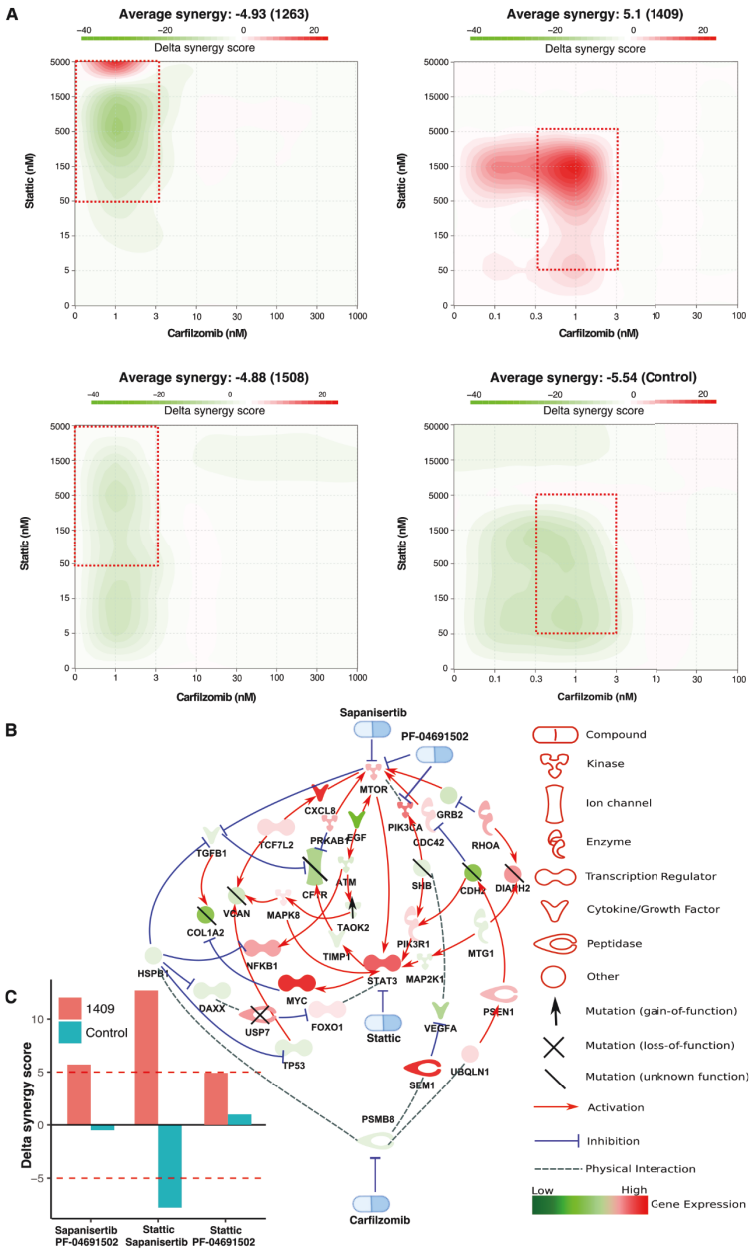

Figure 18: Experimental testing of the selective combination for patient 1409. (A) the 2D synergy plots of the combination of static and carfilzomib tested in all patients and healthy control ex vivo. (B) patient 1409-specific signaling network constructed by connecting the targets of static and carfilzomib with the patient-specific somatic mutations. (C) experimental testing of three additional combinations targeting mTOR and/or PIK3CA pathways. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

13.3 Case 1508

For T-PLL patient 1508, the *ex vivo* single-compound responses were profiled on the treatment-naïve diagnostic sample (Figure 13), showing that patient 1508 had strong resistance towards most of the single compounds. Such resistance may be driven by an activating mutation in β-catenin (CTNNB1, K672Q) that could induce malignant T-cell proliferation by promoting genomic instability. However, the DCPT platform enabled us to identify a combination showing a relatively strong synergistic effect with delta synergy score of 10.14 at relatively low concentrations: VX-11E with megestrol (Figure 19A). VX-11E is a potent ERK2 (MAPK1) inhibitor and megestrol is a hormone therapy progestogen that acts also as an agonist of the glucocorticoid receptor NR3C1. This combination showed only selective synergistic effect on patient 1508, but not on the other patients as well as the control samples.

With DCPT platform, we constructed the patient-customized pathway cross-talk network by connecting four targets of VX-11E and megestrol (MAPK1, PGR, AR, and NR3C1) to the patient-specific genetic and molecular aberrations. This network suggested that the selective synergy may be linked to the IL2 and IL5 pathways (Figure 19B). These two pathways were activated in patient 1508 based on the exome and RNA-seq profiles. The activation of these two pathways may be caused by the gain-of-function mutations in IL2 gamma receptor (IL2RG and K315E) with signal transducer and activator of transcription 5B (STAT5B and P702S) accompanied by overexpressed IL5 and JAK1. Such mutation combination (STAT5B-IL2RG) appeared only in patient 1508, but not in our other T-PLL patients, which may explain why VX-11E with megestrol were not showing synergistic effect on other T-PLL patients in a previous study (Andersson et al., 2017).

