Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
mgr Małgorzata Elżbieta Oleś
born in: Kraków, Poland
Oral-examination: 23 February 2017

Pharmacogenomics of Primary Blood Cancer Cells

Referees: Dr. Jan O. Korbel
Prof. Dr. Michael Boutros

To Laura ♡
my dear daughter

Abstract

Current developments of multi-omics technologies greatly accelerate advancements in the field of cancer research. They not only allow us to unravel the complex biology of tumors, but are also a huge step towards precision oncology, where a personalized treatment is proposed based on the tumor's unique combination of molecular features. Heterogeneity present both within a single tumor and between patients with the same disease pose serious challenges to treatment success and trial design, as many factors influencing drug response in cancer remain unknown.

The work presented here investigated the diversity of drug response profiles and their associations with the underlying molecular features of 273 primary cancer samples. The analysis combined data from high-throughput drug profiling of 90 compounds with multi-omics comprising exome sequencing, RNA sequencing and methylation profiling.

The analysis of various hematological malignancies uncovered a rich landscape of phenotype-genotype relationships. Targeted inhibition of, for example, CHK, SERCA, BCL2 and survivin revealed disease-specific pathway dependencies within B- and T-cell lymphomas. Moreover, the observed similarity of drug response profiles identified unexpected activity of compounds including CHK inhibitors, which implied to act through B-cell receptor (BCR) signaling pathway. By focusing specifically on chronic lymphocytic leukemia, dissection of the molecular fundamentals of known biomarkers was possible. Drug response measured *ex vivo* confirmed the biological relevance of major predictors of patient clinical outcome: the sensitivity of IGHV unmutated samples to BCR inhibition and the resistance of *TP53* mutated samples to chemotherapy and nutlin-3. Susceptibility of trisomy 12 cases to SYK, BTK, PI3K and MEK inhibitors suggested a mode of action through amplification of BCR pathway signaling. Additionally, the study proposed possible targeting options for recurrently mutated genes, such as *BRAF*, *CREBBP* and *PRPF8*. Multivariate analysis estimated the contributions of mutations, RNA expression and DNA methylation to our power to predict drug response. Finally, drug response was shown to be valuable and frequently superior to established biomarkers in predicting patient clinical endpoints.

In summary, the combination of viability screening with multi-omics profiling is a powerful tool for studying disregulated signaling pathways in cells. *Ex vivo* drug profiling performed on primary cancer samples has proved to be a proxy of true biology. Together with faithful reproducibility of measurements this shows great promise in directly advancing precision oncology.

Zusammenfassung

Aktuelle Entwicklungen im Bereich der “omik”-Technologien tragen wesentlich zur Beschleunigung des Fortschritts in der Krebsforschung bei. Diese Entwicklungen ermöglichen nicht nur die Entschlüsselung der komplexen Tumorbiologie, sondern stellen auch einen großen Schritt nach vorn bzgl. der personalisierten Onkologie dar, die eine individuelle Behandlung eines jeden Tumors, basierend auf dessen einzigartigen molekularen Eigenschaften zu ihrem Ziel hat. Heterogenität, sowohl auf der Ebene des einzelnen Tumors, als auch zwischen verschiedenen Patienten mit der gleichen Krankheit stellt eine große Herausforderung sowohl für die Behandlung von Patienten als auch für die Planung wissenschaftlicher Studien dar, da viele Faktoren, die die Wirkung bestimmter Medikamente beeinflussen, nach wie vor unbekannt sind.

Die vorliegende Arbeit untersucht Dosis-Wirkung-Zusammenhänge und deren Wechselwirkung mit molekularen Eigenschaften anhand von 273 primären Tumorproben. Die Analyse verbindet dabei Daten aus der Hoch-Durchsatz-Charakterisierung von 90 Wirkstoffen mit verschieden “-omik”-Messungen: Exome- und RNA-Sequenzierung, sowie Methylierung.

Die Analyse verschiedener hämatologischer Tumore führte dabei zu einer Entschlüsselung eine großen Zahl von Verbindungen zwischen Genotyp und Phänotyp. So konnte beispielsweise durch eine gezielte Hemmung von CHK, SERCA, BCL2 und Survivin gezeigt werden, dass eine je nach Tumorart (B- oder T-Zellen Lymphome) verschiedene Signalweise angesprochen werden. Desweiteren wurde eine durch eine Ähnlichkeitsanalyse der Dosis-Wirkungsprofile oft unerwartete Aktivität bestimmter Wirkstoffe, wie z.B. CHK Inhibitoren, die den B-Zellen Rezeptor (BCR) Signalweg beeinflussen, festgestellt. Im Zentrum der Arbeit steht besonders die chronische lymphatische Leukämie, dabei war es möglich, die Wirkungsweise bekannter Biomarker auf einer fundamentalen Ebene aufzuklären. Wirkungsmessungen bestimmter Medikamente anhand von *Ex vivo* Experimenten bestätigten die biologische Relevanz wichtiger Marker, die zur Vorhersage klinischer Ergebnisse geeignet sind, so z.B. die Sensitivität von Proben ohne IGHV Mutation bzgl. der Inhibition von des B-Zellen Rezeptors (BCR) und das Nichtanschlagen von *TP53*-mutierten Proben auf Chemotherapie und Nutlin-3. Die Anfälligkeit von Proben mit Trisomie 12 bzgl. SYK, BTK, PI3K und MEK Inhibition legte nahe, dass diese durch eine verstärkte Aktivierung des BCR Signalweges wirken. Außerdem liefert die vorliegende Arbeit potentielle Angriffspunkte im Falle immer wieder vorkommender Genmutationen wie *BRAF*, *CREBBP* and *PRPF8*. Mit Hilfe multivariater Analyse wurde der Beitrag

der Mutationen, der RNA Expression und der DNA Methylierung zur Vorhersage der Dosis-Wirkungsbeziehungen abgeschätzt. Zudem konnte gezeigt werden, das die Dosis-Wirkungsbeziehung oft eine wertvolle Vorhersage des klinischen Resultates liefert, die Prognosen auf Basis etablierter Biomarkers häufig überlegen ist.

Zusammenfassend kann festgehalten werden, dass die umfangreiche Analyse von Dosis-Wirkungsbeziehungen in Kombination mit verschieden "-omik"-Daten ein leistungsstarkes Werkzeug zur Untersuchung disregulierter Signalwege in Zellen ist. Es hat sich gezeigt, dass *Ex vivo* Analysen basierend auf Zellen aus primären Tumoren der tatsächlichen Biologie sehr nahe kommen können. Diese Ergebnisse zeigen vielversprechende Wege zur personalisierten Krebsmedizin auf.

Streszczenie

Zauważalny obecnie rozwój technologii multiomicznych znacznie przyczynia się do postępów w badaniach nad rakiem. Pozwala nie tylko na rozwikłanie złożoności biologicznej nowotworów, lecz stanowi także duży krok w kierunku precyzyjnej onkologii, w której to indywidualna terapia proponowana jest na podstawie unikalnej kombinacji cech molekularnych danego guza. Różnorodność obecna zarówno w obrębie pojedynczego nowotworu, jak i pomiędzy pacjentami z tym samym typem nowotworu, stanowi poważne wyzwanie dla terapii i badań klinicznych, ponieważ wiele czynników wpływających na odpowiedź na leki pozostaje wciąż nieznanych.

Przedstawiona praca bada różnorodność profili odpowiedzi na leki oraz ich związki z charakterystyką molekularną 273 próbek nowotworowych. Przeprowadzona analiza łączy dane z eksperymentów wysokiej wydajności testujących 90 leków z danymi multiomicznymi składającymi się z egzomu, transkryptomu oraz metylomu.

Analiza różnych nowotworów hematologicznych ukazała bogaty zbiór relacji pomiędzy fenotypem a genotypem. Ukierunkowana inhibicja, na przykład, CHK, SERCA oraz BCL2 ujawniła specyficzne dla danej choroby zależności w szlakach sygnalizacji komórkowej chłoniaków komórek B i T. Ponadto, podobieństwa pomiędzy profilami odpowiedzi na leki wykazały nieoczekiwane działanie niektórych z nich, włącznie z inhibitorami CHK, które przejawiały aktywność skierowaną na szlak sygnalizacji receptora limfocytu B (BCR). Poprzez skoncentrowanie się jednym typie nowotworu, przewlekłej białaczce limfatycznej, możliwa była dokładna analiza podstaw molekularnych znanych biomarkerów. Działanie leków mierzone *ex vivo* potwierdziło biologiczne znaczenie głównych czynników predykcyjnych w rokowaniu pacjenta: podatność próbek z niezmutowanym IGHV na inhibicję szlaku BCR, oraz odporność próbek z mutacją genu *TP53* na chemioterapię i nutlin-3. Podatność próbek z trisomią 12 na inhibicję SYK, BTK, PI3K i MEK sugerowała tryb działania przez amplifikację sygnału szlaku BCR. Dodatkowo w badaniu zaproponowane zostały możliwe cele terapii ukierunkowanych dla często zmutowanych genów, takich jak *BRAF*, *CREBBP* i *PRPF8*. Analiza wielowymiarowa pozwoliła oszacować wkład mutacji, ekspresji RNA oraz metylacji DNA w możliwość przewidywania odpowiedzi na leki. Na koniec pokazano, że odpowiedź na leki jest cennym i często nawet lepszym narzędziem od klasycznych biomakerów w prognozowaniu rokowań pacjenta.

Podsumowując, połączenie analizy odpowiedzi na leki z charakterystyką multiomiczną jest mocnym narzędziem w badaniu rozregulowanych ścieżek sygnałowych. Profilowanie leków *ex vivo* odzwierciedla rzeczywistą biologię, a wraz z dużą powtarzalnością pomiarów stanowi obiecującą przyszłość w bezpośrednim rozwijaniu precyzyjnej onkologii.

Acknowledgements

First and foremost, I would like to express my deep gratitude to everyone who endorsed my admission to the EMBL International PhD Programme and to those who interacted with me throughout this unforgettable journey. Sincere thanks to my supervisor Wolfgang Huber for not only teaching, advising and supporting me, but also for giving me the opportunity to work on this exciting cutting-edge project. Warm thanks to the whole Huber Group, both the past and the present members. Especially to my friends Simon Anders, Alejandro Reyes, Aleksandra Pękowska, Paul-Theodor Pyl, Bernd Klaus and Mike Smith for their guidance, support, advice and many stimulating discussions, both scientific and casual ones.

Collaboration played a key part in my work, therefore I want to thank all of my collaborators for their great input. In particular Thorsten Zenz for his continuous stream of ideas, which made me not only a better scientist but also an advanced user the *ggplot2* R package. Special thanks go to Sascha Dietrich, Leopold Sellner, Mikołaj Słabicki, Taronish Dubash and Emma Andersson.

I would also like to thank the members of my Thesis Advisory Committee: Jan Korbel, Michael Boutros and Oliver Stegle for their care, guidance and insightful comments on my project. Thanks to Nassos Typas and Bernd Fischer for agreeing to be a part of my thesis defense committee.

Last but not least, I would like to thank all my family. They always believe in my abilities and encourage me to pursue new challenges. Special thanks to my husband Andrzej, who always stood by me even when the days were rough. Warm thanks to my parents, Mirosława and Stanisław, and to my dear brother Tomek, for their constant support, understanding and including me in their daily life despite being over 1000 km apart. Finally, I thank my brother-in-law Tomek, who was a great example of how to lead a fulfilling life—taking good care of your work and responsibilities but at the same time not forgetting about other passions, joy and social interactions. Realizing this made the final stages of writing this thesis much more comfortable and bearable.

Contents

Abstract vii

Zusammenfassung ix

Streszczenie xi

Acknowledgements xiii

Acronyms 1

Preface 3

1 Introduction 5

1.1 Biology and stratification of hematological malignancies 6

1.1.1 B-cell neoplasms 6

1.1.2 T-cell neoplasms 7

1.1.3 Myeloid neoplasms 8

1.2 Standard treatment strategies for hematologic malignancies 8

1.3 Personalized therapy in hematologic malignancies 9

1.4 Pharmacogenomic studies and cancer treatment 11

1.5 Overview of the undertaken approach 12

2 Phenotypic profiling of primary cancer samples by high-throughput drug screening 15

2.1 Performed drug screens 15

2.1.1 Pilot screen 15

2.1.2 Main screen 17

2.2 Processing of raw values obtained from cell viability assay 18

2.2.1 Data normalization and quality control 18

2.3 Data reproducibility 20

2.3.1 Reproducibility within same drug screen 21

2.3.2 Reproducibility between different drug screens 23

2.3.3 Reproducibility between different screening platforms 23

2.4 Parameters characterizing drug response 25

3 Molecular profiling of primary cancer samples by using multi-omics 27
3.1 Genomics . . . 28
3.1.1 Fluorescence in situ hybridization . . . 28
3.1.2 Targeted sequencing . . . 28
3.1.3 SNP arrays . . . 28
3.1.4 Whole exome sequencing . . . 28
3.1.5 Aggregation of genomic information . . . 31
3.2 Transcriptomics . . . 32
3.3 Methylomics . . . 32
3.4 Mutational landscape of patient cohort . . . 33
3.5 Contributions . . . 35

4 *Ex vivo* drug sensitivity in primary cancer cells 37
4.1 Clustering of drug phenotypes . . . 38
4.1.1 Chronic lymphocytic leukemia . . . 38
4.1.2 Mantle cell lymphoma . . . 39
4.1.3 T-cell prolymphocytic leukemia . . . 40
4.1.4 Summary . . . 41
4.2 Functional classification of lymphoproliferative disorders . . . 43
4.3 Influence of cell lineage and disease subtype on drug response . . . 43

5 Molecular factors influencing drug response in CLL 53
5.1 Biomarkers of chronic lymphocytic leukemia . . . 54
5.1.1 IGHV status . . . 54
5.1.2 *TP53* mutation / 17p deletion . . . 56
5.1.3 Trisomy 12 . . . 60
5.1.4 Rare mutational subclones . . . 65
5.2 Summary . . . 67

6 Modulators of drug response and clinical outcome 69
6.1 Multivariate assessment of drug response determinants . . . 70
6.2 Predictors of patient outcome . . . 73

7 Tools facilitating collaboration 79

8 Conclusions and perspectives 83

Appendix A Author's publications 85

Appendix B Patient clinical data 87

Appendix C Characteristics of drugs used in the drug screens 95

Appendix D Concentrations of drugs used in the drug screens 99

Appendix E Influence of batch effect on drug-gene associations 103

Appendix F Dose-response curve fitting **105**

Appendix G Genes used in variant calling **107**

Appendix H Comparison of significant associations of drug response and mutation in the two drug screens **109**

Appendix I Results of multivariate Cox regression **111**

Bibliography **113**

Acronyms

CLL chronic lymphocytic leukemia

MCL mantle cell lymphoma

MZL marginal zone lymphoma

LPL lymphoplasmacytic lymphoma

B-PLL B-cell prolymphocytic leukemia

HCL hairy cell leukemia

HCL-V hairy cell leukemia variant

FL follicular lymphoma

T-PLL T-cell prolymphocytic leukemia

Sézary Sézary syndrome

PTCL-NOS peripheral T-cell lymphoma not otherwise specified

AML acute myeloid leukemia

BCR B-cell receptor

BTK Bruton's tyrosine kinase

SYK spleen tyrosine kinase

PI3K phosphoinositide 3-kinase

ROS reactive oxygen species

CHK checkpoint kinase

HSP heat shock proteins

SERCA sarco/endoplasmic reticulum Ca2+ ATPase

EGFR epidermal growth factor receptor

IGHV immunoglobulin heavy-chain variable

U-CLL CLL with unmutated IGHV region

M-CLL	CLL with mutated IGHV region
VSE	**VisualScreenExplorer**
RMSD	root-mean-square deviation
IC_{50}	half maximal inhibitory concentration
EC_{50}	half maximal effective concentration
AUC	area under curve
PC	principal component
CR	complete response
OS	overall survival
WHO	World Health Organization
DKFZ	German Cancer Research Center
NCT	National Center for Tumor Diseases
EMBL	European Molecular Biology Laboratory
FISH	fluorescence in situ hybridization
WES	whole exome sequencing
SNP	single nucleotide polymorphism
SNV	single nucleotide variant

Preface

Pharmacogenomics is a scientific discipline which aims to use different levels of genomic information to infer a person's response to drugs. Accessibility of high-throughput drug screening and sequencing techniques promoted the field significantly over the last years. Studying patient tumor cells in that fashion is currently considered as the finest available way to advance precision medicine [1].

Until recently high-throughput pharmacogenomics was pursued only by using immortalized cancer cell lines as a model of the disease. The study presented in this dissertation goes one step further. The screening was performed on primary tumor samples derived from a large cohort of patients. Only such cells mirror the real biology of tumors, and only such approach allows to achieve sufficient statistical power of the analysis. The bioinformatic analysis aimed to explore the landscape of drug response determinants in order to: (i) infer similarities in the drugs' mode of action, (ii) classify different diseases into functionally convergent groups, (iii) identify compounds and targets preferably active in the presence of biomarkers or recurrently mutated genes, (iv) explain heterogeneous drug response, (v) and estimate how much of it can be explained by each multi-omics data type, and finally, (vi) assess the importance of drug response in predicting patient's survival. All of these objectives are addressed in this thesis.

The dissertation is organized into eight chapters. **Chapter 1** introduces the different types of studied lymphoproliferative malignancies, describes the concept of personalized therapy and reviews recent pharmocogenomic studies. The last section presents the layout of the conducted study. The following two chapters give an overview of data and its preprocessing, as well as assessments of its quality and integrity. Specifically, the drug screening and multi-omics approaches are described in **Chapter 2** and **Chapter 3**, respectively. The main results are presented in Chapters 4–6. **Chapter 4** focuses exclusively on the drug response data. First, we use the "guilt by association" approach to query the similarities of action between compounds. Next, we undertake a classification of the diseases based on drug response profiles. As the last step, we compare the different diseases to identify disease-specific vulnerabilities. In **Chapter 5** we interrogate individual gene-drug associations and revise the targets of drugs which showed sensitive or resistant profile in the presence of mutation in order to inform pathway dependencies. We confirm clinical observations and that *ex vivo* drug screening is able to discover new associations. **Chapter 6** combines multi-omics data comprising genetics, transcriptomics and methylomics

in order to select groups of features responsible for the variability of response to a given drug. We also show that drug profiling can be used to predict clinical outcome. Tools which facilitated collaboration are shortly reviewed in **Chapter 7**. It mainly concentrates on the developed web application, which allowed everybody in the project irrespective of their training in bioinformatics, to visualize the data and to formulate hypotheses. The last chapter **Conclusions and perspectives** contains closing remarks and outlooks on developments in the studied field. Last but not least, the thesis includes nine appendices which serve as a reference and provide additional information for the sake of completeness of the document. **Appendix A** includes a list of manuscripts (published, submitted for review, and in preparation) on the projects in which the author was involved during her doctoral studies.

The study introduced here is the result of a fruitful collaboration between the groups of Dr. Wolfgang Huber from the European Molecular Biology Laboratory (EMBL) and Prof. Dr. med. Thorsten Zenz affiliated with the National Center for Tumor Diseases (NCT). The experimental part was performed at NCT, whereas bioinformatic analysis was carried out at EMBL. Thanks to close collaboration with physician practitioners the study gained clinical insights. Although the project grew over time and the number of people working on it increased recently, the author of the thesis was the chief bioinformatican providing analysis, as well as managing and processing all of the available data. The presented results were directly obtained by the author unless otherwise stated.

1 Introduction

Doctors are men who
prescribe medicines of which they know little,
to cure diseases of which they know less,
in human beings of whom they know nothing.

— **Voltaire**
Voltaire's Notebooks 1952

Curing cancer in a given patient is not only the responsibility of medical doctors, but mostly, although implicitly, the obligation of cancer research community. First insights into cancer biology and its classification were gained exclusively by medical professionals who had direct contact with the affected patients. These actions, based on histopathological exams of malignant tissue, were hardly ever useful for cancer treatment itself; rather, they were intended to provide speculation on the prognosis and probable course of the disease. With continuous advancements in biological and sequencing technologies over the last two decades, scientists were able to acquire a much deeper knowledge of the biology of cancer. They were able to characterize hallmarks of carcinogenesis, identify malfunctioning signaling pathways and provide genetic markup of malignant cells, determine cells' strategies to fight misbehaving cells (for example, by activation of tumor suppressor genes), redefine classification of tumors, and build models of cancer evolution. And even though the understanding is still far from complete, with some tumor types being much better studied than the others, there are attempts to use this knowledge as feedback to adjust patients' treatment according to the available information. Since every tumor is different, these treatment strategies have to be individualized. This is the concept of personalized medicine which currently starts to blossom. By the end of the day, it is scientists' responsibility to provide medical doctors with specialized and precise tools to treat malignant diseases.

1.1 Biology and stratification of hematological malignancies

The findings described in this dissertation, although we believe can be generalized to a diversity of cancers, are entirely based on examples of hematological malignancies. It is instrumental for the reader to be introduced to the general characteristics of specific entities on which the thesis focuses.

Hematological malignancies are a group of nonepithelial cancers which originate from hematopoietic stem cells of blood-forming tissue, such as bone marrow, or in the cells of immune system [2]. Their extensive classification distinguishes types depending on: (i) the stage of hematopoiesis from which the tumor originates, (ii) the place where tumor cells accumulate (and therefore cause symptoms), (iii) the severity of the disease, (iv) and the molecular biomarkers identified in the cells. The most up-to-date classification is published by the World Health Organization (WHO)[3, 4]. However, it is very complex, still ambiguous and subject to improvements as new facts about hematological malignancies are being discovered.

1.1.1 B-cell neoplasms

B lymphocytes (B-cells) take part in acquired immune system by secreting antibodies upon stimulation. Their maturation process originates from common lymphoid progenitor cells and involves bone marrow and secondary lymphoid organs (spleen and lymph nodes). A mature (or naïve) B-cell, which expresses B-cell receptor (BCR) on its cell surface, finally differentiates into a plasma cell or a memory B-cell. B-cell neoplasms can arise at different stages of the B-cell differentiation processes.

Chronic lymphocytic leukemia

Chronic lymphocytic leukemia (CLL) is the most common leukemia in developed countries. It predominantly affects elderly white people (with 72 years old being the median age at time of diagnosis) and twice as many males than females [5]. CLL remains an incurable disease with remarkably heterogeneous clinical course. Asymptomatic patients may never require treatment. However, if they progress and are in need of treatment, they tend to relapse afterwards and acquire resistance to chemotherapy [6]. So far several factors have been identified to have significant influence on overall survival (OS) and response to standard treatment, with the most relevant being: (i) mutation status of immunoglobulin heavy-chain variable (IGHV) region [7, 8], (ii) chromosomal aberrations such as deletions 17p, 11q, and 13q, and trisomy 12 [9], (iii) and single gene mutations (especially in the *TP53* gene [10]).

Mantle cell lymphoma

Mantle cell lymphoma (MCL) is a relatively rare disease with aggressive clinical course and poor prognosis. Patients (mostly white males) are diagnosed typically at an older age (60–70 years old) in already advanced stages. MCL begins with

accumulation of abnormal cells in lymph nodes causing their enlargement. At the later stages it also affects blood and the gastrointestinal tract. MCL cells harbor t(11,14)(q13;q32) translocation [11].

Hairy cell leukemia

Hairy cell leukemia (HCL) is a rare disease with indolent clinical course. The name comes from microscopic observations in which the malignant cells appear 'hairy'. Hairy cells accumulate in bone marrow, where they prevent normal hematopoiesis, and in the spleen [12]. HCL is preliminary diagnosed during routine complete blood tests ordered due to recurrent infections and weakness. Final diagnosis is achieved after bone marrow biopsy. Two driver mutations of HCL have been identified: *BRAF* V600E (present in all patients) [13] and *CDKN1B* (present in 16% of patients) [14]. About 10% of HCL patients have a more aggressive variant of the disease (HCL-V) which differs mainly by internal morphology and overall appearance of the malignant cells [15].

Other

Other neoplasms originating from the various stages of B-cell differentiation, which were included in the study are: **marginal zone lymphoma (MZL)**, **lymphoplasmacytic lymphoma (LPL)**, **B-cell prolymphocytic leukemia (B-PLL)** and **follicular lymphoma (FL)**. Their individual characteristics are mentioned in the text when needed.

1.1.2 T-cell neoplasms

T lymphocytes (T-cells), the other type of white blood cells, act through cell mediated immunity by activation of destructive processes or cells in response to antigens. They originate in bone marrow from common lymphoid progenitors and maturate in thymus. During the maturation process T-cells learn how to distinguish invasive cells from the healthy ones present in the body. Mature T-cells, which express T-cell receptor on their cell surface, circulate in the blood stream awaiting activation. T-cells, similarly to B-cells, can arise at different stages of their differentiation.

T-cell prolymphocytic leukemia

T-cell prolymphocytic leukemia (T-PLL) is a relatively rare and incurable leukemia with aggressive clinical course. It originates from mature T-cells and affects elderly people, with males outnumbering females [16]. T-PLL is characterized by resistance to conventional treatment, however, recent introduction of alemtuzumab into first therapy regime significantly prolongs patients' survival. Genetic abnormalities involving chromosome 14 are present in around 3/4 of diagnosed cases. These include: inversion, tandem translocation t(14;14) and translocation t(X;14)(q28;q11). Also common is the mutation in *ATM* gene.

Other

Another two T-cell originated diseases which appear later in this dissertation are **Sézary syndrome (Sézary)** and **peripheral T-cell lymphoma not otherwise specified (PTCL-NOS)**. Both malignancies are rare, difficult to treat, and even hard to diagnose, especially the latter one [17].

1.1.3 Myeloid neoplasms

Myeloid cells are derived from common myeloid progenitor cells and can be classified into erythrocytes, granulocytes (neutrophils, eosinophils and basophils) and monocytes. The last two types are involved in the innate immune system. They fight with inflammation by phagocyte bacteria, larger organisms or damaged cells, destroying parasites and secreting substances [18]. Myeloid neoplasms arise due to overproduction of abnormal cells from the common myeloid progenitor.

Acute myeloid leukemia

25% of adult leukemia patients are diagnosed with acute myeloid leukemia (AML) [19]. AML is a disease with poor outcome, which in about 15% of cases is caused by previous chemo- or radiotherapy or environmental factors, such as pesticide exposure or benzene inhalation. The disease is characterized by the heterogeneity of its genomic landscape with abundance of driver mutations which, among others include *NPM1*, *RUNX1*, *ASXL1*, *CEBPA* genes and cytogenic aberrations such as t(15;17)(q22;q12), t(8;21)(q22;q22) and t(X;11q23) [20].

1.2 Standard treatment strategies for hematologic malignancies

A successful treatment of a hematologic malignancy is when the patient achieves complete response (CR). Although it does not mean that the cancer has been cured, the signs of disease can no longer be detected. At this stage the patient is progression-free (or in complete remission). It is not uncommon that the disease returns some time after treatment (the patient relapses). Treatment strategies are being evaluated in clinical trials conducted by health care institutions. During such trials, medical professionals often estimate the superiority of a tested new treatment over the well-established one based on OS parameter. In short, they compare the number of patients which are alive after a given time following the treatment administration.

The decision of treatment strategy depends on the staging of the disease and patient-unique factors [6]. Most indolent diseases (or disease subtypes), such as CLL or HCL, are predominantly detected in an asymptomatic phase accidentally during routine medical checkups. These are usually followed by a "watch and wait" strategy comprising only increased frequency of checkups (without administering any treatment) which is continued until the disease progresses. However, in symptomatic patients the treatment has to be provided quickly. As already mentioned in the previous Section 1.1, hematological malignancies are affecting mostly elderly people,

therefore the intensity of treatment has to be adjusted individually to the patient's fitness level. The more aggressive the treatment, the more efficient it is and the more durable outcome can be expected [5]. However, it is up to the medical professional to balance the treatment decision between: (i) its toxicity and efficacy and (ii) outcome of improving or just maintaining the current patient's quality of life [21]. Clinical trials usually recruit younger (and fitter) patients than it is expected for a disease, which makes these studies irrelevant in prognostication of treatment effectiveness in unfit patients [22]. Last but not least, another important factor in the decision-making process is comorbidity, that is coexistence of multiple diseases, such as blood hypertension or diabetes, in addition to cancer [22].

Standard therapy of hematologic malignancies involves chemoimmunotherapy and/or hematopoietic stem cell transplantation (with or without preceding irradiation). With the latter being a risky procedure (about 40% patients die from complications related to transplantation [23]), it is the field of chemoimmunotherapy where advancements are most desirable. The standard of care in hematopoietic malignancies haven't changed much during the last decades and still includes unspecific drugs selected by observation (for example, from the deadly weapons used to eliminate enemies in the World War I) which are chlorambucil [24], cyclophosphamide or vincristine, rather than targeted compounds. With growing knowledge about underlying biology of cancers, the care for some individual entities was revolutionized [25, 26, 27, 28]. The discoveries of single targeted drugs often propagate on other tumor entities which share similar biology. Although the idea is in principle good, it does not always give expected results. Currently the gold standard in treatment of fit CLL patients is FCR (fludarabine, cyclophosphamide, rituximab), a mixture containing one unspecific drug (cyclophosphamide), and two targeted ones. Rituximab, monoclonal antibody directed against membrane surface protein CD20, has been shown to be effective despite the fact that CLL cells express CD20 at low levels [29]. On the other hand, fludarabine is proved to be ineffective in patients harboring mutation in *TP53* gene, which occurs in 8–15% of cases [10, 30, 31]. Being able to detect significant individual characteristics of a patient before the treatment is administered could spare not only patient's suffering from unnecessary treatment, but also money [32].

1.3 Personalized therapy in hematologic malignancies

Curing blood cancer is like ordering a cocktail in a café. Ideally, each client could have an individual mix tailored to his or her taste. Of course, there are mainstream alternatives that taste the majority of people (patients), just like smoothies in McDonald's (and old untargeted therapies), but in reality they do not satisfy (cure) anyone entirely. In order to have a really delicious (working) cocktail a personalized mixture of fresh, cherry-picked ingredients (compounds that target deregulated

pathways specific to the patient) is needed. Currently, this concept which is called **'personalized therapy'** (or **'precision medicine'**) seems to be the only reasonable way to cure cancer.

Identification of biomarkers, biologic or genetic factors underlying pathologic state of a cell, is the first step towards precision oncology. Biomarkers can be aimed to: determine diagnosis (diagnostic biomarkers), forecast course of disease (prognostic biomarkers), predict response prior to treatment (predictive biomarkers) or outcome after the treatment (surrogate biomarkers), and keep track of the disease (monitoring biomarkers) [33]. They can be anything that captures the biology of tumor cells: from the composition of antigens on cell surface, through DNA methylation or gene expression profiles, to eventually genomic mutations and rearrangements. Advancements in the laboratory and in sequencing techniques over the last two decades let scientists explore the landscape of biomarkers for the most common hematologic malignancies [34, 20].

Biomarker identification followed by targeted treatment already revolutionized the care of individual cancers. In chronic myeloid leukemia, for example, where the BCR-ABL tyrosine kinase is constitutively activated (fusion gene resulted from the translocation between chromosomes 9 and 22), targeted inhibition of BCR-ABL1 put patients into remission [26]. In another study, the same was achieved in HCL patients after treating them with vemurafenib, an inhibitor against specific mutation V600E in the *BRAF* gene [28]. These successes were possible because of the following two aspects. First, virtually all cases of a particular disease harbored the specific targeted biomarker, which was driving the tumor growth. Second, targeted inhibitor of the disregulated pathway was available to use. The situation is completely different for heterogeneous diseases, such as CLL [35], for which patients still receive non-targeted treatments. While a new generation of targeted drugs is emerging [36, 37, 38], the impact of recurrent mutations on drug response is uncertain, including mutations in *ATM*, *NOTCH1*, *BIRC3*, *SF3B1*, *BRAF* and *KRAS*, deletions on 13q, 11q, and trisomy of chromosome 12 [39, 40].

