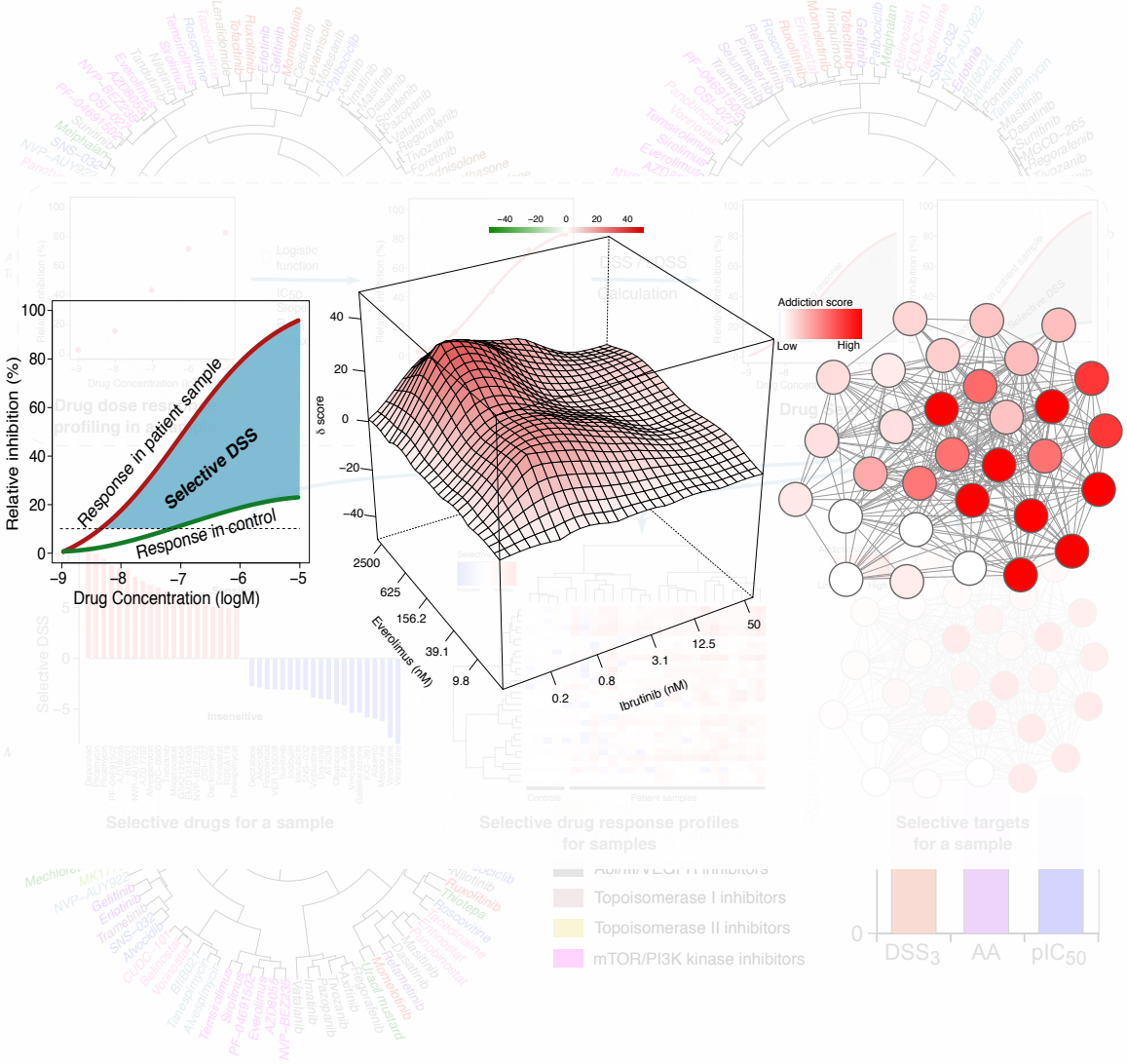
UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

# QUANTITATIVE MODELING AND ANALYSIS OF DRUG SCREENING DATA FOR PERSONALIZED CANCER MEDICINE




BHAGWAN YADAV

Institute for Molecular Medicine Finland, FIMM
Doctoral School in Health Sciences (DSHealth)
Doctoral Program in Biomedicine (DPBM)
University of Helsinki, Finland

# Quantitative modeling and analysis of drug screening data for personalized cancer medicine

**Bhagwan Yadav**

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine, University of Helsinki, for public examination in Lecture Hall 2, Biomedicum Helsinki on Friday, March 24, 2017, at 12 o'clock noon.

HELSINKI, 2017

**Supervised by:**
Prof. Tero Aittokallio, Ph.D.
FIMM-EMBL Group Leader
Institute for Molecular Medicine Finland, FIMM
University of Helsinki
Helsinki, Finland

and

Dr. Krister Wennerberg, Ph.D.
FIMM-EMBL Group Leader
Institute for Molecular Medicine Finland, FIMM
University of Helsinki
Helsinki, Finland

**Reviewed by:**
Dr. Jaakko Hollmén
Aalto University School of Science
Department of Computer Science
Espoo, Finland

and

Dr. Petri Auvinen
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

**Official Opponent:**
Prof. Mats Gustafsson, Ph.D.
Department of Medical Sciences
Uppsala University
Uppsala, Sweden


Cover layout by Bhagwan Yadav
Cover image by Bhagwan Yadav
ISBN 978-951-51-2965-9 (paperback)
ISBN 978-951-51-2966-6 (PDF)
http://ethesis.helsinki.fi/
Helsinki 2017

*Coming together is a beginning; keeping together is progress; working together is a success.*

– Henry Ford

*Do the difficult things while they are easy and do the great things while they are small. A journey of a thousand miles must begin with a single step.*

–Laozi

*Dedicated to my parents and family*

"*Family is the most important thing in the world.*"

– Princess Diana

# TABLE OF CONTENTS

# ABBREVIATIONS

| | |
|---|---|
| AA | Activity area |
| AACR | American association for cancer research |
| ABC | Activated B-cell-like subtype |
| AGCT | Adult granulosa cell tumor |
| ALL | Acute lymphoblastic leukemia |
| AML | Acute myeloid leukemia |
| AUC | Area under the curve |
| CCLE | The Cancer cell line encyclopedia |
| CI | Combination index |
| CLL | Chronic lymphoblastic leukemia |
| CML | Chronic myeloid leukemia |
| DR | Drug response |
| DSRT | Drug sensitivity and resistance testing |
| DSS | Drug sensitivity scoring |
| $EC_{50}$ | Half maximal effective concentration |
| GCT | Granulosa cell tumor |
| GDSC | Genomics of drug sensitivity in cancer |
| $GI_{50}$ | Half maximal growth inhibition |
| GR | Growth rate |
| hGL | Mural granulosa-luteal |
| HSA | Highest single agent |
| HTDS | High-throughput drug screening |
| $IC_{50}$ | Half maximal inhibitory concentration |
| KAR | Kinase addiction ranker |
| KIEN | Kinase inhibitors elastic net |
| KISS | Kinase inhibition sensitivity score |
| MR | Maximum relevance |
| PM | Personalized medicine |

| | |
|---|---|
| ROC | Receiver operator characteristic |
| SVM | Support vector machines |
| TAS | Target addiction scoring |
| WBC | White blood cells |
| ZIP | Zero interaction potency |

# ORIGINAL PUBLICATIONS

This thesis is based on the following original publications:

I. ***Yadav, B.**, Pemovska, T., Szwajda, A., Kulesskiy, E., Kontro, M., Karjalainen, R., Majumder, M.M., Malani, D., Murumägi, A., Hintsanen, P., Knowles, J., Porkka, K., Heckman, C., Kallioniemi, O., Wennerberg, K., Aittokallio, T.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. Scientific Reports. 2014; 4:5193. doi: 10.1038/srep05193

II. ***Yadav, B.**, Gopalacharyulu, P., Pemovska, T., Khan, S.A., Szwajda, A., Tang, J., Wennerberg, K., Aittokallio, T.* From drug response profiling to target addiction scoring in cancer cell models. Disease Model and Mechanism, 2015; 8: 1255-1264; doi: 10.1242/dmm.021105

III. ***Yadav, B.**, Wennerberg, K., Aittokallio, T., Tang, J.* Searching for drug synergy in complex dose-response landscapes using an interaction potency model. Computational and Structural Biotechnology Journal, 2015; 13: 504-513; doi:10.1016/j.csbj.2015.09.001

IV. *Haltia, UM.*, Andersson, N.*, **Yadav, B.***, Färkkilä, A., Kulesskiy, E., Kankainen, M., Tang, J., Bützow, R., Riska, A., Leminen, A., Heikinheimo, M., Kallioniemi, O., Unkila-Kallio, L., Wennerberg, K., Aittokallio, T., Anttonen, M.* Systematic drug sensitivity testing reveals synergistic growth inhibition by dasatinib or mTOR inhibitors with paclitaxel in ovarian granulosa cell tumor cells. Gynecological Oncology 2017; In the press; http://dx.doi.org/10.1016/j.ygyno.2016.12.016

* Equal contribution

These articles are reproduced with permission from the copyright holders.

# ABSTRACT

Despite recent progress in the field of molecular medicine, the treatment and cure of complex diseases such as cancer remains a challenge. Development of resistance to first-line chemotherapy is a common cause of current anticancer treatment failure. To deal with this problem, the personalized medicine (PM) approach has been adapted toward more targeted cancer research and management. The PM approach is based on each patient's genetic, epigenetic and drug response profiling, which is used to design the best treatment option for the given patient. As the PM approach is increasingly being adopted in clinical practice, there is an urgent need for computational models and data mining methods that allow fast processing and analysis of the massive relevant profiling datasets.

High-throughput drug screening enables systematic profiling of cellular responses to a wide collection of oncology compounds and their combinations, hence providing an unbiased strategy for personalized drug treatment selection. However, screening experiments with patient-derived cell samples often results in high-dimensional data matrices, with inherent sources of noise. This complicates many downstream analyses, such as the detection of differential drug activity or understanding the mechanisms behind drug sensitivity and resistance in a given patient.

To meet these challenges, a computational pipeline for drug response profiling was developed in this thesis. The pipeline was based on a novel metric to quantify drug response, called the drug sensitivity score (DSS). Further, by combining the normalized drug response profile of each cancer sample with a global drug-target interaction network, a target addiction score (TAS) was developed to de-convolute the selective protein targets and obtain knowledge on their functional importance. Finally, delta scoring was

developed to quantify drug combination effects and to address the problem of the clonal evolution of cancer, which often leads to resistance to mono therapies.

This novel computational pipeline improves understanding of cancer development and translates compound activities into informed treatment choices for clinicians. As exemplified in two case studies of adult acute myeloid leukemia (AML) and adult granulosa cell tumor (AGCT), the models developed here have the potential to significantly contribute to the effective analysis of data from individual cancer patients and from pan-cancer cell line panels. Hence, these models will play a substantial role in future personalized cancer treatment strategies and the selection of effective treatment options for individual cancer patients.

# INTRODUCTION

Cancer is a complex heterogeneous disease that has inter- and intra-tumor heterogeneity. Inter-tumor heterogeneity exists between tumors, whereas intra-tumor heterogeneity occurs within a tumor. Inter-tumor heterogeneity originates from genetic, epigenetic and phenotypic variations between tumors of different tissue origin, cell types and individuals, as well as variations between individuals with the same tumor type (Burrell et al., 2013; Heppner, 1984; Janku, 2014; Loewe, 1953; Meacham and Morrison, 2013; Vogelstein et al., 2013; Wild, 2012). Although our understanding of tumor biology has advanced, the clinical success of targeted therapies remains limited. The inter- and intra-tumor heterogeneity of a tumor is one of the biggest challenges in cancer treatment (Janku, 2014; Vogelstein et al., 2013). Therefore, clinics urgently need to find new approaches for developing effective anticancer treatments.

High-throughput drug screening (HTDS) is a functional approach to testing large numbers of compounds, to find the best hits for phenotype-based drug discovery or drug repurposing for the personalized treatment of cancer. HTDS can be implemented as an automated operation platform with a highly sensitive compound testing system (Liu et al., 2004; Macarron et al., 2011). Because of its high efficiency and low cost, HTDS provides the opportunity to profile cellular responses to a large collection of compounds, providing an unbiased means for selecting personalized treatment (Pemovska et al., 2013; Tyner et al., 2013). However, this approach often results in high-dimensional datasets, complicating many downstream analyses, such as the detection of differential drug activity or the prediction of drug sensitivity and resistance.

Drug sensitivity is often measured using a single parameter, such as half maximal inhibitory concentration ($IC_{50}$), half maximal effective concentration

($EC_{50}$) or half maximal growth inhibition ($GI_{50}$; Marx et al., 2003). As these are only a point estimation of the drug response curve, such parameters cannot capture the complete information from the complex dose-response curves. In HTDS settings, going through thousands of drug response curves visually to remove inconsistencies is a demanding task. In the personalized medicine setting in particular, drawing conclusions to clinically translate the drug response is difficult, due to potential side effects to normal cells. Thus, we need a better model for the quantification of drug sensitivity, especially for scoring the differential drug response between patient and control samples.

Although the genome sequencing of cancer samples provides a valuable molecular profile, extracting key information from this and translating it for clinics remains a challenge. Compared to genomic or transcriptomic profiles, functional classification of cancer based on drug screening profiling may provide more actionable outcomes for clinical utility. The drug sensitivity and resistance of a cell is highly dependent on a set of cellular targets. HTDS, combined with drug target profiles, enables systematic mapping of the drug response phenotype to the genotypic dependencies of cancer cells, leading to the identification of the cellular targets responsible for the drug response. Such an integrated approach offers insights into the mechanism of the action of the compounds and their efficacies in specific cell or cancer types. This is called target deconvolution (Gujral et al., 2014; Lee and Bogyo, 2013). However, there is currently a lack of systematic approaches that use drug response profiling to reveal target dependencies in an unbiased manner.

Because of inter and intra-tumor heterogeneity, a patient often develops resistance to single drug treatments. It is therefore necessary to develop a combinatorial drug treatment strategy for improved efficacy, overcoming resistance and an improved clinical outcome (Fitzgerald et al., 2006; Flaherty et al., 2012; Lehár et al., 2009). HTDS makes it possible to systematically

evaluate a number of pairwise drug combinations (Borisy et al., 2003). Drug combination classes, such as synergistic or antagonistic, are derived from the difference in observed and expected combination responses. The expected response is calculated on the basis of a selected non-interaction reference model. Most reference models have been developed for low throughput drug combination experiments, making model assumptions invalid for the complex drug interactions of various compound and dose pairs. Therefore, improved reference models are needed, especially for high-throughput combination studies.