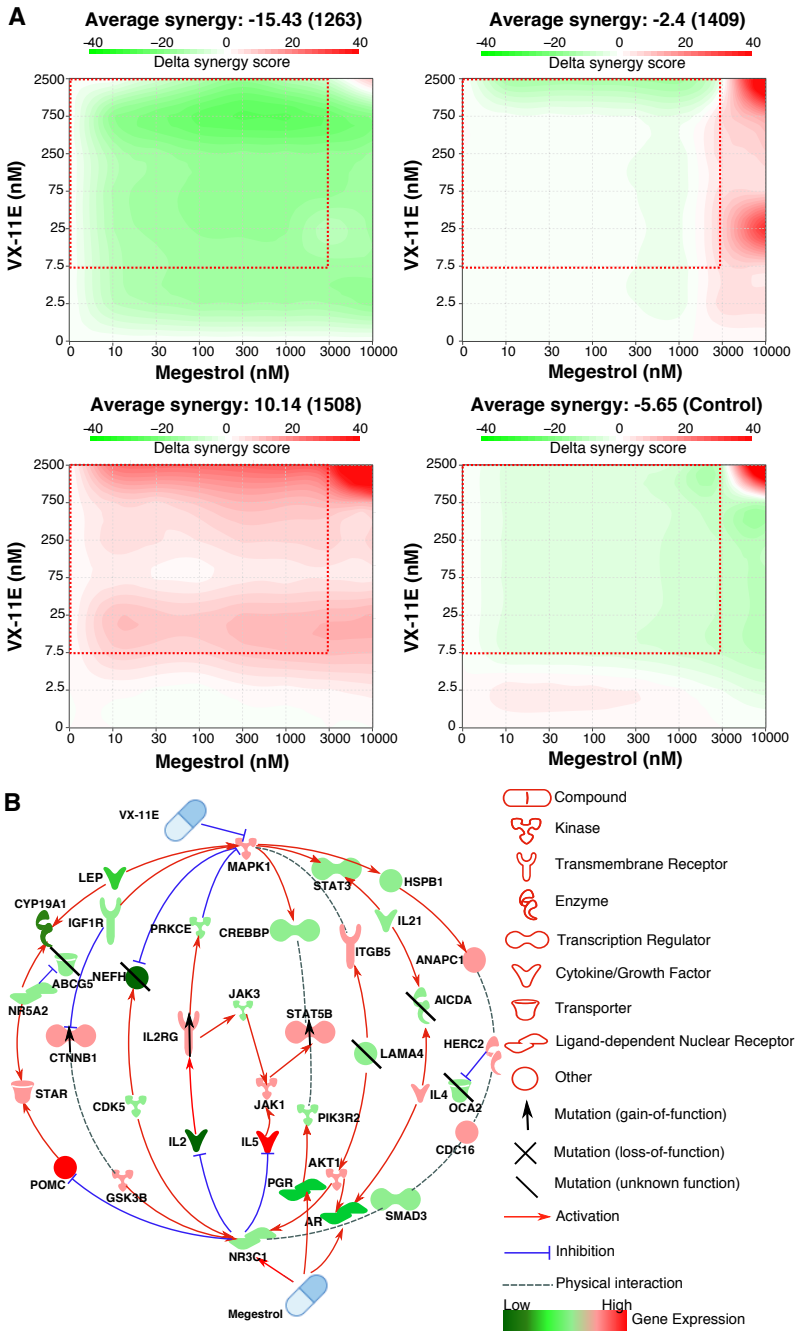

Figure 19: Experimental testing of the selective combination (megestrol and VX-11E) for patient 1508. (A) the 2D synergy plots of the combination of megestrol and VX-11E tested in all patients and healthy control *ex vivo*. (B) patient 1508-specific signaling network constructed by connecting the targets of megestrol and VX-11E with the patient-specific somatic mutations. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

Discussion

In the past decade, personalized medicine and drug combination therapy have gained more attention in medical research. In contrast to traditional single drug therapy, which may easily lead to drug resistance for complex disease such as cancer, drug combination therapy is promising alternative to reduce drug resistance by targeting multiple disease related pathways simultaneously. Such therapy could also reduce the drug toxicity as the dose for each drug when combined is generally lower than using them separately as monotherapy. However, there are currently two main challenges in developing drug combination therapy for cancer patients. The first challenge is how to identify drug combinations more efficiently and less costly for each individual cancer patient. The most straightforward way is to screen all possible drug combinations for each patient sample or cell line. Even with advanced technologies in drug screening platforms, such exhaustive testing is still very time consuming and costly. Further, the scarce patient cells rarely allow for testing hundreds of combinations for each patient. The second challenge is how to quantify the effects of the two combined drugs, which is very closely related to the first challenge. After screening drug combinations in such preclinical drug screening platforms that provide the phenotypic endpoint responses *in vitro* or *ex vivo*, the next step is to evaluate the degree of combination synergy in a systematical manner. Such systematical evaluation is crucial for the prioritization of effective and safe drug combinations for further clinical examination (Yadav et al., 2015). This thesis presents computational methods and easy-to-use software and open-access implementation to address these two challenges by utilizing powerful computational models with integrated genomic and functional data.

Publication I demonstrated that computation models can be used to accelerate the identification of synergistic drug combination in anticancer research by utilizing the integrated drug-target and drug sensitivity profiles, which are becoming increasingly available in many cancer programs. In the study, the TIMMA model was applied to two real case studies, the Tyner study and the Miller study. In the Tyner study, manually-curated drug-target interaction data as well as the drug sensitivity data were available for 151 patient-derived samples, which make it an ideal case to illustrate the full analytical pipeline of the TIMMA-R package. The Tyner study showed that introducing more target classes does not necessarily increase the prediction accuracy, compared to using only binary drug-target interaction data. With high-quality drug-target interaction data, such as those manually curated from experiment results, the classification schema seems to

have little impact on the prediction results but increase the model search space, which makes TIMMA-R take longer time to make predictions. Unlike the Tyner study, the Miller study provided only primary target information on the intended therapeutical actions of the drugs, but provides limited information on the polypharmacological effects of a drug combination. In such a case, a complete analysis pipeline for utilizing TIMMA-R package contains collecting the quantitative drug-target interaction data from public databases or literature (Tang et al., 2018; Tanoli et al., 2018), analyzing the dose-response curves from drug screen to acquire drug sensitivity data, and predicting drug combination responses with the TIMMA-R package.

Publication III provided an overview of our in-house experimental-computational drug combination design and analysis pipeline, including an experimental protocol for drug combination assays and computational tools for combination plate design and scoring. This pipeline is designed for both cancer cell line and patient-derived sample applications. The experimental protocol contains the cell culture protocol, combination playout for a 384-well plate, and the phenotypic readouts. To facilitate the process of drug combination plate preparation, we developed an in-house software, called FIMMCherry, to automatically generate complex pick lists used by the robotic system for automated dispensing and visualize the combination plates. To evaluate the high-throughput drug combination screening, we developed an R-package, SynergyFinder, which provide efficient implementations of commonly used synergy scoring models such as highest single agent (HSA), Loewe, Bliss, and zero interaction potency (ZIP) model. The SynergyFinder web-tool paper (Ianevski et al., 2017) has been cited already >100 times (based on Google Scholars), and there is now also an extended version of SynergyFinder (version 2.0), which implements additional features supporting especially higher-order combination data analytics and exploratory visualization of multi-drug synergy patterns, along with automated outlier detection procedure, extended curve-fitting functionality and statistical analysis of replicate measurements (Ianevski et al., 2020).

Publication II demonstrated the translational use of our computational platform, drug combination prediction and testing (DCPT) platform, which is an extension of the original TIMMA model and makes use of the experimental-computational drug combination analysis pipeline of Publication II. Using DCTP, we could successfully identify synergistic drug combinations for T-PLL patient derived samples. The study used T-PLL as the first case study to show that it could successfully predict at least one synergistic drug combination for each individual

patient, each demonstrating with various resistance patterns and synergy mechanisms. DCPT platform extended the original TIMMA model by using high-throughput genomic profiles, such as exome-seq and RNA-seq profiles, along with single-drug responses from *ex vivo* drug testing to identify patient-specific synergistic drug combinations. More importantly, DCPT included healthy controls to prioritize cancer-selective drug combinations. Regarding the use of healthy controls to avoid toxic effects or effects associated only with the cell phenotype, DCPT shares some degrees of similarities with the Therapeutic Algorithmic Combinatorial Screen (TACS) algorithm (Kashif et al., 2015), which is an iterative experimental-computational algorithm searching for promising drug combinations. As TACS did not focus on less-exhaustive search strategies for large sets of compounds, TACS is limited to relatively small sets (<50) of compounds in practice, whereas DCPT can deal with much larger compound panels (> 200 in the T-PLL case study), thanks to its effective two-stage design, in which single-compound profiling is followed by combination predictions and experimental validation of prioritized combinations. Although it is beneficial to include healthy controls that can improve the cancer-selectivity of the predictions, it should be noted that the genomic profiles or the drug sensitivity profiles of control samples are not always available in clinical applications.