Last but not least, the identification of biomarkers is always performed in population-wise studies mostly due to requirement of obtaining the necessary statistical power in the analysis. Scientists have to make sure they are not conducting (de)personalized studies while focusing on disease-specific rather than on patient-specific characteristics [41]. Moreover, not only every patient a harbors a unique set of biomarkers, but also the individual characteristics of a tumor change over time due to cancer evolution or the administered treatment. This suggests that patient care should involve close monitoring during treatment and follow-up visits. But this only makes sense if we can do something about these changes, meaning if it is possible to translate the findings into treatment strategy.

1.4 Pharmacogenomic studies and cancer treatment

Response to anti-cancer agents is often restricted to subsets of patients, but the recognition of factors underlying this heterogeneity and the identification of the corresponding biomarkers is incomplete [42, 43]. There is a need for platforms that can comprehensively map drug responses, determine associated biomarkers and provide hypotheses for mechanisms underlying the variable response.

Several studies in recent years made an effort to identify determinants of drug response in a high-throughput fashion [44, 45, 46]. They tested hundreds of immortalized cancer cell lines derived from multiple cell origins. In their analysis they confronted cell sensitivity to treatment with a wide range of compounds (including targeted small molecules) with cells' molecular markup. Although these studies were able to prove the usefulness of high-throughput drug testing by recapitulating known gene-drug dependencies, a few problems with them have been noticed.

First, the extensive data resources which have been published alongside Gernett et al. [44] (Genomics of Drug Sensitivity in Cancer, GDSC) and Barretina et al. [45] (Cancer Cell Line Encyclopedia, CCLE) brought up the question of reproducibility. These two studies used overlapping sets of both cell lines and compounds, so direct comparison was possible. The part which caused the most controversy was the drug response data. One report claimed poor correlation of the measurements [47], while the other found the correlation—after adjusting for biologically-relevant analytical factors—satisfactory [48]. The most recent study went one step further by producing yet another drug sensitivity dataset [49]. Their results were more consistent with CCLE than GDSC, however, the agreement with neither was impressive. The sources of discrepancies included: (i) lack of a standardized experimental protocol (for example, the number of cells seeded per well on a screening plate, cell viability assays, drug concentrations used), and (ii) different definitions of drug response (half maximal inhibitory concentration (IC_{50}) *vs.* area under curve (AUC)). Reproducibility of drug screen data is an important issue especially when the findings are to be translated into clinical practice. We found this matter very important, therefore Section 2.3 provides extensive reproducibility checks on the dataset we have been working on.

Second, using cell lines as a model of a disease puts the results far away from being clinically relevant. The immortalized cells are often covering only a small subset of disease heterogeneity and potential resistance mechanisms [50]. Moreover, same mutations in different cancer types could result in diverse drug response phenotypes. Therefore, in order to make the statistically significant discoveries a lot of cell lines per tumor type have to be tested. Finally, developing cell lines not only takes a lot of time but sometimes is just not feasible, as for CLL, for instance.

The shortcomings resulting from using cell lines can be defeated by working with short-term cultures of primary cancer cells instead. Such approach not only avoids clonal selection to occur, but also preserves natural genetic and phenotypic diversity of the sample, including rarer mutations and combinatorial patterns of mutations.

Responses of primary tumor cells to panels of inhibitors *ex vivo* have recently been used to derive individualized therapeutic options for the donating patients [51, 52, 50]. Molecular characterization of individual samples yielded novel genetic markers and led to drug repurposing opportunities [53, 54, 55].

1.5 Overview of the undertaken approach

In the previous section we reviewed the biggest studies published so far, which were linking drug response to the biology of cancer cells. Their approaches differed mainly by the studied disease model. They used either cell lines or a limited number of primary tumor cells. The study which is the topic of this dissertation goes one step forward by testing hundreds of primary tumor samples of one cancer type. This allows to cover the disease heterogeneity in depth. The layout of the study, which involved: obtaining mononuclear cells from patients' peripheral blood, functional and molecular profiling of samples followed by a various types of analyses, is presented in the Figure 1.1.

The main focus of the study was CLL. However, additional samples of other leukemia and lymphoma were also examined (see Appendix B for details). Tumor material was collected from patients during routine blood exams, which made the sampling quite straight-forward. Availability and abundance of this material makes haematological malignancies a great research model. Even in relatively simple culture conditions the cells survive outside the body for several days, which is long enough to perform experiments on them.

The crucial component of conducting a translational research is a well-chosen study cohort. It should mirror the different factors of age, sex, disease stage etc., and even the acquired chromosomal aberrations in a proportion which is usually seen in clinical practice. The presented study satisfies these conditions (see Figure 3.3), which makes the results directly applicable to medicine with potential benefit to participating patients.

The obtained samples were subjected to cell profiling. A high-throughput *ex vivo* viability assay measured the response of cells to drug treatment. This part is discussed in detail in Chapter 2. Additionally, the same samples underwent genomic characterization by: whole exome sequencing, RNA-seq, DNA methylation and SNP-array. The details are described in Chapter 3.

We have mapped drug sensitivity profiles of these tumors to genomic information in order to both understand the biology behind heterogeneous drug response and discover possible novel therapeutic targets. Abundance of samples and compounds tested allowed us to: (i) provide functional classification of the compounds, (ii) define patient functional subgroups, (iii) select determinants of drug response, and together with the available patient metadata (iv) forecast clinical outcomes.

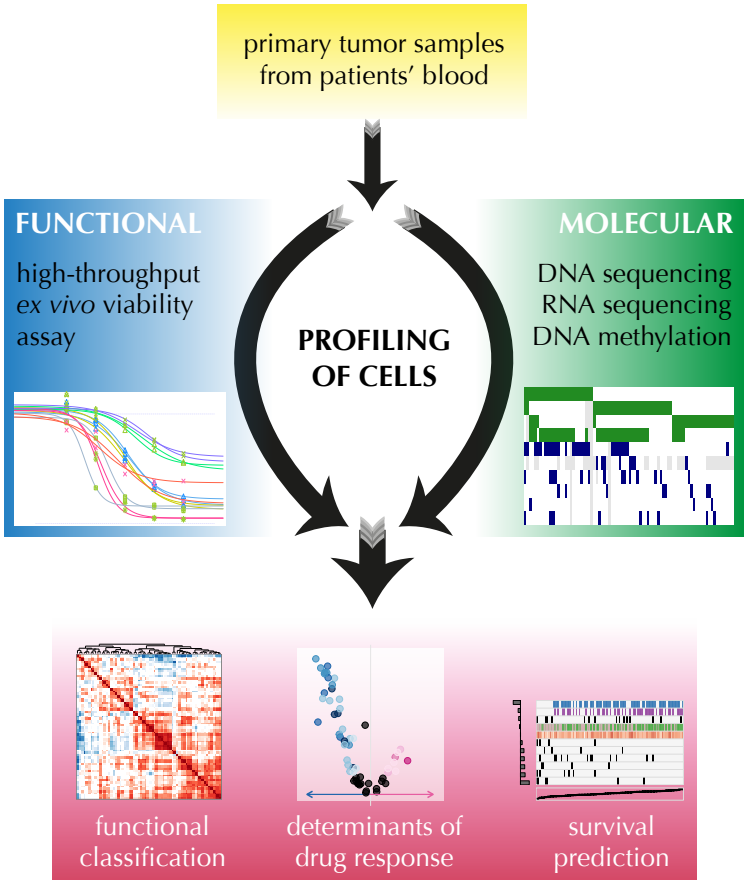


Figure 1.1 Layout of the study.

Primary tumor sample collection was performed by selecting mononuclear cells from patients' blood specimens. These cells where then subjected to functional and molecular profiling. The former was testing drug response simply by using a viability assay, while the latter characterized the cells on multiple levels such as genetic mutation, gene expression and DNA methylation. Analyses included i.a. classification of patients and drugs based on a single dataset, and finding co-dependencies among different kinds of datasets.

Phenotypic profiling of primary cancer samples by high-throughput drug screening

2

The true method of knowledge is experiment.

— **William Blake**
All Religions are One (1788)

2.1 Performed drug screens

Functional profiling of patient samples was done within two main initiatives, called for simplicity—the drug screens. They were performed mainly by Leopold Sellner at the NCT Heidelberg. The compounds were first seeded into 384-well plates and then patient samples were added. One sample was used per plate. Each plate included untreated control wells with DMSO solution instead of a drug. Cell viability was assessed using the ATP-based CellTiter Glo assay (Promega, Fitchburg, WI, USA) and luminescence level was measured with a Tecan Infinite F200 Microplate Reader (Tecan Group AG, Männedorf, Switzerland) with an integration time of 0.2 seconds per well. The compounds used in the drug screens are listed in Appendix C together with their short characteristics. Concentrations of drugs were manually selected and are cataloged in Appendix D. Figure 2.1 shows the number of patient samples stratified by the diagnosis, and the number of compounds used in the two drug screens—pilot screen and main screen—stratified by the annotated target.

2.1.1 Pilot screen

The pilot screen was a preliminary drug screen designed and performed mainly in order to:

Figure 2.1 Characteristics of samples and drugs used in the drug screens.

Panel A shows a summary of patient samples and disease entities analyzed in the pilot screen (top left), the main screen (bottom left) and both screens combined (right). The outer circle characterizes the patient diagnosis and the inner circle indicate the cell lineage.
Panel B shows a summary of drugs used in the pilot and the main screen stratified by the annotated target.
Panels C and D show the overlap of patient samples and drugs, respectively, between the screens.

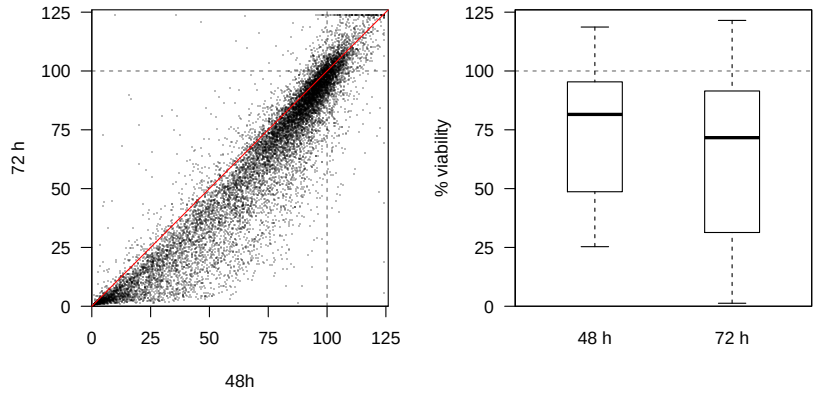


Figure 2.2 Viability of cells after 48 h and 72 h of drug treatment.

Left panel shows a scatter plot which compares cell viability readouts after two incubation periods. The majority of dots (83%) lays under the diagonal, meaning that the cells were in general less viable after 72 h than after 48 h. At the third day of incubation period more of extra-viability effects, where cells appear to be much more viable than controls, were observed as compared to the second day (percentage of cells exhibiting viability above 115%: 2.1% after 48 h and 2.7% after 72 h).
Right panel shows more pronounced spread of viability readouts in the 72 h time point, with 11.1% and 18.8% of values below the 20% viability threshold for 48 h and 72 h, respectively.

- prove feasibility of the screening approach,
- validate clinical observations,
- asses prospectiveness of further drug screens.

To reach these goals we took a variety of compounds ($n = 67$), including clinically used drugs and tool compounds targeting important pathway nodes in lymphoma and cancer, in limited concentration steps (16 drugs with one and 51 drugs with two), and 111 unique patient samples. The cell viability was measured in two time points: 48 h and 72 h after drug treatment. Both screening days produced meaningful readouts, however, the longer the incubation period was, the more spontaneous cell death was observed (Figure 2.2). Therefore, we performed the analysis only for the first screening time point.

Each compound-concentration pair was present in the screening plate twice. Besides that, for each dilution steps there were eight wells containing negative controls placed on the plate.

2.1.2 Main screen

After the pilot screen has proven successful in its assignments, the follow-up, more advanced drug screen was designed and conducted. Fundamental purposes of this screen, called the main screen, were to:

- validate clinical observations,

- find the causes of heterogeneous drug response within groups of patients suffering from the same disease,
- stratify samples according to their drug response profiles,
- define molecular characteristics of samples which modulate drug response,
- describe to which degree those factors are either responsible for or can predict patient outcome.

Compared to the pilot screen there were several changes accommodated in the drug screen design. The number of CLL patient samples was increased from 97 to 184 in order to cover the complex heterogeneity within the disease. Additionally, also the number of samples representing less common than CLL diseases was increased (e.g. T-PLL, from 5 to 25 samples), and some new ones were introduced (e.g. AML, the disease of myeloid origin). The list of compounds was modified to include mainly chemotherapeutics and targeted inhibitors already in clinical use or prospective treatments for cancer which are currently being evaluated in clinical trials. Each compound was screened in 5 concentration steps, giving, in principle, the opportunity to fit the drug-response curve and retrieve the well-known parameters, e.g., IC_{50} and AUC. Moreover, each compound-concentration pair was placed on the screening plate once. Whole two columns on the plate (32 wells in total) contained negative controls, in which the patient sample was incubated with DMSO solution instead of a drug.

2.2 Processing of raw values obtained from cell viability assay

In this section, we provide explanation of how we normalized the raw luminescence measurements into the viabilities, which were later used for analysis. Furthermore, we discuss the types of quality control checks, which are crucial for the assessment of drug screen performance. Additionally, we demonstrate how to calculate other parameters, such as IC_{50} or half maximal effective concentration (EC_{50}), which are potentially useful in the analysis of high-throughput drug screens.

2.2.1 Data normalization and quality control

Differences in plate layout between the pilot and the main screens, especially in the number of negative controls and replicates of drug-concentration pairs available, forced us to take a slightly different approach to data normalization of these screens. In both screens, the viability was calculated using raw luminescence intensities for each drug-concentration pair for each screening plate (sample) separately. In the pilot screen, the mean of the two measurements for a given drug-concentration pair was divided by the mean of the 8 negative control wells. In the main screen, however, one measurement for a given drug-concentration pair was divided by the median of 32 negative control wells present on each plate. These values were then multiplied by 100, resulting in viability scores.

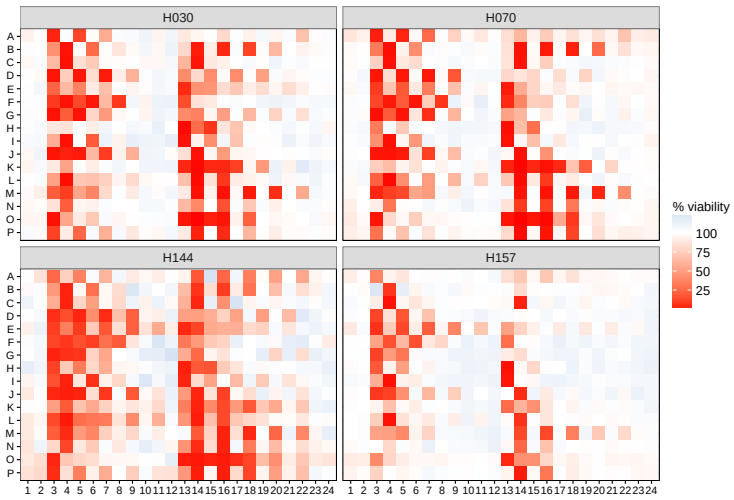


Figure 2.3 Examples of normalized intensities in drug testing plates.

Figure shows viability values in the four example screening plates: two CLL (H030, H070), one MCL (H144) and one T-PLL (H157) samples. Columns 1–2 contained negative controls. Differences in viabilities for samples representing distinct diseases could be observed. Here, in comparison to CLL, MCL and T-PLL samples tend to be more sensitive and resistant, respectively, to the compounds used in the screen.

Performing quality control checks is an essential step before actual data analysis. It includes identification of potential problems with:

- plate spatial effects: both edge and column or row effects,
- batch effects,
- data reproducibility.

We will now elaborate on these by taking the normalized values of the main screen as an example. If any special effects could be identified, we would need to account for them before starting the analysis.

The first kind of quality checks involved plotting the normalized values for each sample as a heat map matching the exact layout of the screening plate. Figure 2.3 presents an example of such plots. We did not detect any alarming problems with edge effects in any of the experiments. The plates project similar within samples of the same disease origin. The differences between plates of divergent sample origin look reasonable, for instance T-PLL patients tent to respond weaker to chemotherapy than CLL patients, therefore the viability values of their cells are closer to 100% what results in plates looking more pale.

In the main screen wells containing negative controls were placed in the first two columns, which gave us one more opportunity to rule out serious spatial effects. Figure 2.4 shows viabilities of the negative controls. For the vast majority of plates the values are close to 100%. The average negative control mean viability per plate

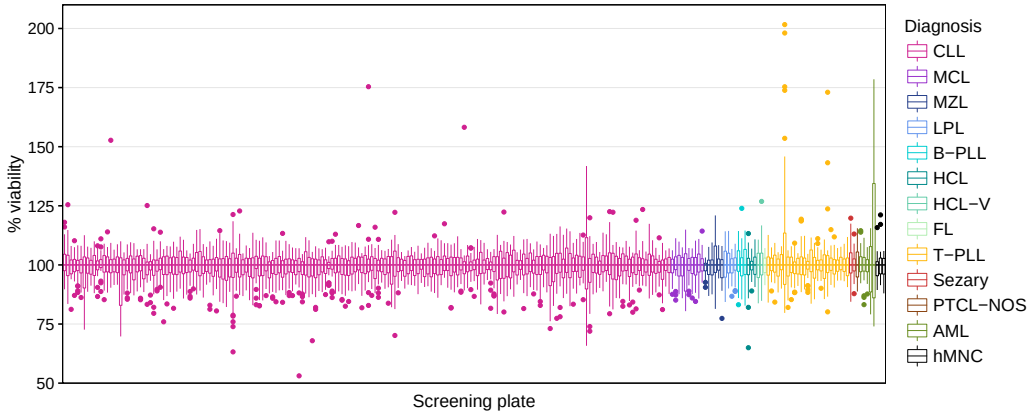


Figure 2.4 Viability of negative controls per screening plate.

The box plot shows the normalized luminescence intensities of the negative control wells per screening plate (sample). The plates were sorted and colored by disease of donating patients.

was 99.63% (range 94.35–112.23%) with standard deviation of 6.08% (range 2.86–33.03%). Although we can observe a few outliers (which are independent from the type of disease), the screen looks good from this angle.

Column and row effects are usually introduced by the machines used for plating drugs and samples on to the screening plates. Both the pilot and the main screens were performed by hand. There were no column or row effects present in neither of them.

Assessment of batch effects is the second step of screen quality checkup. The main screen was performed over a time period of 1.5 years in three batch groups by at least two different people. The batches were divided by samples screened in 2013, in 2014 before August and in 2014 in August and September. In Figure 2.5 we show the mean viability of all the drug-concentration pairs for each plate in the function of time. All the screening days look comparable with no special differences between the batches. This was further evaluated and the results are included in Appendix E.

The third and last part of quality control is the reproducibility of measurements. This is a broad topic which we describe in detail in the following section.

Processing of raw files from the experiment was done with the help of Bioconductor R package *cellHTS2* [56] (version 2.34.1). I wrote bash scripts to construct the input annotation tables, and functions in R for reading in and for normalizing the data.

2.3 Data reproducibility

The important role of conducting reproducible research is being increasingly acknowledged over the last months [47, 48, 49]. Therefore, we feel obliged to report on the reproducibility of the performed drug screens in depth.

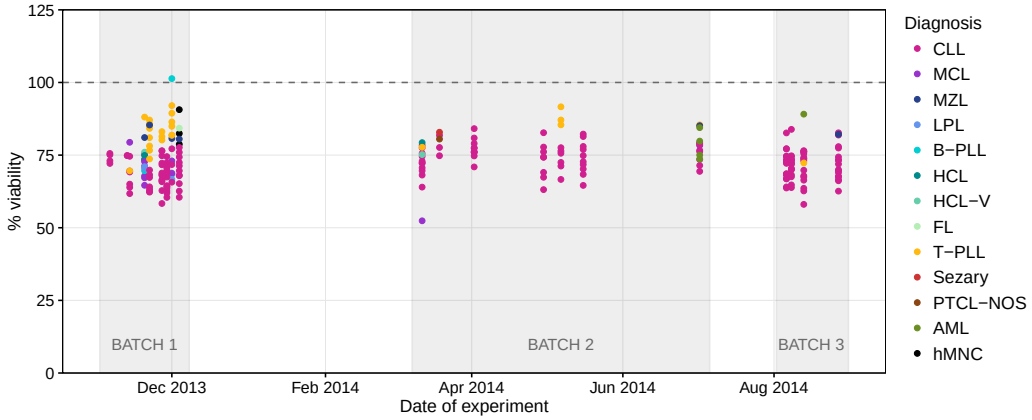


Figure 2.5 Dependency of mean viability on the experiment date.

Each dot in the figure represents the mean viability for a sample. Color codes the diagnosis of the patient from whom the sample was taken. Experiments were conducted in three batches, which mainly differed in the person performing the screen. Batch effects seem to be of marginal importance here. The confirmation is presented in Appendix E.

2.3.1 Reproducibility within same drug screen

The most basic level of data reproducibility assessment is to ask whether the obtained values agree within a given screening platform.

Three selected CLL samples from the pilot screen were assayed twice. The replicates were performed on two different days. Additionally, measurements from both incubation times: 48 h and 72 h were available. Figure 2.6 compares drug responses of the two replicates for each patient after the two incubation periods. The obtained Pearson correlation coefficients ranged from 0.74 to 0.92, with the mean being 0.85. Coefficients calculated for the two incubation periods were not significantly different from each other.

Six samples of the main screen were screened in two replicates, too. Three of them were CLLs, the remaining ones were MCL, HCL and T-PLL. Two of the CLL samples (H030, H073) were assayed on the same day whereas the rest was assayed on different days. Figure 2.7 compares drug responses of the two replicates for each patient. The mean of the obtained Pearson correlation coefficients was 0.91, with a range of 0.84–0.97. Although the coefficients itself are high (which suggest powerful concordance), the spread of individual values is rather big.

In the following analysis we used one out of the two available replicates. The high level of reproducibility within the pilot screen made the selection irrelevant. Only because there was a long 4-month gap between experiments with replicates we decided to go with the one whose results came in earlier. The level of reproducibility of the main screen was a motivation to search for clues of which replicate to choose. The criteria was set based on the goodness of the fit of the dose-response curve (Appendix F). In short, we took the replicate for which a greater number of drugs produced good fits of the sigmoid curve.

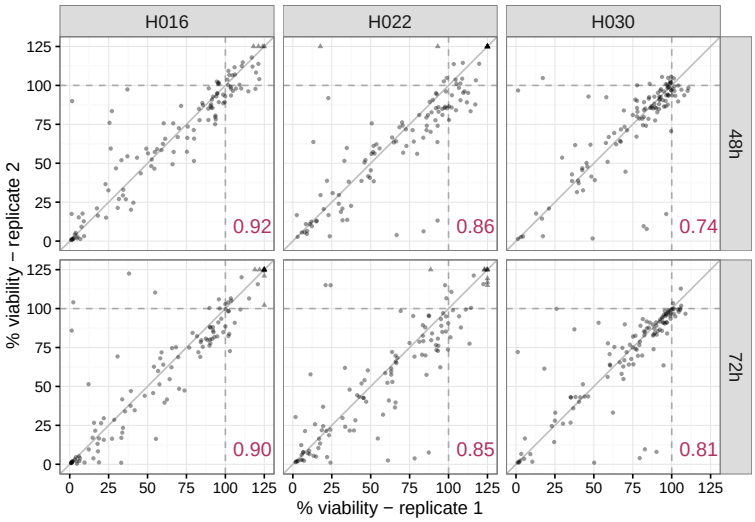


Figure 2.6 Reproducibility of drug response measurements in the pilot screen.

The scatter plots compare drug response measurements in patient samples that were repeatedly assayed at two different time points. The two measurements were performed after 48 h and 72 h of incubation. The calculated Pearson correlation coefficients are shown within each plot. Triangles indicate data points outside the plotting range.

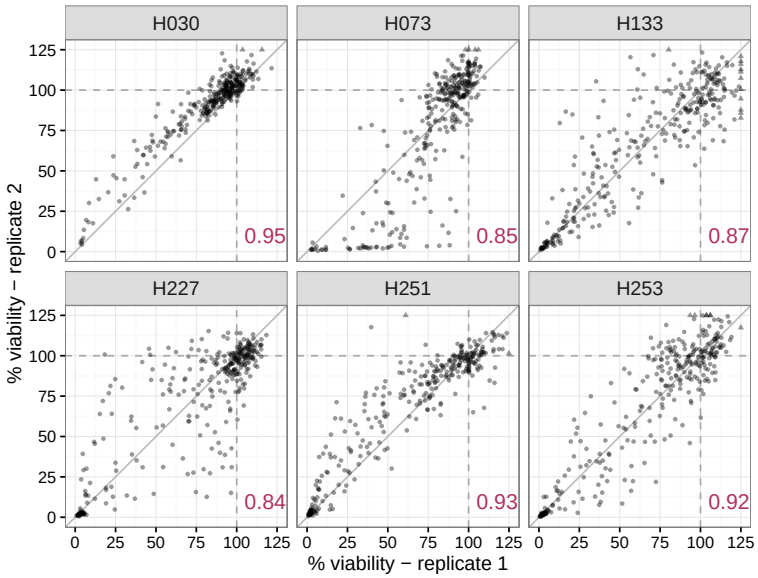


Figure 2.7 Reproducibility of drug response measurements in the main screen.

The scatter plots compare drug response measurements in six patient samples that were done in two replicates. The calculated Pearson correlation coefficients are shown within each plot. Triangles indicate data points outside the plotting range. There are 320 dots within each plot, which correspond to drug responses to 64 compounds in 5 concentrations.

2.3.2 Reproducibility between different drug screens

The next level in judging whether the data is reproducible was to compare drug responses between pilot and main screens. Both experiments were done in the same laboratory and usually by the same person. Although the design differed, there were 67 patient samples and 26 drug-concentration pairs (for 25 compounds) that were overlapping and therefore ready for comparisons, see Figure 2.8. Some drugs showed non-linear (vorinostat) or shifted to one side (for example dasatinib or orlistat) dependency. For such, a useful measure of reproducibility is the Spearman correlation coefficient, which appears to be high within this group. The second type group of drugs is characterized by producing low response phenotype in tested samples (for example everolimus or ralimetinib). The proper measure of reproducibility in such cases is the root-mean-square deviation (RMSD), which basically states how big the spread between each pair of the compared values is.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}, \tag{2.1}$$

where n is the number of compared pairs, while x and y are the viabilities from pilot and main screens, respectively.

In the end, we report both: the Spearman correlation coefficient and the RMSD, because only when considered together they provide a good measure of data reproducibility. To sum up, the reproducibility of drug response measurements was very good, and the only potentially problematic drug identified was bortezomib.

2.3.3 Reproducibility between different screening platforms

Our collaborators from the University of Helsinki, predominantly interested in T-PLL disease, tested three of our samples (H086, H172, H253) on their drug screening platform (Helsinki screen). This gave us additional opportunity for assessing robustness of the main screen data. 30 compounds were overlapping between both drug screens. Unfortunately, the drug concentration steps were different, which made direct comparison of the drug responses impossible.

The solution to this shortcoming was to compare IC_{50} values obtained from the dose-response curve fitting (Figure 2.9). The curves were fitted as explained in Appendix F. The parameter stating the goodness of the fit F.2 was much better for the Heidelberg than for the Helsinki platform (median Γ 45.9 vs. 131.9, respectively). Moreover, for some drugs (for example ibrutinib) the Helsinki platform used too low concentrations to be able to detect the effect of the drug. The Pearson correlation coefficient for the comparison was 0.75. There are two drugs which are clearly outliers: bortezomib and everolimus. In the scatter plot they appear way below the diagonal, meaning that the samples were less sensitive to these drugs when tested in Heidelberg than in Helsinki.

Figure 2.8 Reproducibility between pilot and main screens.

The scatter plots compare drug response measurements of 26 drug-concentration pairs in 67 patient samples that were screened in both, pilot and main screens. Spearman correlation coefficients together with RMSD in the brackets are shown within each plot. Triangles indicate data points outside the plotting range.

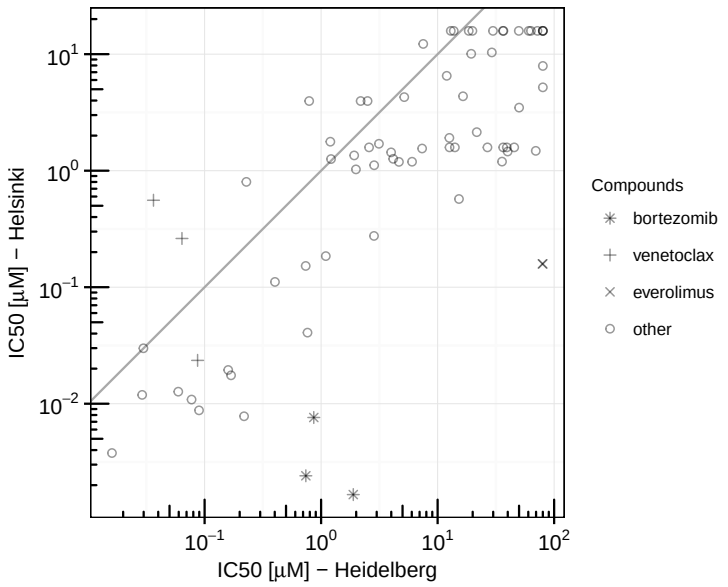


Figure 2.9 Comparison of the two drug screening platforms.

The scatter plot compares IC_{50} values for sample-drug pairs obtained from both screens. For 17 pairs the dose-response curve could not be obtained for at least one platform, and these pairs were removed from the plot. In case when the dose-response curve did reach 50% viability, the values were censored. The three drugs (bortezomib, venetoclax and everolimus) which exhibited poor similarity are plotted with distinct marks. Pearson correlation coefficient for all 73 comparisons was equal to 0.75.

In conclusion, despite the differences in platforms and their sensitivity, the overall drug responses correlate well. Smoother dose-response curves obtained with Heidelberg data confirm the platform to be robust and that the produced data is of good quality.

2.4 Parameters characterizing drug response

Analysis of high-throughput drug screens is lacking standardization [47, 49]. The ways of estimating fundamental parameters of drug response wildly differ between the published studies. However, all of these approaches rely on the fit of a dose-response curve. The drug response was previously characterized by the following parameters:

- drug potency: half maximal inhibitory concentration [44], half maximal effective concentration [57],
- drug efficacy: curve asymptotes,
- drug potency and efficacy: area under curve [45].

Even though these parameters are sufficient to distinguish differences between sensitive and resistant cells, and between effective and ineffective drugs, they have serious shortcomings. In a high-throughput setting, the majority of the dose-response curves

are shallow and do not reach the half-maximal inhibition. This leads to censoring of potency parameters and produces ties which are not applicable to many statistical methods. AUC, however, is relevant only when cells are treated with same range of drug concentrations. If this is not the case, drug response of cells can be compared only within one drug, as the comparison between different drugs is ambiguous. There have been developments on multiparmetric approaches which take into account both the midpoint and the shape of the dose-response curve [58]. However, they are falling for the obstacles of curve fitting as well.