# LITERATURE REVIEW

## 1. Cancer as a complex disease

According to the web definition by the American Association for Cancer Research (AACR), cancer is not a single disease; it encompasses over 200 diseases, and is characterized by uncontrolled division or proliferation of abnormal cells in a part of the body (https://www.aacrfoundation.org/Pages/what-is-cancer.aspx). According to this definition, cancer is a complex and heterogeneous disease, due to the contribution of a combination of predisposing genetic variants and environmental factors (Lindblom et al., 2004). The complexity of cancer is also associated with tumor heterogeneity through morphological and epigenetic plasticity and clonal evolution (Loewe, 1953). According to a study by Anand et al., genetic factors contribute to only 5%–10% of the total cancer risk, and environmental factors are the major contributor to cancer risk, at 90%–95% (Anand et al., 2008). It is also known that only 5%–10% of all cancer is due to an inherited genetic defect (Anand et al., 2008). Gene alteration or mutation causes many cancer phenotypes, and are a result of the interaction of the genes with the environment (Anand et al., 2008). Hence, molecular characterization becomes a real challenge. Despite substantial research and studies on the molecular classification of various cancer types and subtypes, most of them still remain poorly characterized.

Over one hundred types of cancer exist (http://www.cancer.gov/about-cancer/what-is-cancer). Cancer type classifications are usually based on the organ or tissue in which the tumor arises. The two broad types of cancer are solid and liquid, depending on where in the body the cancer grows (https://www.stjude.org/treatment/disease/solid-tumors/what-is-solid-tumor.html). Solid tumors form a mass or lump of tissue in different body parts, such as the ovary or breast. Liquid tumors, such as leukemia or

lymphoma, on the other hand, are not usually localized to any specific area and originate due to alterations in hematopoietic or lymphoid tissues. Some lymphomas, however, may be localized to the lymph node (https://www.verywell.com/cancer-types-4013927). According to the type of cells from which they originate, cancers are also divided into six major groups such as carcinomas, lymphomas, leukemias, myeloma, sarcomas and mixed types (https://training.seer.cancer.gov/disease/categories/classification.html)

Leukemia is a hematological cancer caused by an uncontrolled increase in the number of white blood cells (WBC), which crowd out normal blood cells. This means that the body cannot get enough oxygen to its tissues, control bleeding or fight infections. There are four common types of leukemia: Acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), and chronic lymphoblastic leukemia (CLL).

AML is a heterogeneous hematological cancer, which is characterized by the infiltration of bone marrow, blood, and other tissues, caused by the abnormal differentiation and clonal evolution of the myeloid progenitor cells. AML is cured in only around 40% of adult patients aged under 60. The heterogeneity of AML has been known for a long time, yet no great improvement in the treatment has been seen so far (Longo et al., 2015; Saultz and Garzon, 2016).

Ovarian cancer is a solid type of cancer that originates from abnormal cells in the ovaries of women. It is often not detected until it has spread within the pelvis and abdomen. Ovarian cancer is difficult to treat at the late stage and is therefore frequently fatal. Ovarian cancer is the fifth most common cancer among females. It is also a leading cause of mortality from gynecological cancers. About 70% of ovarian carcinomas are not diagnosed until a late stage. Therefore, early diagnosis and treatment are essential (Argento et al., 2008). Although there are many types of ovarian cancer, about 80% of malignancies arise from the epithelium of the ovary (Shepherd, 2000).

Adult granulosa cell tumor (AGCT) is a type of rare ovarian cancer of non-

epithelial origin, accounting for 2%–5% of ovarian malignancies and 70% of sex chord-stromal tumors (Mancari et al., 2014; Pectasides et al., 2008). AGCTs occur in a young age group and are usually detected at an early stage, often having features of hyperestrogenism (Kottarathil et al., 2013; Mancari et al., 2014). A mutation of the transcription factor FOXL2 is associated with 97% of the AGCT patients. SMAD3, the regulator of CCND2 and the transcription factor GATA4, has also been identified as a potential diagnostic biomarker and therapeutic target (Mancari et al., 2014).

Despite all the technological and molecular advances towards the development of new anticancer treatments, resistance to drug treatments continues to be a major clinical challenge. To overcome resistance to monotherapies, different drugs in different combinations can be tested in pre-clinical assays before clinical trials. For instance, already over fifty years ago it was discovered that testicular cancers can be successfully treated with a different combination of an alkylating agent, an antimetabolite and the antibiotic actinomycin D (Li, 1960). After these first studies, many drug combinations have been approved for the clinical treatment of various cancer types (http://www.cancer.gov/about-cancer/treatment/drugs).

## 2. Personalized medicine: The future of cancer medicine

Personalized medicine is a broad and rapidly growing field of health care. It is an approach to tailoring treatment for each individual patient on the basis of their unique clinical, genetic, epigenetic, and environmental information (Chan and Ginsburg, 2011). It is also sometimes defined as P4 medicine i.e. predictive, preventive, personalized, and participatory (Hood and Flores, 2012). The main goals of personalized medicine are 1) to identify the optimal treatment for each individual patient; 2) to maximize the treatment benefit;

and 3) to minimize adverse effects (Schleidgen et al., 2013). The approach suits the treatment of complex, heterogeneous diseases such as cancers; however, many challenges need to be overcome before it can be actionable in clinics. There is currently no standard way of carrying out personalized treatment optimization. To fully meet the promise of P4, medicine requires the development of standardized experimental platforms and computational models to handle multiscale and multivariate data (Hood and Flores, 2012).

# 3. Functional compound screening

High-throughput compound screening is often divided into two categories: Biochemical assays and cell-based assays (An and Tolliday, 2009; Zang et al., 2012). Biochemical assays involve the specific binding of compounds to the interesting targets in a homogeneous reaction (An and Tolliday, 2009). The limitation of this method is that not all the targets can be purified for biochemical measurements, and also that the assay may not represent a tissue-specific response because of cell complexity (Zang et al., 2012). Cell-based assays are of three types: Second messenger assays, reporter gene assays, and cell proliferation/toxicity assays (An and Tolliday, 2009; Zang et al., 2012). In recent publications, different cell-based assays have been used for high-throughput compound screening (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Pemovska et al., 2013; Shoemaker, 2006). It has also been recently demonstrated that different experimental parameters, such as the storage of compounds, growth medium, cell viability assay, cell plating, and drug concentration range may have a substantial impact on the data generated, and may influence drug testing results (Gautam et al., 2016; Haibe-Kains et al., 2013; Hatzis et al., 2014; Haverty et al., 2016). Standardized guidelines for choosing different experimental parameters and making data more consistent between laboratories are critical if we are to improve data sharing and reusability.

# 4. Quantification of compound activity

When cells are treated with chemical compounds, they respond. The response is often measured as percentage cell viability, death, growth, or inhibition. The compound response across different concentrations is often modeled using the sigmoidal/logistic function, providing fitting parameters such as half maximal inhibitory concentration ($IC_{50}$) (Barretina et al., 2012; Cheng and Prusoff, 1973; Garnett et al., 2012; Sebaugh, 2011), half maximal effective concentration ($EC_{50}$) (Sebaugh, 2011), or half maximal growth inhibition ($GI_{50}$) (Marx et al., 2003). The above metrics are highly sensitive to the number of cell divisions taking place over the time of the drug response assay (Hafner et al., 2016). Perhaps the most commonly used sigmoidal function for dose-response curve fitting is

$$y = d + \left\{\frac{a - d}{1 + 10^{b(c-x)}}\right\} \qquad \text{Eq.1}$$

Here, *a* is the maximal drug response, i.e. top asymptote of the sigmoid curve; *b* is the slope of the curve; *c* is the $IC_{50}$ / $EC_{50}$ / $GI_{50}$; *d* is the minimal drug response, i.e. bottom asymptote of the sigmoid curve; *x* is a variable concentration; and *y* is the observed drug response.

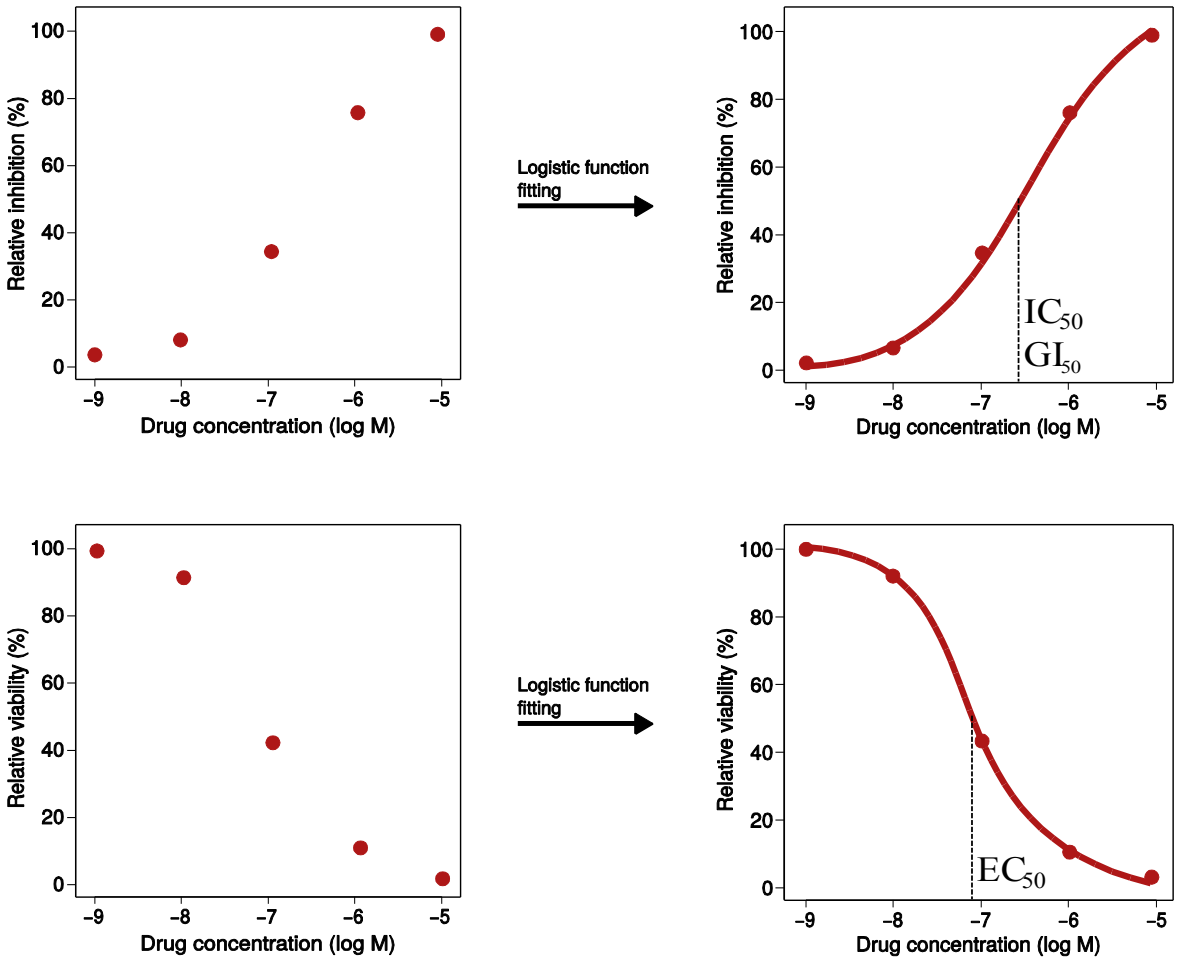



***Figure 1:*** *Drug response before and after logistic curve fitting. Drug response measured as percent inhibition (top left) or percent viability (bottom left) at different concentrations (left panel) and logistic fitting of the drug dose-response curve (right panel). $IC_{50}$, $GI_{50}$ and $EC_{50}$ are the logistic fitting parameters to be estimated.*

The standard response measures, such as $IC_{50}$, $EC_{50}$ and $GI_{50}$, are point estimates of the drug/compound response curve (Figure 1). Drug response measures can be either absolute or relative. The absolute $IC_{50}$ /$EC_{50}$ /$GI_{50}$ measure is a response corresponding to 50% control, whereas the relative measure corresponds to a response halfway between the lower and upper plateaus (Sebaugh, 2011). The $IC_{50}$ estimate can be improved if the slope of

the dose response curve is a property of the cell line and does not change with the specific compound. According to this assumption, the center of the dose response curve is only influenced by each compound in a cell line (Vis et al., 2016). Although these measures are widely used for quantifying drug response, the point estimation of the response curve is not always sufficiently accurate to represent the actual sensitivity of the drug in a given cancer sample.

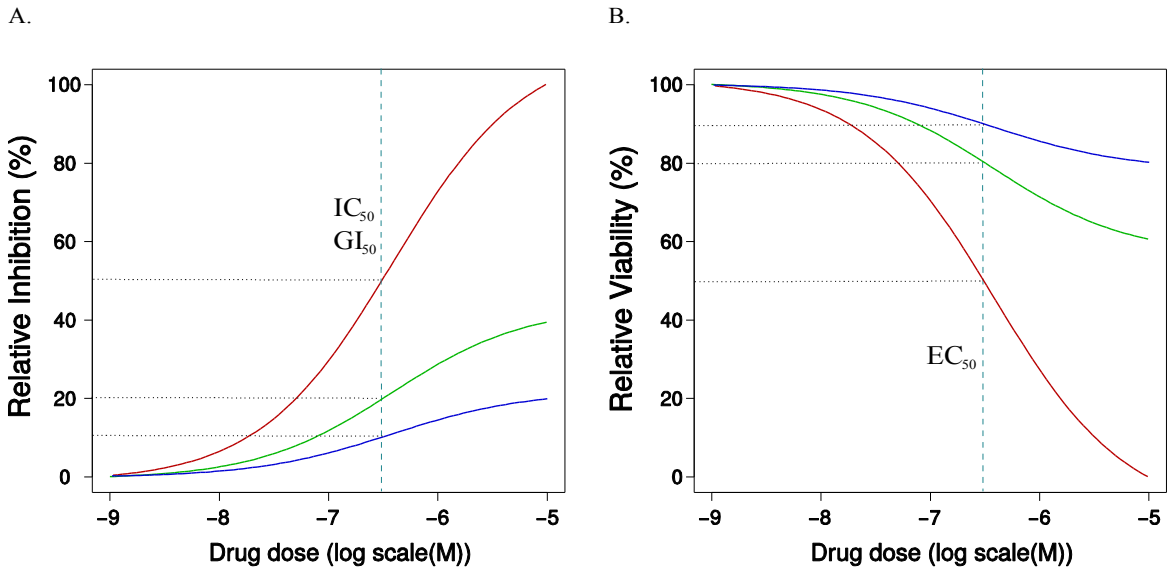



***Figure 2:*** *Example of relative $IC_{50}$ / $GI_{50}$ / $EC_{50}$ for different response curve shapes, shown in different colors. A) In this plot, the $IC_{50}$ / $GI_{50}$ for the three completely different drug responses is equal; B) $EC_{50}$ is equal for the three different responses.*

Different response curve shapes can have the same relative $IC_{50}$ value (Figure 2). This also applies to the relative $EC_{50}$ and relative $GI_{50}$ calculation. To avoid such problems, a model that considers the whole response curve, such as activity area (AA) or area under the curve (AUC), is used for the drug response quantification. Activity area is computed as a sum of the difference

between compound activity and reference activity over the concentration points (Barretina et al., 2012):

$$Activity\ Area = \sum_{i=1}^{8} A_i - A_{ref} \qquad \text{Eq.2}$$

Here, $A_i$ is activity at concentration $i$, $A_{ref}$ is reference activity of either 0, or max (0, Alow), and $A_{low}$ is the activity at its lowest concentration.