In the T-PLL case study in Publication III, 42% (10 out of 24) of the model predictions were confirmed to show synergistic effects in experiments (Figure 20). Such result is significantly better than random sampling of drug pairs, which has around 4% to 14% drug combinations to be synergistic in high-throughput drug combination screens (Jaeger et al., 2017; Zhang et al., 2007; Borisy et al., 2003; Miller et al., 2013). More importantly, the study showed that DCPT platform can make patient-specific predictions, which is more challenging than merely identifying generally toxic combinations that not only kill most cancer cells but also damage normal cells. In order to make patient-specific predictions, those DCPT predicted combinations were evaluated in term of differential synergy, which compares the effect not only to healthy samples but also to other T-PLL patient samples. In such a way, it is more likely to have better clinical results. The study emphasized the importance of validating the predicted synergies by replacing one drug or both of drugs with agents that inhibit the same patient-specific targets or pathways. Although such combinations are targeting the same primary targets as the predicted combinations, different modes of actions may lead to drastically different combination effects.

Combinations of already approved drugs with predicted selective synergies are the most straightforward drug repositioning cases for T-PLL. In publication III, such cases included combinations of tacrolimus and temsirolimus with the highly selective BCL-2 inhibitor venetoclax, which has been approved for the treatment of patients with relapsed/refractory chronic lymphocytic leukemia (CLL) and acute myeloid leukemia (AML) (Juárez-Salcedo et al., 2019). In another recent study, venetoclax has already shown clinical responses towards two late stage refractory T-PLL patients as single drug treatment (Boidol et al., 2017). Other combinations with selective synergies which may be more easily to be applied soon in clinic are those combinations including late-stage investigational drugs that are approved as chemotherapies, such as proteasome inhibitor carfilzomib, a second-generation proteasome inhibitor (PI) approved by FDA for relapsed and or refractory multiple myeloma as a single agent (Groen et al., 2019). With standard chemotherapeutics, most T-PLL patients develop drug resistance with various resistance mechanisms, which makes it critical to develop patient-customized combination therapies for T-PLL patients. On the other hand, due to the rarity if the disease, it is difficult to make randomized or stratified clinical trials in a large enough cohort of T-PLL patients, in which the selective combinations could be validated clinically. Therefore, developing personalized precision medicine is critically needed for rare cancers such as T-PLL.

For complex diseases such as cancer, there are many complex biological factors involved in causing the drug resistance, such as extensive pathway cross-talks, network rewiring, and pathway redundancy (Dancey and Chen, 2006). Such complex mechanisms demand polypharmacologic approaches to target multiple pathways simultaneously. As a polypharmacologic approach, our DCPT platform makes predictions based on target profiles of hundreds of drugs. Those drugs target various cancer-related pathways and most of them have overlapping target profiles with each other. Since single-compound functional testing is becoming increasingly popular, it is possible to model the systems-wide drug-target interaction in the network setting at individual patient level. In DCPT platform, signaling network analysis is performed as the last step by constructing patient-specific network after prediction have been made, which differs from those approaches taking pathway cross-talks into consideration already in the prediction step (Jaeger et al., 2017; Chen et al., 2016). To date, signaling pathways and metabolic networks are still not completely mapped in most cancer contexts. Such incomplete information could lead to less accurate predictions. Therefore, such incomplete information was not used to make predictions in the current DCPT platform. However, once more comprehensive and accurate enough network information becomes available for certain cancer types, it is

possible to utilize the network information already in the prediction phase by extending from target proteins to target pathways in the DCPT platform.

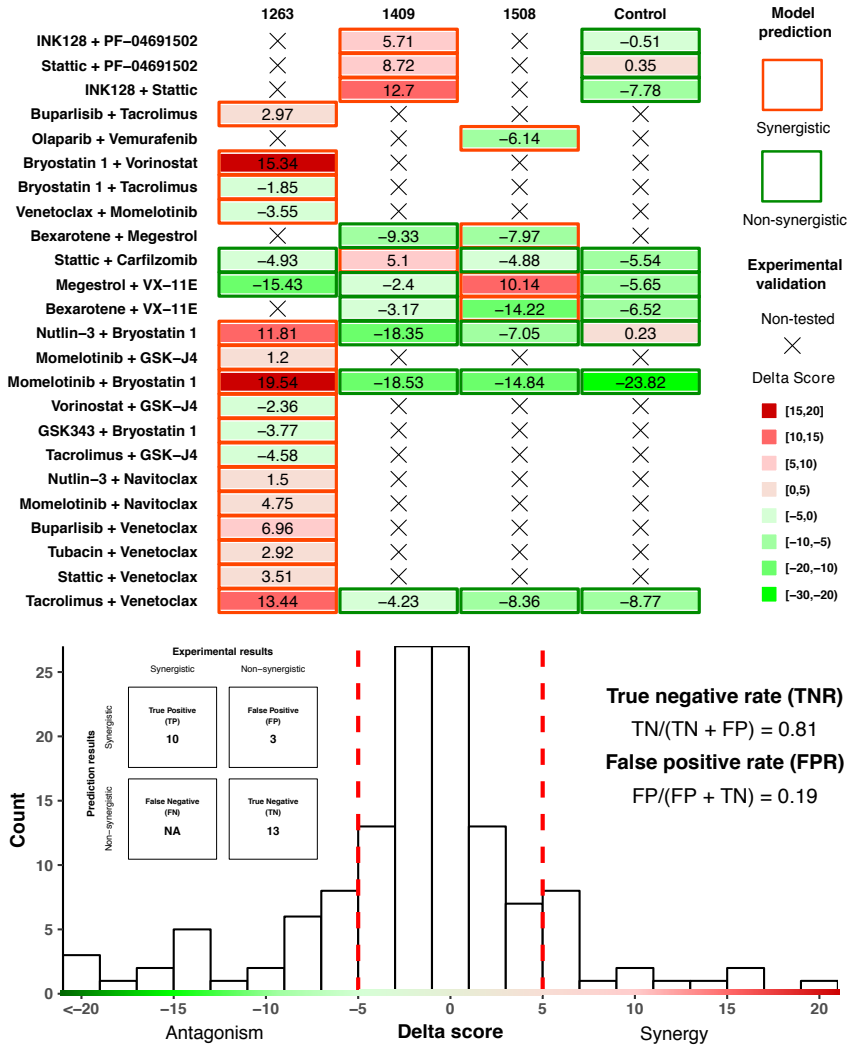

Figure 20: Comparison of the DCPT-based synergy predictions with ex vivo testing in the patient-derived cells. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