The use of raw measurements of cell viability normalized by negative controls allowed us to overcome the weaknesses mentioned above. This simple approach prevented us from data alteration in cases where viabilities over 100% were observed. The reasons for such an effect could be explained by slower spontaneous dying of cells in the presence of the drug in comparison to negative controls. Censoring the measurements to 100% in such a case will strip the data from the potential biological characteristics right from the beginning. Moreover, with such an approach it is easier to account for the differences in drugs' mode of action. It is known that inhibitors of certain important signaling pathways, for example AKT/PI3K/mTOR, are producing shallow dose-response curves which prevents them from correct interpretation if usual metrics for drug response are used [57]. In summary, we believe that by performing the analysis on separate cell viabilities we can test our data in a robust fashion while still preserving all the biological effects.

3

Molecular profiling of primary cancer samples by using multi-omics

If you think big, then it's going to be big.

— **Emeril Lagasse**

Rapid advancements of sequencing technology since its introduction in 1977 by the pioneers: Frederick Sanger, Allan Maxam and Walter Gilbert, let scientists gain insights on the fundamental processes which hold the key to understanding identity and functioning of the living study subjects.

Techniques of targeted sequencing focused only on a few selected genes of interest allow to detect genetic variants in a cost-effective manner. Whole exome and whole genome sequencing techniques provide an overview of the wide landscape of genetic variants and structural rearrangements in the studied material. Array-based technologies detecting RNA expression and epigenetic markup, such as DNA methylation, complement the genetics with crucial additional information. They give follow up information of gene regulation and expression, which together with the genetic structure show all levels of functional activity of genes. Nowadays these so called 'multi-omics' techniques are fast, efficient and affordable enough to be applied in a high-throughput manner. Strong competition between the companies developing these technologies led to their high accessibility thanks to the simplification of the design, yet leaving enough freedom to accommodate specific usage.

In this study we used several of these state-of-the-art techniques to thoroughly characterize the collected patient samples. The principal aim was to utilize these different blocks of information as features and measure their impact on drug response of the samples.

3.1 Genomics

We used fluorescence in situ hybridization (FISH), targeted sequencing and whole exome sequencing (WES) to characterize genetic aberrations of primary tumor samples. We don't have full coverage across all patients for each of these techniques. Details on data availability are mentioned in the following sections. Analysis was carried out on the harmonized and combined data, which contained simplified binary information of whether a given gene (or cytoband in case of FISH) was mutated or not.

3.1.1 Fluorescence in situ hybridization

FISH is a technique which allows to locate the sequence of interest in chromosomes by using fluorescent probes. The FISH experiment was conducted for six genomic regions, which could identify mutations such as deletion 11q22 ($n = 184$), deletion 17p13 ($n = 182$), deletion 13q14 ($n = 177$), trisomy 12 ($n = 174$), deletion 6q21 ($n = 148$) and gain 8q24 ($n = 140$). Samples were considered mutated even when only a small mutated subclone was identified.

3.1.2 Targeted sequencing

Sequencing was performed on a GS Junior benchtop sequencer (Roche, Penzberg, Germany) as described in [59]. This technology allows to sequence the specific loci of interest in an effective and cost-efficient manner. It uses pyrosequencing, a method in which nucleotides are detected during synthesis of the strand complementary to the single DNA strand which is being sequenced. Targeted sequencing was performed for the following genes: *BRAF* ($n = 253$), *NOTCH1* ($n = 253$), *TP53* ($n = 252$), *SF3B1* ($n = 252$), *MYD88* ($n = 252$), *KRAS* ($n = 199$), *NRAS* ($n = 198$), *EZH2* ($n = 196$), *PIK3CA* ($n = 196$).

3.1.3 SNP arrays

DNA copy numbers and single nucleotide polymorphisms (SNPs) were determined with Illumina CytoSNP-12 and HumanOmni2.5-8 microarrays ($n = 169$; Illumina, San Diego, CA, USA). The results were verified by exome sequencing data for a subset of patients ($n = 121$).

3.1.4 Whole exome sequencing

High-throughput exome sequencing of paired control-tumor DNA samples was performed on Illumina HiSeq 2000 machines (Illumina, San Diego, CA, USA). Material from 121 patients was analyzed.

The pair-end reads were processed by two separate platforms. The first one, a well-established pipeline located at the German Cancer Research Center (DKFZ) in Heidelberg, provided a list of genome-wide single nucleotide variants (SNVs). This

previously described platform [60, 61] uses SAM tools *mpileup* and *bcftools* with parameter adjustments to allow for calling of somatic variants with heuristic filtering. The variants were annotated with RefSeq (version September 2013) using ANNOVAR [62] and only those of high confidence were selected. However, this platform is typically used for the analysis of solid tumors. To confirm the validity of applying it to hematologic malignancies, we developed a second platform which is described in details below. Our pipeline used lenient SNV calling strategy that allowed us to manually and visually inspect the mutated genome regions. Additionally, it was also used to confirm the copy number estimates obtained from FISH and SNP arrays.

Data processing

The processing pipeline consisted of six steps. These led from raw sequencing data stored in text-based FASTQ files to summarized per base nucleotide tallies saved to a highly flexible and access-efficient HDF5 file format. For each step either a Python script (run in Python 2.6.6) or an R script (run in R 3.1.1) together with a bash launcher script were written. These scripts automatically created lists of inputs and passed them to the processing functions, which were executed on the EBI computing cluster. A detailed description of processing steps follows below. In some of them Picard tools version 1.106 (http://broadinstitute.github.io/picard/) were used.

Align to reference GSNAP version 2013-08-14 [63] was used to align reads to the reference genome GRCh37 available in ENSEMBL database release 72. GSNAP software provides means of parallelizing the alignment procedure which were employed due to a relatively large number of processed files. The files were split per sequencing lane into 10 parts (`-parts` parameter) and 24 sub-processes were used for each of these parts (`-nthreads=24`). The software reported one single alignment per read (`-npaths=1`) and wrote them to an output SAM file [64] (`-A sam`) in a format compatible with Picard tools (`-sam-use-0M`).

Merge SAM files Picard tools was used to merge SAM files and to convert them to a compressed binary BAM file format on the single sample level. The Java tool `MergeSamFiles` took as input all 10 parts created during the alignment step across all sequencing lanes for a given sample and created one BAM file as output.

Remove duplicates `MarkDuplicates` from Picard tools was used to identify and remove duplicated reads. These duplicates can arise from either the library preparation process or the sequencing process. In either case, the safest approach is to count each read mapped to unique coordinates only once.

Locally realign reads Alignment of a single sample is prone to artifacts, mainly due to the existence of indels (insertions or deletions). This causes many bases to mismatch the reference near the place of misalignment, which in turn leads to confusing these phenomena with SNP. The local realignment step is crucial in assuring

that the number of mismatching bases is minimized across all the reads and samples. This step consisted of two parts, both of which were performed using the Genome Analysis Toolkit (GATK) version 2.5-2 [65, 66].

The first part identified genomic intervals which needed realignment. This was done by running `RealignerTargetCreator` on an input consisting of BAM files of all the samples as well as a VCF file containing an comprehensive list of short human variations downloaded from the dbSNP database [67] (file 00-All.vcf.gz dated 06-08-2013). This computationally intensive part was called on genomic intervals 1000000 bases long, which were then merged into one file by a bash command.

Output of the previous part served as input to the second part in which a realigner processed the suspicious intervals identified in the first part. This was done by running `IndelRealigner` for each patient separately, using as input tumor-control sample pairs and the intervals created in part 1. To accelerate the process, each chromosome was processed separately resulting in 24 output BAM files per patient sample.

Merge BAM files This step utilized the same function as step 2 (`MergeSamFiles` from Picard tools) to merge the alignment for all chromosomes split across separate BAM files for each patient sample.

Create nucleotide tallies During this step the processed reads, which were saved in compressed binary BAM files, were converted into easily accessible and reduced in size nucleotide tallies saved in hdf5 file format. This new format allows for fast access to the data for all samples, while at the same time enables easy exploration of the data. One tally file contains information about count, coverage, deletion, insertion and reference sequence per nucleotide base. The conversion process requires time and computational resources, so to speed it up separate tally files for each chromosome were created, additionally dividing each of them into smaller intervals of a maximum length of 10^5 bases.

In order to achieve this a whole bunch of R/Bioconductor packages was utilized:

- *GenomicRanges* [68] (version 1.18.1), *exomeCopy* [69] (version 1.12.0) and *GenomeInfoDb* [70] (version 1.2.2) to efficiently represent and manipulate genomic annotations and alignments;
- *rhdf5* [71] (version 2.10.0) and *h5vc* [72] (version 2.0.5) to create nucleotide tallies and to interface between R and HDF5 file format;
- *BSgenome.Hsapiens.UCSC.hg19* [73] (version 1.4.0) and *TxDb.Hsapiens.UCSC.hg19.knownGene* [74] (version 3.0.0) to use publicly available resources of annotations and genome sequences.

Parallel processing in this step was enabled by R package *BatchJobs* [75] (version 1.4). The created data structure allowed for efficient further analysis of the data.

Variant calling

Variant calling was performed with `callVariantsPaired` function from Bioconductor R package *h5vc*. We identified 88 genes which were either interesting from our current research perspective or have been identified as related to cancer (mostly lymphoproliferative disorders) in previously publish studies. A complete list of genes can be found in Appendix G.

We used fairly week thresholds for calling somatic variants. This was motivated by the overall purpose of our pipeline, which was to visualize and guide the variant calling done by the standard DKFZ pipeline. We decided to be more inclusive, because further filtering of the results was possible. Being exclusive could limit the possibility to spot the variant appearing at low frequency rate. Each control-tumor sample pair was taken separately. In order to call the variant, the following criteria had to be fulfilled:

- coverage per strand in either sample should be ≥ 2,
- support for the alternative allele per strand in the tumor sample should be ≥ 2,
- support for the alternative allele per strand in the control sample should be ≤ 4,
- support for the deletion per strand in the control sample should equal 0.

The output of `callVariantsPaired` function gives rich information about each variant found.

The analysis provided both: a list of the SNVs called and customized mismatch plots (Figure 3.1) for the visual inspection of somatic variants. The latter was plotted per patient-gene pair spanning all exons of the given gene, with genomic ranges extended by 100 bases on each side of every exon. Annotation of exons was taken from Bioconductor R package *TxDb.Hsapiens.UCSC.hg19.knownGene*.

Copy number estimation

Copy numbers were estimated with the help of Bioconductor R packages: *h5vc* and *DNACopy* [76], in a similar way as described previously [77]. In short, overlapping genomic ranges for annotated exons were first merged into fragments. Then, for each fragment of each sample the coverage depth was calculated with `binnedCoverage` function from package *h5vc*. The coverage was further adjusted by accounting for the GC-dependent bias. Finally, the resulting data was smoothed (`smooth.CNA`) and regions of estimated equal copy number were joined using circular binary segmentation [78]. An example visualization of the outcome can be appreciated in Figure 3.2.

3.1.5 Aggregation of genomic information

Genomic information collected by using all the different sequencing techniques explained above was manually merged and curated. The process involved well-established genetic markers. This approach allowed us to eliminate inconsistencies within

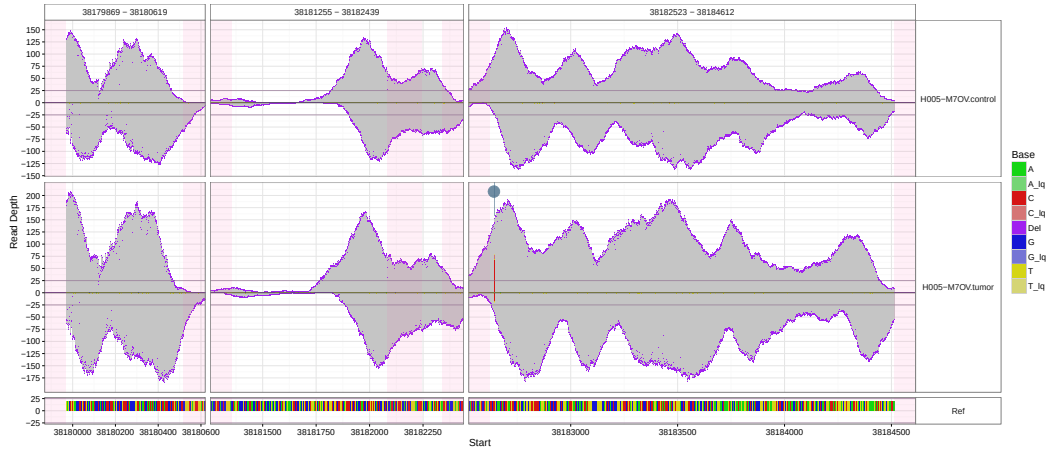


Figure 3.1 Mismatch plot.

Example of mutation in *MYD88* gene for one of the studied patients visualized by a mismatch plot. The two, top and bottom panels correspond to control and tumor samples, respectively. Grey shadows show the coverage on base resolution level. The multi-colored line at the bottom represents the reference strand. White background defines exon ranges and pink area represents parts of introns by which the exons were extended. The SNV is marked by the blue dot at the top of the bottom plot. The horizontal lines are guides which mark the coverage of 25 reads.

datasets and to come up with a comprehensive set of genomic characteristics of studied samples. As a result, a binary table (0/1 standing for unmutated and mutated gene, respectively) of 264 samples and 89 genomic features was created. Even though for most of the features the information about frequency of the mutated allele was available, in the analysis we included this information only for *TP53* and *BRAF* genes.

3.2 Transcriptomics

RNA sequencing was performed for 103 patient samples on Illumina HiSeq 2000. Output reads were aligned to the human reference genome (GRCh 37.1 / hg 19). Read counts were calculated with *HTSeq* [79] and differential expression analysis was performed in Bioconductor R package *DESeq2* [80].

3.3 Methylomics

196 CLL patient samples were subjected to DNA methylation profiling using Illumina Infinium HumanMethylation450 BeadChip arrays (Illumina, San Diego, CA, USA). Each specimen was categorized to one of the three groups (highly-/interemdiate-/low-programmed, HP/IP/LP in short) according to reference [81].

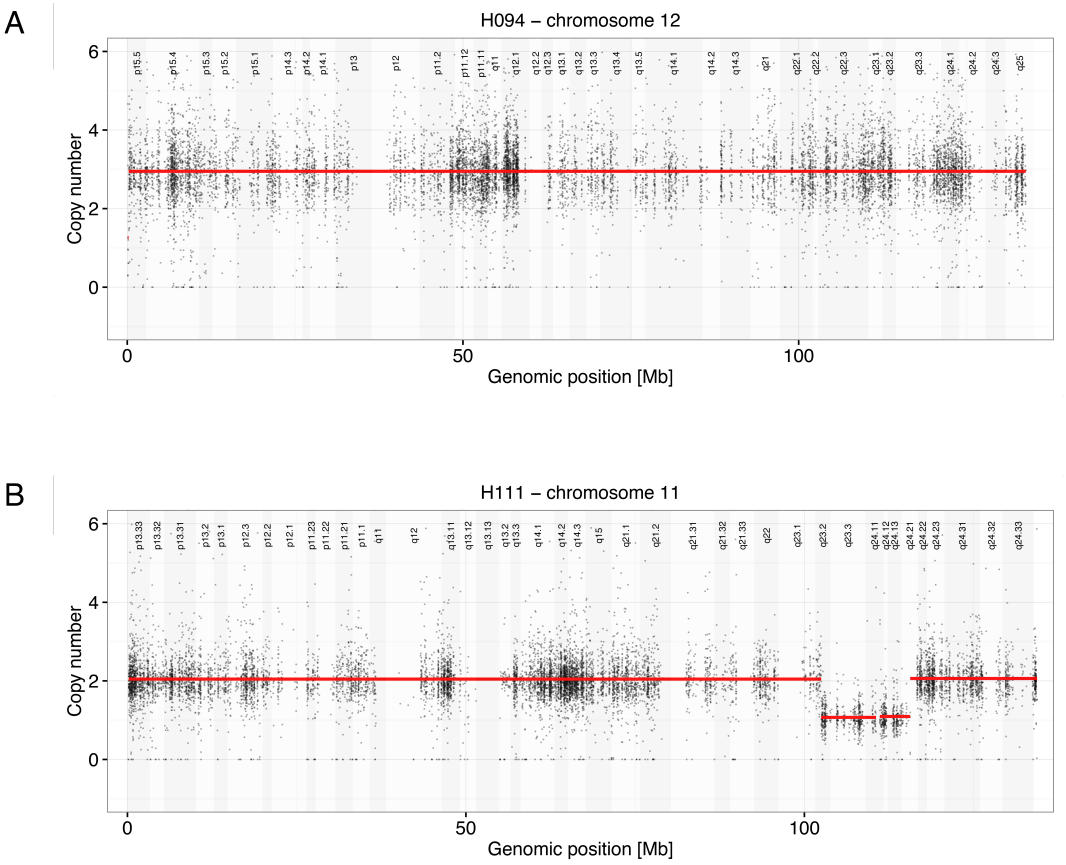


Figure 3.2 Segmentation of copy number estimates.

Two examples of estimated copy number changes. Pink and blue vertical stripes color-code different cytobands. Dots stand for the genomic fragments which were the resolution of calculation. Red horizontal lines are segments characterized by the same copy number.
Panel A shows the complete gain of an extra copy of chromosome 12 observed in patient H094.
Panel B shows partial deletion of chromosome arm 11q observed in patient H111.

3.4 Mutational landscape of patient cohort

The genetic landscape of studied CLL patients was heterogeneous as shown in Figure 3.3. This recapitulates the real characteristics of the disease seen in clinical setting and is a first step towards translation and applicability of the results.

During carcinogenesis CLL cells acquire multiple mutations, with some of them being more frequent then the others. In the investigated CLL cohort the most frequently appearing mutations were: del13q14 (61%), *TP53* (19%), del11q22.3 (17%), *SF3B1* (15%), trisomy 12 (15%), del17p13 (14%), *ATM* (12%) and *NOTCH1* (10%). These numbers agree with the fractions observed in other studies: 55% del13q14, 15% *TP53*, 10-18% del11q22.3, 5-17% *SF3B1*, 15-20% trisomy 12, 7-10% del17p13, 5-17% *ATM* and 4-12% *NOTCH* [82, 83, 84].

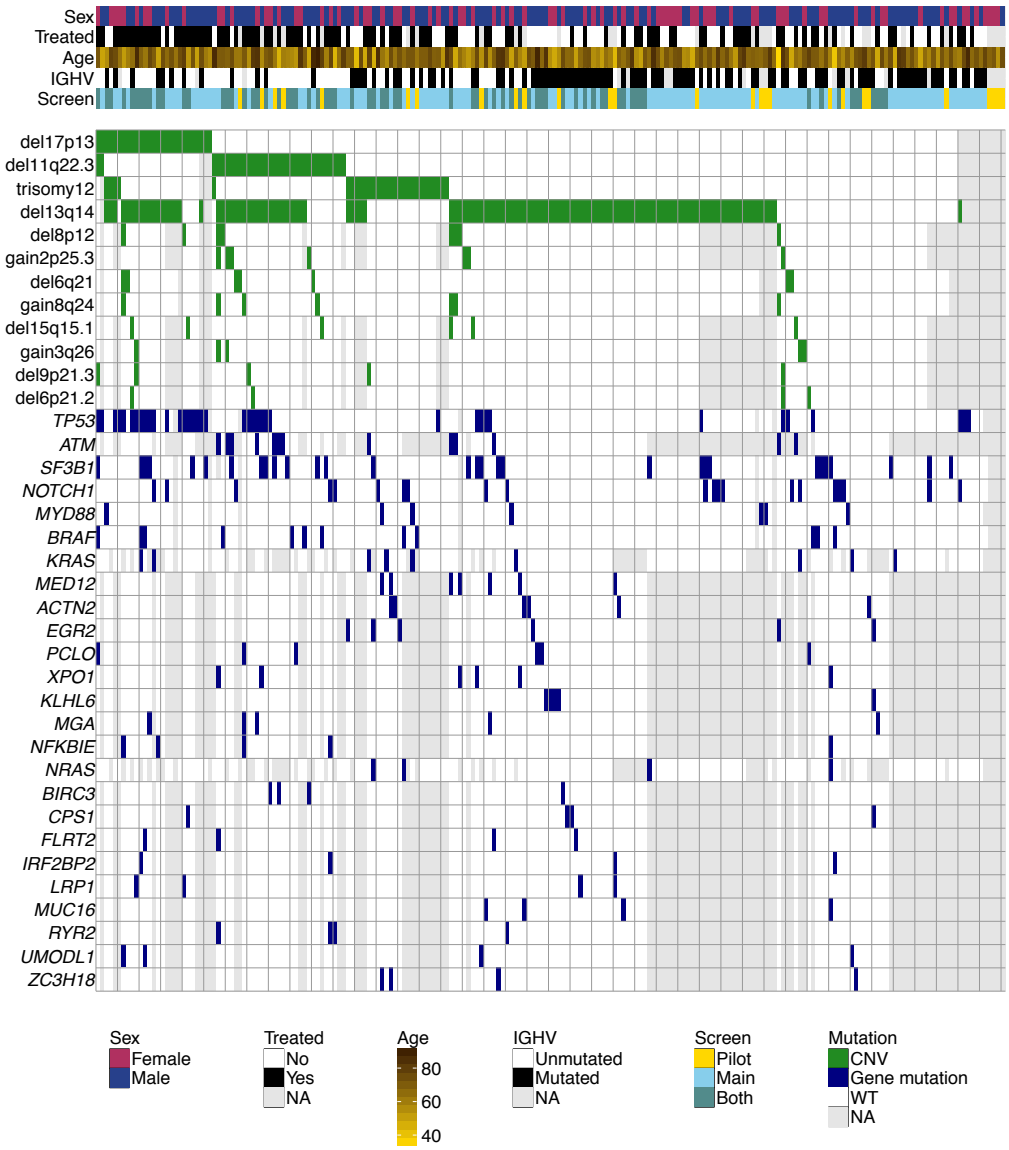


Figure 3.3 Genetic landscape and clinical characteristics of the CLL cohort.

The figure shows coexistence of a selection of clinical factors (top part) and mutations (bottom part) for the total of 211 CLL patient samples (x-axis) present in drug screens. Gray color codes unavailable data (NA).

Some mutations tend to occur together. That is a case for i.a. *TP53* and del17p13, or del13q14 and all other mutations. However, mutations of del17p13, del11q22.3 and trisomy 12 have a tendency for exclusiveness. This has been already observed before and the groups have been correlated with clinical outcomes [9]. It has been shown that the prognosis is the worst for carriers of 17p deletion, followed by those with 11q deletions, and better for those with trisomy 12 or none of the mentioned earlier. Surprisingly, patients with 13q deletions have favorable prognosis as compared to the other four groups. However, predicting patient's clinical outcome based

on the presence of a given biomarker is strongly dependent on the available treatment strategies. Differences in prognosis mentioned above are likely to change in the upcoming years due to emergence of new targeted treatments.

3.5 Contributions

The author of this thesis was responsible for the processing pipeline described in the current chapter. In essence, she processed raw sequencing data from the whole exome sequencing, performed variant calling and copy number estimation. She controlled genetic data consistency and made mismatch plots, which guided the manual process of mutation annotation of the samples. During many steps she received guidance from a former colleague, Paul-Theodor Pyl, the developer of Bioconductor R package *h5vc*, which she was using. However, all the scripts she wrote and run on the cluster solely by herself. DKFZ pipeline was run by Marc Zapatka. Manual aggregation of genomic information (Section 3.1.5) was performed by Bian Wu. Transcriptomic data was processed by Sascha Dietrich, Simon Anders and Sophie Rabe, and methylomics data by Christopher Oakes.

4

Ex vivo drug sensitivity in primary cancer cells

[Cancer treatment] is almost-not quite, but almost-as hard as finding some agent that will dissolve away the left ear, say, and leave the right ear unharmed.

— **William Woglom**
The Emperor of All Maladies:
A Biography of Cancer

Our drug screens measure the individual phenotypes produced by the compounds acting on patient primary material by comparing the ATP content of treated cells to the value obtained for non-treated controls. This measure is assumed to be a good surrogate of drug effectiveness. We tested a great variety of drugs. Although some of them have the same main target assigned (see Appendix C), they all differ in chemical structure and therefore they interact with their target in an unique way. Efficacy of binding to the target is modulated on different levels and can be influenced by, for example, receptor-ligand binding affinity, ligand efficacy, drug concentration and/or individual characteristics of the patient's cells. Other effects, directly connected to drug concentration, are called off-target effects. The higher the concentration, the higher the chance that a drug will modulate other targets (different from the one it was designed for), which are not necessarily biologically related to the target of interest. The choice of proper concentrations for each drug in the screen was a very important step especially in the context of following the analysis part.

4.1 Clustering of drug phenotypes

In the analysis we tried to minimize the influence of the off-target effects by focusing on the lowest drug concentrations. Therefore, we used the mean of drug response over the two lowest concentrations from the main screen to compare the response profiles between different patient samples. The Pearson correlation of all drug response profiles for CLL, MCL and T-PLL samples was calculated separately. We did it for the three biggest patient subgroups according to the diagnosis ($\geq$ 10 cases). The correlation coefficients were subjected to hierarchical cluster analysis and visualized as symmetrical heatmaps (Figures 4.1, 4.2, and 4.4).

4.1.1 Chronic lymphocytic leukemia

Although CLL is a very heterogeneous disease, we could observe groups of drugs which exhibit similar drug response profiles. The correlation coefficients of the unique pairs of drugs ranged from -0.49 to 0.91, and 3% of them showed a relatively high correlation > 0.6. We can distinguish seven meaningful drug clusters in Figure 4.1. These include drugs with identical or related targets, and drugs of converging pathway dependence.

The most prominent **cluster I** is formed mainly by inhibitors of the BCR signaling pathway, which target such kinases as Bruton's tyrosine kinase (BTK), spleen tyrosine kinase (SYK) or phosphoinositide 3-kinase (PI3K). Several drugs within this cluster produced response patterns similar to that of BCR inhibitors, consistent with the roles of their annotated targets being downstream of BCR signaling (MEK, AKT, LYN or SRC). Additionally, some other compounds showed unexpected phenotypic similarity to BCR inhibitors. These include checkpoint kinase (CHK) inhibitors (AZD7762; PF 477736) and the heat shock proteins (HSP) inhibitor (AT13387). **Cluster II** contains all three inhibitors of redox signaling / reactive oxygen species (ROS) present in the drug screen. **Cluster III** comprises drugs which target two serine/threonine protein kinases, which are taking part in DNA damage response. **Cluster IV** includes the drugs of converging p53 pathway dependence including nutlin-3 (inhibits the interaction between MDM2 and tumor suppressor p53) and two chemotherapeutics: fludarabine and doxorubicin. Although these drugs show more activity in higher concentrations, we can already see their activity in low drug concentrations. **Cluster V** consists of BH3-mimetics drugs, and quite unexpectedly, the drug designed to combat obesity. **Cluster VI** comprises inhibitors of MAPK/ERK signaling pathway, which plays a role in inhibiting the uncontrolled growth of cancer. The last **cluster VII** contains drugs targeting CHK and JAK inhibitors.

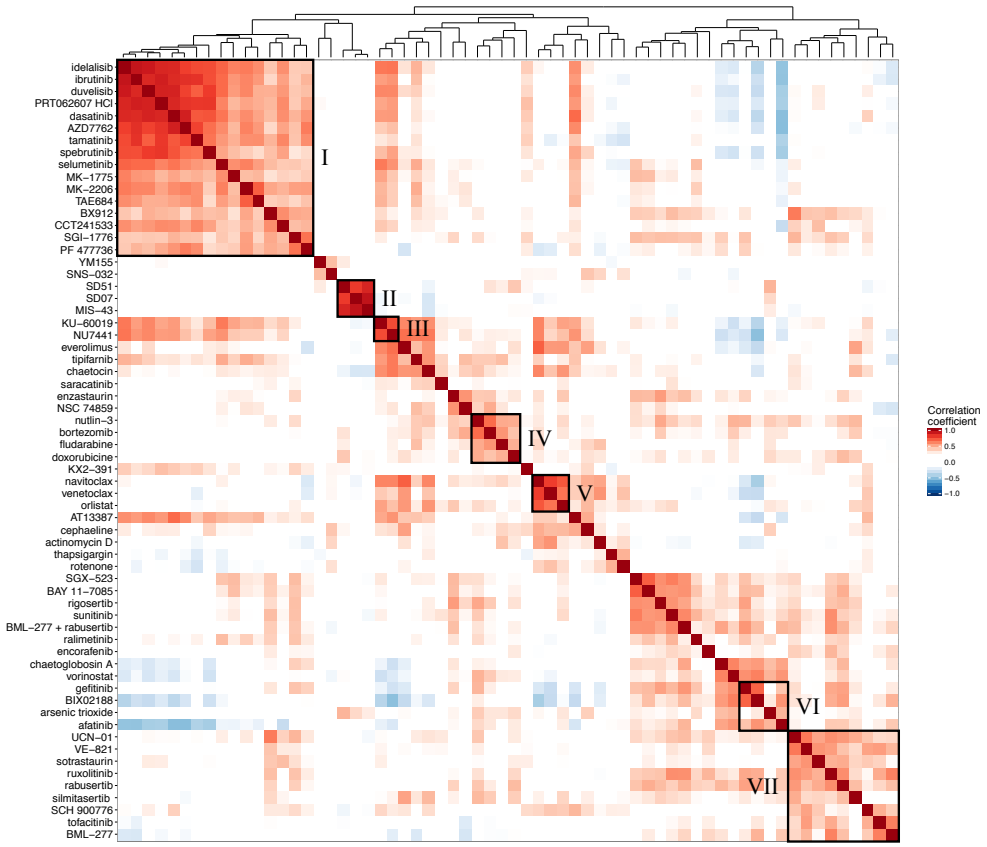


Figure 4.1 Correlation of drugs based on response profile in CLL samples.

The symmetric heat map shows Pearson correlation coefficients for all pairs of drugs for CLL. Major clusters include: (**I**) B-cell receptor signaling inhibitors i.a. idelalisib (PI3K), ibrutinib (BTK), and duvelisib (PI3K), (**II**) inhibitors of reactive oxygen species (SD51, SD07, MIS-43), (**III**) inhibitors of ATM (KU-60019) and DNAPK (NU7441), a complementary couple in DNA double strand break repair, (**IV**) modulators of p53 pathway and chemotherapeutics (fludarabine, nutlin-3, doxorubicin), (**V**) BH3-mimetics drugs (navitoclax, venetoclax), (**VI**) inhibitors of MAPK/ERK pathway (gefetinib, BIX02188, afatinib), and (**VII**) drugs which interfere with key players in cell cycle regulation and DNA damage response.

4.1.2 Mantle cell lymphoma

In the case of MCL, due to a common disease origin with CLL, we would expect similar drug response profiles and hence comparable clusters to appear. The dynamic range of correlation coefficient was spanning from -0.85 to 0.97, which made the heat map in Figure 4.2 look more prominent than for CLL. 18% of all correlation coefficients for unique drug pairs were strong, either < -0.6 or > 0.6. MCL and CLL share a similar level of heterogeneity, therefore the most probable explanation for such relatively strong correlations between drugs resulting in less specific clusters, was the low number (10) of MCL patient samples available for the analysis. This is not enough for catching multiple drug response phenotypes.