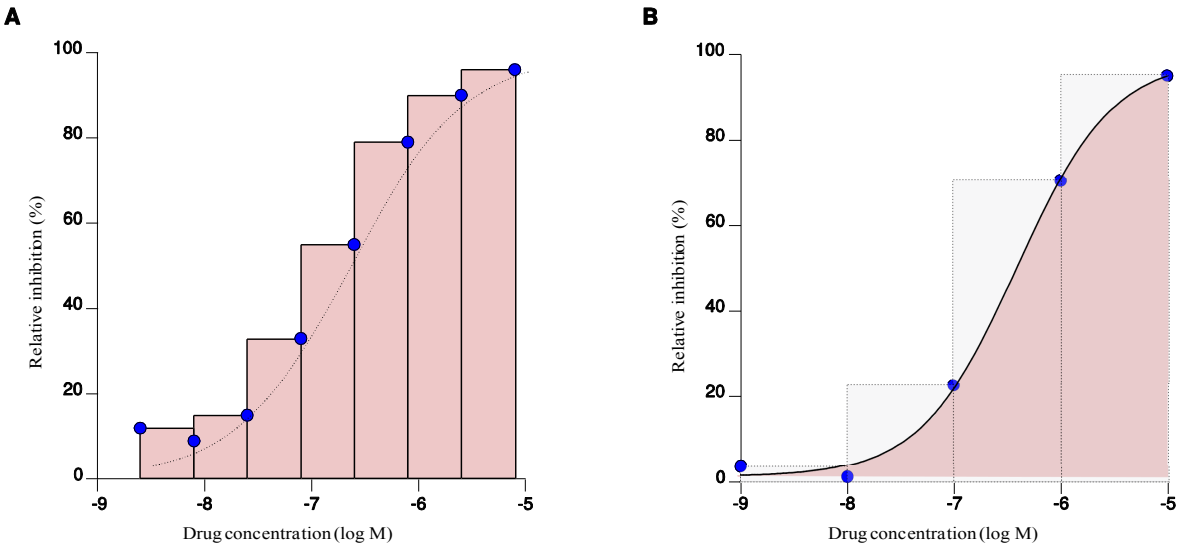



***Figure 3*****:** *Example showing area under the curve calculation of percent inhibition data. The blue dots represent the percent inhibition value measured at a given drug concentration. (A) Histogram approximation of area under drug response curve. (B) Model-based integrated area under drug response curve calculation.*

In addition to such a histogram approximation, the area under the dose-response curve can be numerically integrated in various other ways, such as via the trapezoidal rule (Atkinson, 1989). A histogram approximation such as

AA (Figure 3A) works well with a narrow concentration window and high number of tested concentrations (Barretina et al., 2012), whereas AUC, computed using a trapezoidal approximation or more accurate integrals (Figure 3B), works well even when the concentration window is wide and the number of concentration points are few, or when some data points are missing (Basu et al., 2013; Garnett et al., 2012; Seashore-Ludlow et al., 2015).

Recently, growth rate inhibition (GR) metrics, such as $GR_{50}$ and $GR_{max,}$ were used to estimate the magnitude of the drug-induced growth rate inhibition, with endpoint or time course assay (Hafner et al., 2016). The GR value at time *t* in the presence of a drug at concentration *c* is computed as

$$GR(c,t) = 2^{k(c,t)/k(0)} - 1 \qquad \text{Eq.3}$$

Here, *k(c,t)* is the growth rate of drug-induced cells and *k*(0) is the growth rate of untreated cells. In other words, the GR values are simply the ratio of treated to untreated growth rates, normalized to a single cell division (Hafner et al., 2016).

# 5. Computational target deconvolution

The cellular target of a molecularly-targeted compound is typically a protein (e.g. kinase) hit by a compound to modulate its functionality. In phenotype-based compound screening, identification of the underlying molecular signal is very important for understanding a compound's mode of action (Terstappen et al., 2007). Target deconvolution refers to the identification of molecular targets (in Figure 4, for example) that underlie the observed phenotypic response, which is an important part of the phenotype-based drug discovery

approach (Lee and Bogyo, 2013). Target deconvolution is important not only for elucidating the underlying mechanism of action of the compound, but also for better understanding biological processes such as cell differentiation, cell proliferation, apoptosis, and cell motility in leukemia (Lee and Bogyo, 2013; Terstappen et al., 2007). The advantage of using a phenotype-based approach is that it is an unbiased way in which to find functionally active compounds in the context of whole biological systems (Lee and Bogyo, 2013; Terstappen et al., 2007). Normal signaling from target proteins is important to maintain cellular homeostasis. Body tissue has mechanisms that constantly repair DNA damage to maintain homeostasis. Although all mutations are not harmful, some alterations or mutations may lead to a cancer. The emergence of powerful genome sequencing techniques has made it possible to explore the whole target protein space in a high-throughput setup (Baselga and Arribas, 2004; Fleuren et al., 2016; Kim et al., 2014).

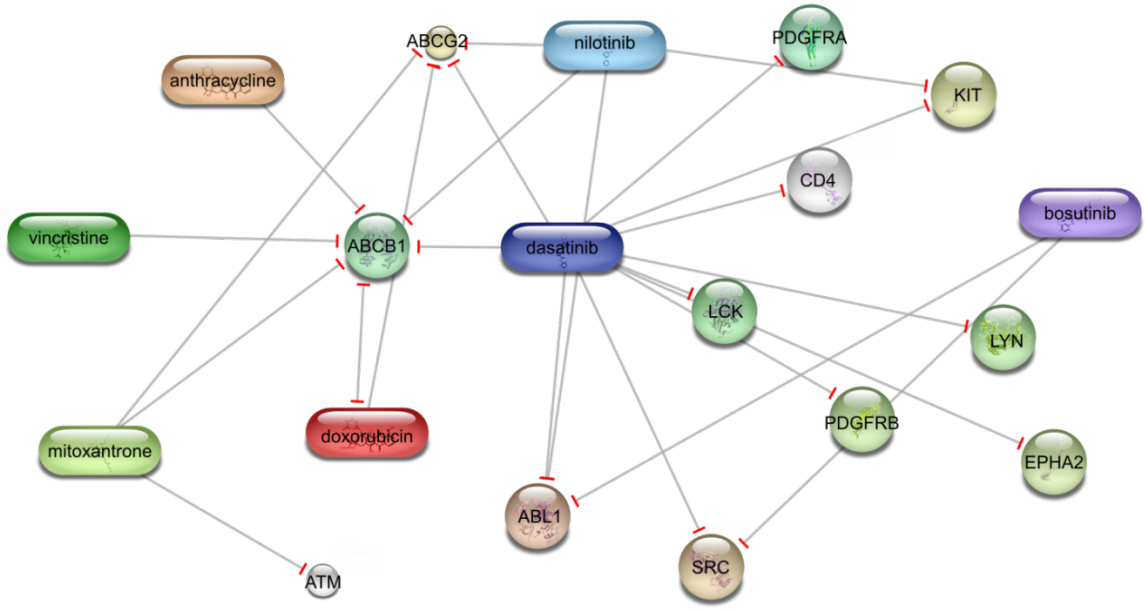



***Figure 4:*** *Example of compound-target interaction network for some drugs in drug screening profiling panel created using STITCH 4.0.*

Advances in high-throughput drug screening has made it possible to monitor and quantify cellular phenotypic activity against hundreds or thousands of compounds (Pemovska et al., 2013, 2015; Tyner et al., 2013). Such information can be further integrated to identify pathway dependencies in hundreds of different cancer cell lines of various types (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Heiser et al., 2012; Seashore-Ludlow et al., 2015). For instance, Al-Ali et al. used target-based screening, which is an efficient technique for identifying potent modulators of individual, unknown targets (Al-Ali et al., 2015). Machine learning models, such as support vector machines (SVM) and the maximum relevance (MR) algorithm, have been used to identify relevant targets (Al-Ali et al., 2015), and one study (Gujral et al., 2014) used the elastic net regression model to identify important kinases. The estimate $\beta_{en}$ of the elastic net model is defined by

$$\beta_{en} = (1 + \lambda_2)\left\{ arg_{\beta} min \, \| Y - X\beta \|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\| \right\} \quad \text{Eq.4}$$

Here, *X* is the *n x p* predictor matrix; *Y* is the response vector, λ is the shrinkage parameter, and $\beta$ are the coefficients of the elastic net regression model. The $L_1$ penalty of the model favors sparse models, whereas the quadratic part of the penalty enables variable grouping affects and stabilizes the $L_1$ norm regularization path (Zou and Hastie, 2005).

Another model called Kinase inhibitors elastic net (KIEN), also based on elastic net regularization, integrates information contained in drug-kinase networks with *in vitro* drug response screening (Tran et al., 2014). KIEN identifies a set of kinases associated with drug sensitivities (Tran et al., 2014). One study used a weighting-based method on drug screening data (Tyner et

al., 2013), and assigned a positive weight to the targets of effective drugs, and a negative weight to the targets of ineffective drugs. The final score was the sum of the effective and ineffective drug target scores. A similar approach, called the Kinase addiction ranker (KAR) has also been developed, in which the model integrates kinase inhibition data and gene expression data with drug screening data (Ryall et al., 2015).

# 6. Compound synergy scoring

A drug combination is the use of more than one compound, with the goal of improving efficacy and overcoming resistance in the treatment of complex and refractory diseases such as cancer. Usage of a single chemical agent for the treatment of disease has limitations, such as inadequate efficacy and potential development of resistance to the compound, which leads to a poor clinical outcome. To overcome these limitations, a combination of two or more compounds is often used (Jia et al., 2009). The effect of a drug combination can be synergistic, additive or antagonistic, depending on whether the combined effect is greater than, equal to, or less than the summed effects of the partner drugs. The combination effects are often quantified by comparing the observed combination response to the expected response under the assumptions of non-interaction predicted by a reference model. An effect is defined as synergistic if the combination produces a greater effect than expected, and antagonistic if the combination produces a smaller effect than expected. An additive effect is when there is no change in effect after the drug combination (Chou, 2006).

There are three popular reference models: The Loewe additivity model (Loewe, 1953), the Highest single agent (HSA) model (Berenbaum, 1989), and the Bliss independence model (Bliss, 1939). The Loewe additivity model defines the expected effect as a self-combination (Figure 5). It assumes that

the two drugs target the same molecule or pathway (Loewe, 1953). The model states that the expected effect $Y_{Loewe}$ satisfies

$$\frac{x_1}{X^1_{Loewe}} + \frac{x_2}{X^2_{Loewe}} = 1 \qquad \text{Eq.5}$$

Here, $X^1_{Loewe}$ and $X^2_{Loewe}$ are concentrations of Compound 1 and 2. $x_1$ and $x_2$ are concentrations of Compound 1 and 2 in mixture (Greco et al., 1995). The combination index (CI) is derived from the Loewe additive model as

$$CI = \frac{x_1}{X_1} + \frac{x_2}{X_2} \qquad \text{Eq.6}$$

Here, CI>1 means that the combination effect is synergistic, CI<1 means that the combination effect is antagonistic, and CI=1 means that the combination effect is additive (Greco et al., 1995).

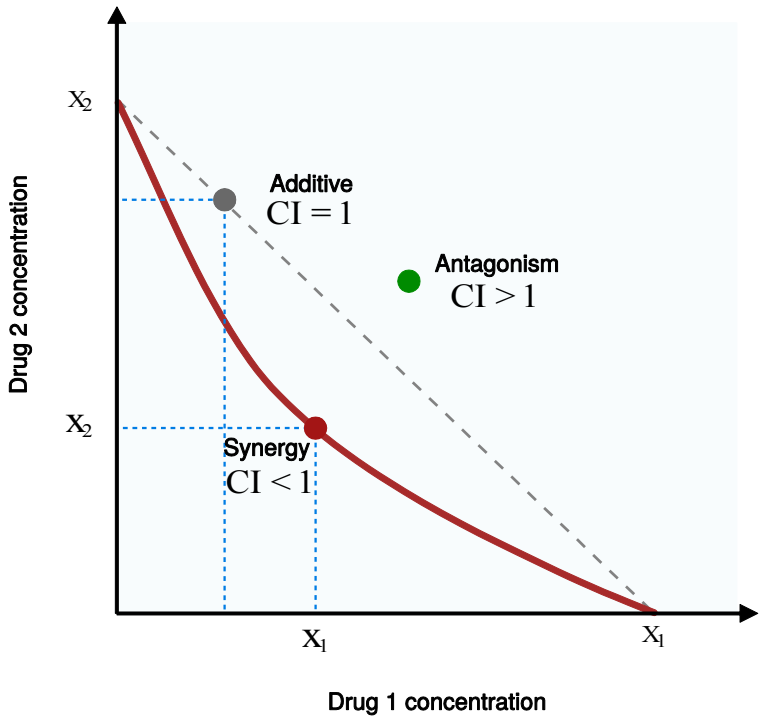



***Figure 5:*** *Illustration of Loewe additivity model. The dotted line is the additivity line where CI=1. The red line is the synergy line where CI<1; the smaller than 1 the CI value is, the better the synergy between the compounds. If CI>1, then the combination is classified as antagonistic.*

The HSA model (or Gaddum's non-interaction model) assumes that the expected effect is simply the maximum of the single agent responses at corresponding concentrations (Berenbaum, 1989). This means that if the effect of the combination exceeds that of a single compound response, there must be some form of drug interaction. The HSA model expects a combination of compounds with only positive effects, the targets of which are not functionally related (Lehár et al., 2007).

$$Y_{HSA} = \max(Y_1, Y_2) \qquad \text{Eq.7}$$

Here, $Y_1$ and $Y_2$ are the effects of the first and second compound, respectively. If the combination effect is higher than the expected effect ($Y_{HSA}$), then it is defined as synergistic; if lower than the expected effect, antagonistic. Otherwise, the effect is defined as additive.