As healthy controls are not always available in clinical applications, we also tested the DCPT approach in an additional leukemia case study in which genomic and functional profiles are only available for acute myeloid leukemia (AML) patients, but not for healthy controls. The goal was to find those common

synergistic drug combinations across all the patient samples or most of the patient samples. In this case study, predictions for three given combinations in eight AML patient samples from DCPT were compared against the experimental results from our previous study (Karjalainen et al., 2017). As shown in Figure 17, combination of ruxolitinib and venetoclax was found to overcome stroma-induced resistance to BCL2 inhibitor venetoclax in AML patients, while combination of ruxolitinib and quizartinib was found not to overcome stroma-induced resistance to FLT3 inhibitor quizartinib. In addition, synergism was detected between WEHI-539 and venetoclax in conditioned medium (CM) cultured cells, but not in mononuclear cell medium (MCM) cultured cells. Among the 7 predicted synergistic drug combinations, 5 were confirmed to show synergy in the experiments, which results in a high true positive rate (TPR, 0.625). Meanwhile, DCPT also maintained a low false negative rate (FNR, 0.375). Due to lack of exome-seq and RNA-seq data from healthy bone marrow (BM) samples, it should be noted that these results are only patient-specific but not cancer selective. As shown in Figure 21, the degree of synergistic effects of these established synergistic combinations varies among these eight patients, further suggesting that personalized drug combinations predictions are needed for leukemia patients.

As with the most combination prediction algorithms (Gayvert et al., 2017; Bulusu et al., 2016; Madani Tonekaboni et al., 2016), DCPT currently predicts only the overall synergy, which is a summary of the synergy/antagonism over the whole dose-response matrix, instead of the optimal concentrations for two compounds to be combined *ex vivo*. To predict the optimal concentrations for two compounds to show synergy, it is usually required to use some global optimization search algorithms (Kashif et al., 2015; Weiss et al., 2015; Nowak-Sliwinska et al., 2016). Such algorithms increase not only the computational complexity, especially when the compound sets are very large, but also the experimental costs because of the iterative rounds of combination validation. However, such more accurate predictions at concentration level make it easier to find more clinically tolerable low-concentration drug combinations by testing them in parallel in healthy control samples. Although our previous studies have shown that *ex vivo* drug responses from leukemia patient-derived samples correlate with the clinical responses (Pemovska et al., 2013; Malani et al., 2017; Pemovska et al., 2015), it should be noted that the *ex vivo* predicted concentrations may not translate to relevant concentrations *in vivo*. In timewise, it takes up to four weeks' time to apply the whole DCPT pipeline to a new clinical application. The current bottle-neck is gathering and analyzing the required sequencing data, which can be accelerated by sequencing multiple patient samples in parallel and streamlining

the exome-seq analysis. Once the process of gathering sequencing data is optimized, the modelling component of DCPT can be used for predicting targeted combinations for cancer patients in less than one week.

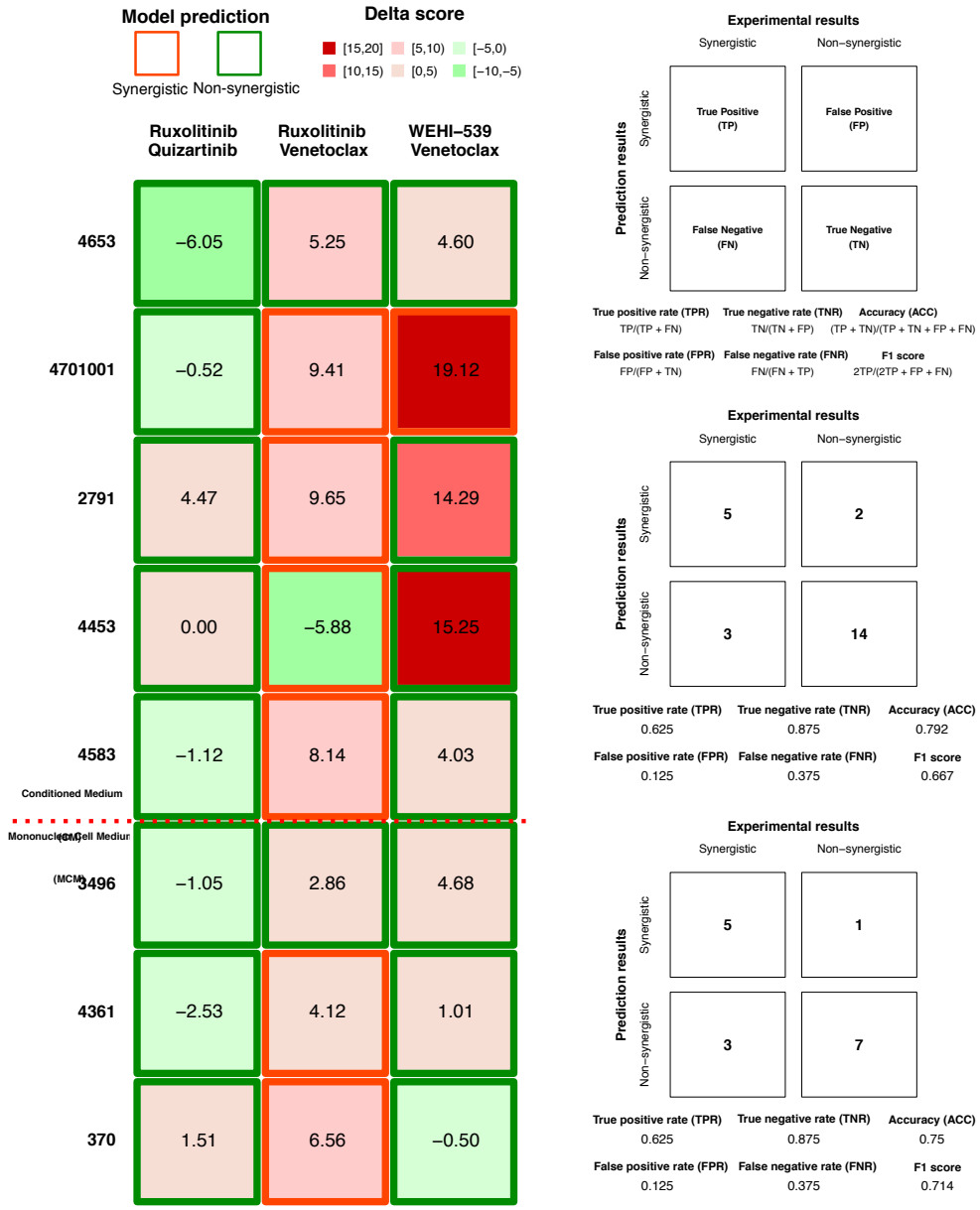

Figure 21: prediction and experimental validation of three combinations in 8 acute myeloid leukemia (AML) patients. Reproduced with permission from American Association for Cancer Research (He et al., 2018b).