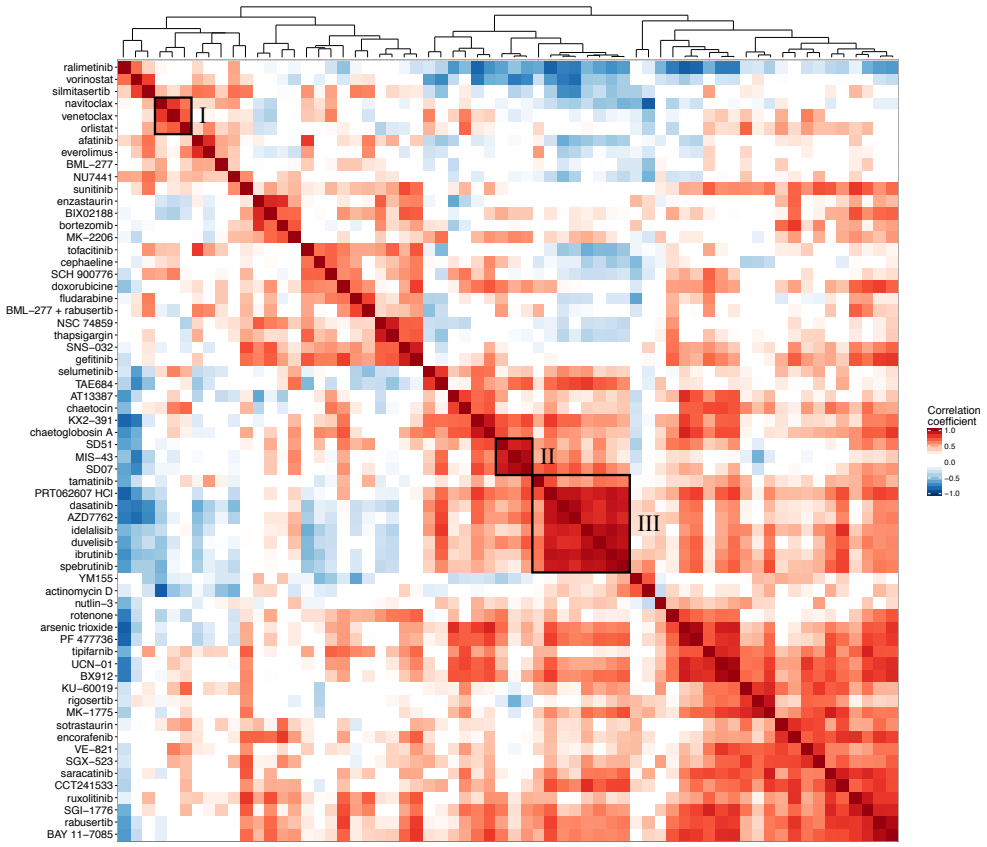


Figure 4.2 Correlation of drugs based on response profile in MCL samples.

The symmetric heat map shows Pearson correlation coefficients for all pairs of drugs for MCL. Major clusters include: (**I**) BH3-mimetics drugs (navitoclax, venetoclax), (**II**) inhibitors of reactive oxygen species (SD51, SD07, MIS-43), (**III**) B-cell receptor signaling inhibitors i.a. idelalisib (PI3K), ibrutinib, spebrutinib (both BTK), duvelisib (PI3K).

We identified 3 both crisp and meaningful clusters. They all can be related to specific clusters also seen in CLL, which confirms the similarity of these two diseases. **Cluster I** includes small molecule compounds active through BH3-mimetics mechanism. Again, surprisingly, these two drugs cluster together with the anti-obesity drug orlistat. **Cluster II** comprises all three tested inhibitors of redox signaling / ROS. The last one, **cluster III** contains mostly inhibitors of BTK, PI3K, SYK, SRC and other elements of BCR signaling pathway.

4.1.3 T-cell prolymphocytic leukemia

For T-PLL the situation is quite different as it is also commonly seen in the clinical setting. Patients are in general less responsive to standard treatments, which is reflected in the drug screen. The disease is characterized by more resistant phenotype than CLL, as shown in Figure 4.3. The list of drugs used in the screen was enriched for compounds interfering with components of the BCR signaling pathway (11/64 drugs; 17%). These are not relevant in the course of T-cell disease.

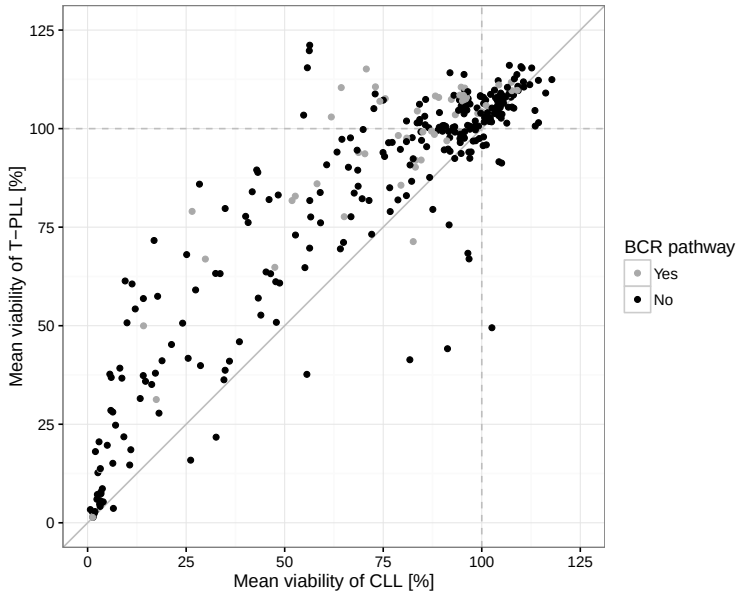


Figure 4.3 Mean drug response of CLL *vs.* T-PLL.

The scatter plot shows the mean drug response for each drug and concentration calculated over all CLL (x-axis) and T-PLL (y-axis) samples available in the main screen. 277/320 (87%) dots are located above the diagonal, which indicates resistance of T-PLL samples to the drug treatment as compared to CLL. The increased resistance of T-PLL holds true irrespective of drug targets.

The Pearson correlation coefficients for unique drug pairs ranged from -0.59 to 0.93. 6% of them exhibited a relatively high correlation > 0.6. Figure 4.4 visualizes this correlation together with clustering of the highly-correlated drug pairs. As the result of common weak activity of compounds in T-PLL, the emerging clusters are sparse and scattered. There are only two clusters which share same or similar targets: **Cluster I** comprising drugs which interfere with cell cycle regulation and DNA damage response mechanisms, and **cluster II** consisting of all three ROS inhibitors. The absence of a cluster with inhibitors of up- and down-stream components of the BCR pathway is reassuring that they have no impact on the T-cell disease. Most drugs cluster just because they exhibit low activity in general.

4.1.4 Summary

We observed distinct and heterogeneous viability effects produced by drugs across samples. Clusters of phenotypic similarities profoundly identify drugs with related targets. The most noticeable examples are BCR inhibitors such as ibrutinib, idelalisib, tamatinib, spebrutinib, duvelisib and PRT062607 HCl, targeting BTK, PI3K and SYK. These were highly correlated with each other, and additionally, showed similar response profiles to drugs such as selumetinib, MK-2206, and dasatinib, which are inhibitors of the downstream BCR targets: MEK, AKT, LYN or SRC. We also

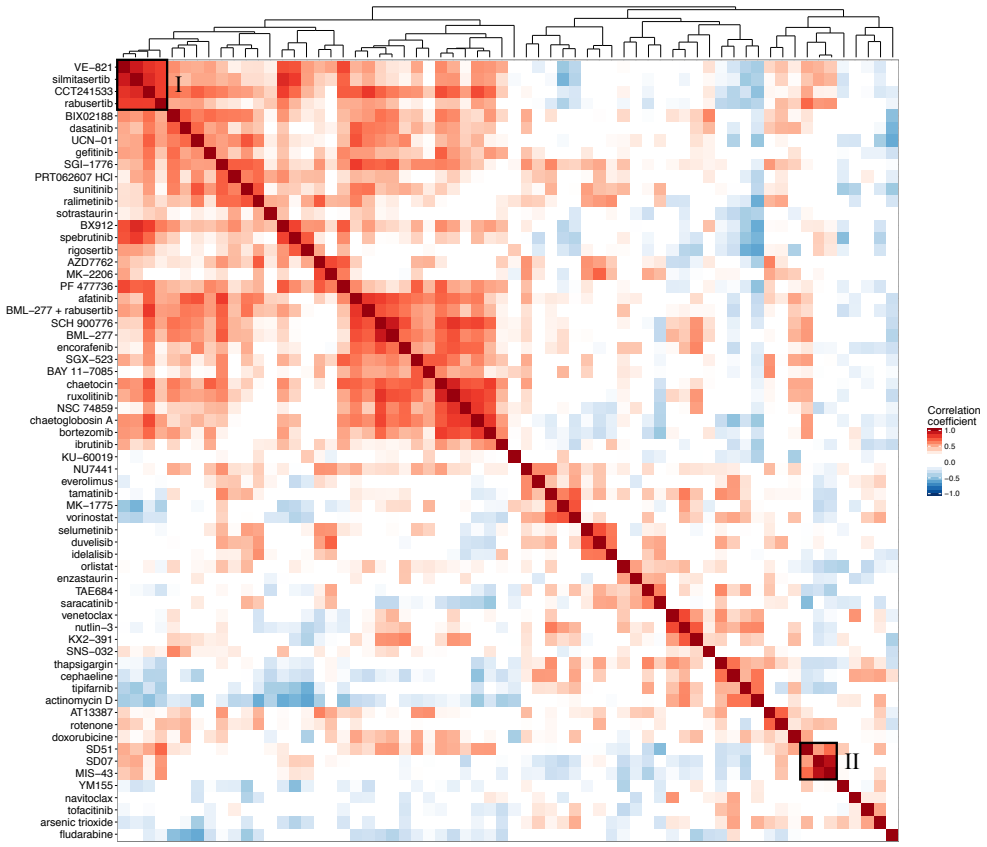


Figure 4.4 Correlation of drugs based on response profile in T-PLL samples.

The symmetric heat map shows Pearson correlation coefficients for all pairs of drugs for T-PLL. The two major clusters include: (**I**) drugs which interfere with a key players in cell cycle regulation and DNA damage response (VE-821, silmitasertib, CCT241533, rabusertib), and (**II**) inhibitors of reactive oxygen species (SD51, SD07, MIS-43).

identified compounds with unexpected phenotypic resemblance to BCR inhibitors, including AZD7762 and PF 477736 (targeting CHK [85, 86]) and AT13387 (targeting HSP). This BCR-like response profile was presumably caused by off-target effects directed towards the BCR pathway. Moreover, drugs of converging pathway dependence, e.g. *TP53*: nutlin-3 and fludarabine, also produced distinctively similar response patterns.

Drug clustering varied depending on the selection of samples, which supports the biological relevance of the results. For instance, the cluster of BCR inhibitors was absent for T-PLL samples, whereas for the other groups of drugs (e.g. targeting ROS or working through BH3-mimetics) clusters were present, which is consistent with the observations in CLL and MCL.

We showed that survival dependencies of tumor cells can be decoded by the analysis of drug response profiles. “Guilt by association” strategy allows for discovery of unanticipated drug targets.

4.2 Functional classification of lymphoproliferative disorders

To gain a global overview of the response variation across patients, we clustered samples by the similarities of their drug response profiles (Figure 4.5).

We used a model-free approach that clustered both samples and drugs based on signatures produced by separate drug concentrations and thus, allowed to account for a dose-dependent target specificity.

The most prominent cluster, which can be characterized by the red-to-blue color gradient within CLL, is driven by a group of drugs including BCR inhibitors. Same pattern emerged for other kinase inhibitors, and we called their response to be "BCR-like". This gradient separates CLL with unmutated IGHV region (U-CLL) and CLL with mutated IGHV region (M-CLL) almost perfectly. Our data also indicates that dependence on BCR signaling varies on a continuous scale between different tumors and is most pronounced in the majority of U-CLL, while weaker in M-CLL. This observation is in line with increased BCR pathway activity in U-CLL [87, 88]. Within M-CLL reliance on other signaling pathways could be observed. The most outstanding example, the dependency on mTOR pathway in a subgroup of M-CLL, was high and exclusive, therefore independent of BCR signaling. A similar pattern of variable response as seen in CLL was apparent in MCL, with a subset of samples showing sensitivity to BCR inhibitors, consistent with previous reports [89, 90]. Pronounced sensitivity to BRAF/MEK inhibition was seen in hairy cell leukemia (HCL), reflecting the known role of the *BRAF* V600E mutation and clinical observations [28]. T-PLL samples again exhibited an extremely resistant phenotype to almost all tested compounds. However, they were also the only disease sensitive to multiple concentrations of thapsigargin.

Together, these results provide a fine-grained classification of disease based on response phenotype and give insight into essential disease-specific signaling signatures of blood cancer as e.g. BRAF/MEK/ERK signaling in HCL, and BCR signaling across a set of B-NHL, including CLL and MCL.

4.3 Influence of cell lineage and disease subtype on drug response

In clinical setting, similarities of response to drugs between patients suffering from same cancer type can be observed. In order to understand the role of cell lineage and disease subtype on drug response, we compared drug sensitivity profiles across diseases at single compound resolution. The analysis included all available diseases for which at least three samples have been collected and tested. The most abundant cohort of CLL samples was used as reference. We compared responses to single drugs between the reference and other diseases using Student t-test (two-side; equal variance). p-values were adjusted for multiple testing, and these were used to call significant associations. The results are presented in Figure 4.6. Predominance of the pink color in the figure indicates general resistance to drugs (rather than sensitivity) for virtually all tested diseases as compared to CLL.

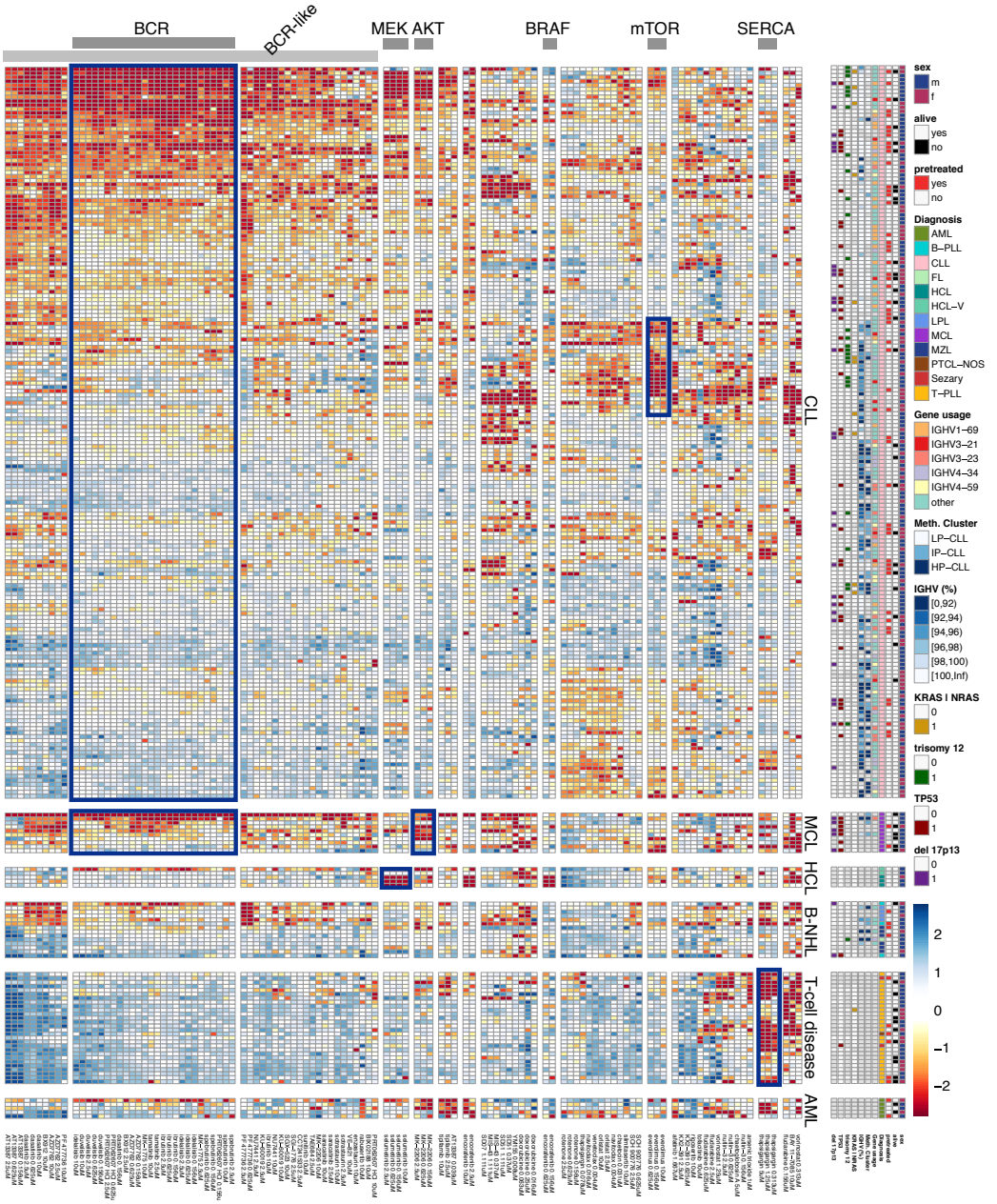


Figure 4.5 Landscape of drug response.

The heat map gives an overview of the measured viabilities for all the samples (rows) under drug treatment. For each drug-concentration pair (column) the viability readouts were centered and presented on a z-score scale. We show 204 drug-concentration pairs (for 53 drugs) which exhibited the most variable response. The most important features of the samples are presented on the right hand side of the figure. Hierarchical clustering (Euclidean metric, complete linkage) of the samples divided the heat map into functional disease subgroups.
Figure created by Wolfgang Huber.

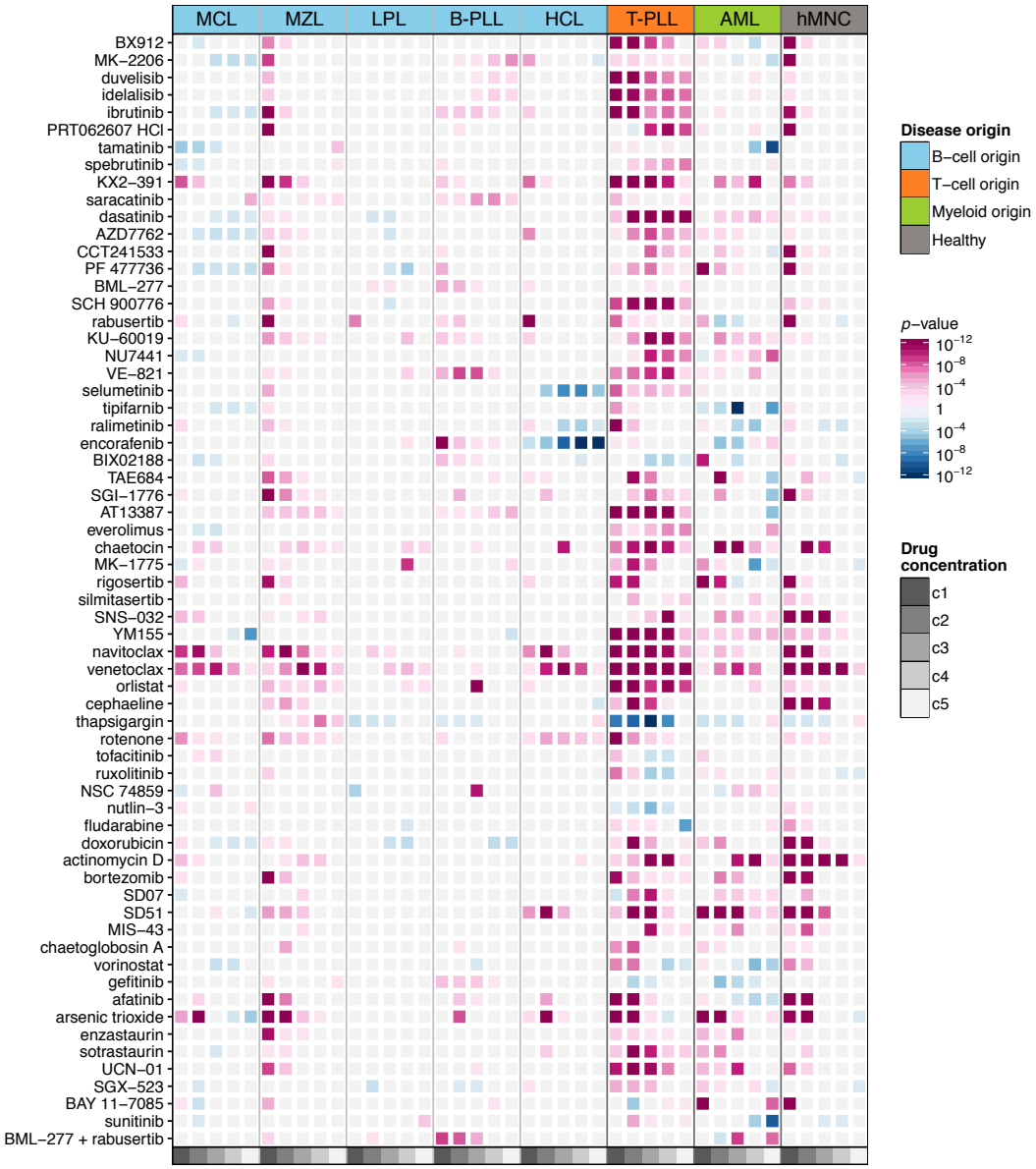


Figure 4.6 Summary of disease-specific drug effects.

The heat map shows the significant differences in drug responses of seven diseases and control mononuclear cells (hMNC) as compared to CLL. The color indicates the direction of difference (pink: less, blue: more sensitive compared to CLL) and p-value (two-sided t-test; 10% FDR). The five columns within each block correspond to the five concentrations tested (c1: highest, c5: lowest). Not significant differences are shown in light gray. FDR of 10% was used.

T-PLL is characterized by the remarkably resistant phenotype, which is in line with clinical observations (see also Figure 4.3). This is clearly evident not only for inhibitors of the BCR signaling pathway, but also for broadly active compounds including the BH3-mimetics navitoclax and venetoclax, the HSP inhibitor AT13387, and agents interfering with reactive oxygen species (ROS). Nutlin-3 (10 μM); $p = 0.001$) was more active in T-PLL (Figure 4.7), which could be potentially reflecting the ab-

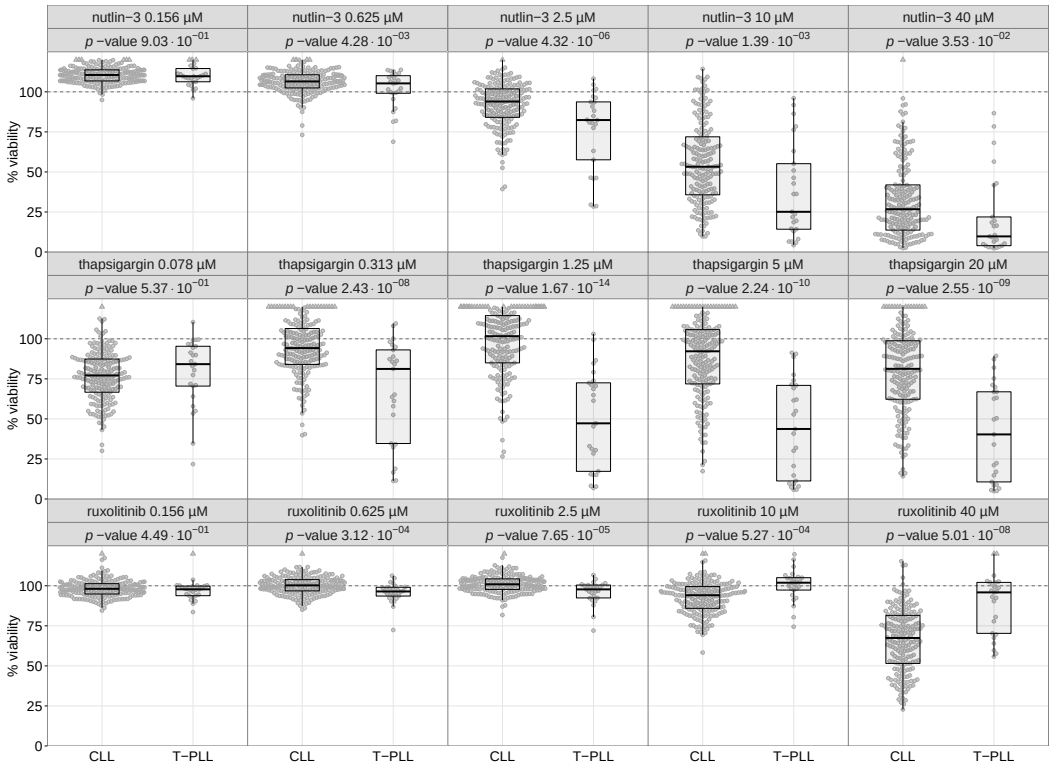


Figure 4.7 Drug response differences between CLL and T-PLL.

T-PLL samples were more sensitive to nutlin-3 and thapsigargin.

sence of *TP53* mutations in T-PLL. We identified thapsigargin, a non-competitive inhibitor of the sarco/endoplasmic reticulum Ca2+ ATPase (SERCA) with preferential activity in T-PLL for 4 of 5 drug concentrations ($p < 0.001$, Figure 4.7). Additionally, JAK inhibitors ruxolitinib and tofacitinib showed increased activity in T-PLL in lower drug concentrations ($p < 0.001$). Both findings offer potential repurposing opportunities.

In comparison to our representative of myeloid-originated disease, AML, a set of drugs showed increased activity. Those included inhibitors of: SYK (tamatinib), farnesyltransferase (tipifarnib), FLT3 (sunitinib) and epidermal growth factor receptor (EGFR) (gefitinib) (Figure 4.8). Their potential in treating AML has already been noticed in other studies [91, 92, 93, 94]. Clinical trials conducted for sunitinib yielded promising results, however, tipifarnib and gefitinib proved not to be an appropriate therapy for AML patients.

Although all B-cell lymphoma subtypes seem to respond to treatment in a similar fashion, several distinct features could be observed. In order to further zoom in into these differences we repeated the analysis, this time using U-CLL and M-CLL separately as references (Figure 4.9). By doing so we redirected the focus from the biggest biological divider of CLL (meaning IGHV status) into more subtle pathway dependencies.

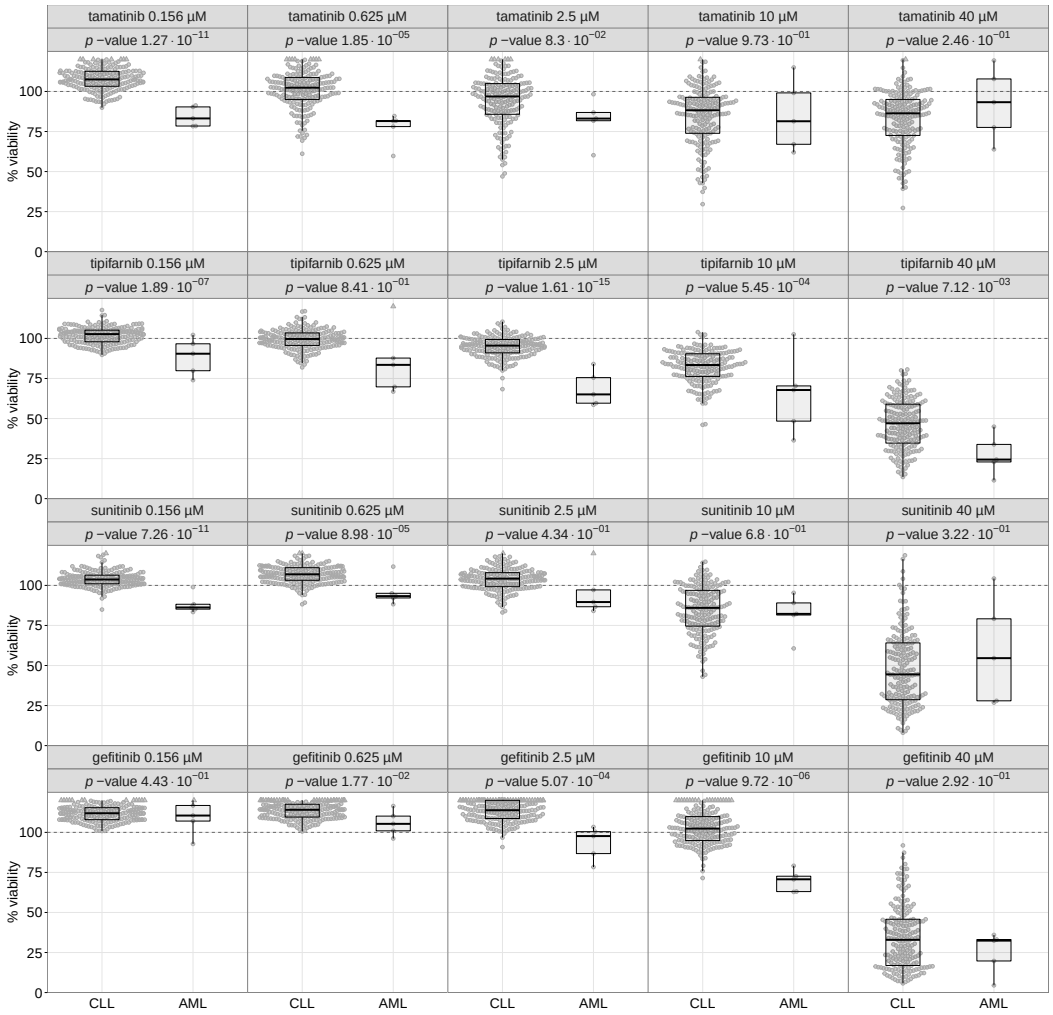


Figure 4.8 Drug response differences between CLL and AML.

AML samples were more sensitive to tamatinib, tipifarnib, sunitinib and gefitinib.

MCL showed novel and specific sensitivity to the survivin inhibitor YM155 (8 nM; $p < 0.001$). On the other hand, it was less sensitive than CLL to BH3-mimetic drugs. Similar to clinical observations, MCL were sensitive to BCR inhibitors [89], showing a similar response pattern to U-CLL and the mTOR inhibitor everolimus (2.5 μM; $p = 0.004$; Figures 4.9A and 4.10). Marginal zone lymphoma samples in this study were distinctly resistant to inhibitors of the BCR including a broader set of kinase inhibitors (dasatinib, PRT062607), and the HSP inhibitor AT13387 (Figures 4.9B and 4.11). Individual drugs, e.g. PF 477736, dasatinib, showed preferential activity in LPL (Figures 4.9C and 4.12). Hairy cell leukemia samples were exquisitely sensitive to a BRAF and MEK inhibition (encorafenib, selumetinib; Figures 4.9E and 4.13). This finding is based on the presence of *BRAF* V600E mutation in HCL, which results in activation of the MEK/ERK pathway [12]. Moreover, CLL samples harboring same *BRAF* mutation tend to respond better than *BRAF* wild-type CLL, although still not as effectively as HCL cases (Figure 4.14).