The Bliss independence model assumes that the individual chemical agents are acting independently and that there is no interaction between their actions; each drug independently targets different molecules or pathways (Bliss, 1939; Goldoni and Johansson, 2007; Greco et al., 1995). This means that the chemical agents in the mixture always have a different mode and possibly a different site of action (Goldoni and Johansson, 2007).

$$Y_{Bliss} = Y_1 + Y_2 - Y_1 Y_2 \qquad \text{Eq.8}$$

Here, $Y_1$ and $Y_2$ are the effects of the first and second compound, respectively. $Y_{Bliss}$ is the expected combination effect. If the combination effect is higher than $Y_{Bliss}$, the interaction is defined as synergistic. If the combination effect is lower than $Y_{Bliss}$, it is defined as antagonistic. Otherwise, the combination effect is defined as additive, which means there is no interaction (Goldoni and Johansson, 2007). The Bliss independence model is not accepted as a universal reference model, because the occurrence of Bliss independence in complex biological systems is rare (Greco et al., 1995). The synergy of the compounds can also be quantified as Beta and Gamma parameters. Beta ($\beta$) is the interaction parameter that minimizes the deviance between the observed combination effects and Bliss independence over all the dose combinations; whereas Gamma ($\gamma$) is the interaction parameter that

minimizes the deviance between the observed combination effect and the HSA model (Griner et al., 2014).

The Loewe model and the Bliss independence model work well when both the chemical agents have positive effects, and also when the effects monotonically increase as a function of the concentrations. The HSA model only makes sense when both chemical agents have positive effects at all concentrations (excluding noise). With the negative effect levels, it becomes unclear which effect is "higher" when the single agents go in different directions (Lehár et al., 2007). The Loewe additivity model is computationally more intensive than the Bliss independence model or the HSA model.

# AIMS OF THE STUDY

The overall aim of the study was to develop and implement computational models for facilitating the individualized treatment of cancer patients on the basis of *ex vivo* HTDS experiments. The models are needed for system-level understanding of cancer development and treatment. Computational models in particular allow for the functional investigation of cellular addictions and other vulnerabilities in individual cancer samples, as well as the systematic stratification of cancer patients and disease subtypes on the basis of their treatment responses. The specific aims were:

1) To develop and implement models for improving the quantification of drug responses (sensitivity and resistance) in the HTDS setting.
2) To develop and implement models for predicting and better understanding the target signal networks behind individual drug response profiles.
3) To develop and implement models for drug combination scoring (synergy and antagonism) in high-throughput screening settings.
4) To apply these models and methods, and suggest new targeted treatment alternatives for patients with AML and AGCT.

# MATERIALS AND METHODS

## 7. Study specimens

Project I used 14 AML patient samples and 4 bone marrow samples from healthy donors as controls as described in Pemovska et al. (2013). Most of the samples were from relapsed and refractory patients. After informed consent with approval was obtained (No. 239/13/03/00/2010, 303/13/03/01/2011), the samples were collected in EDTA-treated tubes in accordance with the ethical standards of the Helsinki University Central Hospital (HUCH), approved by the HUCH Institutional Review Board (Dnro 60/2011). The consent form covers the use of the drug sensitivity data to guide therapeutics with approved drugs. According to Finnish legislation, approved drugs in an off-label mode can be used as therapeutics. All the samples included in this study were freshly processed.

Project IV used six adult granulosa cell tumor (AGCT) patient samples (two primary and four recurrent). Primary cultures of AGCT cells from tumors were established as previously described (Kyrönlahti et al., 2010). The cells were cultured for four to seven days to increase the cell number in Dulbecco's modified Eagle's medium (DMEM)/Ham's F-12 medium, supplemented with 10% FCS, 1% penicillin/streptomycin and L-glutamine (Gibco, Grand Island, NY, USA) in a humidified environment of 37 °C and 5% $CO_2$. The study was approved by the ethical committee of Helsinki University Central Hospital and the National Supervisory Authority for Welfare and Health in Finland. Informed consent was obtained from the patients whose fresh AGCT tissue was used in this study. The KGN and COV434 cell lines were obtained from the Riken BioResource Center. Both cell lines tested negative for mycoplasma infection, and passaged in less than months after receipt or resuscitation of a frozen cell vial. KGN and COV434 cells were cultured in

DMEM:F12 and DMEM, respectively (Gibco), supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% L-glutamine.

The ACGT study used three pooled samples of mural granulosa-luteal (hGL) as a control. Each of the three hGL samples consisted of granulosa-luteal cells derived from 78–122 mature ovulatory follicles from 4–7 different patients, and were wild type for the FOXL2 402C->G mutation. The hGL cells were isolated as previously described (Shi et al., 2009). These cells were obtained from women undergoing IVF treatment at the Women's Clinic, Department of Obstetrics and Gynecology, Helsinki University Hospital. hGL cells are mural granulosa cells from preovulatory follicles that further differentiate into corpus luteum after ovulation, and represent the best available model of healthy granulosa cells. Prior to the drug sensitivity and resistance testing (DSRT) screening, freshly isolated hGL cells were suspended into DMEM/F12 growth medium supplemented with 2.5% Nu-serum I, ITS+TM Premix (both from BD Biosciences, Bedford, MA, USA), 1% penicillin/streptomycin, and L-glutamine (Gibco). Bone marrow mononuclear cells from seven healthy donors were used as controls in the AML study, as previously described (Pemovska et al., 2013).

As additional material in Project I, cancer cell line data for drug sensitivity and mutation and copy number profiles were obtained from the Cancer cell line encyclopedia (CCLE), in which 504 cell lines were tested for 24 drugs (Barretina et al., 2012). In Project II, cell line data for drug sensitivity, gene expression, mutation and copy number for 127 cell lines were obtained from the Genomics of drug sensitivity in cancer (GDSC) (Garnett et al., 2012), whereas the drug sensitivity profiles, drug-target and clinical information of 151 leukemia patients were obtained from a study by Tyner (Tyner et al., 2013). Project III used data from a drug screening study of 466 drug

combinations for the activated B-cell-like subtype (ABC) (Griner et al., 2014). Project IV used the drug screening data of a panel of 68 cell lines derived from other solid cancers, including breast (n=19), ovarian (n=31), lung (n=4), pancreas (n=9), and prostate cancer (n=5), as a reference set.

# 8. Drug sensitivity and resistance testing (DSRT)

The DSRT platform is used to test the sensitivity of compounds in high-throughput settings. It consists of about 500 drugs, including both approved and investigational drugs. The compounds were preprinted on tissue culture-treated 384-well plates (Corning 3707) using an acoustic liquid handling device (Echo 550; Labcyte Inc.). Each compound was diluted five times into five different concentrations with 10-fold dilution. The majority of the compound concentrations ranged between 1 nM and 10000 nM. The ready-made plates were stored in nitrogen pressurized StoragePods (Roylan Developments Ltd.). Each plate also included wells with 16 positive and 16 negative controls. The positive control consisted of 100 µM of benzethonium chloride, and the negative control of DMSO. The controls were used to normalize the data. DSRT is done with freshly prepared mononuclear cells derived from patients or healthy donors. Before adding the cells to the plate, the ready-made plate is suspended in 5 µl MCM and placed on a shaker for 30 minutes. A single cell suspension of mononuclear cells (10000 cells/well in 20 µl) was transferred to every well, using a MultiDrop Combi (Thermo Scientific) peristaltic dispenser. The plates were then incubated for 72 hours at 37 $^{o}$C and 5% $CO_2$. After incubation for 72 hours, cell viability was measured using the CellTiter Glo luminescent assay (Promega).

# 9. Development of drug response models

## 9.1. Drug sensitivity scoring (DSS)

The existing methods for the quantification of drug response are either a single point approximation or area approximation. A new model, called DSS, was developed on the basis of the continuous numerical integration of the area under the curve to summarize the drug-response relationship. Given the logistic function (Eq.1), the area under the curve (AUC) can be computed as

$$AUC = \int_{R \geq A_{min}} R(x)dx = I(IC_{50}, Slope, R_{min}, R_{max}, A_{min}) \quad \text{Eq.9}$$

Here, *R(x)* is the normalized drug response at concentration *x*, *I* is an integral response, and $A_{min}$ is the minimum activity level. Three steps of normalization are followed to compute three different versions of DSS, denoted by $DSS_1$, $DSS_2$ and $DSS_3$. $DSS_1$ is calculated by normalizing the area under the drug response curve with the total area, as in Eq.10.

$$DSS_1 = \frac{AUC - t(x_2 - x_1)}{(100 - t)(C_{max} - C_{min})} \quad \text{Eq.10}$$

Here, $x_1$ and $x_2$ are the concentration range at threshold drug response *t*, and $C_{min}$ and $C_{max}$ are the minimum and maximum treated concentration, respectively. $DSS_1$ was further normalized with $\log_{10}$ of max response, $R_{max}$, to normalize the effect of the maximal compound response, which may be due to off-target effects:

$$DSS_2 = \frac{DSS_1}{log_{10}(R_{max})} \quad \text{Eq.11}$$

If a compound shows a response at low concentrations, then that compound is biologically more potent than the compounds that only show a response at higher concentrations. $DSS_2$ was further optimized using a concentration range to emphasize the drugs that obtain their response over a relatively wide dose window rather than the drugs that show a response at only the higher end of the concentration range:

$$DSS_3 = DSS_2 * \frac{x_2 - x_1}{C_{max} - C_{min}} \quad \text{Eq.12}$$

To deal with the off-target effects of the drugs, which are most likely clinically irrelevant, DSS is set to 0 for cases in which the estimated $IC_{50}$ is at or beyond the maximal tested dose $C_{max}$. Differential or selective DSS (sDSS) is calculated by subtracting the average of the healthy controls from the samples.

## 9.2. Target addiction scoring (TAS)

The sensitivity of the cancer cells to inhibiting a particular protein target is estimated using a target addiction score. Once the drug sensitivity profile of the cells in the form of DSS, AUC or $IC_{50}$ is calculated, the target addiction scoring (TAS) for each protein target, *t,* is computed by averaging the observed drug response (DR) over all the compound inhibitors $n_t$ that target protein *t*. This can be expressed mathematically as

$$TAS_t = \sum_{i=1}^{n_t} \frac{DR_i}{n_t} \quad \text{Eq.13}$$

TAS transforms the drug response space to the target space, mapping the observed drug responses to their target addictions. TAS transformation enables the direct comparative study of drug sensitivity with their molecular signatures. It also provides an insight into individual drug vulnerabilities and signal addictions. The same concept was also used in Publication I, and was called the Kinase inhibition sensitivity score (KISS), since that analysis only used kinase targets.

## 9.3. Drug combination scoring

The Loewe additivity model considers the dose-response curves of individual drugs, whereas Bliss independence and the HSA model use point estimates. The zero-interaction potency (ZIP) model developed here considers drug response curves for both individual drugs and their combinations. The drug response curves are described using a four-parameter fitting function (Eq.1), which can be rewritten as

$$y = \frac{E_{min} + E_{max}\left(\frac{x}{m}\right)^{\lambda}}{1 + \left(\frac{x}{m}\right)^{\lambda}} \quad \text{Eq.14}$$

Here, $E_{min}$ and $E_{max}$ are the minimal and maximal effects of the compounds; m is the relative $IC_{50}$ and λ is the shape parameter (λ>0). If there is complete inhibition of the cell growth, then $E_{min}$ = 0 and $E_{max}$ = 0. If zero potency shifts for non-interaction are considered, then the above equation for Compound 1,

for example, can be simplified as

$$y = \frac{\left(\frac{x}{m}\right)^{\lambda}}{1 + \left(\frac{x}{m}\right)^{\lambda}} \quad \text{Eq.15}$$

If x2 dose of Compound 2 is added to Compound 1, and there is no interaction, then $m_1$ and $\lambda_1$ will remain unchanged. The drug response curve for Compound 1 can then be written as

$$y_{2\to1} = \frac{y_2 + \left(\frac{x}{m_1}\right)^{\lambda_1}}{1 + \left(\frac{x}{m_1}\right)^{\lambda_1}} \quad \text{Eq.16}$$

Similarly, the above equation could be written for $y_c^{1\to2}$ , $y_{ZIP}^{1\to2}$, $y_c^{2\to1}$ and $y_{ZIP}^{2\to1}$.Then the delta score is computed as the average of the deviation between $y_c$ and $y_{ZIP}$.

$$\delta = \frac{y_c^{1\to2} - y_{ZIP}^{1\to2}}{2} + \frac{y_c^{2\to1} - y_{ZIP}^{2\to1}}{2} \quad \text{Eq.17}$$

The delta (δ) score centers around 0, which means that there is no interaction. If δ is greater than 0 or less than 0, this means that the interaction is either synergistic or antagonistic, respectively.