In conclusion, this thesis demonstrated that computational platforms, such as TIMMA and DCPT platform, utilizing effective integration of genomic and functional data can be used for patient-customized drug combination prediction and testing of selective synergies. As it is not feasible to experimentally test all the possible drug combinations in practice for clinical applications, due to the exponential number of combinations and scarce patient cells, our computational approach is aimed at prioritization of the most promising drug combinations to narrow down the number of combinations for further experimental testing. Careful *ex vivo* validation of the *in silico* predictions is required before clinical translation. This thesis also demonstrated that computational tools, such as FIMMCherry, enables faster and less error-prone approaches to perform high-throughput drug combination screening for precision oncology applications. In addition, this thesis also demonstrated that network centric investigation of the underlying mechanisms of experimentally-validated synergistic combinations in the patient's cellular context can help us to better understand the functional cross-talks in cancer signaling pathways which are driving the caner progress and drug resistance. Such knowledge can be utilized to develop more effective drug combination therapies also for other patients with similar genomic background.

As the cancer heterogeneity often leads to patients developing monotherapy resistance, it is extremely important to model cancer heterogeneity and develop more effective patient-customized therapy (Dagogo-Jack and Shaw, 2018). New next-generation sequencing (NGS) technologies, such as single-cell sequencing and multi-region sequencing, provide more accurate assessments of the complex clonal architecture of cancers, which can be used in the future versions of DCPT pipeline for predicting targeted combinations inhibiting resistance-driving cancer subclones. However, it is challenging to obtain comprehensive drug sensitivity profiling at subclone-level. Therefore, computational tools are needed to analyze drug sensitivities for different clones with the help of identified drug response biomarkers. One of the future developments of DCPT pipeline will be focusing on subpopulation drug combination predictions, once clone-level drug sensitivity profiling becomes available in routine cancer studies. The open-source implementations made available for the computational tools presented in this thesis will make it possible for others to extend them to new applications and datasets, as well as to integrate the tools as part of in-house drug combination pipelines.

Acknowledgement

The work for this dissertation was carried out in the Computational Systems Medicine group, Institute for Molecular Medicine Finland (FIMM), University of Helsinki from 2015 to 2020. I would like to thank Integrative Life Sciences (ILS) doctor program for providing financial support for me to complete my dissertation. I would like also to thank Cancer Society Finland and Doctor School of Health, University of Helsinki, for supporting me to attend scientific conferences and courses.

My sincere and deepest thanks go to my supervisor Professor Tero Aittokallio. I was so lucky to be your student. Without your consistent support, inspiration, and your insights in this field, this thesis would not have been possible. Thank you for being patient with me since I had pure computational background when I started. Your supervision helped me learn how to make good plans, how to improve research skills, and how to think independently during the years. You always told me to relax and have a balanced life, which made my PhD life less stressful. Thank you also for the amazing Nepal adventure!

My sincere and deepest thanks also go to my co-supervisor Assistant Professor Jing Tang. Thank you for your dedicated support, guidance, and encouragement on my PhD projects, career development, and personal life over the years. Without your mentorship, I would not have made it through the transitioning from a computer scientist to a computational biologist. Thank you also for helping me grow as a scientist.

I would like to thank my thesis committee members: Professor Satu Mustjoki and Dr. Henri Xhaard for your constructive advice on my PhD projects. I would like to also thank official reviewers of my thesis: Adjunct Professor Jorrit Enserink and Professor Olivier Taboureau for your valuable comments on my thesis.

I am very grateful to all my co-authors for their collaborative effort. Thank you Prof. Krister Wennerberg for the nice discussion and your expertise on drug targets. Thank you Dr. Emma Andersson for your genetics expertise on the T-PLL project and performing the experiments. Thank you Sanna Timonen for performing the experimental validations. Thank you Dr. Jani Saarela, Laura Turunen, and Evgeny Kulesskiy from the FIMM HTB unit for your drug screening expertise. Thank you Professor Satu Mustjoki and Professor Steffen Koschmieder for the nice collaboration and your expertise on the T-PLL project.

Our research group has always nice activities such as group retreats and group lunches. Here I want to say thanks to current and former group members for making my PhD journey so interesting: Dr. Bhagwan Yadav, Dr. Alok Jaiswal, Dr. Balaguru Ravikumar, Dr. Anna Cichonska, Dr. Abhishek Gupta, Dr. Gopal Peddinti, Dr. Suleiman Ali Khan, Dr. Anil Kumar, Dr. Laxman Yetukuri, Dr. Ammad-ud-din Muhammad, Dr. Ziaurrehman Tanoli, Dr. Milla Kibble, Dr. Agnieszka Szwajda, Sanna Timonen, Mehreen Ali, Aleksandr Ianevski, Yevhen Akimov, and Zaid Alam. Thank you all for the academic and no-academic discussion, laughter, and help in my PhD life.

I would like to express my sincere thanks to Dr. Gretchen Repasky, research training coordinator at FIMM, for making my PhD training experience so unique. You were always there to help and provide wise advice for my PhD training and career development. Tack så mycket!

Other current and former FIMM members I would like to thank: Dr. Juho Miettinen, Dr. Markus Vähä-Koskela, Dr. John Patrick Mpindi, Dr. Päivi Östling, Dr. Riikka Karjalainen, FIMM HTB unit, and FIMM IT unit.

Many thanks to my dearest Chinese friends at biomedicum: Minxia Liu, Pu Chen, Jie Bao, Wenyu Wang, Yinyin Wang, Kaiyang Zhang, Shuyu Zheng, Tianduanyi Wang, Yu Fu, Ping Jiang, Zuyue Chen, and Kecheng Zhou. Thank you for all the fun moments!

Finally, my deepest gratitude goes to my family members: my parents, my sister, and my brother-in-law. You are always my greatest supporters. Your unconditional love and support keep me moving forward!

Liye He
Helsinki, 2020

References

Abbas Z and Rehman S (2018) An overview of cancer treatment modalities. In: Shahzad H (ed) *Neoplasm*.

Abbas-Aghababazadeh F, Li Q and Fridley BL (2018) Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing. *PLOS ONE* 13(10): e0206312.

Al-Lazikani B, Banerji U and Workman P (2012) Combinatorial drug therapy for cancer in the post-genomic era. *Nature Biotechnology* 30: 679–692.

Andersson EI, Pützer S, Yadav B, et al. (2017) Discovery of Novel Drug Sensitivities in T-prolymphocytic leukemia (T-PLL) by High-Throughput Ex Vivo Drug Testing and Mutation Profiling. *Leukemia*.

Bao R, Huang L, Andrade J, et al. (2014) Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer informatics* 13(Suppl 2): 67-82.

Barretina J, Caponigro G, Stransky N, et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603.

Bayat Mokhtari R, Homayouni TS, Baluch N, et al. (2017) Combination therapy in combating cancer. *Oncotarget* 8(23): 38022-38043.