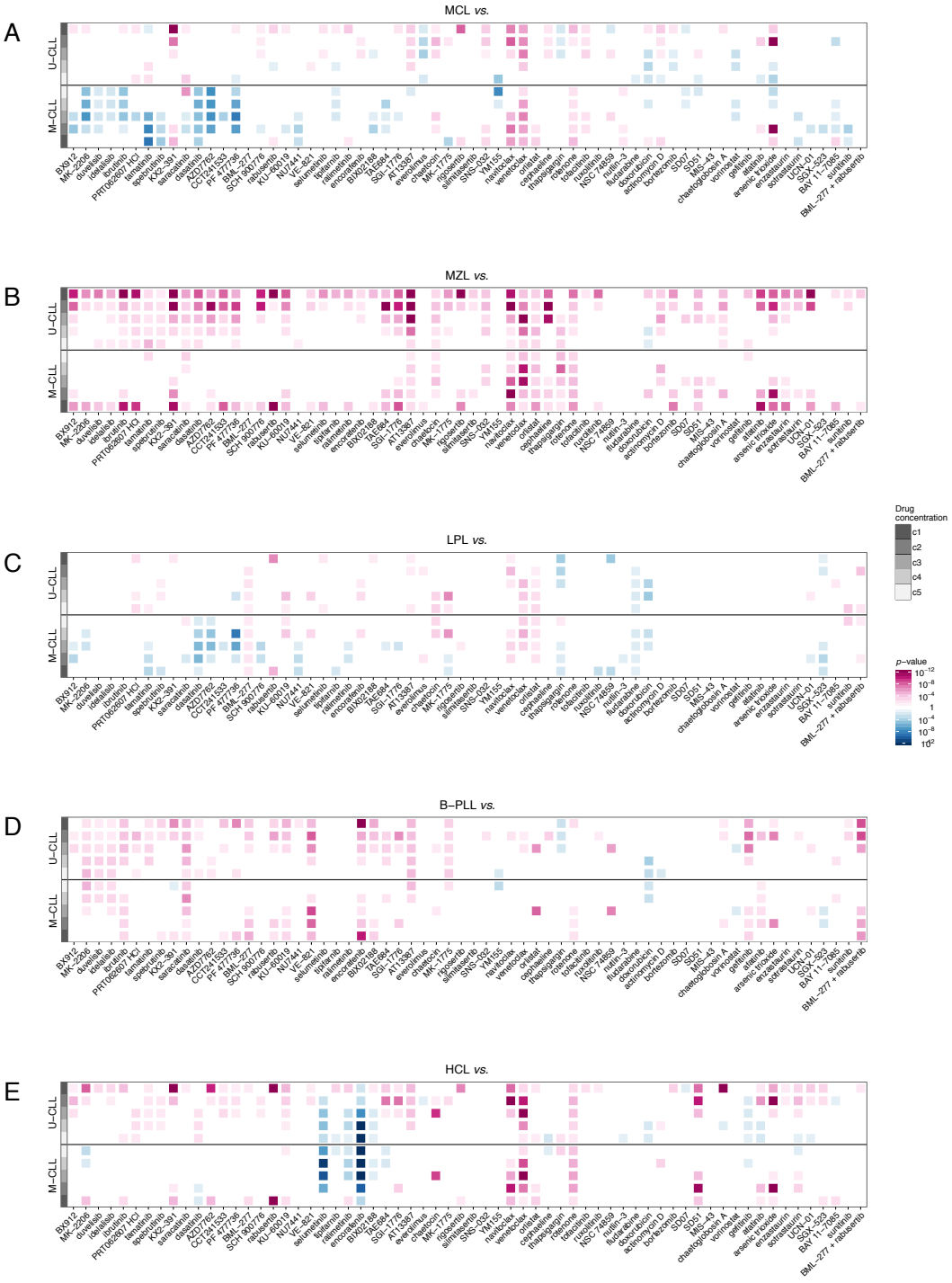


Figure 4.9 Drug response differences within B-cell cancers.

Drug response in MCL (**Panel A**), MZL (**Panel B**), LPL (**Panel C**), B-PLL (**Panel D**), and HCL (**Panel E**) as compared to two references: U-CLL (top heat map) and M-CLL (bottom heat map). The pink and blue colors indicate a more resistant or sensitive drug response profile than the reference, respectively. The bar in different shades of gray on the left hand side of each heat map indicates the drug concentrations.

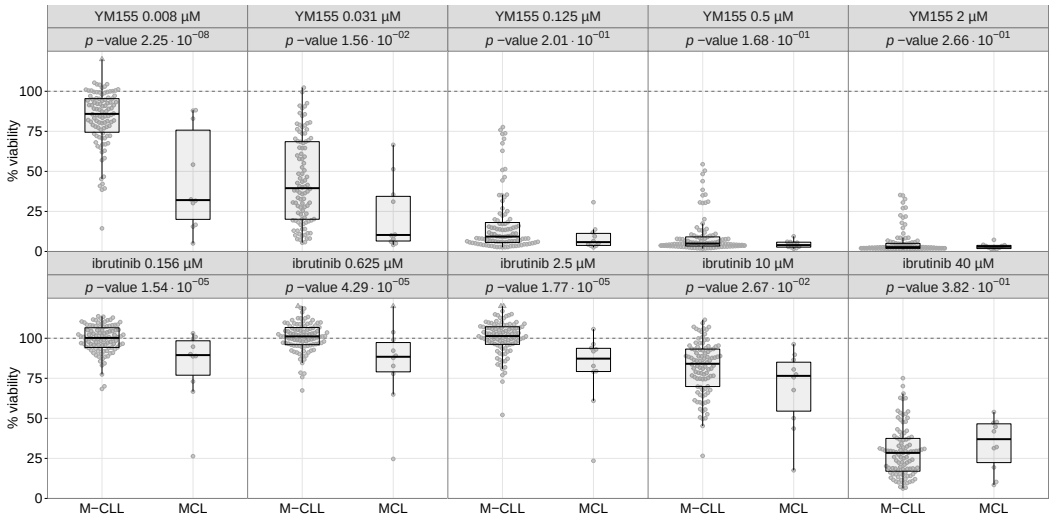


Figure 4.10 Drug response differences between M-CLL and MCL.

MCL samples were sensitive to survivin inhibitor YM155 and BCR inhibitor ibrutinib.

These data demonstrate that distinct and disease-specific sensitivity profiles of blood cancer subtypes can be uncovered by *ex vivo* drug response profiling. Additionally, new disease-specific effects could be observed, which in principle can be exploited clinically or used for improving the not-yet-perfect disease classification.

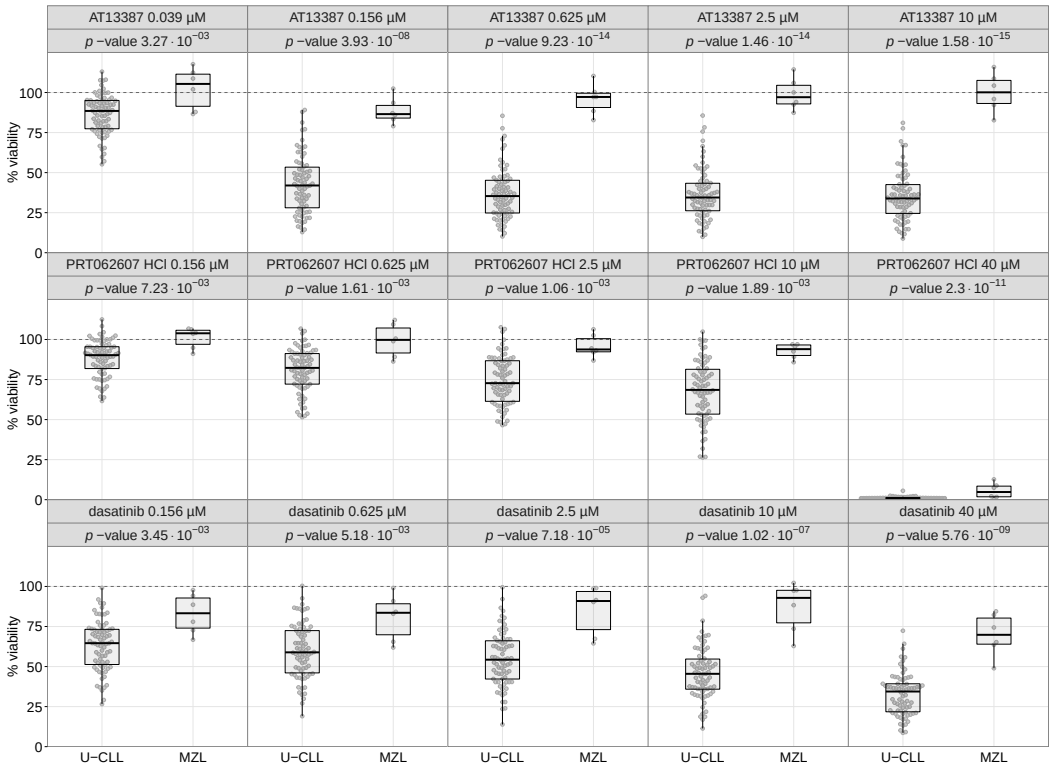


Figure 4.11 Drug response differences between U-CLL and MZL.

MZL samples were resistant to the HSP inhibitor AT13387 and kinase inhibitors (dasatinib, PRT062607).

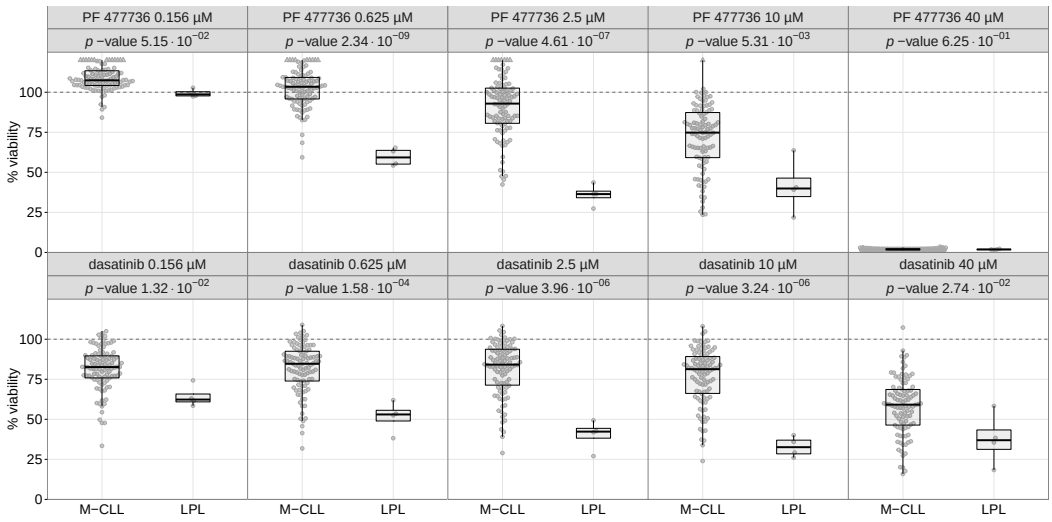


Figure 4.12 Drug response differences between M-CLL and LPL.

LPL samples were more sensitive to PF 477736 and dasatinib

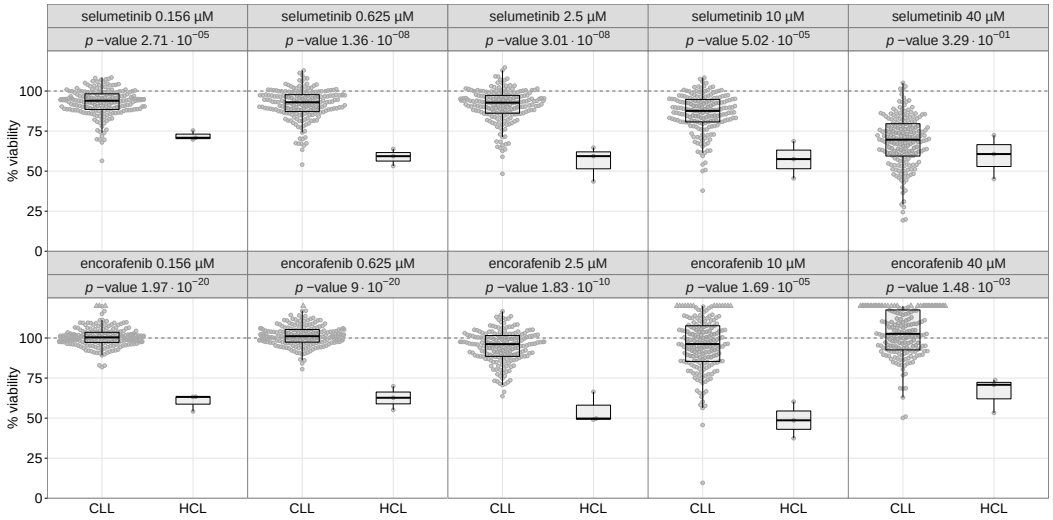


Figure 4.13 Drug response differences between CLL and HCL.

HCL samples were more sensitive to inhibitors of MEK (selumetinib) and BRAF (encorafenib).

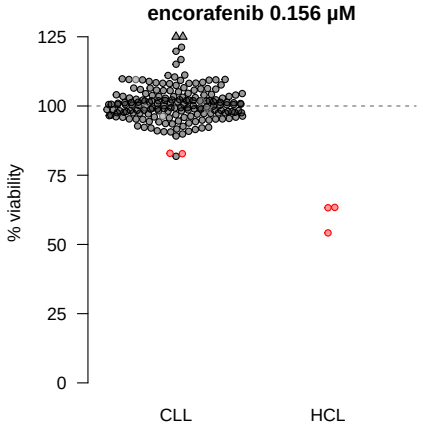


Figure 4.14 Drug response in samples with V600E *BRAF* mutation.

Bee swarms show drug response for all CLL and HCL samples to BRAF inhibitor encorafenib. The red color indicates the presence of *BRAF* V600E mutation with an allele frequency $> 10\%$.

Molecular factors influencing drug response in CLL

5

The statistician is no longer an alchemist expected to produce gold from any worthless material offered him. He is more like a chemist capable of assaying exactly how much of value it contains, and capable also of extracting this amount, and no more.

— **Sir Ronald A. Fisher**

Knowledge about the presence of single gene mutations is becoming increasingly valuable to inform treatment approaches. Therefore, we combined information on molecular aberrations for each sample and evaluated their associations with drug sensitivity profiles in CLL samples. The matrix of molecular aberrations comprised: somatic mutations (aggregated at gene level), copy number aberrations and IGHV mutation status.

We univariantly tested each genomic feature (43 features for the pilot screen and 63 for the main screen) for their associations with the drug response by using Student t-test (two-sided, with equal variance). Each concentration was tested separately, and the minimal size of the compared groups was set to 3. For each screen the p-values were adjusted for multiple testing by applying the Benjamini-Hochberg procedure. Adjusted p-values were then used for setting the significance threshold of 10% false discovery rate. The distinction between monoallelic and biallelic deletion 13q was discarded from the results [95].

Before going into any detailed analysis, we checked again for the potential confounding of the results with the fact that the main screen was performed in three separate batches (see Figure 2.5). We repeated the drug-feature association tests using batch group as a blocking factor in a two-way ANOVA test, and then com-

pared the p-values from both tests (see Appendix E). Only one drug, bortezomib, showed discrepant p-values, and further exploration of its data suggested that it lost its activity during storage. All the remaining associations yielded equivalent results regardless of whether tested with or without batch as a blocking factor. Therefore, all the reported p-values for associations come from the t-test without blocking for batch effects.

5.1 Biomarkers of chronic lymphocytic leukemia

Within the main screen we identified 273 highly significant associations between drug response and molecular aberration (see Figure 5.1). All 64 compounds were associated with at least one mutation, and similarly, 48 (79%) of mutations were associated with at least one compound. The detected associations were supported by one (157; 57.5%) or multiple drug concentrations (54, 27, 23 or 12 associations corresponding to 19.8%, 9.9%, 8.4%, 4.4% of all associations for 2–5 concentrations, respectively). IGHV was the factor which modulated the response to the most number of compounds (46; 72%), including all 10 drugs which are targeting either the BCR pathway directly or downstream from it—the PI3K/AKT pathway. Excluding IGHV, in majority of cases (141; 62%) the mutated samples exhibited a more resistant phenotype as compared to wild type ones. Mutation caused increased sensitivity to drug treatment in 85 (37%) of all associations. There was one association where the mutation caused divergent effect on the drug response, depending on the drug concentration. Moreover, we compared the number of significant drug-gene associations between the pilot and the main screens and observed their increase coupled with more samples tested (for details see Appendix H). Given the fact that we were able to detect many single gene-drug associations across a variety of gene mutations, the data suggests that those biomarkers could explain therapeutic selectivity of the drugs.

Mutations, which were identified to be associated with the drug response targeting diverse molecular processes including: DNA damage (TP53), MEK/ERK signalling (*BRAF*, *RAS*), epigenetic modification (*CREBBP*), pre-mRNA-processing / splicing (*PRPF8*, *SF3B1*) and nuclear export (*XPO1*). Some of the associations included mutations with less well-defined function such as *ABI3BP*, *MED12*, *UMODL1* and gain 8q24. Several unexpected associations including rare variants will be followed up in Section 5.1.4.

5.1.1 IGHV status

IGHV status is a well-established marker of prognosis in CLL. Variable regions of immunoglobulin, which are part of the B-cell receptor, have unique characteristics in malignant cells. They allow to identify the stage of B-cell differentiation and maturation from which the tumor originates [96]. The B-cell receptor plays a crucial role in B-cells by regulating key signaling pathways responsible for cell survival, apoptosis,

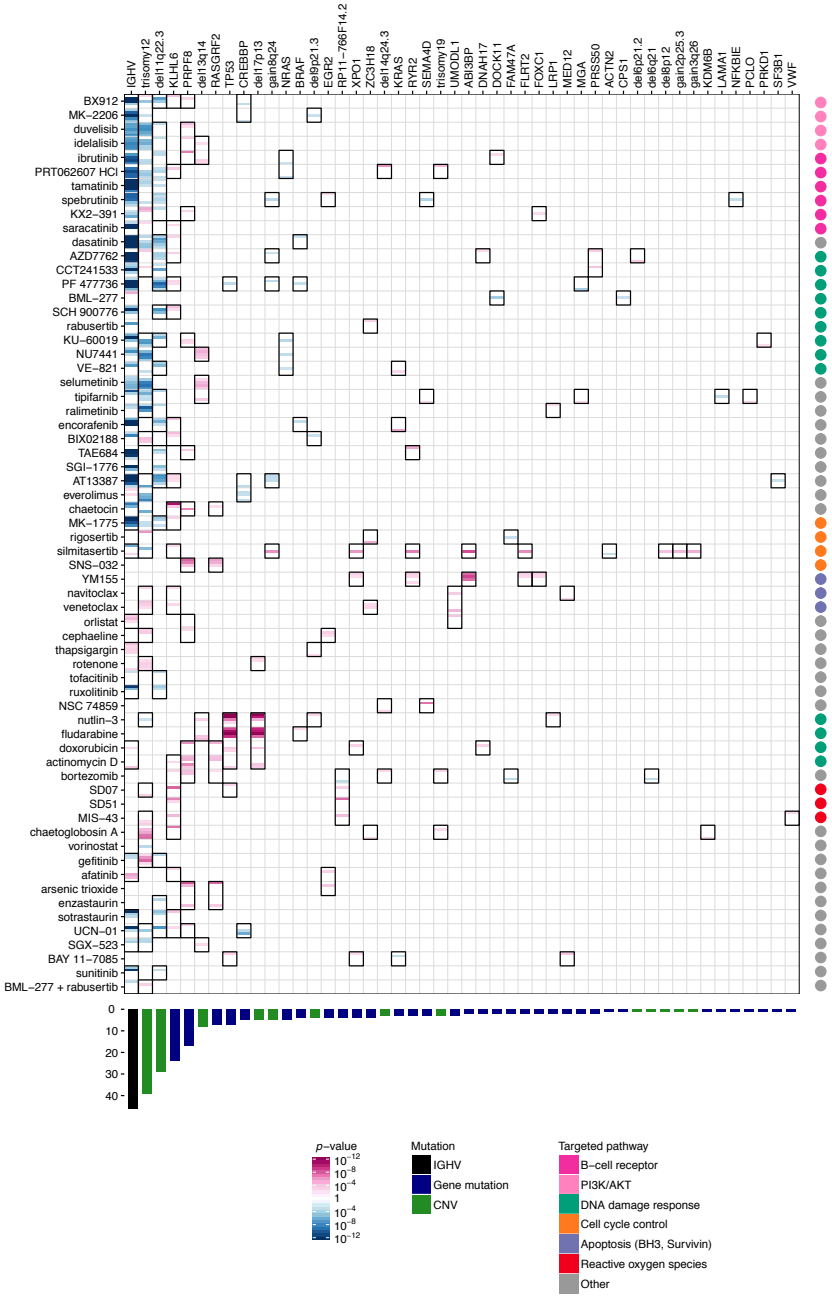


Figure 5.1 Summary of drug-mutation associations found in the main screen.

The heat map shows associations between drug responses (y-axis) and gene mutations, structural aberrations or IGHV status (x-axis) identified within the CLL cohort. Student's t-test was used to compare responses of the drug (in each concentration step separately) between samples with and without a given mutation. A false discovery rate of 10% was used. p-values and direction of the effect are color coded, with pink and blue indicating resistant and sensitive phenotypes, respectively, in the presence of the tested mutation. For IGHV status, pink indicates sensitivity of M-CLL and blue indicates sensitivity of U-CLL. Each gray box consists of 5 horizontal bars, which correspond to five concentration steps (highest to lowest concentration going from top to bottom of a square). The bar plot on the bottom summarizes the number of associations found per gene (on the drug level; color coded by type of mutation). Dots on the right indicate basic characteristics of drugs.

proliferation, migration, and, only in case of healthy B-cells, differentiation. The level of antigens present on the surface of B-cells is assumed to be the main factor modulating the response to BCR activation [97] (most presumably self-activation [98]). In the case of U-CLL expression of these immunoglobulins is retained to a greater extent than in M-CLL (B-cells which went through somatic hypermutation in germinal center) [97]. Therefore, it is believed that the pro-survival signals received by malignant B-cells can be suppressed by targeting kinases downstream of BCR.

Indeed, IGHV mutation status was a major determinant of CLL response to the six BCR inhibitors (ibrutinib, saracatinib, tamatinib, KX2-391, spebrutinib, and PRT062607 HCl; Figure 5.2). Their robust differences were observed over multiple drug concentrations (Figure 5.3), including the lowest ones tested. The obtained effect sizes were comparable to the ones observed in the previous smaller studies like Guo et al. [88]. The effectiveness of BCR inhibitors in patients is currently being extensively tested in clinical trials. Some of them are already showing promising results in improving the survival of patients [36, 37].

Moreover, U-CLL were sensitive to SRC inhibitor dasatinib, similarly to what already has been noticed in a modest cohort of patients [99]. Previous studies with this drug suggested a potential benefit when used in combination with drugs whose target is outside of BCR signaling [100]. Such approach could disturb pro-survival signals originating from BCR-independent pathways which play a significant role, especially in M-CLL.

Several compounds exhibited a "BCR-like" profile. These included CHK inhibitors AZD7762 [85], PF477736 [86] (Figure 5.4) and AT13387 targeting HSP90 (Figure 5.5). We further investigated the effect of AZD7762 by looking at the changes in gene expression under treatment (Figure 5.6). The results suggest that the CHK inhibitor acts on the BCR signaling pathway through off-target effects. Similarly, AT13387 most probably interferes with components of BCR-signalosome, such as BTK and SYK [101].

5.1.2 *TP53* mutation / 17p deletion

TP53 gene is a tumor suppressor, whose function is frequently impaired in cancer due to mutation. It encodes a whole family of proteins, which are regulating gene expression through binding to DNA. *TP53* gene is sometimes called the "guardian of the genome", because it influences cell fate under stress signals. It guides the cell into DNA repair, cell-cycle arrest, senescence or apoptosis [105]. In CLL, *TP53* gene is a well-known biomarker. If it is mutated, patients poorly respond to standard front line fludarabine-based therapy [10]. After many recommendations to include the *TP53* mutation status into treatment decision [10, 106], it was finally accepted as a first, and so far the only one, genetic factor which guides treatment regimen

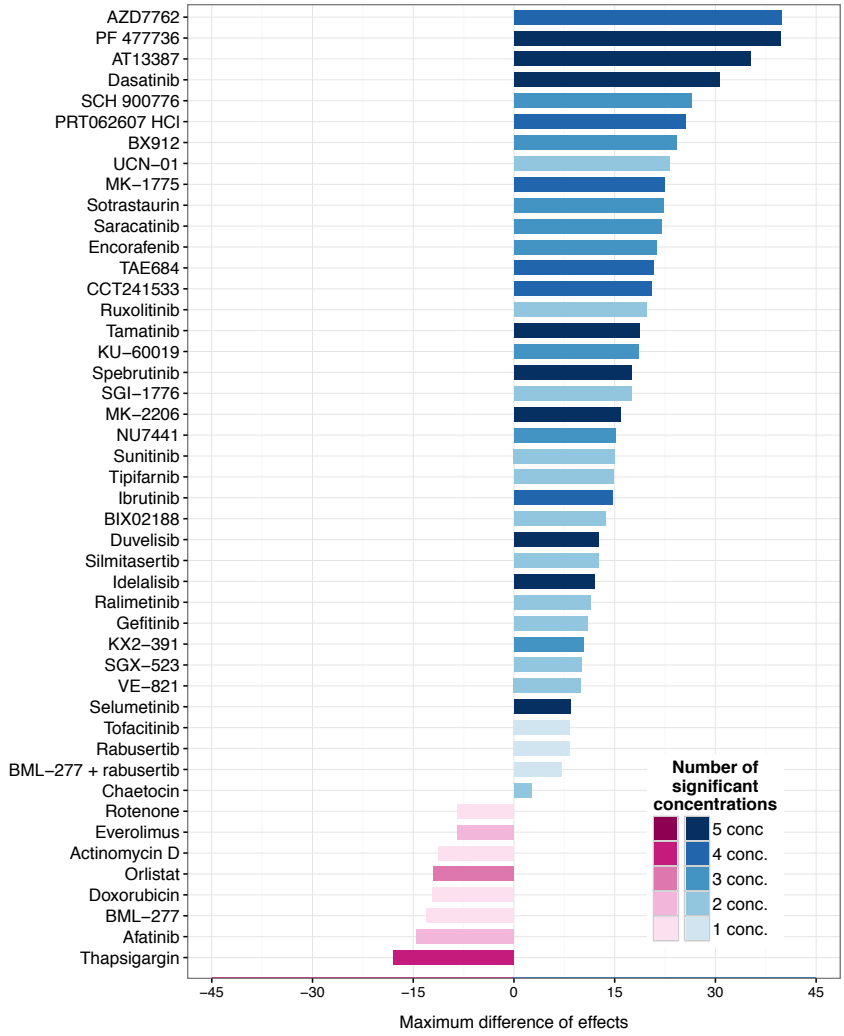


Figure 5.2 Drug responses associated with IGHV status.

Bar plot shows drugs (y-axis) which produced significantly different responses between U-CLL and M-CLL groups. Positive difference (blue) indicates higher sensitivity of U-CLL as compared to M-CLL. Conversely, negative difference (pink) indicates higher sensitivity of M-CLL as compared to U-CLL. The five concentrations of each drug were tested separately and the color intensity of the bars encodes for how many of them the difference in viability was statistically significant (FDR 10%). The length of a bar shows the greatest effect produced within the tested concentration steps.

applied in the hospital [5]. *TP53* mutation often occurs together with the deletion 17p, and its frequency increases with subsequent treatments from around 5-15% [107] to approximately 44% [108].

Our approach markedly recapitulated clinical observations. Response to fludarabine and doxorubicin—the two chemotherapeutics used in the study, which both cause cell death through apoptosis [109, 110]—were highly dependent on the mutation status of *TP53* gene and deletion 17p (Figure 5.7). These associations were very strong, thus yielded significant with multiple drug concentration steps. A similarly robust effect of resistant phenotype within mutated sample was observed for nutlin-3.

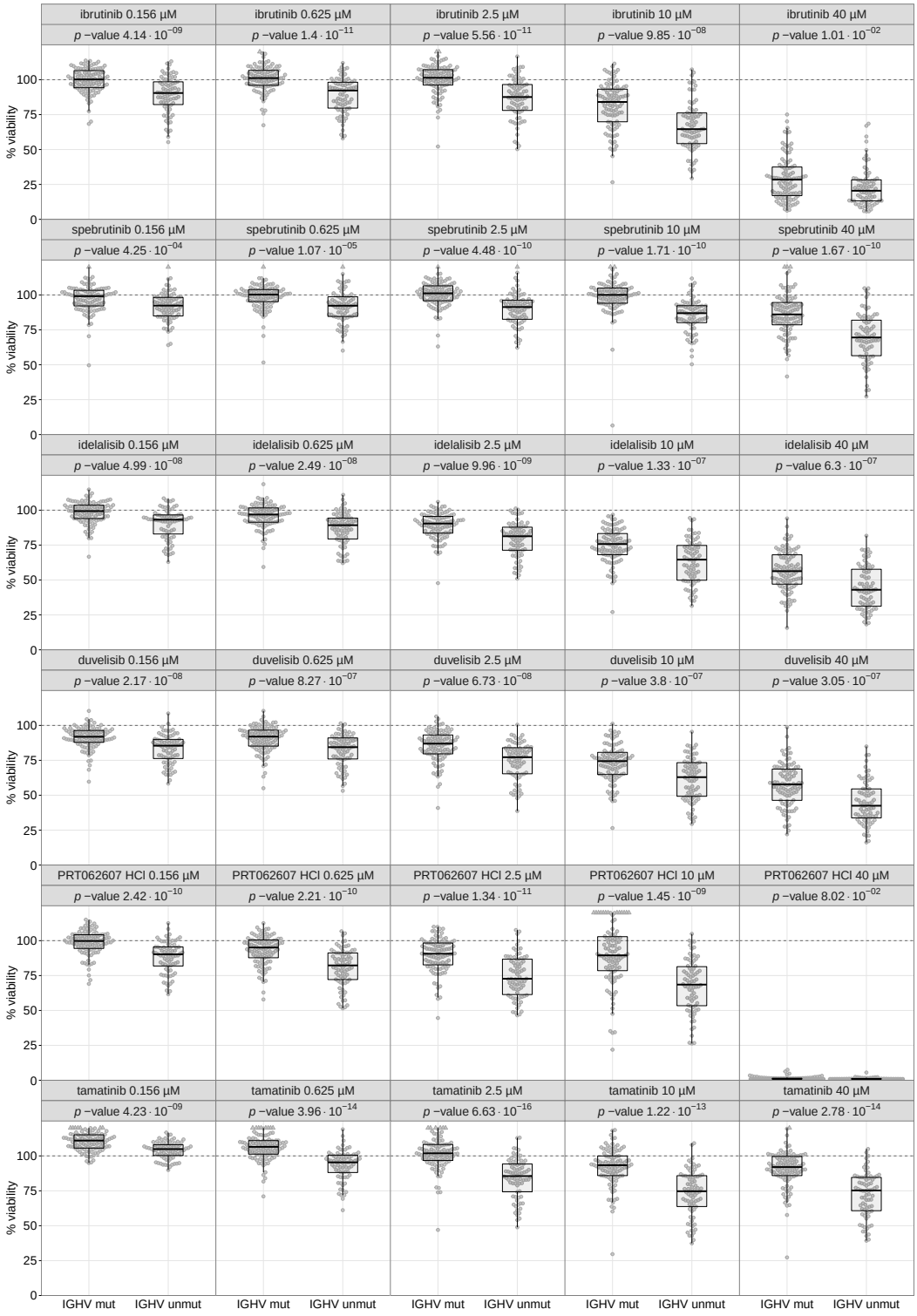


Figure 5.3 Drug response to BCR inhibitors stratified by IGHV status.

Effect of treatment with BCR inhibitors targeting BTK (ibrutinib, spebrutinib), PI3K (idelalisib, duvelisib), and SYK (PRT062607, tamatinib) is significantly different between M-CLL ($n = 98$) and U-CLL ($n = 74$) samples.