# RESULTS

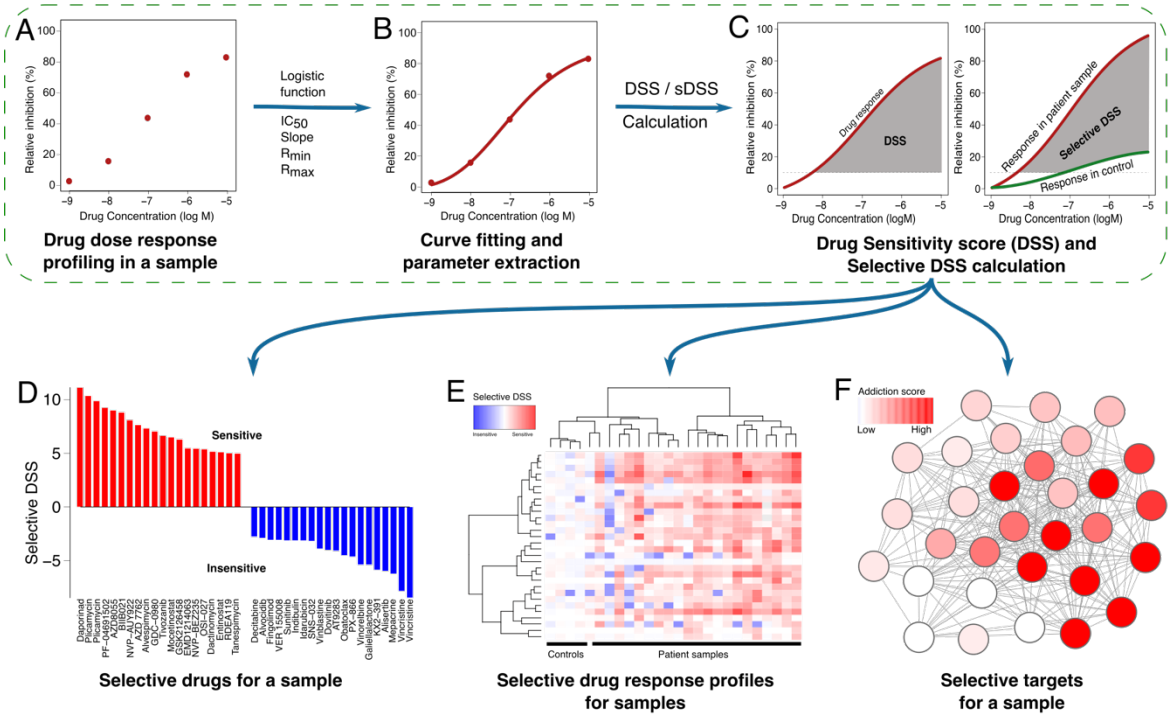



***Figure 6:*** *Systematic overview of drug sensitivity and target addiction analysis pipeline. (A) Drug response of sample at different concentrations; (B) 4-parameteric logistic curve fitting to extract fitting parameters such as $IC_{50}$, slope. $R_{min}$ and $R_{max}$; (C) Computation of drug sensitivity score (DSS) or selective DSS (sDSS) using extracted parameters; (D) Waterfall plot showing top sensitive and insensitive drugs for patient samples; (E) Heatmap of drug sensitivity profiles across samples; and (F) Target addiction scoring and network visualization.*

In this thesis, three novel models were developed and implemented to analyze drug screening data. In Publication I, a model to quantify drug sensitivity using an integrative compound response model was developed. Publication I also introduced the concept of target addiction scoring (TAS), focusing on a set of kinase targets only, and was called the Kinase inhibition

sensitivity score (KISS). In Publication II, the concept of TAS was extended in pan-cancer cell line analysis, to take both kinases and non-kinase protein targets into account. In Publication III, the ZIP model was developed as a reference model for drug combination response scoring. In Publication IV, I applied developed models such as DSS and ZIP to AGCT patient samples, with the aim of identifying targeted treatments. Overall, in this thesis, I developed an integrative bioinformatics approach (Figure 6) to analyzing and understanding the underlying biology behind cancer progression, aiming toward the eventual goal of personalized cancer treatment.

# 10. Drug sensitivity scoring (DSS)

A robust method for quantifying a complex drug-response relationship to a single metric was developed on the basis of an integrative quantitative model. The selective DSS was calculated by subtracting the control response from the patient DSS (Figure 7A). The method was systematically evaluated using AML patient samples to test the quantitative performance together with other metrics such as $IC_{50}$ and Activity Area (AA) (Figure 7B),. For the evaluation, experts visually categorized 795 compound response curves into five different classes, namely very active (18), active (30), semi-active (65), low active (70) and inactive (612), and the predictive accuracy of each response score was evaluated using the receiver operator characteristic (ROC) analysis. ROC analysis (Figure 7B) showed that DSS could better distinguish between active and inactive drug responses than $IC_{50}$ ($p < 10^{-5}$, DeLong's test, $DSS_3$ vs. $IC_{50}$) and AA ($p = 6.6 \times 10^{-9}$, DeLong's test, $DSS_3$ vs. AA).

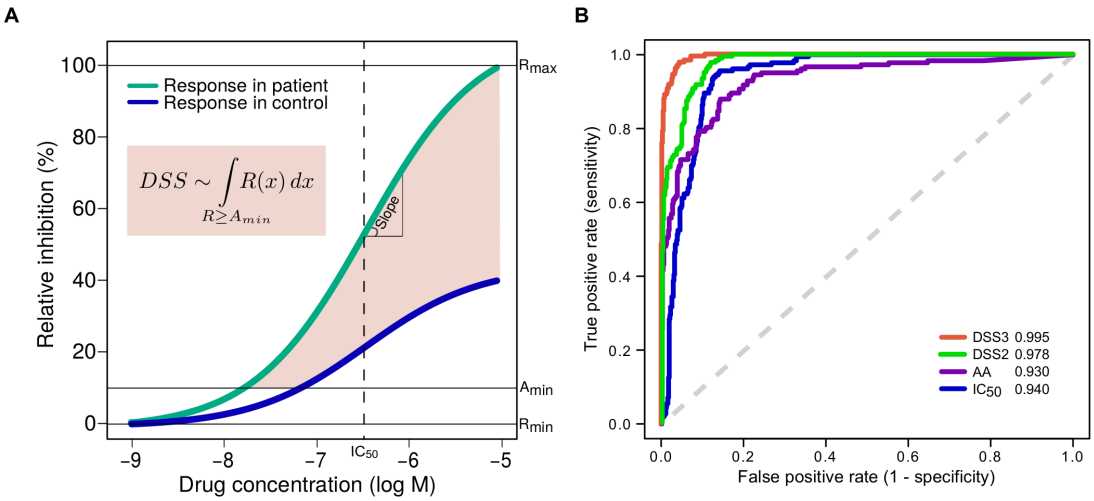



***Figure 7:*** *Calculation and evaluation of drug sensitivity score (DSS). (A) Selective DSS calculation using healthy bone marrow controls, and (B) evaluation of predictive accuracy of drug scoring metric using receiver operator characteristic (ROC) analysis to distinguish between 612 inactive and 183 active cases of drug response. The dotted line corresponds to random classification.*

Twenty-two bone marrow aspirates were obtained from 14 leukemia samples and 4 bone marrow controls, which were screened against 204 compounds, using DSRT. It was assumed that the compounds with a similar mode of action would have similar compound response profiles. To evaluate this, the drugs were classified into 13 different classes and unsupervised clustering of compounds was mapped to the classes using Rand Index (Rand, 1971). The Rand Index analysis clearly showed that DSS provides better clustering of drugs than AA and $IC_{50}$.

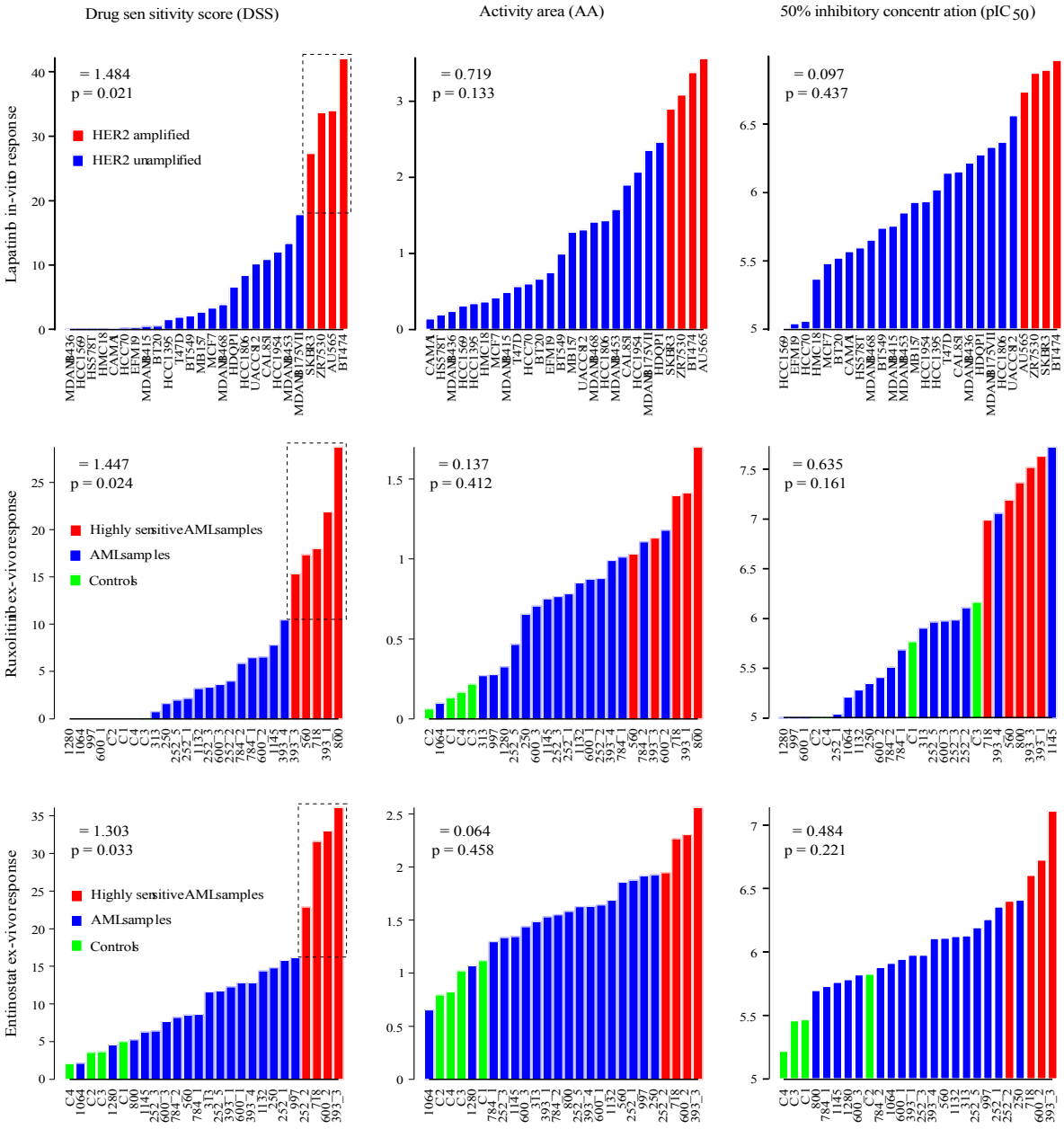



***Figure 8:*** *Distribution analysis of CCLE cell lines and AML patient samples. Highly sensitive cell lines or samples are in red; other cell lines or samples are in blue; Bone marrow controls are in green in cases of patient samples. In the first row, the distribution of lapatinib sensitivity in HER2 amplified and unamplified CCLE cell lines. The 2nd and 3rd row show the distribution of ruxolitinib and entinostat sensitivity in AML patients.*

The comparative cluster analysis of the compounds with their mode of action class using the Rand index demonstrated that DSS systematically improved compound clustering compared to $IC_{50}$ ($p = 1 \times 10^{-4}$, permutation test) and AA ($p < 5 \times 10^{-4}$, permutation test). A distribution analysis of the CCLE cell line, as well as AML patient drug screening data across compounds and across samples/cell lines showed that DSS provided more distinct sub-groupings of the cell lines or samples (Figure 8). DSS also helped identify drug responsive and non-responsive groups of patients or cell lines, enabling further investigation of the underlying biology to explain the groupings. I presented three examples of such cases in Publication I. The first example presented the drug response to PLX4720 (a selective RAF kinase inhibitor) across BRAF wild type and BRAF-V600E mutant cell lines; the second example, presented the drug response to PD-0325901 across wild type RAS and mutated RAS in hematopoietic and lymphoid cells; and the third example presented the lapatinib drug response across breast cancer cell lines. All these examples showed that DSS improved the drug response scoring resolution in cancer cell lines as well as in *ex vivo* patient samples. Since DSS, as well as AA, are based on the area under the curve, their performance was comparable than that of $IC_{50}$. However, DSS computes the exact area under the curve instead of using a trapezoidal type of approximation. Further normalization (see MATERIALS AND METHODS section for details), using maximal drug response and concentration ranges in the case of DSS, also improved the quantification compared to other response metrics. An R-package for DSS was implemented and freely distributed as open source software.

# 11. Target addiction scoring (TAS)

TAS is a method of transforming a drug response phenotypic profile into a target addiction profile, using a drug screening profile of the sample and a collection of the drugs' protein targets as input. TAS provides a deep insight into the ranking of targets according to functional importance in a given sample. The TAS approach (Figure 9) was systematically tested and evaluated in cell line data from GDSC, the largest publicly available database for drug sensitivity in cell lines, as well as their molecular profiles (Yang et al., 2013).

First, 107 cell lines were used with complete drug sensitivity profiles for all 138 compounds to evaluate the model. The predictive power of the TAS signatures was assessed using an additional 20 cell lines from GDSC. The 107 cell lines were clustered using an unsupervised classification method based on their gene expression, drug sensitivity and TAS profile. The comparative analysis of the clusters shows that tissue origins are the drivers of the gene expression, whereas TAS profile clusters were relatively independent of the tissue origin. This means that the information-encoded TAS profiles were only partly in agreement with their genomic signatures, providing complementary information in comparison to genomic profiling.

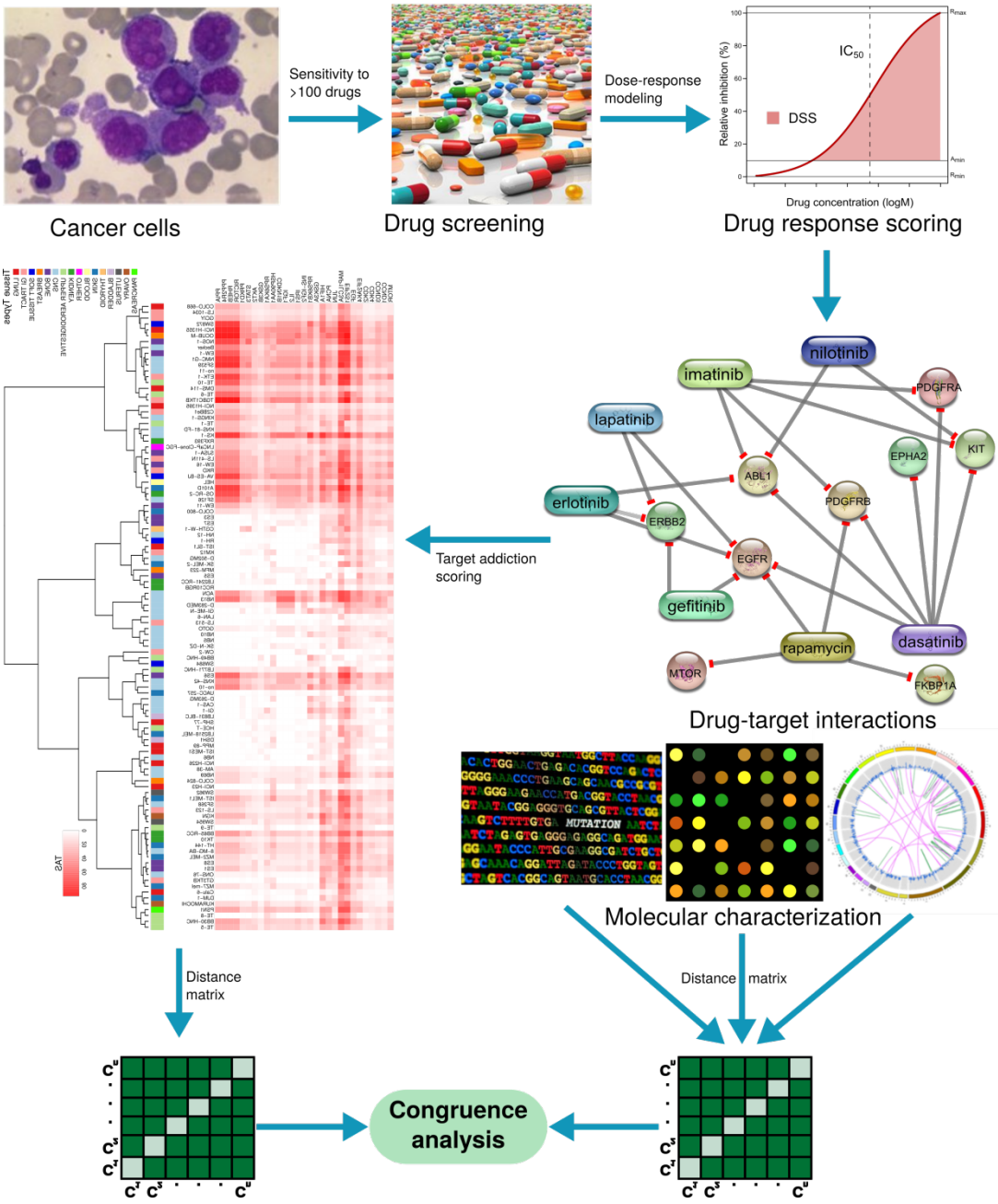



***Figure 9:*** *Schematic illustration of target addiction scoring and its integrative analysis with molecular profiles such as gene expression, copy number and mutation. The approach takes as input the drug response profiles of 107 cell lines and drug-target interaction information on 138 compounds to compute the target addiction score (TAS). The resulting TAS profiles of the cell lines were compared with the respective molecular and genomic profile using congruence analysis.*

A bootstrap analysis was performed to check the robustness of the clusters of 107 cell lines based on $IC_{50}$, AUC, DSS, and their TAS. The result showed that regardless of the drug response metrics, TAS always produced more robust sub-clusters. Of all the TAS-based clustering, TAS based on DSS had notably higher robustness in the sub-clusters.