Behjati S and Tarpey PS (2013) What is next generation sequencing? *Archives of disease in childhood. Education and practice edition* 98(6): 236-238.

Belkadi A, Bolze A, Itan Y, et al. (2015) Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America* 112(17): 5473-5478.

Bento AP, Gaulton A, Hersey A, et al. (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Research* 42(D1): D1083-D1090.

Berenbaum MC (1989) What is synergy? *Pharmacological Reviews* 41: 93-141.

Bertier G, Hétu M and Joly Y (2016) Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC medical genomics* 9(1): 52-52.

Bill KLJ, Garnett J, Meaux I, et al. (2016) SAR405838: A Novel and Potent Inhibitor of the MDM2:p53 Axis for the Treatment of Dedifferentiated Liposarcoma. *Clinical Cancer Research* 22(5): 1150.

Bliss CI (1939) The toxicity of poisons applied jointly. *Annals of Applied Biology* 26: 585-615.

Boidol B, Kornauth C, van der Kouwe E, et al. (2017) First in human response of BCL-2 inhibitor venetoclax in T-cell prolymphocytic leukemia. *Blood* 130: 2499–2503.

Borisy AA, Elliott PJ, Hurst NW, et al. (2003) Systematic discovery of multicomponent therapeutics. *Proceedings of the National Academy of Sciences* 100(13): 7977–7982.

Breiman L (2001) Random Forests. *Mach. Learn.* 45(1): 5–32.

Bulusu KC, Guha R, Mason DJ, et al. (2016) Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discovery Today* 21(2): 225–238.

Carlson MJ, Thiel KW and Leslie KK (2014) Past, present, and future of hormonal therapy in recurrent endometrial cancer. *International journal of women's health* 6: 429-435.

Casper J, Zweig AS, Villarreal C, et al. (2018) The UCSC Genome Browser database: 2018 update. *Nucleic acids research* 46(D1): D762-D769.

Chen D, Zhang H, Lu P, et al. (2016) Synergy evaluation by a pathway-pathway interaction network: a new way to predict drug combination. *Molecular BioSystems* 12(2): 614–623.

Dagogo-Jack I and Shaw AT (2017) Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology* 15: 81.

Dagogo-Jack I and Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology* 15(2): 81-94.

Dancey JE and Chen HX (2006) Strategies for optimizing combinations of molecularly targeted anticancer agents. *Nature Reviews Drug Discovery* 5: 649–659.

Davis MI, Hunt JP, Herrgard S, et al. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology* 29(11): 1046-1051.

Dearden C (2012) How I treat prolymphocytic leukemia. *Blood* 120(3): 538–551.

Dutta D, Barr VA, Akpan I, et al. (2017) Recruitment of calcineurin to the TCR positively regulates T cell activation. *Nat Immunol* 18(2): 196–204.

Duşa A (2008) A mathematical approach to the boolean minimization problem. *Quality & Quantity* 44(1): 99.

EBCTCG (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *The Lancet* 365(9472): 1687-1717.

Escoufier Y (1973) Le Traitement des Variables Vectorielles. *Biometrics* 29(4): 751-760.

Fang H-B, Chen X, Pei X-Y, et al. (2015) Experimental design and statistical analysis for three-drug combination studies. *Statistical Methods in Medical Research* 26(3): 1261-1280.

Figueroa-Jiménez LA, Cabrera-Márquez AL, Báez-Díaz L, et al. (2016) A patient with white blood cell counts more than a million: A diagnostic and therapeutic challenge. *El Bisturi* 2016: 12-16.

Finn OJ (2012) Immuno-oncology: understanding the function and dysfunction of the immune system in cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* 23 Suppl 8(Suppl 8): viii6-viii9.

Fisher R and Larkin J (2012) Vemurafenib: a new treatment for BRAF-V600 mutated advanced melanoma. *Cancer management and research* 4: 243-252.

Fitzgerald JB, Schoeberl B, Nielsen UB, et al. (2006) Systems biology and combination therapy in the quest for clinical efficacy. *Nature Chemical Biology* 2: 458.

Fleifel D, Rahmoon MA, AlOkda A, et al. (2018) Recent advances in stem cells therapy: A focus on cancer, Parkinson's and Alzheimer's. *Journal, genetic engineering & biotechnology* 16(2): 427-432.

Forrest S (1993) Genetic algorithms: principles of natural selection applied to computation. *Science* 261(5123): 872.

Gao Y, Davies Stephen P, Augustin M, et al. (2013) A broad activity screen in support of a chemogenomic map for kinase signalling research and drug discovery. *Biochemical Journal* 451(2): 313.

Gaulton A, Bellis LJ, Bento AP, et al. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40(D1): D1100-D1107.

Gayvert KM, Aly O, Platt J, et al. (2017) A Computational Approach for Identifying Synergistic Drug Combinations. *PLOS Computational Biology* 13(1): e1005308.

Gleeson MP, Hersey A, Montanari D, et al. (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nature Reviews Drug Discovery* 10(3): 197-208.

Goldsmith C (2011) *Leukemia.* Twenty-First Century Books.

Greco W, Unkelbach HD, Pöch G, et al. (1992) Consensus on concepts and terminology for combined action assessment: the saariselkä agreement. *Arch. Complex Environ. Stud.* 4: 65-69.

Greco WR, Bravo G and Parsons JC (1995) The search for synergy: a critical review from a response surface perspective. *Pharmacological Reviews* 47(2): 331.

Groen K, van de Donk N, Stege C, et al. (2019) Carfilzomib for relapsed and refractory multiple myeloma. *Cancer management and research* 11: 2663-2675.

Halling-Brown MD, Bulusu KC, Patel M, et al. (2012) canSAR: an integrated cancer public translational research and drug discovery resource. *Nucleic Acids Research* 40(D1): D947-D956.

Hassanpour SH and Dehghani M (2017) Review of cancer from perspective of molecular. *Journal of Cancer Research and Practice* 4(4): 127-129.

He L, Kulesskiy E, Saarela J, et al. (2018a) Methods for High-throughput Drug Combination Screening and Synergy Scoring. In: von Stechow L (ed) *Cancer Systems Biology: Methods and Protocols*. New York, NY: Springer New York, pp.351–398.

He L, Tang J, Andersson EI, et al. (2018b) Patient-Customized Drug Combination Prediction and Testing for T-cell Prolymphocytic Leukemia Patients. *Cancer Research* 78(9): 2407.

He L, Wennerberg K, Aittokallio T, et al. (2015) TIMMA-R: an R package for predicting synergistic multi-targeted drug combinations in cancer cell lines or patient-derived samples. *Bioinformatics* 31(11): 1866-1868.

Hennessey VG, Rosner GL, Bast RC, Jr., et al. (2010) A Bayesian approach to dose-response assessment and synergy and its application to in vitro dose-response studies. *Biometrics* 66(4): 1275-1283.

Herbst RS (2005) Role of novel targeted therapies in the clinic. *British Journal of Cancer* 92(1): S21-S27.