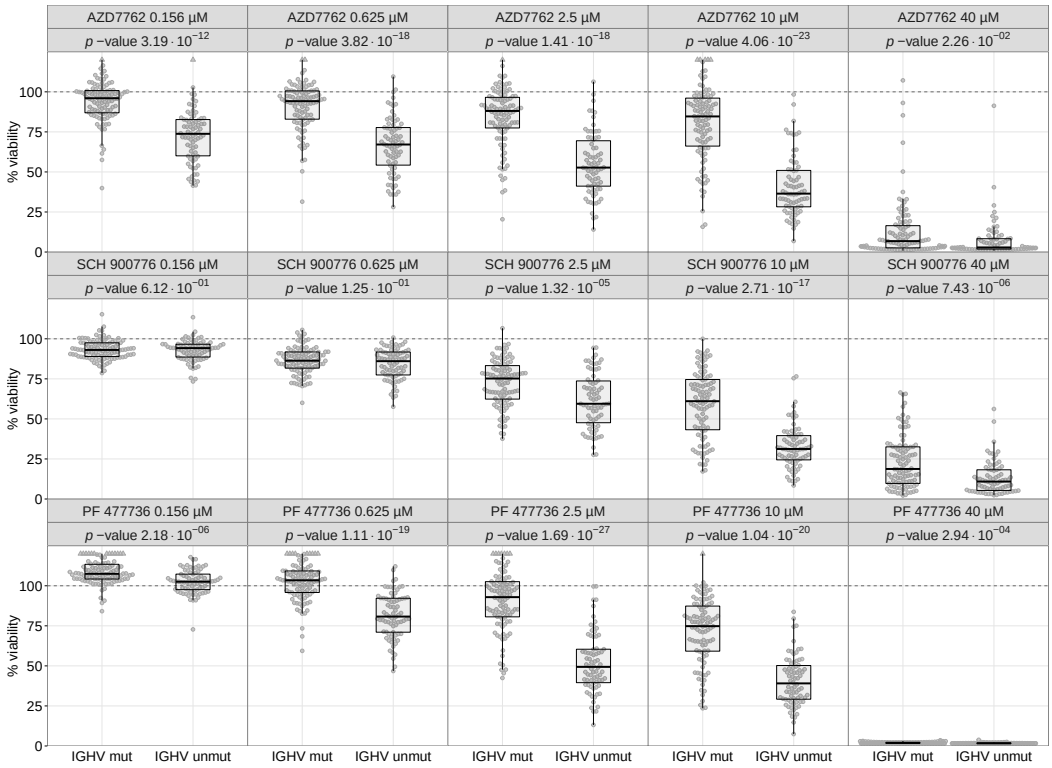


Figure 5.4 Drug response to CHK inhibitors stratified by IGHV status.

Effect of treatment with CHK inhibitors AZD7762, SCH 900776 and PF 477736 is significantly different between M-CLL ($n = 98$) and U-CLL ($n = 74$) samples.

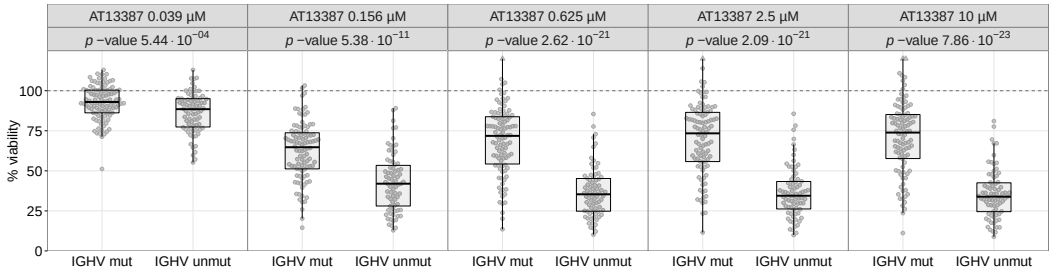


Figure 5.5 Drug response to HSP inhibitor stratified by IGHV status.

Effect of treatment with HSP inhibitor AT13387 is significantly different between M-CLL ($n = 98$) and U-CLL ($n = 74$) samples.

This drug inhibits MDM2, a negative regulator of p53 proteins, and therefore facilitates cell apoptosis. Additionally, in agreement with clinical observations, samples which are harboring neither *TP53* mutation nor 17p deletion were highly susceptible to treatment with all three drugs. Knowing the approximate mode of action of both fludarabine and nutlin-3, we were able to confirm the dependency of the response on the size of the *TP53* mutated clone (Figure 5.8).

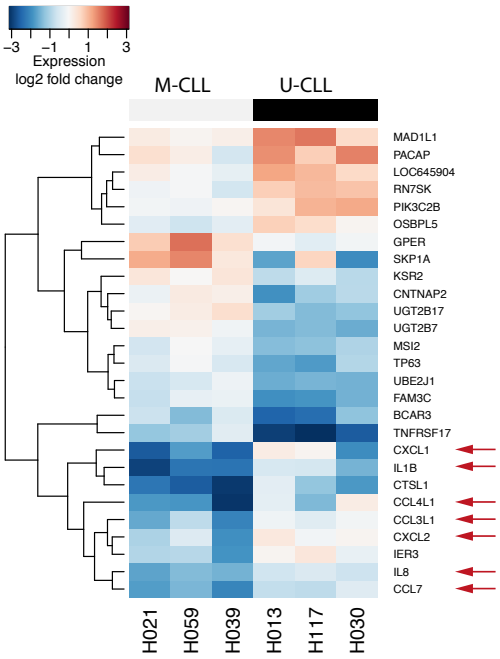


Figure 5.6 Changes in gene expression under treatment with AZD7762.

Gene expression profiling was performed on three patient samples within each of the two groups: M-CLL and U-CLL, before and after treatment with CHK inhibitor AZD7762. For each sample, we divided the gene expression level of the treated sample by the untreated one, and log2 transformed the obtained ratios. For each gene, we compared log2 fold changes between M-CLL and U-CLL using moderated t-test statistics. In order to increase power, we used independent filtering, which limited our analysis to 1000 most variable probes [102]. Multiple testing was controlled by applying Benjamini-Hochberg procedure (FDR 20%; $p = 0.0058$).
Figure shows down-regulation of cytokines and chemokines (arrows) exhibited by M-CLL samples, which suggests off-target effects of AZD7762 on the BCR pathway [103, 104].

Clinical observations similar to the described above were also reported in MCL [111]. Although we tested only 10 representative cases of the disease (seven of which acquired a mutation in TP53 gene), we found the *TP53* mutation status to be significantly associated with reduced response to both nutlin-3 and fludarabine (Figure 5.9).

5.1.3 Trisomy 12

Trisomy of chromosome 12 is an important independent predictor of patient's clinical outcome [9]. This one of the most frequently observed mutations in CLL is almost always characterized by a complete duplication of one chromosome [82]. Relatively good prognosis for patients harboring trisomy 12 is likely to be a reason why we are still lacking knowledge about its impact on the activity of the cell.

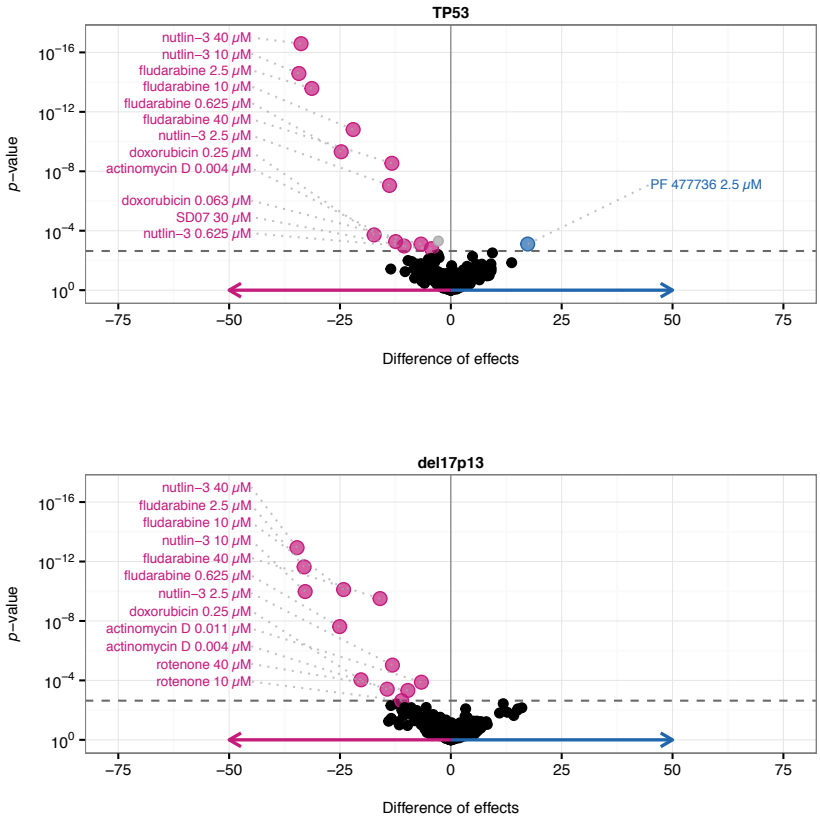


Figure 5.7 Mutation in gene *TP53* and deletion 17p13 are associated with resistance to chemotherapy and nutlin-3.

Volcano plots summarize all drug responses which were modulated by *TP53* mutation (top panel) and 17p13 deletion (bottom panel). Significance was tested using the Student t-test. p-values (y-axis) were adjusted using Benjamini-Hochberg procedure. 10% FDR was used. x-axis shows difference in the mean viability between wild type and mutated groups. Presence of either mutation (both often co-occur) make cells more resistant (pink) to chemotherapy and nutlin-3.

We observed that the response to 39 drugs was significantly modulated by the presence of trisomy 12 (Figure 5.10). In the majority of cases, the chromosomal aberration was sensitizing the cells to the drug (62%). We found trisomy 12 to be positively associated with the response to the inhibition of: SYK (tamatinib, PRT062607 HCl), BTK (ibrutinib, spebrutinib), PI3K (duvelisib, idelalisib), mTOR (everolimus), AKT (MK-2206) and MEK (selumetinib). Separate analysis confirmed that these findings were not confounded by IGHV mutation status. Figure 5.11 presents the effect of several inhibitors stratified by IGHV status based on raw viability data. It has been shown previously that patients with trisomy 12 uniquely respond to treatment with ibrutinib [112], and present elevated phosphorylation of ERK [113] which could be stimulated by BCR. All this suggests that trisomy 12 causes amplification of the BCR signaling, which affects its downstream pathways, such as PI3K, mTOR and MEK/ERK.

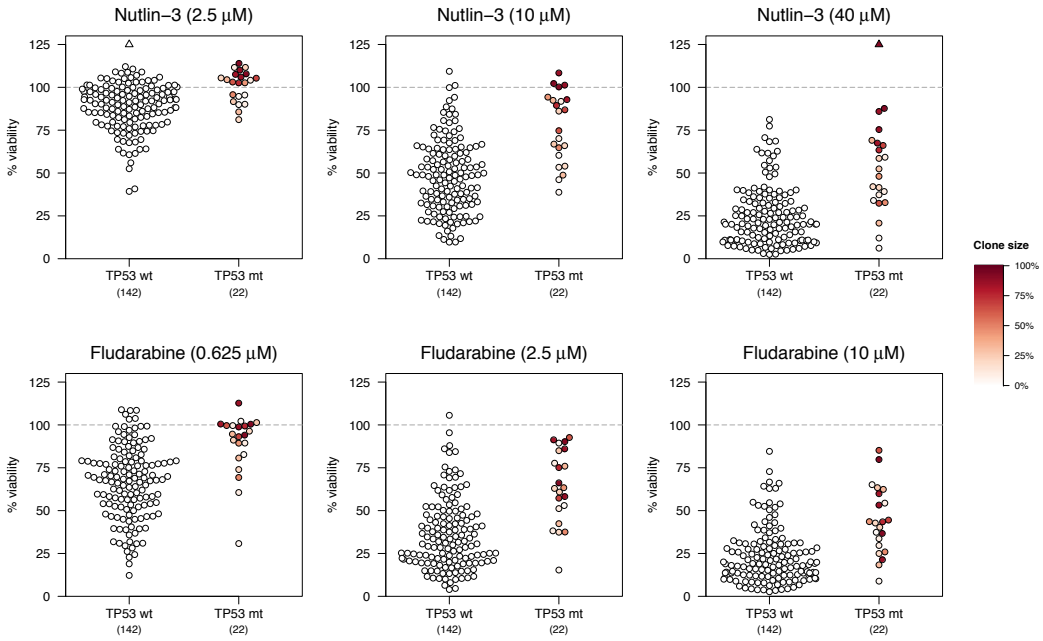


Figure 5.8 Drug response dependence on *TP53* mutated subclone.

Within *TP53* mutated samples ($n = 22$), the level of sensitivity to treatment with nutlin-3 and fludarabine is inversely proportional to the size of the mutated clone (Pearson correlation ranged from 0.53 to 0.76 for nutlin-3 and from 0.28 to 0.48 for fludarabine).

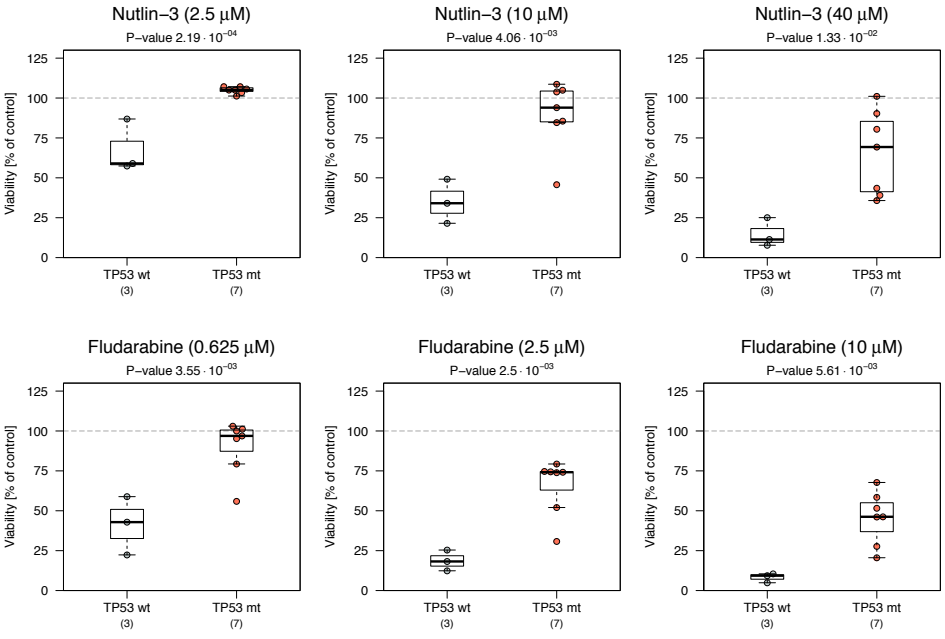


Figure 5.9 TP53 mutation status modulates drug response in MCL.

Viability of MCL cells under treatment with nutlin-3 and fludarabine was significantly influenced by the presence of mutation in the *TP53* gene. Strong effect could be observed even with small sample sizes ($n = 3$ and $n = 7$ for *TP53*wt and *TP53*mt, respectively).

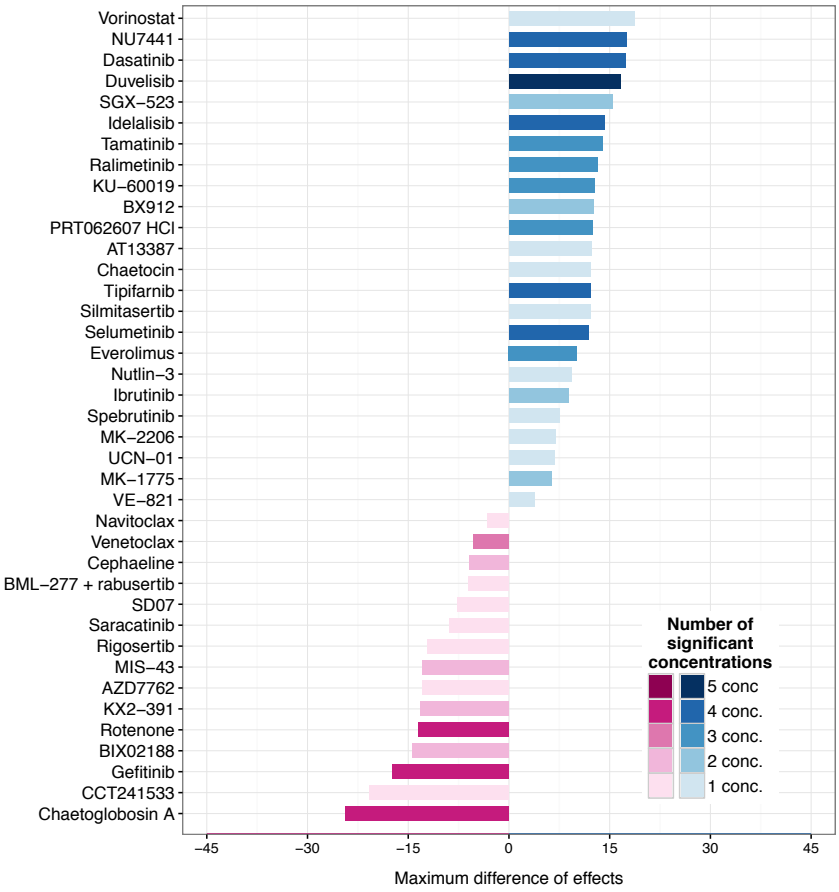


Figure 5.10 Drug responses associated with trisomy 12.

Bar plot shows drugs (y-axis) which produced significant difference in the mean viability (x-axis) between CLL without and with trisomy 12. Blue/pink color indicates that the cells with trisomy 12 were more sensitive/resistant to the drug treatment, respectively. The five concentrations of each drug were tested separately and the color intensity of the bars encodes for how many of them the difference in viability was statistically significant (FDR 10%). The length of a bar shows the greatest effect produced within the tested concentration steps.

Trisomy 12 exhibited a resistant phenotype to 15 of the tested compounds. These included two ROS inhibitors: SD07 and MIS-43, and two drugs acting through BCL2: venetoclax and navitoclax. Particularly interesting is the second pair. BCL-2 protein family is involved in a pro-survival pathway. A whole novel class of drugs, the so-called 'BH3-mimetics', like venetoclax, were developed to directly target the members of BCL2 and initiate programmed cell death [114]. It has been shown that these molecules are active no matter if the mutation of p53, an upstream initiator of apoptosis, is present [115]. Although we see the impact of trisomy 12 on the response to BH3-mimetics, the resistant phenotype is yet to be confirmed during clinical studies.

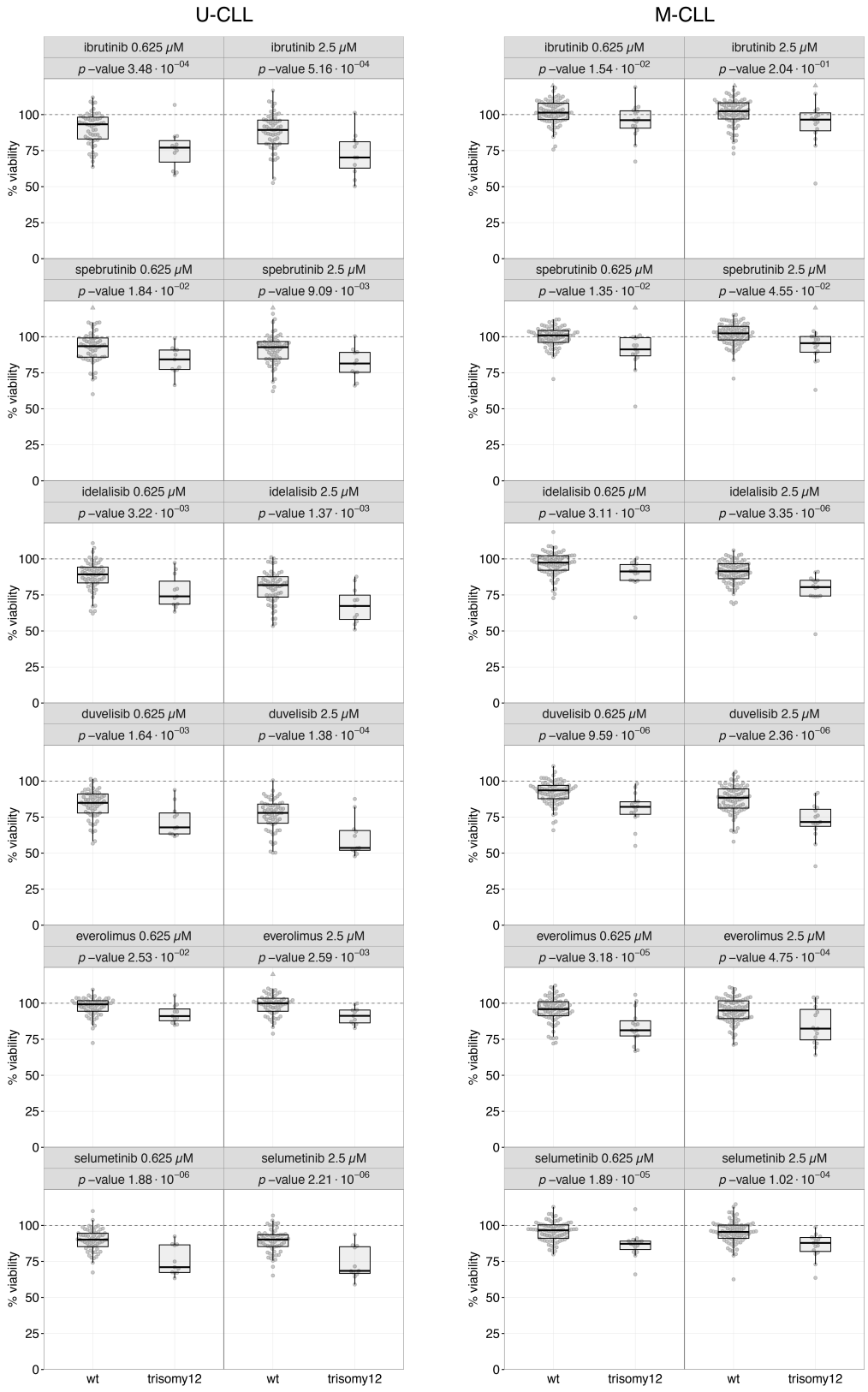


Figure 5.11 Drug response to BCR, mTOR and MEK inhibitors stratified by the presence of trisomy 12 and IGHV status.

Irrespective of IGHV mutation status, samples with trisomy 12 were more sensitive to treatment with BTK (ibrutinib, spebrutinib), PI3K (idelalisib, duvelisib), mTOR (everolimus) and MEK (selumetinib) inhibitors than samples with unmutated chromosome 12.

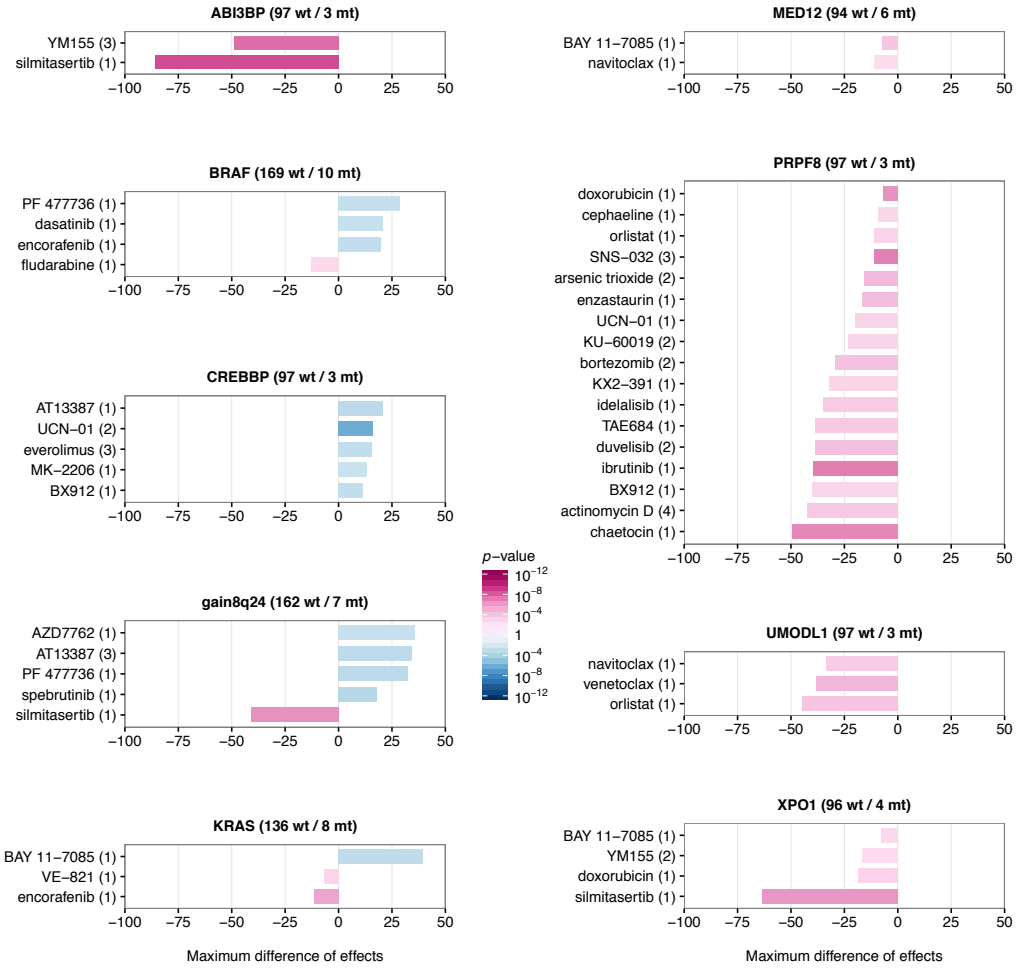


Figure 5.12 Rare mutations as modulators of drug response in CLL.

Each plots shows all drugs (y-axis), whose response was influenced by the presence of a given rare mutation (plot titles; alphabetical order). The effect of mutation (calculated as the difference in mean viabilities between wild type and mutated groups) is represented by the length of the bar. The color of the bar encodes the direction of the effect, where blue and pink indicates that samples harboring the mutation were more sensitive and resistant than wild type samples, respectively. Intensity of a bar indicates p-value. Numbers in brackets in y-axis labels indicate how many concentrations of a given drug showed significant association. Only the effects and the p-values of the drug concentration producing the strongest effect are shown. The brackets in titles indicate sizes of the two compared groups.

5.1.4 Rare mutational subclones

Almost 90% of CLL samples in our study harbor at least one mutation in the assessed genes. Although each has a potential influence on the heterogeneity of drug response, majority of them is very rare, which prevents the analysis from providing statistically significant results. We included rare mutations in our analysis only if they were present in at least 3 samples of the studied cohort. This allowed us to observe some trends for which only a limited evidence is present (Figure 5.12).

BRAF is an oncogene, which when mutated causes tumorgenic transformation through activation of MEK/ERK signaling [116]. We noticed encorafenib, a BRAF inhibitor targeting specific V600E mutation, to show favorable effects in the *BRAF* mutated samples. However, the very weak effect observed (consistent with other studies [117]) is due to the low frequency of V600E mutated subclone in CLL.

Mutations in oncogene KRAS occur frequently in many malignancies. We found the NF-κB inhibitor (BAY 11-7085) to be more effective in KRAS mutated samples. On the other hand, the same mutation caused resistance to encorafenib. This observation is in line with the counter-intuitive fact that MAPK/ERK signaling gets activated under treatment with BRAF inhibitor [118, 119].

CREBBP is a transcriptional coactivator which plays a role in hematopoiesis [120]. It has been shown that mutations in *CREBBP* influence epigenetics by causing reduced histon acetylation. We observed samples harboring the mutation in *CREBBP* to exhibit a phenotype sensitive to compounds targeting HSP90 (AT13387), PKC (UCN-01), mTOR (everolimus), and two components of AKT pathway (MK-2206, BX912).

PRPF8 gene encodes one of many proteins which are involved in splicing of pre-mRNA. We detected resistance of *PRPF8* mutated samples to many compounds. These included inhibitors of BCR pathway, such as ibrutinib, duvelisib, idelalisib and KX2-391, and compounds playing part in DNA damage response—actinomycin D and doxorubicin.

XPO1 gene is recurrently found mutated in CLL [121]. It encodes proteins which are responsible for nuclear export, whose activity is affected when *XPO1* is mutated [122]. Our *XPO1* mutated samples were highly resistant to silmitasertib, a small-molecule inhibitor of CK2.

The amplification of q24 fragment of chromosome 8 usually involves a protein coding gene called MYC and has been associated with a number of hematopoietic tumors. Specifically, it was associated with short time to first treatment clinical endpoint in CLL [123]. Similarly to the *XPO1* mutation explained above, samples with gain of 8q24 were resistant to silmitasertib. However, presumably the more important fact was that the mutated samples were sensitive to BTK inhibitor spebrutinib, and two CHK inhibitors (AZD7762 and PF 477736), which are acting on BCR pathway by off-target effects, as suggested in Section 5.1.1.

Yet another mutation which caused strong resistance to silmitasertib was affecting *ABI3BP* gene. The same effect was also noticed for survivin inhibitor YM155. Although both effects are strong, the exact impact of the mutation on cancer is still unknown.

Mutation in *MED12* reoccurs in CLL patients and so far was associated with markers of poor prognosis [124]. In our study it modulated the response to inhibitors targeting NF-κB (BAY 11-7085) and BLC2 (navitoclax) pathways .

The last mutation which showed interesting associations is affecting *UMODL1*. Although little is known about its function, we obtained consistent results indicating that BH3-mimetics drugs (venetoclax and navitoclax) are less active in the mutated samples.

5.2 Summary

Unsupervised univariate analysis identified a number of gene-drug associations in CLL. The main purposes of the study comprised:

- validation of clinically relevant biomarkers of drug response and patient outcome,
- determination of cellular pathways affected by a given mutation,
- identification of vulnerabilities of recurrent but rare mutations.

IGHV status was by far the greatest modulator of drug response. Sensitivity of U-CLL samples to treatment with targeted BCR inhibitors confirmed their stronger (than in M-CLL) dependency on BCR signaling pathway. The introduction of such compounds into standard clinical care is likely to improve the, so far inferior, prognosis of U-CLL patients. Mutation in *TP53* produced a phenotype resistant to chemotherapy and nutlin-3. Unexpectedly, the activity of many compounds was influenced by the presence of trisomy 12, a biomarker associated with moderate prognosis. Those drugs include a wide range of targeted inhibitors of BCR pathway components suggesting a role of trisomy 12 in the amplification of BCR signaling. We found a number of individual examples of drug-gene associations within the studied group of compounds, however, their true relevance could be only assessed by a more comprehensive screen including compounds covering a wider range of possible targets in a cell.

Modulators of drug response and clinical outcome

6

The only relevant test of the validity of a hypothesis is comparison of its predictions with experience.

— **Milton Friedman**
Essays in Positive Economics
University of Chicago Press (1953), 1970, 3–43

In the previous chapter we investigated the influence of a single molecular feature on multiple drug response profiles. This approach allows to identify deregulated pathways and to relate them to a specific feature of interest, for example gene mutation. However, the real power of multi-omics studies lays in the ability to consolidate the information from different biological layers (genomics, transcriptomics, etc.) in order to look at them simultaneously. The resulting big picture opens new opportunities to study codependency and interplay between those layers and to look at their combined impact on, for example, drug response or patient's clinical outcome. So far, multi-omics data, such as gene mutation, copy number estimates and gene expression, were integrated into machine learning techniques in order to understand the drug response of cell lines [45, 46]. Although such approach has great potential, a few suggestions on boosting the power of analysis were made. The two main concerns included: (i) sufficient sample size [125] and (ii) focus on one tumor entity at a time [45]. In our study we address both these issues by concentrating on a large cohort of CLL tumor samples. By using primary material we were able to integrate not only different omics but also patient clinical characteristics. To our best knowledge this is the first attempt of such a comprehensive analysis.