A clinical application was shown using 151 primary leukemia patient samples profiled for 66 kinase inhibitors (Tyner et al., 2013). The TAS was applied to 151 leukemia patients' drug response data. This patient case study demonstrated patient-specific addiction patterns that were not explained by their diagnostic markers. The TAS package was implemented using R programming language and was freely distributed as open source software.

# 12. Delta scoring of drug combinations

Delta scoring was based on the ZIP model (Figure 10A) and applied to the published data (Griner et al., 2014), in which 466 drugs were screened in combination with Ibrutinib. The ROC curve analysis (Figure 10B) of the delta score demonstrated better accuracy in the classification between synergistic and antagonistic effects than other combination scoring methods such as HSA, Gamma, Beta, and CI. The delta score range is between -1 (antagonistic effect) and +1 (synergistic effect), which makes the data useful and easy to interpret.

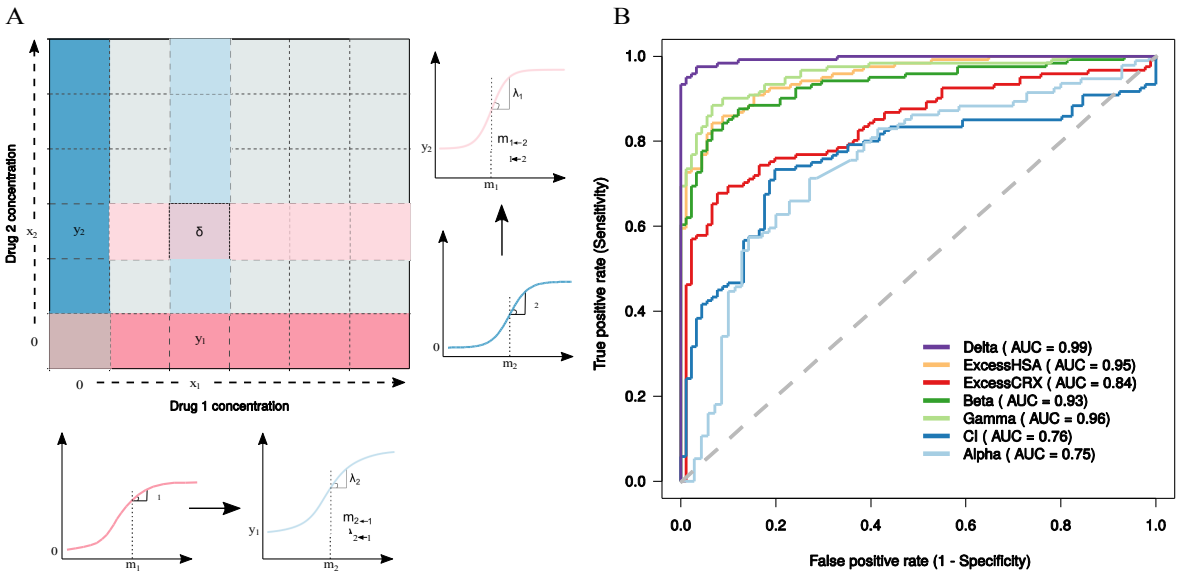



***Figure 10:*** *(A) Illustration of ZIP model formulation for delta scoring in a drug response matrix. The degree of interaction at a dose combination (x1, x2) is evaluated by comparing midpoint m and shape parameter λ from the individual drug responses (first column and last row) as well as their combined effects at Column x1, Row x2. Delta scoring considers the changes in m and λ for the dose-response curves between Drug 1 alone (bottom row) and the combination after adding x2 (Row x2), as well as between Drug 2 alone (first column) and the combination after adding x1 (Column x1). (B) Accuracy of classification of different drug combination scoring methods using ROC (receiver operator characteristics) analysis. The area under the ROC curve for each drug combination scoring method is shown for the classification of 112 synergistic and 91 antagonistic drug combinations.*

Even though antagonistic effects are often neglected in drug-drug interaction analysis in cancer studies, the antagonistic effects might be important for understanding the cross-talk between cancer signaling pathways, which may in turn provide clinical guidelines to avoid the administration of toxic drug combinations.

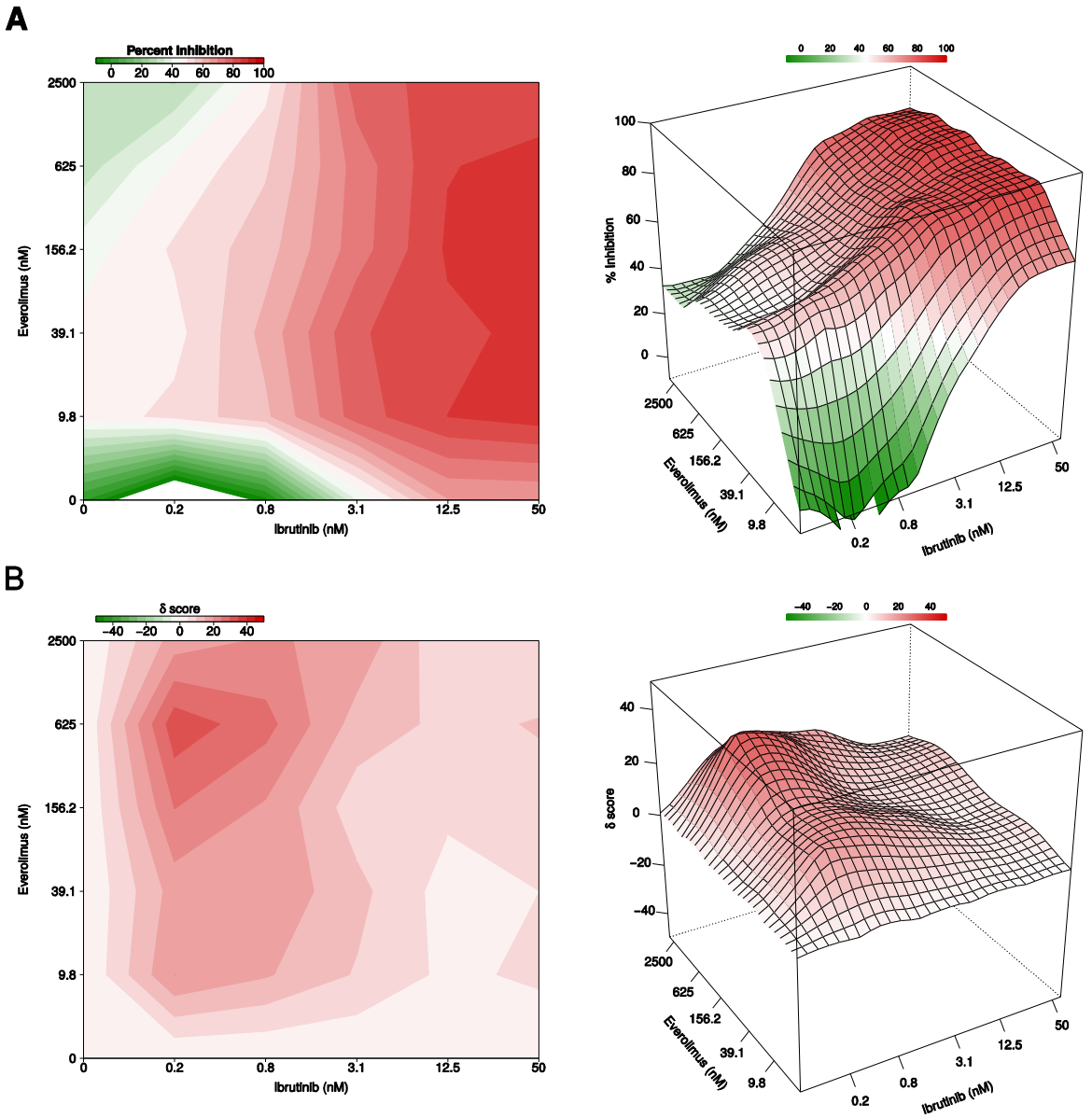



***Figure 11:*** *Illustration of two-dimensional (2D) and three-dimensional (3D) visualization of drug combination data. (A) Visualization of percent inhibition data matrix in 2D and 3D. The color is scaled in red (100% inhibition), white (50% inhibition) and green (no inhibition). (B) Visualization of drug combination effects in 2D and 3D. The color is scaled in red (synergistic, $\delta$>0), white (no interaction, $\delta$=0) and green (antagonistic, $\delta$<0).*

A 2D and 3D visualization (Figure 11) of the drug combination landscape was

also created, to give a better understanding of drug-drug interactions over the full concentration ranges, instead of just providing a summary interaction score. The drug interaction landscape provides a clear picture of the concentration range at which the drugs are synergistic or antagonistic. The package was implemented using R programming language and is freely distributed as open source software.

# 13. Application to AGCT data

In this work, I applied the methods developed in this thesis, including DSS and delta scoring of drug combinations, to analyze AGCT datasets, with the aim of suggesting novel, targeted treatments for AGCT.

Granulosa cell tumor (GCT) is a rare ovarian tumor, and occurs in around 5% of total ovarian cancer patients. It has two subtypes, of which the adult granulosa cell tumor (AGCT) type is the most common, whereas the juvenile type is rare. Although AGCTs can be diagnosed at Stage I, they recur in about 30% of patients, and about 60% of refractory patients die from recurrent disease. Even though radical surgery is the primary treatment of AGCT, chemotherapy is needed in the advanced stages and in recurrent disease. A platinum-based drug such as cisplatin or carboplatin is commonly used in combination with bleomycin and etopside or paclitaxel. However, the response rate of the treatment is very low (around 22%). As AGCT is a rare cancer subtype, performing randomized clinical trials is challenging. Hence, other approaches for revealing potential therapeutic targets for AGCT are needed.

In this analysis, I integrated different omics profiles of patients, such as drug screening, gene expression and mutation profiles. A total of seven AGCT

patient samples and three hGL (healthy granulosa-luteal cells) controls underwent drug screening across 230 compounds. Two GCT cell lines, KGN and COV434, were also screened for the same collection of compounds, and gene expression profiling was carried out for the seven AGCTs. The drug responses were quantified using DSS. The correlation analysis (Figure 12) shows that the correlation of average AGCT with the KGN cell line is higher (Spearman correlation r = 0.83) than that with the COV434 cell lines (Spearman correlation r = 0.74) and hGL (Spearman correlation r = 0.68). This means that the KGN cell line can be used as a cell line model, as it mimics the AGCT response profile.

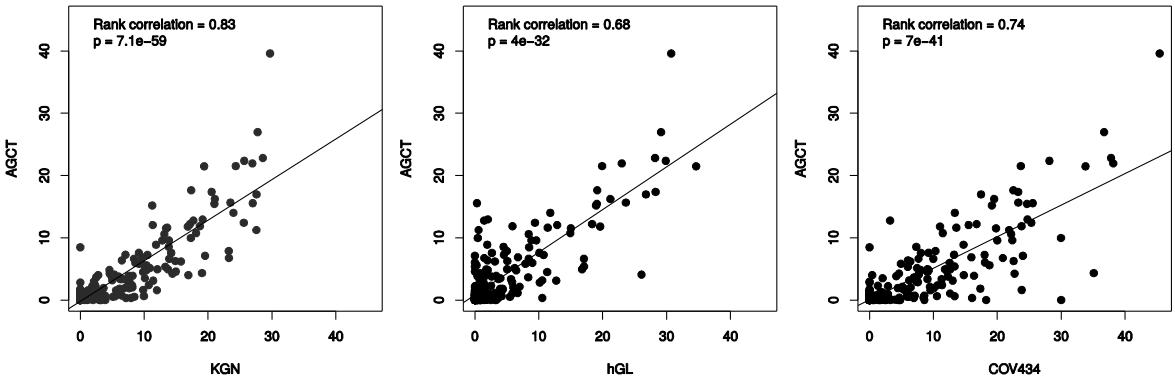



***Figure 12:*** *Correlation analysis of AGCT samples with KGN, COV484, and hGL cells. The drug screening profiles of AGCT were compared to the FOXL2 C134W mutated GCT cell line KGN, the FOXL2 wild-type GCT cell line COV434, and the FOXL2 wild-type hGL pool. The analysis used an average of seven AGCT samples and an average of three hGL controls.*

All AGCT samples, as well as the KGN cell line, had *FOXL2* mutation and were sensitive to the tyrosine kinase inhibitor (dasatinib), whereas *FOXL2* wild-type cells, such as COV434 and hGL, were resistant. The GCT cell lines, such as COV434 and KGN, were more sensitive to conventional drugs than

the AGCT samples. The selective DSS (sDSS) of AGCT samples against hGL controls showed dasatinib (a tyrosine kinase inhibitor) as one of the highly selective compounds (mean sDSS =11.3). The AGCT drug response profiles were compared to the bone marrow (BM) mononuclear cell profiles to evaluate sensitivity to healthy proliferating cells. Comparison with BM showed that both dasatinib and YM155 were selective.