Herzog R and Papenfort K (2018) Chapter Thirteen - Transcriptomic Approaches for Studying Quorum Sensing in Vibrio cholerae. In: Carpousis AJ (ed) *Methods in Enzymology*. Academic Press, pp.303-342.

Herzyk P (2014) Chapter 8 - Next-Generation Sequencing. In: Padmanabhan S (ed) *Handbook of Pharmacogenomics and Stratified Medicine*. San Diego: Academic Press, pp.125-145.

Hitzemann R, Darakjian P, Walter N, et al. (2014) Chapter One - Introduction to Sequencing the Brain Transcriptome. In: Hitzemann R and McWeeney S (eds) *International Review of Neurobiology*. Academic Press, pp.1-19.

Holle L, Pruemer J and Kim H (2015) Overview of Cancer Therapy. *Cancer Therapy: Prescribing and Administration Basics*.

Housman G, Byler S, Heerboth S, et al. (2014) Drug resistance in cancer: an overview. *Cancers* 6(3): 1769-1792.

Huang L, Li F, Sheng J, et al. (2014) DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics (Oxford, England)* 30(12): i228-i236.

Ianevski A, Giri AK and Aittokallio T (2020) SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Research*.

Ianevski A, He L, Aittokallio T, et al. (2017) SynergyFinder: a web application for analyzing drug combination dose–response matrix data. *Bioinformatics* 13: 504–505.

IARC (2014) World Cancer Report 2014.

Jaeger S, Igea A, Arroyo R, et al. (2017) Quantification of Pathway Cross-talk Reveals Novel Synergistic Drug Combinations for Breast Cancer. *Cancer Research* 77: 459–469.

Jovanović KK, Roche-Lestienne C, Ghobrial IM, et al. (2018) Targeting MYC in multiple myeloma. *Leukemia*.

Juárez-Salcedo LM, Desai V and Dalia S (2019) Venetoclax: evidence to date and clinical potential. *Drugs in context* 8: 212574-212574.

Karjalainen R, Pemovska T, Popa M, et al. (2017) JAK1/2 and BCL2 inhibitors synergize to counteract bone marrow stromal cell-induced protection of AML. *Blood* 130: 789–802.

Kashif M, Andersson C, Hassan S, et al. (2015) In vitro discovery of promising anti-cancer drug combinations using iterative maximisation of a therapeutic index. *Scientific Reports* 5: 14118.

Kasteng F, Sobocki P, Svedman C, et al. (2007) Economic evaluations of leukemia: A review of the literature. *International Journal of Technology Assessment in Health Care* 23(1): 43-53.

Kim D, Pertea G, Trapnell C, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14(4): R36.

Kirouac DC, Saez-Rodriguez J, Swantek J, et al. (2012) Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. *BMC Systems Biology* 6(1): 29.

Koskela HLM, Eldfors S, Ellonen P, et al. (2012) Somatic STAT3 Mutations in Large Granular Lymphocytic Leukemia. *New England Journal of Medicine* 366: 1905–1913.

Kosugi S, Momozawa Y, Liu X, et al. (2019) Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biology* 20(1): 117.

Kukurba KR and Montgomery SB (2015) RNA Sequencing and Analysis. *Cold Spring Harbor protocols* 2015(11): 951-969.

Kulesskiy E, Saarela J, Turunen L, et al. (2015) Precision Cancer Medicine in the Acoustic Dispensing Era: Ex Vivo Primary Cell Drug Sensitivity Testing. *Journal of Laboratory Automation* 21(1): 27-36.

Kummar S, Gutierrez M, Doroshow JH, et al. (2006) Drug development in oncology: classical cytotoxics and molecularly targeted agents. *British journal of clinical pharmacology* 62(1): 15-26.

Lamb J, Crawford ED, Peck D, et al. (2006) The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* 313(5795): 1929.

Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357-359.

Law V, Knox C, Djoumbou Y, et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42(Database issue): D1091-D1097.

Lederer S, Dijkstra TMH and Heskes T (2018) Additive Dose Response Models: Explicit Formulation and the Loewe Additivity Consistency Condition. *Frontiers in pharmacology* 9: 31-31.

Lehár J, Krueger AS, Avery W, et al. (2009) Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nature Biotechnology* 27(7): 659-666.

Lheureux S, Denoyelle C, Ohashi PS, et al. (2017) Molecularly targeted therapies in cancer: a guide for the nuclear medicine physician. *European journal of nuclear medicine and molecular imaging* 44(Suppl 1): 41-54.

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25(14): 1754-1760.

Li P, Huang C, Fu Y, et al. (2015) Large-scale exploration and analysis of drug combinations. *Bioinformatics* 31(12): 2007-2016.

Li WV and Li JJ (2018) Modeling and analysis of RNA-seq data: a review from a statistical perspective. *Quantitative biology (Beijing, China)* 6(3): 195-209.

Liu J, Dang H and Wang XW (2018) The significance of intertumor and intratumor heterogeneity in liver cancer. *Experimental &Amp; Molecular Medicine* 50: e416.

Loewe S (1953) The problem of synergism and antagonism of combined drugs. *Arzneimittel-Forschung* 3: 285-290.

Loewe S and Muischnek H (1926) Effect of combinations: mathematical basis of the problem. *Naunyn Schmiedebergs Arch Exp Pathol Pharmakol* 114: 313-326.

Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12): 550-550.

Madani Tonekaboni SA, Soltan Ghoraie L, Manem VSK, et al. (2016) Predictive approaches for drug combination discovery in cancer. *Briefings in bioinformatics* 19(2): 263-276.

Malani D, Murumägi A, Yadav B, et al. (2017) Enhanced sensitivity to glucocorticoids in cytarabine-resistant AML. *Leukemia* 31: 1187–1195.

Marx V (2013) The big challenges of big data. *Nature* 498: 255.

Mathews Griner LA, Guha R, Shinn P, et al. (2014) High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 111(6): 2349-2354.

Miller ML, Molinelli EJ, Nair JS, et al. (2013) Drug Synergy Screen and Network Modeling in Dedifferentiated Liposarcoma Identifies CDK4 and IGF1R as Synergistic Drug Targets. *Science Signaling* 6(294): ra85.

Normand EA, Alaimo JT and Van den Veyver IB (2018) Exome and genome sequencing in reproductive medicine. *Fertility and Sterility* 109(2): 213-220.

Nowak-Sliwinska P, Weiss A, Ding X, et al. (2016) Optimization of drug combinations using Feedback System Control. *Nat. Protocols* 11(2): 302–315.

Pemovska T, Johnson E, Kontro M, et al. (2015) Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. *Nature* 519: 102–105.

Pemovska T, Kontro M, Yadav B, et al. (2013) Individualized Systems Medicine (ISM) strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discovery* 3(12): 1416–1429.