The work described in this chapter was done jointly by myself, Britta Velten and Sascha Dietrich.

6.1 Multivariate assessment of drug response determinants

The key challenge in identifying biomarkers of drug response lays in the multifactorial nature of the underlying biology. Multiple genes play a role, but each one is responsible only for a part of the variability, and their effects can be interdependent. Intratumor heterogeneity only adds to the complexity of this convoluted information.

We dissected the data dependencies by using multivariate linear regression modeling with L1-penalty (i.e., lasso regression). The analysis was performed in two steps. First, the regularization parameter which controls the strength of the penalty imposed on the predictor features was tuned by using 10-times repetition of 10-fold cross-validation implemented in function `cvr.glmnet` from R package *ipflasso* (version 0.1) [126]. Optimized parameter was further used for modeling performed by function `glmnet` from R package *glmnet* (version 2.0) [127]. The set of features which were used as predictors comprised: gene mutations, copy number estimates, IGHV status, DNA methylation, gene expression and demographic covariates (including sex and age). Depending on the question asked, we used a different subset of the above. However, the high dimensional nature of gene expression and DNA methylation data made the model prone to overfitting and hard to interpret. To overcome this, we performed principal component analysis based on 5000 of the most variable features from each experiment and selected only the first 20 principal components (PCs) to use in a model. All the other data types were binarized depending on the absence or presence of the given feature in a sample. All gene mutations and copy number estimates which were detected in less than 6 samples were removed from the analysis. Moreover, we also used summarized DNA methylation data [81]. LP, IP and HP clusters were marked as 0, 0.5 and 1, respectively. Overall, we included only features for which the data was complete in at least 90% of samples. All features were scaled to unit variance before supplying them to a model.

We modeled the response to chemotherapeutics and a selection of targeted drugs. Depending on compound activity we used responses to different concentration steps. For doxorubicin, fludarabin and nutlin-3 we used mean of all concentrations, whereas for ibrutinib, idelalisib, selumetinib, everolimus and PRT062607 we used mean over the two lowest concentrations. We asked two fundamental questions: (i) what are the modulators of drug response, and (ii) what data types explain best the variability of drug sensitivity profiles.

To answer the first question we supplied gene mutations, copy number estimates, IGHV status and methylation cluster information as features to the model. Those which played a significant role in influencing drug responses are shown in Figures 6.1 and 6.2.

We observed consistence in dependence on *TP53* mutation and 17p deletion in the group comprising chemotherapeutics. Response to fludarabine was almost equally strongly driven by those two factors. For nutlin-3, mutation in *TP53* gene was the leading predictor followed by deletion 11q and a much weaker contribution from deletion 17p. Multiple predictors of comparable effect were detected for doxorubicin. In

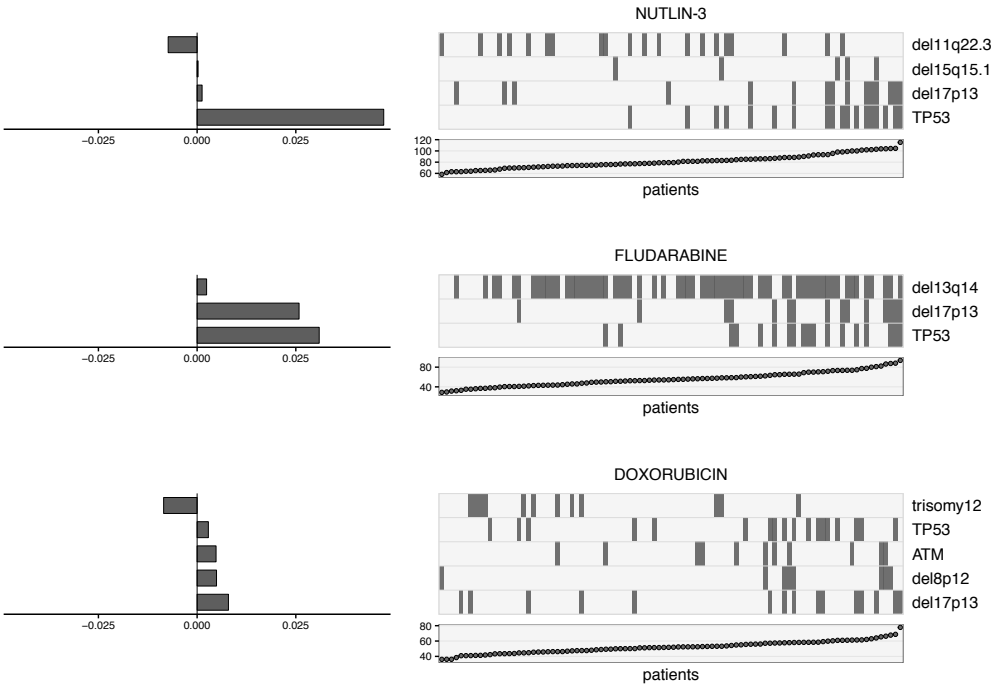


Figure 6.1 Modulators of drug sensitivity to chemotherapeutic agents.

Results of the multivariate analysis assessing the influence of different molecular features (rows) on the response to nutlin-3 and chemotherapeutics: fludarabine and doxorubicin. Patient samples (columns) are sorted according to decreasing viability induced by a given compound (bottom scatter plots). Horizontal bars on the left are model coefficients, which when negative indicate increased sensitivity to drug treatment in the presence of the feature.

this case the most important one was trisomy 12, which when present was sensitizing cells to the drug. Moreover, the results suggested a link between the mutation in *ATM* gene and response to doxorubicin.

Response to the kinase inhibitors was predominantly modulated by DNA methylation cluster, IGHV status and the presence of trisomy 12. The first two were indicators of cells' resistance to the drug, whereas trisomy 12 was making the cells more sensitive. It is worth noting that our model tends to pick correlated features randomly (mixing parameter alpha = 1), therefore the presence of both IGHV status and DNA methylation cluster simultaneously as key factors was unexpected. For BCR inhibitors, DNA methylation cluster was a dominant predictor of drug response. The more downstream in the BCR signaling pathway the target was, the more complex was the constructed model. This was observed for the MEK inhibitor selumetinib. Moreover, we found gain 8q24 as a modulator of response to kinase inhibitors mentioned above. The most modest model was determined for mTOR inhibitor everolimus. In this case both genetic features: trisomy 12 and mutation in *SF3B1* gene, are sensitizing cells to the drug.

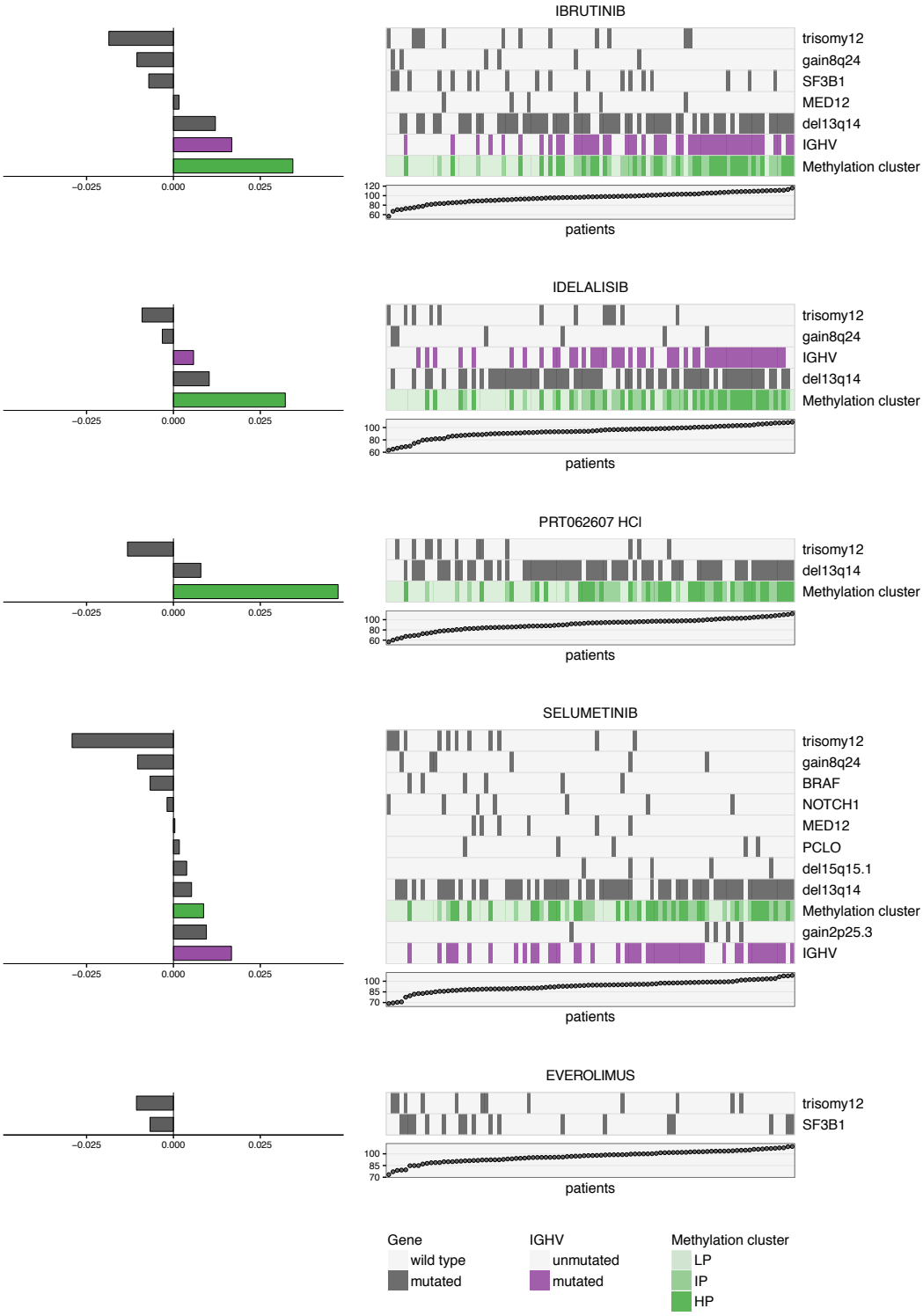


Figure 6.2 Modulators of drug sensitivity to targeted kinase inhibitors.

Results of the multivariate analysis assessing the influence of different molecular features (rows) on the response to targeted kinase inhibitors: ibrutinib (BTK), idelalisib (PI3K), PRT062607 HCl (SYK), selumetinib (MEK) and everolimus (mTOR). Patient samples (columns) are sorted according to decreasing viability induced by a given compound (bottom scatter plots). Horizontal bars on the left are model coefficients, which when negative indicate increased sensitivity to drug treatment in the presence of the feature.

To address the second question we took: genetic and demographic features, IGHV status, 20 PCs of DNA methylation and RNA expression as explained previously, and asked about their both separate and joint contribution to the drug response variability. Results are summarized in Figure 6.3. Again, sensitivity to drug treatment with chemotherapeutics was almost exclusively modulated by genetic factors (gene mutation and copy number changes), with small contributions from RNA expression for doxorubicin and nutlin-3. Variability of drug sensitivity profiles of BCR inhibitors was explained separately by RNA expression, IGHV status and DNA methylation to a comparable degree. A combination of these three did not improve prediction. Yet again, MEK inhibitor selumetinib was characterized by the most complex model. Here every tested feature except demographics contributed to the drug response, with the most significant one being RNA expression and genetics. Although in general a subset of tested features played an important role in modulating drug response profiles, that was not the case for doxorubicin and everolimus. Low prediction power in these two cases suggests that there might be some additional not-yet-known factors.

In summary, our analysis supported the finding of trisomy 12 being the fundamental modulator of response to BCR inhibitors in CLL. We comprehensively provided confirmation and extended the list of biomarkers influencing sensitivity profiles [87, 128, 39, 81]. We showed that depending on the compound, drug response can be predicted by different multi-omics datasets, and that multiple multi-omics datasets share a lot of redundancy.

6.2 Predictors of patient outcome

Primary tumor samples used in the study could be linked to specific patients and their clinical data. We extracted the available information on dates of treatments and death, if applicable. These were then used to create two clinical endpoints: from the sampling date to the date of the following treatment (time to treatment, TTT), and from the sampling date to the time of death (overall survival, OS), for which we conducted the survival analysis.

First we asked which of the most frequent aberrations in CLL have impact on both TTT and OS. We used log-rank test and obtained hazard ratios for the tested features (Figure 6.4). Results are consistent both between the two clinical endpoints and with previous studies [8, 7, 10, 39]. Presence of mutated IGHV was an indicator of better prognosis. On the other hand, presence of deletions 11q and 17p, mutations in *TP53*, *BRAF*, *SF3B1* genes significantly worsened the patient outcome. In this case, patients have a more aggressive disease—they need treatment earlier and also die sooner. Although we observed a negative tendency towards both OS and TTT for patients harboring trisomy 12 or mutation in *NOTCH1* gene, the difference was not statistically significant. It might be due to the low fraction of patients with these mutations present in the study.

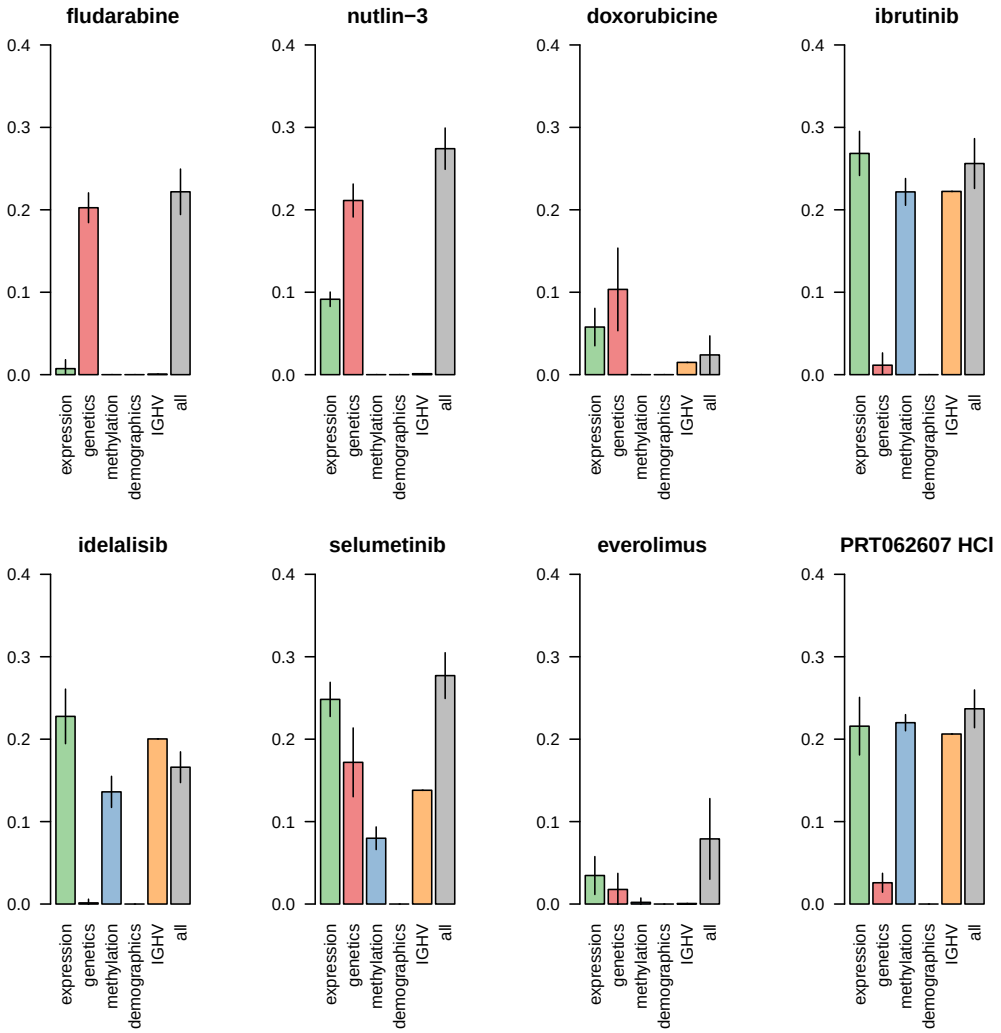


Figure 6.3 Predictors of drug response.

The figure shows how much variance in drug response of the samples (y-axis) to 8 compounds can be explained by different multi-omics data. In this analysis, for each data type and drug combination, 10-fold cross-validation was first conducted in order to optimize the penalty parameter using function `cv.glmnet` from *glmnet* R package. The obtained parameter was used in a model with 100 repetitions of cross-validation from which the mean and standard error of the explained fraction of variance were calculated. *Figure created by Britta Velten*

Encouraged by the results confirming established genetic biomarkers, we checked if *ex vivo* drug response could also predict patient survival (Figure 6.4). As previously, we focused on the two groups of compounds: chemotherapeutics (doxorubicin, fludarabine and nutlin-3) and targeted kinase inhibitors (ibrutinib, idelalisib, PRT062607 HCl, selumetinib and everolimus). Yet again, we took a mean viability over five and two lowest concentration steps for the first and the second group, respectively. Univariate Cox regression analysis evaluated hazard ratios for OS and TTT, which corresponded to good responders to the above drugs.

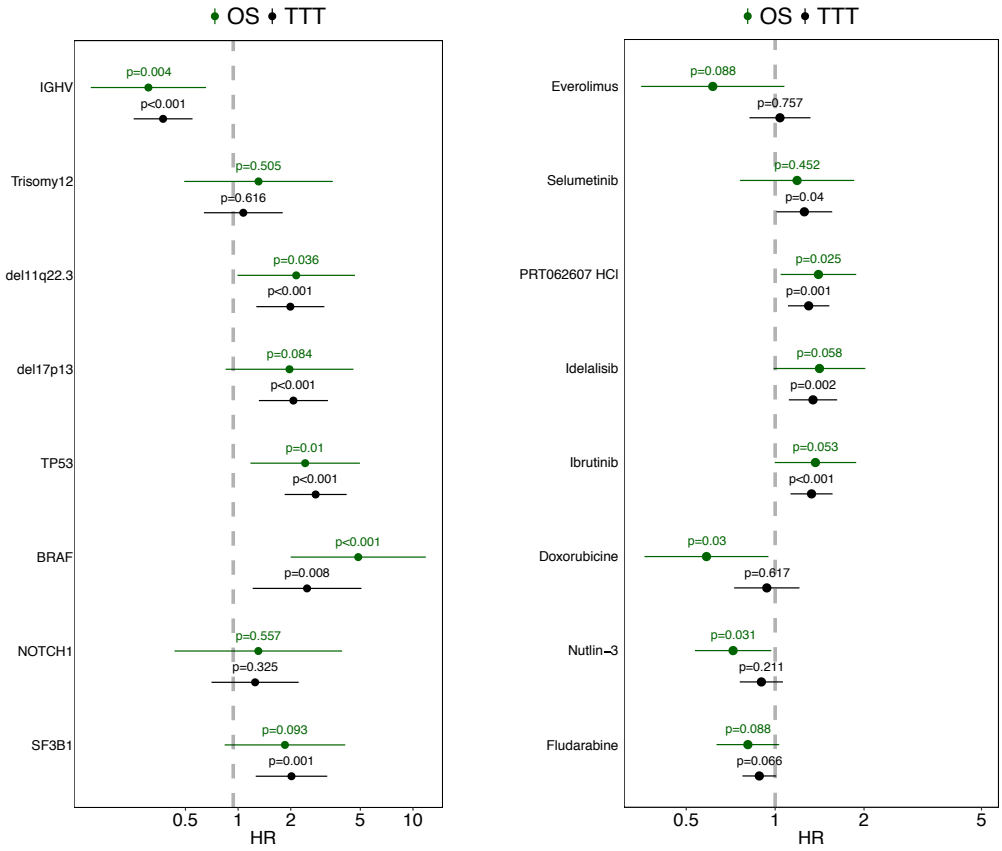


Figure 6.4 Biomarkers of patient outcomes.

Assesment of dependence of molecular factors (left panel) and eight drugs (ibrutinib, idelalisib, selumetinib, everolimus, PRT062607, fludarabine, doxorubicin and nutlin-3; right panel) on two clinical endpoints: from sampling date to treatment (TTT; $n = 162$) and from sampling date to death (OS; $n = 172$). Figure shows hazard ratios (HR) obtained from either log-rank test (left panel) or univariate Cox regression analysis (right panel). The horizontal bars correspond to 95% confidence intervals.
Figure created by Sascha Dietrich

Positive response to chemotherapy and nutlin-3 showed favorable OS. *TP53* mutation was partially responsible for this observation. We further divided samples into responder and non-responder groups using maximally selected rank statistics [129]. Survival analysis based on these groups showed even greater hazard ratio associated with drug response and clinical endpoints than the traditional biomarker *TP53* gene mutation (Figure 6.5). Additionally, we used multivariate Cox model which tested all frequent and established biomarkers (i.e. trisomy 12, deletions 11q and 17p, IGHV status and *TP53* mutation) together with response to doxorubicin. The unique information included in the drug response profile of doxorubicin significantly increased the prediction accuracy (see Appendix I).

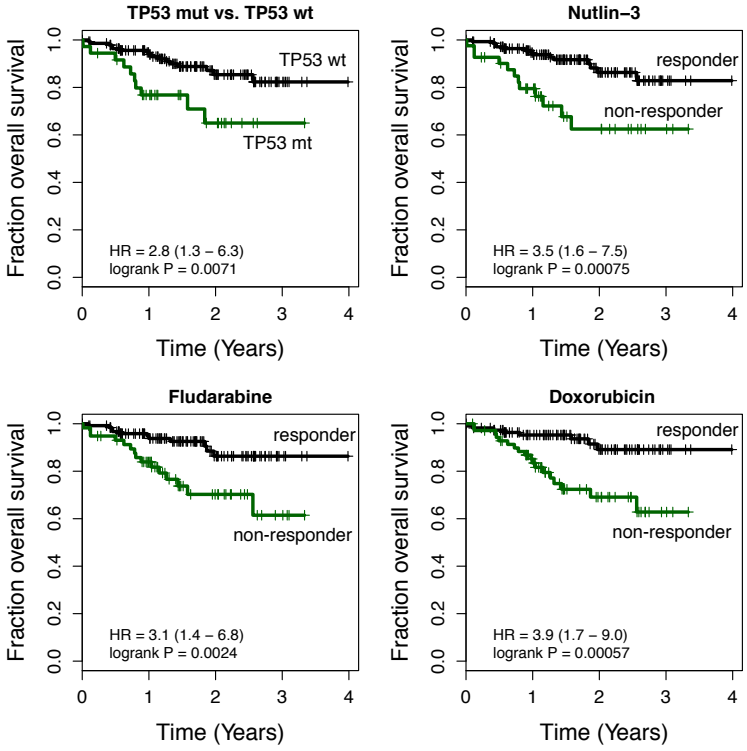


Figure 6.5 Chemotherapeutics and patient survival.

Kaplan-Meier estimates of overall survival according to *TP53* mutation status and response to nutlin-3, fludarabine and doxorubicin. Patients were assigned to the responder and non-responder groups based on maximally selected rank statistics [129].
Figure created by Sascha Dietrich

Better response to BCR inhibitors is connected with worse prognosis of both TTT and OS, with greater impact on the former. The closer the target to the BCR receptor, the better the observed correlation. Drug response measured *ex vivo* simulated the effects of the most important biomarker in CLL, meaning IGHV mutation status (Figure 6.6). To make this point even stronger, we first divided patients into responders and non-responders to ibrutinib and idelalisib, and only then calculated Kaplan-Meier estimates. Those BTK and PI3K inhibitors significantly separated the compared groups and gave comparable hazard ratios to the ones observed for IGHV status. Targeted treatment towards BCR is not yet a front line therapy in CLL. Unmutated IGHV status is still being connected to worse prognosis. We show that this can dramatically change if the therapies inhibit activity of the BCR signaling pathway. However, in order to estimate the benefits of treating U-CLL with targeted therapies aiming for BCR kinases we have to be patient. Survival studies need extensive follow up data to prove the beneficial effects.

In summary, we showed that *ex vivo* drug response can predict clinical endpoints as good as or sometimes even better than the well-established biomarkers. Our analysis suggests that the biomarker landscape should be updated with drug response

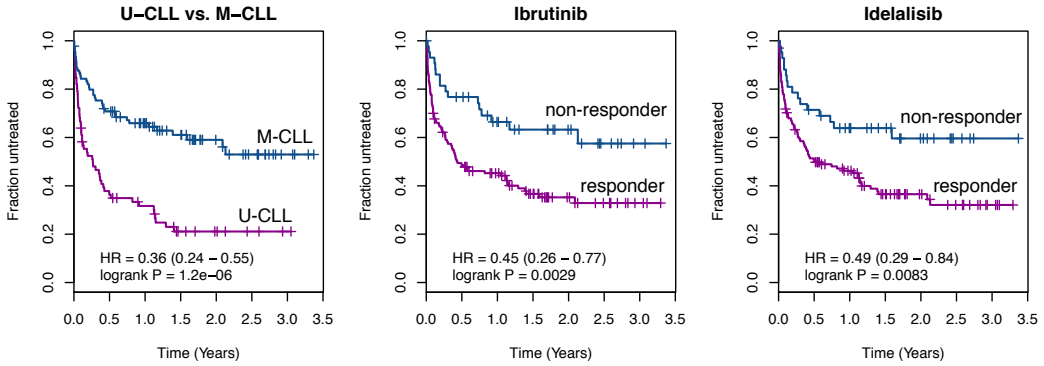


Figure 6.6 BCR inhibitors and patient survival.

Kaplan-Meier estimates of time between sampling and treatment according to IGHV mutation status and response to ibritinib and idelalisib. Patients were assigned to the responder and non-responder groups based on maximally selected rank statistics [129].
Figure created by Sascha Dietrich

signatures. Their continuous nature contains more information than, for example, single mutation, therefore we obtain much more fine-grained insight into the biology of the tumor. We are convinced that *ex vivo* drug screens have the potential to become a daily routine for medical professionals. A tool which will make them more confident in diagnostics and treatment decisions, as well as in prognostication.

Tools facilitating collaboration

7

Collaboration isn't about giving up our individuality; it's about realizing our greater potential.

— **Joseph Rain**
Facebook, 8 April 2016

Nowadays, close scientific collaboration can be improved by a variety of tools which facilitate both communication between scientists and formulation of hypotheses; by tools, which streamline the flow of ideas and the process of invention. Bioinformatic data analysis usually needs far-reaching understanding of the underlaying biology of the model object. In our project, besides bioinformatics competence, clinical expertise and assessment of therapeutic relevance were necessary. Jointly with Prof. Dr. med. Thorsten Zenz's group (whose many members are active medical doctors) from the NCT Heidelberg we brought these two parts together in a fruitful collaboration.

The old-fashioned, yet very popular method of data exploration is to create numerous figures or tables which satisfy each combination of possible features. In this way, one can produce advanced reports (using for example *ggplot2*, *lattice*, *ggvis* or *ReportingTools* in R) in a high-throughput fashion. However, this usually leads to either multi-page PDFs or long HTML documents, which then have to be comprehended in a rather inefficient manual fashion. Such approach was used in the project several times, but quickly became bothersome.

In last years significant development of the RStudio Shiny framework was made. This framework allows to build interactive web applications on top of the powerful R engine. Depending on the complexity of the programmed interface, the user can interact with the data by her- or himself, which stimulates the formulation of hypotheses that can be tested within the same application. Additionally, there are no requirements for the user to have any programming experience.

One of such applications which I implemented **VisualScreenExplorer** (VSE), turned out to be particularly helpful (Figure 7.1). It uses drug responses of the main screen, patient genetic information, and metadata as input. The output is a plot of responses of two drug-concentration pairs across patients. The user can modify the information which is displayed in the plot by:

1. choosing both the drug and the concentration for each axis separately,
2. color-coding the points according to the features of the samples: diagnosis, sex, IGHV status, and selected important genetic markers,
3. subsetting the samples (points) according to: diagnosis, IGHV status, and presence of mutation in *TP53* gene,
4. modifying the appearance of the plot, such as axis ranges and point transparency.

Hovering with the mouse cursor over a point in the plot shows the ID and characteristics of the given sample. Additionally, the current state of the app is encoded in an URL which can be shared with others, or saved for future reference.

By using the VSE app one can quickly visualize some of the conclusions of the study. In Figures 7.1 and 7.2A drugs which share the same target show almost identical profile of drug responses. Figure 7.2B shows the resistant phenotype of T-PLL in response to BCR inhibition in comparison to the sensitive phenotype exhibited by CLL. Figure 7.2C, however, shows sensitivity of *BRAF* mutated HCL samples to BRAF inhibitor encorafenib, whereas all the other diseases stay resistant to this drug.

In order to enable other scientists to construct and to test their own hypotheses using the valuable dataset we collected, an open-access version of the application will be distributed together with the published article. It will be available on a dedicated EMBL server under: `http://pace.embl.de/VisualScreenExplorer`.

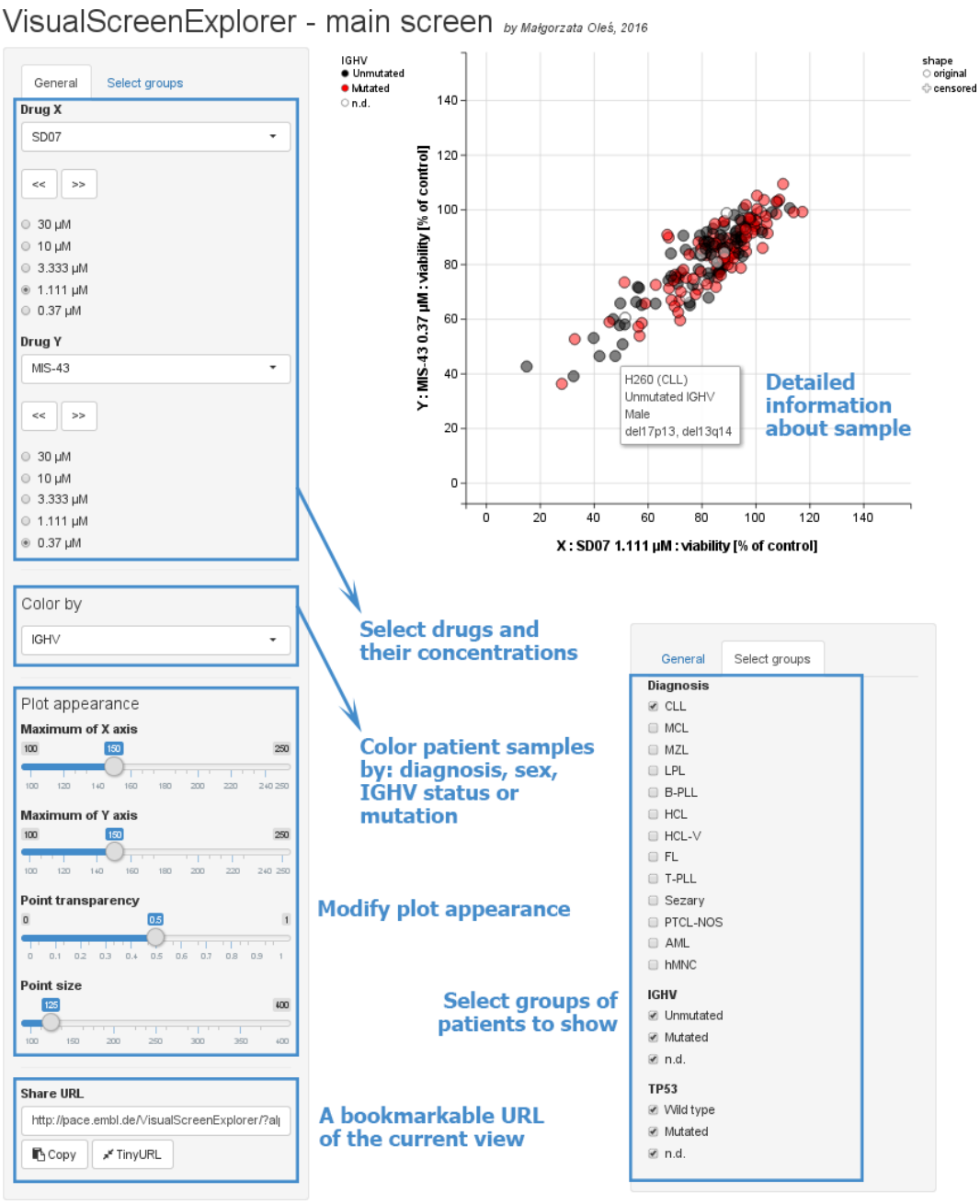


Figure 7.1 Interface of the VisualScreenExplorer.