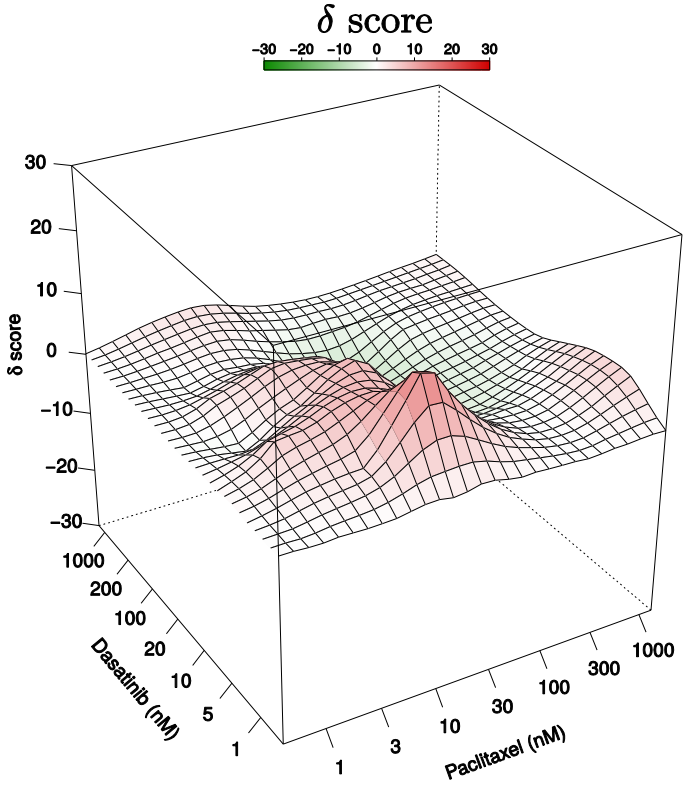



***Figure 13:*** *Drug combination of dasatinib and paclitaxel in KGN cell line. The color is scaled in red (synergistic, $\delta$>0), white (no interaction, $\delta$=0) and green (antagonistic, $\delta$<0).*

The efficacy of dasatinib in combination with conventional therapeutic paclitaxel was evaluated. When dasatinib was combined with paclitaxel, the

drug combination result showed synergistic effects in the KGN cell line, suggesting potential use of this combination for therapeutic purposes (Figure 13). No synergy was observed in the COV434 cell line.

This case study shows the general added value of the computational pipeline developed in this thesis for high-throughput drug screening for personalized oncology.

# DISCUSSION

The term “personalized medicine”, also known as “individualized medicine” or “precision medicine” is almost a buzzword today. After the sequencing of the human genome, hopes have been high for improvements in health care for many complex diseases such as cancer. Anticancer research and clinical practice are currently undergoing a paradigm shift in the field of medicine from conventional to personalized treatment (Kaur and Petereit, 2012). Personalized medicine is a broad field of health care that is informed by the individual’s unique clinical, molecular genomic, epigenomics and their associated environmental information. In order to fully make personalized medicine a reality, several barriers must be overcome, such as education, accessibility, regulation, and reimbursement. (Chan and Ginsburg, 2011). As personalized medicine involves a huge amount of data integration and analysis, we must address technical challenges such as the development of storage, and analytical and interpretive methods for translating or transforming raw information content for proactive, predictive, preventive and participatory health (Overby and Tarczy-Hornoch, 2013). In addition to these technical challenges, we also face ethical, legal challenges (Cordeiro, 2014), and economic challenges (Jakka and Rossbach, 2013).

This work dealt with the problems of large-scale omics as well high-throughput drug screening data analysis and mining. The development, advancement and usage of high-throughput technologies have led to an exponential amount of data growth that requires new tools and techniques to meet the challenges of data acquisition, storage, distribution, and analysis (Frelinger, 2015; Marx, 2013; Stephens et al., 2015; Zhang et al., 2015). So far, no state-of-the-art methods or pipelines exist that can serve all these purposes. Since the datasets are unique, the existing bioinformatics pipelines and methods must be customized. New models and efficient computational

pipelines are also greatly needed for analyzing and understanding the huge profiling datasets to gain meaningful insights that can be translated in clinics (Alyass et al., 2015; Frelinger, 2015; Marx, 2013; Ocana and Pandiella, 2010).

In this thesis, I have developed, implemented and applied novel bioinformatics models and approaches for the mining and analysis of high-throughput drug screening data. These new models enable 1) better quantification of drug sensitivity, 2) scoring and understanding of the target addiction of cancer cells, 3) the integration of target addiction with molecular profiling data, and 4) the quantification and analysis of drug combinations in high-throughput settings. The thesis includes the original publications of the individual methods developed here, and their integrative application to patient samples. The models have been carefully tested and found to be useful in the high-throughput setting, by minimizing the effort of manual inspection of such large datasets.

Publication I describes the development of a quantification model for drug response, through integrating dose-response parameters such as the $IC_{50}$, slope, and minimum and maximum response. The traditional approach of drug response quantification uses the $IC_{50}$, which is just one parameter and single point of information from the response curve (Sebaugh, 2011). A recent analysis has demonstrated the importance of the multiple parameters of drug response relationships (Fallahi-Sichani et al., 2013), by showing that using only the $IC_{50}$ as a measure of drug sensitivity causes other valuable information in the response curve to be lost. To capture all the information contained in the dose-response curve, I have developed an integrative quantitative model that considers all the parametric information in computing the optimized area under the curve. The availability of in-house AML patient sample data, as well as public CCLE and GDSC data, enabled systematic

comparative analysis with existing methods such as $IC_{50}$ (Garnett et al., 2012) and AA (Barretina et al., 2012). DSS was found to be robust against sources of technical variability, such as the interpolation of intermediate missing data points. DSS has shown improved reproducibility compared to $IC_{50}$, because $IC_{50}$ depends on the concentration range as well as the curve fitting methods and their inherent assumptions, whereas DSS is not affected. DSS has also improved patient sub-groupings and the stratification of cell lines. DSS calculation could be further improved by updating the drug response curve fit using the upper bend point of the dose response curve (Pozdeyev et al., 2014) and by incorporating growth rate differences (Hafner et al., 2016) or drug-induced proliferation changes (Harris et al., 2016). DSS could be adopted for siRNA screening or antibody screening, or any other screening in which the response is approximately sigmoid.

A recent comparison between CCLE and GDSC drug screening data revealed inconsistencies between the drug screening results of different laboratories (Haibe-Kains et al., 2013). The inconsistencies originate from either experimental or computational factors, or both. To enhance the reproducibility of drug screening data, it is important to follow standardized guidelines of experimental design, as well as the best computational pipelines and models (Hatzis et al., 2014; Haverty et al., 2016). Recent works show that calculating the drug response on the basis of the area under the curve model enhances the reproducibility of the drug screening data within and between laboratories (Bouhaddou et al., 2016; Geeleher et al., 2016; Mpindi et al., 2016).

Publication II expanded on the concept of the kinase inhibition sensitivity score (KISS), which was already presented in Publication I. In the case of target addiction scoring (TAS), the focus was extended to non-kinase targets. In this study, I applied a cell-based pan-cancer approach, showing that TAS

provides information that is independent of the tissue origin. This complementary information could provide functional relationships in the addiction patterns across different subtypes of cancer. The TAS profile contained primary and secondary druggable targets, which made the results more pharmaceutically actionable and clinically highly relevant, compared to the use of genomic profiling, which cannot provide actionable therapeutics for many different cancer types. The current model could be extended using an additional level of genomic and molecular information, including chromosome rearrangement, protein activities and epigenomic data, to obtain a more in-depth insight of the observed addiction patterns. The analysis showed that in comparison to drug screening profiles and genomic profiles, hierarchical clustering of TAS profiles provided highly robust clusters. However, the pipeline could be improved by using probabilistic model-based clustering and including a functional annotation of the genes to guide clustering and take into account the effects of the many compounds on the multiple targets.

In Publication III, I developed the ZIP reference model for drug combination interactions, which is compatible with the high-throughput experiments. I also systematically evaluated the limitations of the currently used drug combination reference models. The ZIP model incorporates the advantageous features of both the Loewe additivity model and the Bliss independence model, which enable systematic assessment of the various drug interaction patterns that arise from large-scale drug combination testing studies. The model computes the expected effect as the zero interaction between two drugs, which means that the potency of one drug is not altered by adding another drug. The delta score is calculated by subtracting the observed effect from expected effects. The analysis showed that merely having a combination score does provide the full picture of the interaction landscape. Therefore, I implemented the 2D and 3D interaction plots, which illustrate a more complete landscape of the drug interaction. The delta score

is centered at 0 and is symmetric with the finite boundary, making scoring data more directly interpretable than that in the Loewe additivity model. The current model utilizes the curve fitting of drug response relationships. In the future, this model could be extended to quantify the degree of interaction potency or potency shift between individual drugs and drug combinations.

Publication IV describes a case study of the application of DSS, TAS, and delta scoring in AGCT patient samples. The case study shows that the developed methods are well suited to the data analysis and integration of drug screening profiles, molecular profiles and drug combination profiles for high-throughput and personalized medicine settings. The results from these methods are clinically and pharmaceutically interesting and relevant, and warrant many future follow-up and confirmation studies.

# CONCLUSION

The massive growth of high-throughput data demands new models to better understand and interpret high-dimensional data and to extract the key information of most clinical and biological relevance. To meet some of these challenges, I have developed three novel models: (i) DSS, to quantify drug response, (ii) TAS, to capture the addiction signature of the cells, and (iii) the ZIP model, for scoring drug combination effects. The above models have been systematically evaluated, and demonstrated that these models were robust and perform better than existing models. All these models, i.e. DSS, TAS, and delta scoring using the ZIP model, are implemented in the R programing language and are freely distributed as open source R packages. The clinical and pharmacological relevance of the results are shown in both cancer cell lines and patients samples. The methods will significantly contribute to the effective analysis of high-throughput data from cancer cell lines and patient-derived samples, promote better understanding of cancer progression and develop individualized treatment options.

Despite the successful applications and relevant findings, some key challenges remain to be tackled in future research. For instance, the drug screening data of different laboratories are not consistent, an issue which needs to be addressed by a broader array of assays and approaches. Recent computational approaches, such as growth rate inhibition matrices (GR) or drug-induced proliferation (DIP), may help reduce some of these inconsistencies (Hafner et al., 2016; Harris et al., 2016). Other key challenges arise from the inter- and intra-heterogeneity of tumors and the clonal evolution of cancer cells, which requires development of novel models in the future.

# ACKNOWLEDGEMENTS

The work presented in this PhD thesis was performed at FIMM, between 2011 and 2016. In spite of my hard work, my PhD would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

My sincere appreciation goes to Susanna Rosas for helping me with practical matters and also for providing temporary housing at the beginning of my work at FIMM. I also thank Research Training Coordinator Gretchen Repasky for her continuous monitoring and guidance of the research training at FIMM.

I would like to express my thanks to FIMM and its administration for providing such outstanding research facilities, infrastructure and an enjoyable environment. I would also like to thank the Biocenter Finland International Visitor Fund, Doctoral Programme in Biomedicine (DPBM), the University of Helsinki Chancellor's travel grant and Cancer Society Finland's travel grant for their financial support. I am also thankful to DPBM for providing interesting courses, training and symposiums.

My PhD work was part of a large collaborative project on Personalized Cancer Medicine for hematological malignancies, and involved several research groups (Aittokallio, Kallioniemi, Wennerberg, Heckman/Knowles) at FIMM and the Helsinki Hematology Research Unit, as well as the Department of Hematology, University Helsinki Hospital Comprehensive Cancer Center. I thank all the members of these groups. All the models developed in this project were tested on data from these large collaborative projects. I would also like to thank all the principal investigators, Tero Aittokallio, Kirster Wennerberg, Olli Kallioniemi, Caroline Heckman, Jonathan Knowles, Kimmo

Porkka, Satu Mustjoki, and Mikko Anttonen, for their visions, valuable comments and suggestions. I would also like to express my gratitude to all my collaborators, who have helped me work on a diverse number of projects.

I am very grateful to my supervisor Prof. Tero Aittokallio for providing me with the opportunity to obtain my PhD and work on fascinating projects for Personalized Medicine. You constantly motivated me to complete the project. I thank you for all your help and input throughout my research; for your support and encouragement. I am also thankful to my co-supervisor, Dr. Krister Wennerberg, for his valuable time, comments, suggestions, and fruitful collaboration during my PhD research.

I am thankful to my thesis committee members Dr. Sergey Kuznetsov and Dr. Tapio Pahikkala for their guidence during PhD. I am also grateful to Dr. Jaakko Hollmén and Dr. Petri Auvinen for accepting invitation to serve as reviewers of this thesis and their valuable comments which indeed improved the thesis a lot.

I am also grateful to my fellow PhD students Tea Pemovska, Agnieszka Szwajda, John Patrick Mpindi, Mika Kontro, Riikka Karjalainen, Muntasir Mamun Majumder, Ashwini kumar, Disha Malani, Poojitha Ojamise, and Ulla-Maija Haltia.

I extend my thanks to all my past and present group members, Himanshu Chheda, Anna Cichonska, Abhishekh Gupta, Liye He, Alok Jaiswal, Suleiman Khan, Milla Kibble, Mehreen Ali, Teemu Daniel Laajala, Gopal Peddinti, Balaguru Ravikumar, Zia Ur Rehman, Jing Tang, Sanna Timonen, Laxmana Yetukuri, Bogdan Iancu, Agnieszka Szwajda, and Sushil Kumar Shakyavar,

for being so interactive and for all our wonderful discussions. Special thanks go to Dr. Peddinti Gopalacharyulu, Dr. Suleiman A. Khan, and Dr. Jing Tang for their help and guidance during my PhD projects.

I thank my friends and colleagues at FIMM: Swapnil Potdar, Saeed Khalid, Prson Gautam, Poojitha Ojamise, Vishal Sinha, and Ashwini Kumar for being so friendly and cooperative when I needed help. I extend special thanks to Ashwini Kumar and Disha Malani for providing me with help and support when I needed it the most, and for all the fun we had together.

I am grateful to the FIMM Technology Center; Timo Miettinen, Olle Hansson, Kari Tuomainen, Kyösti Sutinen, and Teemu Perheentupa, for their IT support. I also thank Dr. Milla Kibble and University of Helsinki language center for language corrections.

I express my sincere gratitude to my father Late Ram Janma Yadav and mother Sita Yadav. I am also grateful to my brother Ram Prakash, brother-in-law Deb Narayan and sister Sujan for their unconditional love, support and encouragement. I also thank my wife Nirmala for all her love and care, and for tolerating me at the crucial stages of my PhD studies.