Pennock GK and Chow LQM (2015) The Evolving Role of Immune Checkpoint Inhibitors in Cancer Treatment. *The oncologist* 20(7): 812-822.
Polychronakis I, Dounias G, Makropoulos V, et al. (2013) Work-related leukemia: a systematic review. *Journal of occupational medicine and toxicology (London, England)* 8(1): 14-14.
Quine WV (1952) The Problem of Simplifying Truth Functions. *The American Mathematical Monthly* 59(8): 521-531.
Quine WV (1955) A Way to Simplify Truth Functions. *The American Mathematical Monthly* 62(9): 627-631.
Rai MF, Tycksen ED, Sandell LJ, et al. (2018) Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society* 36(1): 484-497.
Rao MS, Van Vleet TR, Ciurlionis R, et al. (2019) Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Frontiers in Genetics* 9: 636.
Reddy AS and Zhang S (2013) Polypharmacology: drug discovery for the future. *Expert review of clinical pharmacology* 6(1): 41-47.
Ritz C, Baty F, Streibig JC, et al. (2015) Dose-Response Analysis Using R. *PLOS ONE* 10: e0146021.
Rixe O and Fojo T (2007) Is Cell Death a Critical End Point for Anticancer Therapies or Is Cytostasis Sufficient? *Clinical Cancer Research* 13(24): 7280.
Robert P and Escoufier Y (1976) A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 25(3): 257-265.
Roell KR, Reif DM and Motsinger-Reif AA (2017) An Introduction to Terminology and Methodology of Chemical Synergy-Perspectives from Across Disciplines. *Frontiers in Pharmacology* 8: 158-158.
Sebaugh JL (2011) Guidelines for accurate EC50/IC50 estimation. *Pharmaceutical Statistics* 10(2): 128-134.
Shannon P, Markiel A, Ozier O, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13(11): 2498-2504.
Sherry ST, Ward MH, Kholodov M, et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 29(1): 308-311.

Sun Y, Sheng Z, Ma C, et al. (2015) Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nature communications* 6: 8481-8481.

Swerdlow SH, Campo E, Pileri SA, et al. (2016) The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127(20): 2375–2390.

Tang J, Karhinen L, Xu T, et al. (2013) Target Inhibition Networks: Predicting Selective Combinations of Druggable Targets to Block Cancer Survival Pathways. *PLOS Computational Biology* 9(9): e1003226.

Tang J, Szwajda A, Shakyawar S, et al. (2014) Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis. *Journal of Chemical Information and Modeling* 54(3): 735-743.

Tang J, Tanoli Z-U-R, Ravikumar B, et al. (2018) Drug Target Commons: A Community Effort to Build a Consensus Knowledge Base for Drug-Target Interactions. *Cell chemical biology* 25(2): 224-229.e222.

Tang J, Wennerberg K and Aittokallio T (2015) What is synergy? The Saariselkä agreement revisited. *Frontiers in Pharmacology* 6: 181.

Tanoli Z, Alam Z, Vähä-Koskela M, et al. (2018) Drug Target Commons 2.0: a community platform for systematic analysis of drug–target interaction profiles. *Database* 2018.

Thorvaldsdóttir H, Robinson JT and Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14(2): 178-192.

Trapnell C, Pachter L and Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* 25(9): 1105-1111.

Tyner JW, Yang WF, Bankhead A, 3rd, et al. (2013) Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. *Cancer Research* 73(1): 285-296.

Vafaee F, Dashti H and Alinejad-Rokny H (2019) Transcriptomic Data Normalization. In: Ranganathan S, Gribskov M, Nakai K, et al. (eds) *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press, pp.364-371.

Vempati UD, Chung C, Mader C, et al. (2014) Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). *Journal of Biomolecular Screening* 19(5): 803-816.

Ventola CL (2017) Cancer Immunotherapy, Part 2: Efficacy, Safety, and Other Clinical Considerations. *P & T : a peer-reviewed journal for formulary management* 42(7): 452-463.

Vesely MD, Kershaw MH, Schreiber RD, et al. (2011) Natural Innate and Adaptive Immunity to Cancer. *Annual Review of Immunology* 29(1): 235-271.

Vivekanandarajah A, Atallah JP and Gupta S (2013) T-cell prolymphocytic leukaemia (T-PLL): a rare disease with a grave prognosis. *BMJ case reports* 2013: bcr2013009808.

Vlot AHC, Aniceto N, Menden MP, et al. (2019) Applying synergy metrics to combination screening data: agreements, disagreements and pitfalls. *Drug Discovery Today* 24(12): 2286-2298.

Wadapurkar RM and Vyas R (2018) Computational analysis of next generation sequencing data and its applications in clinical oncology. *Informatics in Medicine Unlocked* 11: 75-82.

Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* 10(1): 57-63.

Warr A, Robert C, Hume D, et al. (2015) Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda, Md.)* 5(8): 1543-1550.

Weiss A, Berndsen RH, Ding X, et al. (2015) A streamlined search technology for identification of synergistic drug combinations. *Scientific Reports* 5: 14508.

Wu H-C, Chang D-K and Huang C-T (2006) Targeted Therapy for Cancer. *Journal of Cancer Molecules* 2(2): 57-66.

Wu M, Sirota M, Butte AJ, et al. (2015) Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials.gov. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*: 68–79.

Yadav B, Pemovska T, Szwajda A, et al. (2014) Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Scientific Reports* 4: 5193.

Yadav B, Wennerberg K, Aittokallio T, et al. (2015) Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. *Computational and Structural Biotechnology Journal* 13: 504–513.

Yan L, Rosen N and Arteaga C (2011) Targeted cancer therapies. *Chinese journal of cancer* 30(1): 1-4.

Yang J, Tang H, Li Y, et al. (2015) DIGRE: Drug-Induced Genomic Residual Effect Model for Successful Prediction of Multidrug Effects. *CPT: pharmacometrics & systems pharmacology* 4(2): e1-e1.

Yang W, Soares J, Greninger P, et al. (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research* 41(D1): D955-D961.

Yung-Chi C and Prusoff WH (1973) Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology* 22(23): 3099-3108.

Zaman N, Li L, Jaramillo Maria L, et al. (2013) Signaling Network Assessment of Mutations and Copy Number Variations Predict Breast Cancer Subtype-Specific Drug Targets. *Cell Reports* 5(1): 216-223.

Zhang L, Yan K, Zhang Y, et al. (2007) High-throughput synergy screening identifies microbial metabolites as combination agents for the treatment of fungal infections. *Proceedings of the National Academy of Sciences* 104(11): 4606–4611.

Zhao M, Wang Q, Wang Q, et al. (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14(11): S1.

Zhao S, Fung-Leung W-P, Bittner A, et al. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLOS ONE* 9(1): e78644-e78644.

Zhu F, Shi Z, Qin C, et al. (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Research* 40(D1): D1128-D1136.

Zinner RG, Barrett BL, Popova E, et al. (2009) Algorithmic guided screening of drug combinations of arbitrary size for activity against cancer cells. *Molecular Cancer Therapeutics* 8(3): 521.