Bhagwan Yadav
Helsinki, 2017

# REFERENCES

**Al-Ali, H., Lee, D.-H., Danzi, M. C., Nassif, H., Gautam, P., Wennerberg, K., Zuercher, B., Drewry, D. H., Lee, J. K., Lemmon, V. P., et al.** (2015). Rational Polypharmacology: Systematically Identifying and Engaging Multiple Drug Targets To Promote Axon Growth. *ACS Chem. Biol.* **10**, 1939–1951.

**Alyass, A., Turcotte, M. and Meyre, D.** (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med. Genomics* **8**,.

**An, W. F. and Tolliday, N. J.** (2009). Introduction: Cell-Based Assays for High-Throughput Screening. In (ed. Clemons, P. A.), Tolliday, N. J.), and Wagner, B. K.), pp. 1–12. Totowa, NJ: Humana Press.

**Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., Sung, B. and Aggarwal, B. B.** (2008). Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Pharm. Res.* **25**, 2097–2116.

**Argento, M., Hoffman, P. and Gauchez, A.-S.** (2008). Ovarian cancer detection and treatment: current situation and future prospects. *Anticancer Res.* **28**, 3135–3138.

**Atkinson, K. E.** (1989). *An introduction to numerical analysis*. New York: Wiley.

**Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehar, J., Kryukov, G. V., Sonkin, D., et al.** (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–307.

**Baselga, J. and Arribas, J.** (2004). Treating cancer's kinase "addiction." *Nat. Med.* **10**, 786–787.

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

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

**Bliss, C. I.** (1939). The Toxicity of Poisons Applied Jointly1. *Ann. Appl. Biol.* **26**, 585–615.

**Borisy, A. A., Elliott, P. J., Hurst, N. W., Lee, M. S., Lehar, J., Price, E. R.,**

**Serbedzija, G., Zimmermann, G. R., Foley, M. A., Stockwell, B. R., et al.** (2003). Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci.* **100**, 7977–7982.

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

**Burrell, R. A., McGranahan, N., Bartek, J. and Swanton, C.** (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345.

**Chan, I. S. and Ginsburg, G. S.** (2011). Personalized medicine: progress and promise. *Annu. Rev. Genomics Hum. Genet.* **12**, 217–244.

**Cheng, Y. and Prusoff, W. H.** (1973). Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108.

**Chou, T.-C.** (2006). Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacol. Rev.* **58**, 621–681.

**Cordeiro, J. V.** (2014). Ethical and legal challenges of personalized medicine: Paradigmatic examples of research, prevention, diagnosis and treatment. *Rev. Port. Saúde Pública* **32**, 164–180.

**Fallahi-Sichani, M., Honarnejad, S., Heiser, L. M., Gray, J. W. and Sorger, P. K.** (2013). Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat. Chem. Biol.* **9**, 708–714.

**Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. and Sorger, P. K.** (2006). Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol.* **2**, 458–466.

**Flaherty, K. T., Infante, J. R., Daud, A., Gonzalez, R., Kefford, R. F., Sosman, J., Hamid, O., Schuchter, L., Cebon, J., Ibrahim, N., et al.** (2012). Combined {BRAF} and {MEK} inhibition in melanoma with {BRAF} V600 mutations. *N. Engl. J. Med.* **367**, 1694–1703.

**Fleuren, E. D. G., Zhang, L., Wu, J. and Daly, R. J.** (2016). The kinome “at large” in cancer. *Nat. Rev. Cancer* **16**, 83–98.

**Frelinger, J. A.** (2015). Big Data, Big Opportunities, and Big Challenges. *J. Investig.*

*Dermatology Symp. Proc.* **17**, 33–35.

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

**Gautam, P., Karhinen, L., Szwajda, A., Jha, S. K., Yadav, B., Aittokallio, T. and Wennerberg, K.** (2016). Identification of selective cytotoxic and synthetic lethal drug responses in triple negative breast cancer cells. *Mol. Cancer* **15**,.

**Geeleher, P., Gamazon, E. R., Seoighe, C., Cox, N. J. and Huang, R. S.** (2016). Consistency in large pharmacogenomic studies. *Nature* **540**, E1–E2.

**Goldoni, M. and Johansson, C.** (2007). A mathematical approach to study combined effects of toxicants in vitro: Evaluation of the Bliss independence criterion and the Loewe additivity model. *Toxicol. Vitr.* **21**, 759–769.

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

**Griner, L. A. M., Guha, R., Shinn, P., Young, R. M., Keller, J. M., Liu, D., Goldlust, I. S., Yasgar, A., McKnight, C., Boxer, M. B., 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. *Proc. Natl. Acad. Sci.* **111**, 2349–2354.

**Gujral, T. S., Peshkin, L. and Kirschner, M. W.** (2014). Exploiting polypharmacology for drug target deconvolution. *Proc. Natl. Acad. Sci.* **111**, 5048–5053.

**Hafner, M., Niepel, M., Chung, M. and Sorger, P. K.** (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* **13**, 521–527.

**Haibe-Kains, B., El-Hachem, N., Birkbak, N. J., Jin, A. C., Beck, A. H., Aerts, H. J. W. L. and Quackenbush, J.** (2013). Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393.

**Harris, L. A., Frick, P. L., Garbett, S. P., Hardeman, K. N., Paudel, B. B., Lopez, C. F., Quaranta, V. and Tyson, D. R.** (2016). An unbiased metric of antiproliferative drug effect in vitro. *Nat. Methods* **13**, 497–500.

**Hatzis, C., Bedard, P. L., Birkbak, N. J., Beck, A. H., Aerts, H. J. W. L., Stern, D. F., Shi, L., Clarke, R., Quackenbush, J. and Haibe-Kains, B.** (2014). Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward? *Cancer Res.* **74**, 4016–4023.

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

**Heiser, L. M., Sadanandam, A., Kuo, W.-L., Benz, S. C., Goldstein, T. C., Ng, S., Gibb, W. J., Wang, N. J., Ziyad, S., Tong, F., et al.** (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl. Acad. Sci.* **109**, 2724–2729.

**Heppner, G. H.** (1984). Tumor heterogeneity. *Cancer Res.* **44**, 2259–2265.

**Hood, L. and Flores, M.** (2012). A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *N. Biotechnol.* **29**, 613–624.

**Jakka, S. and Rossbach, M.** (2013). An economic perspective on personalized medicine. *Hugo J.* **7**, 1.

**Janku, F.** (2014). Tumor heterogeneity in the clinic: is it a real problem? *Ther. Adv. Med. Oncol.* **6**, 43–51.

**Jia, J., Zhu, F., Ma, X., Cao, Z. W., Li, Y. X. and Chen, Y. Z.** (2009). Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discov.* **8**, 111–128.

**Kaur, J. S. and Petereit, D. G.** (2012). Personalized Medicine: Challenge and Promise. *J. Cancer Educ.* **27**, 12–17.

**Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., et al.** (2014). A draft map of the human proteome. *Nature* **509**, 575–581.

**Kottarathil, V. D., Antony, M. A., Nair, I. R. and Pavithran, K.** (2013). Recent Advances in Granulosa Cell Tumor Ovary: A Review. *Indian J. Surg. Oncol.* **4**, 37–47.

**Kyrönlahti, A., Kauppinen, M., Lind, E., Unkila-Kallio, L., Butzow, R., Klefström, J., Wilson, D. B., Anttonen, M. and Heikinheimo, M.** (2010). GATA4 protects granulosa cell tumors from TRAIL-induced apoptosis. *Endocr. Relat. Cancer*

**17**, 709–717.

**Lee, J. and Bogyo, M.** (2013). Target deconvolution techniques in modern phenotypic profiling. *Curr. Opin. Chem. Biol.* **17**, 118–126.

**Lehár, J., Zimmermann, G. R., Krueger, A. S., Molnar, R. A., Ledell, J. T., Heilbut, A. M., Short, G. F., Giusti, L. C., Nolan, G. P., Magid, O. A., et al.** (2007). Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.* **3**, 80.

**Lehár, J., Krueger, A. S., Avery, W., Heilbut, A. M., Johansen, L. M., Price, E. R., Rickles, R. J., Short 3rd, G. F., Staunton, J. E., Jin, X., et al.** (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat. Biotechnol.* **27**, 659–666.

**Li, M. C.** (1960). Effects of Combined Drug Therapy on Metastatic Cancer of the Testis. *JAMA* **174**, 1291.

**Lindblom, A., Zhou, X.-L., Liu, T., Liljegren, A., Skoglund, J. and Djureinovic, T.** (2004). Colorectal cancer as a complex disease: defining at-risk subjects in the general population – a preventive strategy. *Expert Rev. Anticancer Ther.* **4**, 377–385.

**Liu, B., Li, S. and Hu, J.** (2004). Technological advances in high-throughput screening. *Am. J. Pharmacogenomics Genomics-Related Res. Drug Dev. Clin. Pract.* **4**, 263–276.

**Loewe, S.** (1953). The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* **3**, 285–290.

**Longo, D. L., Döhner, H., Weisdorf, D. J. and Bloomfield, C. D.** (2015). Acute Myeloid Leukemia. *N. Engl. J. Med.* **373**, 1136–1152.

**Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V. S., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., et al.** (2011). Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10**, 188–195.

**Mancari, R., Portuesi, R. and Colombo, N.** (2014). Adult granulosa cell tumours of the ovary. *Curr. Opin. Oncol.* **26**, 536–541.

**Marx, V.** (2013). Biology: The big challenges of big data. *Nature* **498**, 255–260.

**Marx, K. A., O'Neil, P., Hoffman, P. and Ujwal, M. L.** (2003). Data mining the NCI cancer cell line compound GI(50) values: identifying quinone subtypes effective

against melanoma and leukemia cell classes. *J. Chem. Inf. Comput. Sci.* **43**, 1652–1667.

**Meacham, C. E. and Morrison, S. J.** (2013). Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337.

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

**Ocana, A. and Pandiella, A.** (2010). Personalized therapies in the cancer “omics” era. *Mol. Cancer* **9**, 202.

**Overby, C. L. and Tarczy-Hornoch, P.** (2013). Personalized medicine: challenges and opportunities for translational bioinformatics. *Per. Med.* **10**, 453–462.

**Pectasides, D., Papaxoinis, G., Fountzilas, G., Aravantinos, G., Pectasides, E., Mouratidou, D., Economopoulos, T. and Andreadis, C.** (2008). Adult granulosa cell tumors of the ovary: a clinicopathological study of 34 patients by the Hellenic Cooperative Oncology Group (HeCOG). *Anticancer Res.* **28**, 1421–1427.

**Pemovska, T., Kontro, M., Yadav, B., Edgren, H., Eldfors, S., Szwajda, A., Almusa, H., Bespalov, M. M., Ellonen, P., Elonen, E., et al.** (2013). Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3**, 1416–1429.

**Pemovska, T., Johnson, E., Kontro, M., Repasky, G. A., Chen, J., Wells, P., Cronin, C. N., McTigue, M., Kallioniemi, O., Porkka, K., et al.** (2015). Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. *Nature* **519**, 102–105.

**Pozdeyev, N., Yoo, M., Mackie, R., Schweppe, R. E., Choon Tan, A. and Haugen, B. R.** (2014). Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget*.

**Rand, W. M.** (1971). Objective Criteria for the Evaluation of Clustering Methods. *J. Am. Stat. Assoc.* **66**, 846.

**Ryall, K. A., Shin, J., Yoo, M., Hinz, T. K., Kim, J., Kang, J., Heasley, L. E. and Tan, A. C.** (2015). Identifying kinase dependency in cancer cells by integrating high-throughput drug screening and kinase inhibition data. *Bioinformatics* btv427.

**Saultz, J. and Garzon, R.** (2016). Acute Myeloid Leukemia: A Concise Review. *J. Clin. Med.* **5**, 33.

**Schleidgen, S., Klingler, C., Bertram, T., Rogowski, W. H. and Marckmann, G.** (2013). What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Med. Ethics* **14**, 55.

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

**Sebaugh, J. L.** (2011). Guidelines for accurate EC50/IC50 estimation. *Pharm. Stat.* **10**, 128–134.

**Shepherd, J. E.** (2000). Current strategies for prevention, detection, and treatment of ovarian cancer. *J. Am. Pharm. Assoc. (Washington,D.C. 1996)* **40**, 392–401.

**Shi, F.-T., Cheung, A. P. and Leung, P. C. K.** (2009). Growth differentiation factor 9 enhances activin a-induced inhibin B production in human granulosa cells. *Endocrinology* **150**, 3540–3546.

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

**Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S. and Robinson, G. E.** (2015). Big Data: Astronomical or Genomical? *PLOS Biol.* **13**, e1002195.

**Terstappen, G. C., Schlupen, C., Raggiaschi, R. and Gaviraghi, G.** (2007). Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov* **6**, 891–903.

**Tran, T. P., Ong, E., Hodges, A. P., Paternostro, G. and Piermarocchi, C.** (2014). Prediction of kinase inhibitor response using activity profiling, in vitro screening, and elastic net regression. *BMC Syst. Biol.* **8**, 74.

**Tyner, J. W., Yang, W. F., Bankhead, A., Fan, G., Fletcher, L. B., Bryant, J., Glover, J. M., Chang, B. H., Spurgeon, S. E., Fleming, W. H., et al.** (2013). Kinase Pathway Dependence in Primary Human Leukemias Determined by Rapid Inhibitor Screening. *Cancer Res.* **73**, 285–296.

**Vis, D. J., Bombardelli, L., Lightfoot, H., Iorio, F., Garnett, M. J. and Wessels, L. F.** (2016). Multilevel models improve precision and speed of $IC_{50}$ estimates.

*Pharmacogenomics* **17**, 691–700.

**Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A. and Kinzler, K. W.** (2013). Cancer Genome Landscapes. *Science (80-. ).* **339**, 1546–1558.

**Wild, C. P.** (2012). The Role of Cancer Research in Noncommunicable Disease Control. *JNCI J. Natl. Cancer Inst.* **104**, 1051–1058.

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

**Zang, R., Li, D., Tang, I.-C., Wang, J. and Yang, S.-T.** (2012). Cell-based assays in high-throughput screening for drug discovery. *Int. J. Biotechnol. Wellness Ind.* **1**, 31.

**Zhang, Y., Zhu, Q. and Liu, H.** (2015). Next generation informatics for big data in precision medicine era. *BioData Min.* **8**,.

**Zou, H. and Hastie, T.** (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **67**, 301–320.

THE END

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

- Sir Winston Churchill

ISBN 978-951-51-2965-9 (PAPERBACK)
ISBN 978-951-51-2966-6 (PDF)
HTTP://ETHESIS.HELSINKI.FI/
HELSINKI 2017