VSE interface is split into two panels: the sidebar panel on the left, and the main panel on the right. The sidebar panel (gray area) consists of two tabs: "General" (displayed on the left) and "Select groups" (copied to the bottom-right corner for visualization purposes). The user can interact with different input widgets (described by blue labels within the figure) in order to modify the output plot located in the main panel. The plot is automatically updated each time the state of the widgets changes. The results can be further explored by hovering the mouse cursor over the points in the plot, which will show a label with detailed characteristics of the sample of interest.

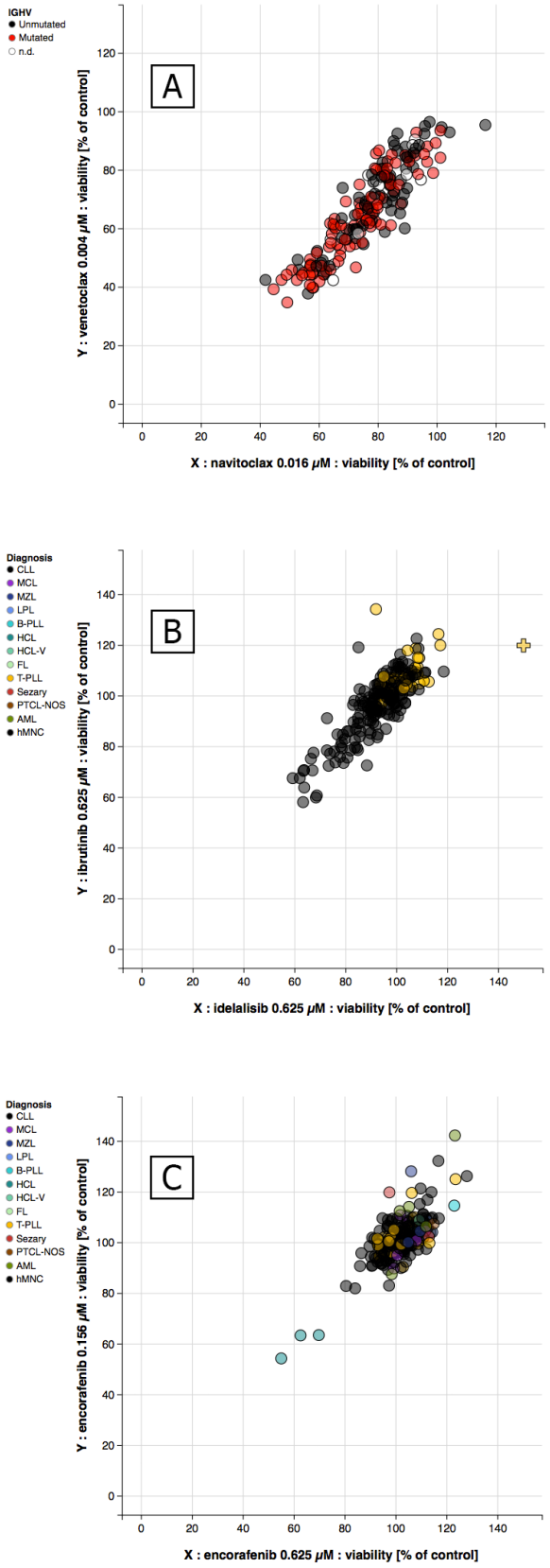


Figure 7.2 Selected findings visualized in VisualScreenExplorer.

The figure shows three example observations of the presented study, which could be noticed by using VSE. See the main text for details.

Conclusions and perspectives

8

The success of precision medicine will depend on our ability to translate large compedia of genomic, epigenomic, and proteomic data into clinically actionable predictions.

— **Costello et al.**
Nature 2014

The presented study is, to our knowledge, the first one which comprehensively and on a large-scale investigates drug sensitivities together with multi-omics characteristics of a panel comprising primary tumor samples. Throughout the study we paid particular attention to clinical relevance and applicability of the results. We closely examined reproducibility of our approach and confirmed its scalability both within same screening platform and between two different platforms.

We identified unexpected disease-specific sensitivities. Profoundly resistant to chemotherapy T-cell prolymphocytic leukemia exhibited sensitivity to SERCA inhibition. Mantle cell lymphoma, however, was sensitive to YM155 drug targeting survivin. This inhibitor of apoptosis has already shown anti-tumor activity in prostate and pancreatic cancers studied *in vitro* [130, 131]. Moreover, our approach allowed us to unravel mechanisms of action of the known biomarkers. For example, sensitivity of CLL samples harboring trisomy 12 to a wide range of BCR inhibitors suggested the amplification of BCR signaling in those cells. This observation is in line with previous studies which concluded short progression free survival [9], high pERK levels [113] and general overrepresentation of trisomy 12 in B-cell lymphomas. We discovered additional unknown targets of known drugs which offers drug repurposing opportunities. Specifically, CHK and HSP inhibitors exhibited both similar response profile to BCR inhibitors, and strong and selective activity in U-CLL, concluding that they target components of the BCR pathway. Furthermore, comparison of drug response profiles provided the identification of a new subgroup within M-CLL. It was

molecularly distinct by sensitivity to mTOR inhibition, suggesting a potential new targetable pathway. Last but not least, we show great potential in predicting clinical outcome from *ex vivo* drug profiling. We believe that together with thoughtful selection of compounds and standardized screening techniques such approach can benefit the care of patients.

Our methodology was independent of the ambiguous interpretation of the dose-response curve parameters. This made us not only bypass the heated discussion on that topic, but also allowed for robustness which stands behind raw viability measurements. Moreover, we avoided parametric models in the analysis and removed unnecessary assumptions in order to see the true complexity of the studied system.

The understanding of the exact influence of all discovered key cancer biomarkers on the deregulation of signaling pathways in a cell is still full of unknowns. Further dissection of mechanisms of action with *ex vivo* drug screens and multi-omics technologies is pending. However, in order to make it informative, the field of drug discovery has to come up with new targeted compounds. Only after specific treatment approaches can be guided by viability assays, introduction of such screens into standard clinical care will be reasonable. Nevertheless, in order to make it available for all patients suffering from cancer, the method has to be optimized and standardized, and the tested compound group has to be selected in such a way that it provides maximum information at a minimum library size. Needless to say, this assay should be distributed in an easy-to-use testing kit. Last but not least, *ex vivo* drug screens have the potential of revolutionizing and accelerating clinical trial studies, which are currently designed to benefit larger groups of patients rather than an individual. Since so many biomarkers play a role in drug response, specific groups on which the testing could be done will significantly shrink, taking away the statistical power of the analysis. Although the vision of redesigning the process of clinical trials is exciting, one has to keep in mind that short and long term side effects of the drugs can not be assessed based on viability assays. These effects need additional testing and careful evaluation.

Author's publications

A

Manuscript which includes work presented in this thesis:

1 S Dietrich*, **M Oleś***, L Sellner*, S Anders, B Wu, J Lu, B Velten, J Hüllein, AK Oleś, M Słabicki, M Lukas, T Walther, CC Oakes, S Rabe, V Kim, M Sill, A Benner, A Jauch, LA Sutton, E Young, R Rosenquist, X Liu, A Jethwa, KS Lee, J Lewis, K Putzker, C Lutz, D Rossi, A Mokhir, K Zirlik, M Herling, F Nguyen-Khac, C Plass, EI Andersson, Satu Mustjoki, C von Kalle, AD Ho, M Hensel, J Dürig, I Ringshausen, M Zapatka, W Huber, T Zenz. Drug Perturbation-Based Molecular Stratification of Blood Cancer (manuscript in preparation)

Additionally, the author of this dissertation was involved in multiple collaborative projects during her PhD studies. Two of them have already been published.

2 S Dietrich, J Hüllein, SC Lee, B Hutter, D Gonzalez, S Jayne, MJ Dyer, **M Oleś**, M Else, X Liu, M Słabicki, B Wu, X Troussard, J Dürig, M Andrulis, C Dearden, C von Kalle, M Granzow, A Jauch, S Fröhling, W Huber, M Meggendorfer, T Haferlach, AD Ho, D Richter, B Brors, H Glimm, E Matutes, O Abdel Wahab, T Zenz (2015) Recurrent CDKN1B (p27) mutations in hairy cell leukemia. *Blood*, **126**, 1005–1008.

3 M Słabicki, K S Lee, A Jethwa, L Sellner, F Sacco, T Walther, J Hüllein, S Dietrich, B Wu, D B Lipka, C C Oakes, S Mamidi, B Pyrzyńska, M Winiarska, **M Oleś**, M Seifert, C Plass, M Kirschfink, M Boettcher, J Gołąb, W Huber, S Fröhling and T Zenz (2016) Dissection of CD20 regulation in lymphoma using RNAi. *Leukemia*

4 A Pekowska, B Klaus, W Xiang, J Severino, N Daigle, F Klein, **M Oleś**, J Ellenberg, L Steinmetz, P Bertone, W Huber. Gain of chromatin loops marks the exit from naive pluripotency (manuscript in review)

5 EI Andersson, S Pützer, O Dufva, B Yadav, S Khan, L He, J Tang, L Sellner, A Schrader, G Crispatzu, **M Oleś**, H Zhang, S Adnan, S Lagström, D Bellanger, JP Mpindi, S Eldfors, T Pemovska, P Pietarinen, A Lauhio, K Tomska, CC Mateos, E Faber, S Koschmider, T Brümmendorf, S Kytölä, P Ellonen, O Kallioniemi, C Heckman, K Wennerberg, K Porkka, W Ding, MH Stern, W Huber, S Anders, T Aittokallio, T Zenz, M Herling, S Mustjoki. Discovery of Novel Drug Sensitivities in T-prolymphocytic leukemia (T-PLL) by High-Throughput Ex Vivo Drug Testing and Mutation Profiling (manuscript submitted)

6 M Lukas, B Velten, ..., **M Oleś** et al. Assessment of drug combination effects in Chronic Lymphocytic Leukemia by *ex vivo* pharmacogenomics screening (manuscript in preparation)

* equal contribution

Patient clinical data

B

Characteristics of samples used in the pilot screen (P) and in the main screen (M); n.d. = no data available.

Patient ID	Diagnosis	Age	Sex	IGHV	Treated	Alive	Screen
H001	hMNC	n.d.	n.d.	n.d.	n.d.	n.d.	P,M
H002	hMNC	n.d.	n.d.	n.d.	n.d.	n.d.	P,M
H003	hMNC	n.d.	n.d.	n.d.	n.d.	n.d.	P,M
H004	CLL	68	female	unmutated	yes	yes	P
H005	CLL	75	male	mutated	yes	yes	P,M
H006	CLL	81	male	mutated	no	yes	P
H007	CLL	79	female	unmutated	yes	yes	P
H008	CLL	56	male	unmutated	no	yes	P
H009	B-PLL	57	male	unmutated	yes	yes	P,M
H010	CLL	73	female	unmutated	no	yes	P,M
H011	CLL	73	female	mutated	no	yes	P,M
H012	CLL	62	female	unmutated	yes	no	P,M
H013	CLL	77	male	unmutated	yes	no	P,M
H014	CLL	86	female	unmutated	yes	no	P,M
H015	CLL	62	female	unmutated	no	yes	P,M
H016	CLL	55	male	mutated	no	yes	P,M
H017	CLL	56	male	unmutated	no	no	P,M
H018	CLL	50	female	mutated	no	yes	P
H019	CLL	70	female	unmutated	yes	no	M
H020	CLL	64	male	mutated	no	yes	P,M
H021	CLL	50	male	mutated	no	yes	P,M
H022	CLL	59	male	mutated	no	yes	P
H023	CLL	71	female	unmutated	yes	no	P,M
H024	CLL	55	male	mutated	no	yes	P
H025	T-PLL	73	male	n.d.	yes	no	P,M
H026	LPL	59	male	mutated	no	yes	P,M
H027	CLL	58	male	unmutated	no	yes	P,M
H028	CLL	73	female	mutated	no	no	P,M
H029	CLL	75	female	mutated	yes	yes	P,M
H030	CLL	53	male	unmutated	no	yes	P,M

H031	CLL	62	female	mutated	no	yes	P,M
H032	CLL	67	male	unmutated	yes	no	M
H033	CLL	63	female	mutated	no	yes	P,M
H035	CLL	79	female	mutated	yes	yes	P,M
H036	CLL	75	female	mutated	no	yes	P,M
H037	CLL	71	male	mutated	no	yes	M
H038	CLL	74	male	mutated	no	yes	M
H039	CLL	55	female	mutated	no	yes	P,M
H040	CLL	84	female	mutated	no	no	P,M
H041	CLL	76	male	mutated	no	yes	P,M
H042	CLL	72	female	unmutated	yes	no	P,M
H043	CLL	44	female	unmutated	yes	yes	P,M
H044	CLL	61	male	unmutated	yes	yes	P,M
H045	CLL	91	male	unmutated	yes	no	P,M
H046	CLL	88	male	mutated	no	yes	P,M
H047	CLL	69	male	unmutated	yes	no	P,M
H048	CLL	65	female	unmutated	yes	yes	M
H049	CLL	58	male	mutated	no	yes	M
H050	CLL	63	female	mutated	no	yes	M
H051	CLL	79	female	unmutated	yes	no	P,M
H052	CLL	82	male	mutated	no	yes	P
H053	CLL	83	female	mutated	no	yes	P,M
H054	CLL	50	female	mutated	no	yes	P,M
H055	CLL	65	male	mutated	no	yes	P,M
H056	CLL	83	male	mutated	no	yes	P,M
H057	CLL	67	male	mutated	no	yes	P,M
H058	CLL	75	female	mutated	no	no	P,M
H059	CLL	55	male	mutated	no	yes	P,M
H060	CLL	75	male	unmutated	no	yes	P,M
H062	CLL	53	male	mutated	no	yes	M
H063	CLL	49	female	mutated	no	yes	P,M
H064	CLL	71	male	n.d.	yes	yes	P,M
H065	CLL	77	female	unmutated	yes	no	P,M
H066	CLL	47	male	unmutated	yes	no	P,M
H067	CLL	77	female	mutated	no	yes	M
H068	CLL	63	male	unmutated	no	yes	P
H069	CLL	77	female	unmutated	yes	no	P,M
H070	CLL	71	male	n.d.	no	yes	M
H071	FL	60	male	mutated	no	yes	M
H072	CLL	58	male	unmutated	no	yes	P,M
H073	CLL	65	male	mutated	yes	yes	P,M
H074	CLL	62	male	unmutated	no	yes	P

H075	CLL	72	male	unmutated	no	yes	P
H076	MCL	67	male	mutated	no	yes	P,M
H077	CLL	70	female	unmutated	no	yes	P,M
H078	CLL	68	male	unmutated	yes	yes	P,M
H079	CLL	48	male	unmutated	no	yes	P,M
H080	CLL	82	male	unmutated	yes	yes	P,M
H081	CLL	64	female	mutated	no	yes	P,M
H082	CLL	82	male	mutated	no	yes	P,M
H083	CLL	69	male	n.d.	no	yes	P,M
H084	CLL	88	male	mutated	no	yes	M
H085	CLL	62	male	unmutated	yes	yes	P
H086	T-PLL	64	male	n.d.	no	yes	P,M
H087	CLL	70	male	unmutated	yes	yes	P
H088	CLL	60	female	mutated	no	yes	P,M
H089	CLL	55	female	mutated	no	yes	P,M
H090	CLL	70	female	mutated	yes	yes	P,M
H091	CLL	69	female	mutated	no	yes	P
H092	MZL	82	male	mutated	no	yes	M
H093	CLL	76	female	unmutated	no	yes	P,M
H094	CLL	46	male	mutated	no	yes	P,M
H095	CLL	53	female	unmutated	no	yes	P,M
H096	CLL	62	female	n.d.	no	yes	P,M
H097	CLL	59	male	unmutated	yes	yes	P
H098	MCL	79	male	unmutated	no	no	P,M
H099	CLL	54	female	mutated	no	yes	M
H100	CLL	74	male	mutated	no	yes	P,M
H101	CLL	73	female	mutated	no	yes	P,M
H102	CLL	78	female	unmutated	no	yes	P,M
H103	CLL	71	male	mutated	no	yes	M
H104	CLL	79	male	unmutated	no	yes	P,M
H105	CLL	49	male	mutated	no	yes	P,M
H106	CLL	71	male	mutated	no	yes	M
H107	CLL	43	male	unmutated	no	yes	P,M
H108	CLL	57	male	mutated	no	yes	P,M
H109	CLL	85	male	unmutated	no	yes	M
H110	CLL	66	male	mutated	no	yes	M
H111	CLL	55	male	unmutated	yes	no	M
H112	CLL	65	male	mutated	no	yes	P
H113	CLL	70	male	mutated	no	yes	P,M
H114	CLL	43	female	unmutated	yes	no	P
H115	CLL	72	male	mutated	no	no	P,M
H116	CLL	54	male	unmutated	yes	yes	P

H117	CLL	51	female	unmutated	yes	yes	P,M
H118	CLL	49	male	mutated	yes	yes	P,M
H119	CLL	60	male	unmutated	no	yes	P
H120	MZL	74	female	mutated	no	yes	P,M
H121	CLL	57	male	mutated	yes	yes	P
H122	LPL	53	male	mutated	no	yes	P,M
H123	CLL	61	male	unmutated	no	yes	P
H124	CLL	66	male	n.d.	no	yes	P
H125	CLL	71	male	n.d.	yes	yes	P
H126	T-PLL	68	male	n.d.	no	yes	P,M
H127	T-PLL	69	female	n.d.	no	no	P,M
H128	T-PLL	78	female	n.d.	no	yes	P,M
H129	CLL	68	female	n.d.	n.d.	n.d.	P
H130	CLL	72	female	n.d.	n.d.	n.d.	P
H131	CLL	53	female	n.d.	n.d.	n.d.	P
H132	CLL	75	male	n.d.	n.d.	n.d.	P
H133	CLL	69	male	n.d.	no	yes	M
H134	Sezary	67	male	n.d.	yes	yes	M
H135	CLL	76	female	mutated	yes	no	M
H136	CLL	66	male	unmutated	yes	yes	M
H137	CLL	53	male	mutated	no	yes	M
H140	HCL	55	female	n.d.	no	yes	M
H141	MCL	46	female	n.d.	yes	no	M
H142	MCL	67	male	n.d.	no	yes	M
H143	HCL-V	n.d.	male	n.d.	n.d.	n.d.	M
H144	MCL	67	male	n.d.	yes	no	M
H145	HCL	45	male	n.d.	no	yes	M
H146	MCL	n.d.	male	n.d.	n.d.	no	M
H147	MCL	59	male	n.d.	no	yes	M
H148	CLL	34	female	unmutated	yes	no	M
H149	T-PLL	83	male	n.d.	no	no	M
H150	T-PLL	75	female	n.d.	no	yes	M
H151	T-PLL	63	female	n.d.	no	no	M
H152	T-PLL	42	female	n.d.	no	no	M
H153	T-PLL	52	female	n.d.	yes	no	M
H154	T-PLL	65	female	n.d.	no	no	M
H155	T-PLL	74	female	n.d.	yes	no	M
H156	B-PLL	61	female	n.d.	no	no	M
H157	T-PLL	70	female	n.d.	no	no	M
H158	MZL	60	male	n.d.	yes	yes	M
H159	LPL	58	female	n.d.	no	yes	M
H160	LPL	62	male	n.d.	yes	yes	M

H161	T-PLL	83	male	n.d.	no	yes	M
H162	MZL	50	female	n.d.	no	yes	M
H163	CLL	65	male	mutated	no	yes	M
H164	CLL	73	female	unmutated	no	yes	M
H165	CLL	58	female	unmutated	no	yes	M
H166	CLL	63	female	unmutated	no	yes	M
H167	CLL	64	female	unmutated	no	yes	M
H168	CLL	58	male	n.d.	yes	yes	M
H169	CLL	42	female	mutated	no	yes	M
H170	CLL	75	female	mutated	yes	yes	M
H171	CLL	73	male	unmutated	yes	yes	M
H172	T-PLL	45	male	n.d.	no	yes	M
H173	CLL	74	female	mutated	yes	yes	M
H174	CLL	64	female	unmutated	yes	yes	M
H175	CLL	62	male	unmutated	yes	yes	M
H176	CLL	70	male	mutated	no	yes	M
H177	CLL	70	male	unmutated	no	yes	M
H178	CLL	71	male	unmutated	yes	yes	M
H179	CLL	61	male	mutated	no	yes	M
H180	CLL	86	male	unmutated	no	yes	M
H181	CLL	76	female	mutated	no	yes	M
H182	CLL	72	female	mutated	no	yes	M
H183	CLL	70	male	unmutated	no	yes	M
H184	CLL	75	male	mutated	no	no	M
H185	CLL	87	female	mutated	no	no	M
H186	CLL	73	female	mutated	no	yes	M
H187	CLL	60	male	unmutated	no	yes	M
H188	T-PLL	71	male	n.d.	no	no	M
H189	T-PLL	71	male	n.d.	yes	yes	M
H190	MCL	66	male	n.d.	no	yes	M
H191	CLL	39	male	n.d.	yes	yes	M
H192	CLL	72	female	mutated	no	yes	M
H193	CLL	81	male	mutated	no	yes	M
H194	CLL	76	male	mutated	no	yes	M
H195	T-PLL	76	male	n.d.	no	yes	M
H196	CLL	85	male	mutated	no	yes	M
H197	CLL	74	female	mutated	yes	yes	M
H198	CLL	58	female	mutated	no	yes	M
H199	CLL	83	male	mutated	no	yes	M
H200	CLL	83	female	unmutated	yes	yes	M
H201	CLL	72	female	n.d.	no	yes	M
H202	CLL	80	male	mutated	n.d.	yes	M

H203	CLL	83	female	mutated	no	yes	M
H204	CLL	66	female	n.d.	no	yes	M
H205	CLL	66	female	unmutated	no	yes	M
H206	CLL	69	male	mutated	no	yes	M
H207	CLL	65	male	mutated	no	yes	M
H208	CLL	73	male	mutated	no	yes	M
H209	CLL	72	female	mutated	yes	yes	M
H210	CLL	73	female	mutated	no	yes	M
H211	CLL	51	male	unmutated	no	yes	M
H212	CLL	74	male	mutated	no	yes	M
H213	CLL	63	male	mutated	no	yes	M
H214	CLL	65	male	unmutated	no	yes	M
H215	CLL	47	male	unmutated	no	yes	M
H216	CLL	48	female	mutated	no	yes	M
H217	CLL	65	male	mutated	no	yes	M
H218	CLL	50	male	n.d.	yes	yes	M
H219	CLL	74	female	mutated	no	yes	M
H220	CLL	75	female	mutated	no	yes	M
H221	CLL	55	male	mutated	no	yes	M
H222	CLL	56	male	mutated	no	yes	M
H223	CLL	47	female	mutated	no	yes	M
H224	CLL	47	male	unmutated	no	yes	M
H225	CLL	47	female	mutated	no	yes	M
H226	HCL	64	male	n.d.	no	yes	M
H227	MCL	64	male	n.d.	no	no	M
H228	CLL	65	male	unmutated	no	yes	M
H229	CLL	75	female	mutated	yes	no	M
H230	CLL	71	male	unmutated	yes	no	M
H231	CLL	47	male	unmutated	no	yes	M
H232	T-PLL	62	male	n.d.	no	yes	M
H233	CLL	55	male	unmutated	no	yes	M
H234	CLL	68	male	unmutated	no	yes	M
H235	CLL	73	male	mutated	no	yes	M
H236	CLL	67	male	mutated	no	yes	M
H237	CLL	73	female	mutated	no	yes	M
H238	CLL	75	male	unmutated	no	no	M
H239	CLL	70	female	unmutated	no	yes	M
H240	CLL	83	male	mutated	no	yes	M
H241	Sezary	59	male	n.d.	no	yes	M
H242	CLL	49	male	unmutated	no	yes	M
H243	CLL	80	male	unmutated	no	yes	M
H244	B-PLL	80	male	n.d.	n.d.	yes	M

H245	PTCL-NOS	80	male	n.d.	yes	yes	M
H246	CLL	75	male	unmutated	no	yes	M
H247	CLL	47	female	mutated	no	yes	M
H248	CLL	63	female	mutated	no	yes	M
H249	CLL	83	male	unmutated	no	yes	M
H250	CLL	52	male	unmutated	no	yes	M
H251	HCL-V	73	male	n.d.	no	yes	M
H252	CLL	70	male	unmutated	no	yes	M
H253	T-PLL	57	female	n.d.	no	yes	M
H254	CLL	75	male	mutated	no	yes	M
H255	CLL	67	male	unmutated	yes	yes	M
H256	CLL	63	female	n.d.	yes	yes	M
H257	CLL	66	female	unmutated	no	yes	M
H258	CLL	65	male	mutated	no	yes	M
H259	CLL	60	male	unmutated	yes	yes	M
H260	CLL	63	male	unmutated	yes	yes	M
H261	T-PLL	49	female	n.d.	no	yes	M
H262	T-PLL	86	male	n.d.	no	n.d.	M
H263	T-PLL	77	male	n.d.	n.d.	n.d.	M
H264	CLL	77	male	mutated	yes	yes	M
H265	CLL	59	male	unmutated	yes	yes	M
H266	CLL	74	male	mutated	yes	yes	M
H267	T-PLL	68	male	n.d.	no	no	M
H268	CLL	83	male	n.d.	no	yes	M
H269	MCL	46	female	n.d.	no	no	M
H270	CLL	67	female	mutated	no	yes	M
H271	CLL	65	male	mutated	no	yes	M
H272	CLL	56	male	unmutated	yes	yes	M
H273	MZL	77	male	n.d.	yes	yes	M
H274	AML	61	female	n.d.	yes	yes	M
H275	AML	77	female	n.d.	no	no	M
H276	AML	62	male	n.d.	no	yes	M
H277	AML	83	male	n.d.	no	no	M
H278	AML	75	male	n.d.	yes	yes	M
H279	T-PLL	58	male	n.d.	no	yes	M
H280	MZL	63	male	n.d.	no	yes	M

Characteristics of drugs used in the drug screens

C

Drug ID	Name	Target
D_001	navitoclax	BCL2, BCL-XL, BCL-W
D_002	ibrutinib	BTK
D_003	idelalisib	PI3K delta
D_004	SNS-032	CDK2/7/9
D_005	olaparib	PARP1/2
D_006	fludarabine	Purine analogue
D_007	vorinostat	HDAC I/IIa/IIb/IV
D_008	bortezomib	Proteasome
D_009	entinostat	HDAC I/III
D_010	nutlin-3	MDM2
D_011	enzastaurin	PKC
D_012	selumetinib	MEK1/2
D_013	afatinib	EGFR, ERBB2
D_014	deforolimus	mTOR
D_015	MK-1775	WEE1
D_016	vismodegib	SMO
D_017	AT13387	HSP90
D_018	RO4929097	Gamma-secretase
D_019	XAV-939	WNT
D_020	AZD7762	CHK1/2
D_021	rigosertib	PLK
D_022	SP600125	JNK
D_023	ralimetinib	p38 MAPK
D_024	SGI-1776	PIM
D_025	NSC 74859	STAT
D_026	tozasertib	Aurora A/B/C, FLT3, ABL1, JAK2
D_027	TAME	Anaphase-Promoting Complex (APC)
D_028	galunisertib	TGF-beta
D_029	TAE684	ALK
D_030	MK-2206	AKT1/2 (PKB)
D_031	pomalidomide	-
D_032	NU7441	DNAPK
D_033	tipifarnib	Farnesyltransferase (FNTA)
D_034	chaetocin	Lysine-specific histone methyltransferase
D_035	saracatinib	SRC, ABL1

D_036	tamatinib	SYK
D_037	2-methoxyestradiol	Microtubules, HIF-1
D_038	serdemetan	MDM2
D_039	thapsigargin	Sarco/endoplasmic reticulum Ca2+ ATPase (SERCA)
D_040	YM155	Survivin
D_041	BAY 11-7085	NFkB
D_042	curcumin	COX2, NFkB
D_043	SGX-523	MET
D_044	MLN9708	Proteasome
D_045	KU-60019	ATM
D_046	ZM 336372	RAF1
D_047	sorafenib	BRAF, RAF1, FLT3, KIT, PDGFRA, PDGFRB
D_048	orlistat	LPL
D_049	chaetoglobosin A	Actin
D_050	dasatinib	ABL1, KIT, LYN, PDGFRA, PDGFRB, SRC
D_051	PKI-402	Pan PI3K
D_052	PIK-75	PI3K alpha
D_053	sunitinib	VEGFR, PDGFRA/B, FLT3, KIT
D_054	gefitinib	EGFR
D_055	lapatinib	EGFR, HER2
D_056	actinomycin D	RNA synthesis
D_057	crizotinib	ALK, MET
D_058	ochratoxin A	Oxidative DNA damage
D_059	lasalocid A	Cation transport
D_060	cephaeline	40S ribosomal subunit
D_061	oligomycin A	ATP synthase
D_062	fumonisin B1	Ceramide synthase
D_063	everolimus	mTOR
D_064	spliceostatin A	SF3b
D_065	ATRA	Retinoic acid and retinoid X receptor agonist
D_066	arsenic trioxide	-
D_067	rotenone	Electron transport chain in mitochondria
D_071	KX2-391	SRC
D_074	VE-821	ATR
D_075	rabusertib	CHK1
D_077	SCH 900776	CHK1, CDK2
D_078	PF 477736	CHK1, CHK2
D_079	spebrutinib	BTK
D_081	venetoclax	BCL2
D_082	duvelisib	PI3K gamma, PI3K delta
D_083	encorafenib	BRAF
D_084	ruxolitinib	JAK1/2/3
D_127	SD07	ROS
D_141	SD51	ROS
D_149	MIS-43	ROS
D_159	doxorubicin	DNA intercalation, Topoisomerase II
D_162	BML-277	CHK2
D_163	CCT241533	CHK2

D_164	BX912	PDK1
D_165	silmitasertib	CK2
D_166	PRT062607 HCl	SYK
D_167	UCN-01	PKC, MK2, CHK1
D_168	sotrastaurin	PKC
D_169	tofacitinib	JAK3
D_172	BIX02188	MEK5
D_CHK	BML-277 + rabusertib	-

Concentrations of drugs used in the drug screens

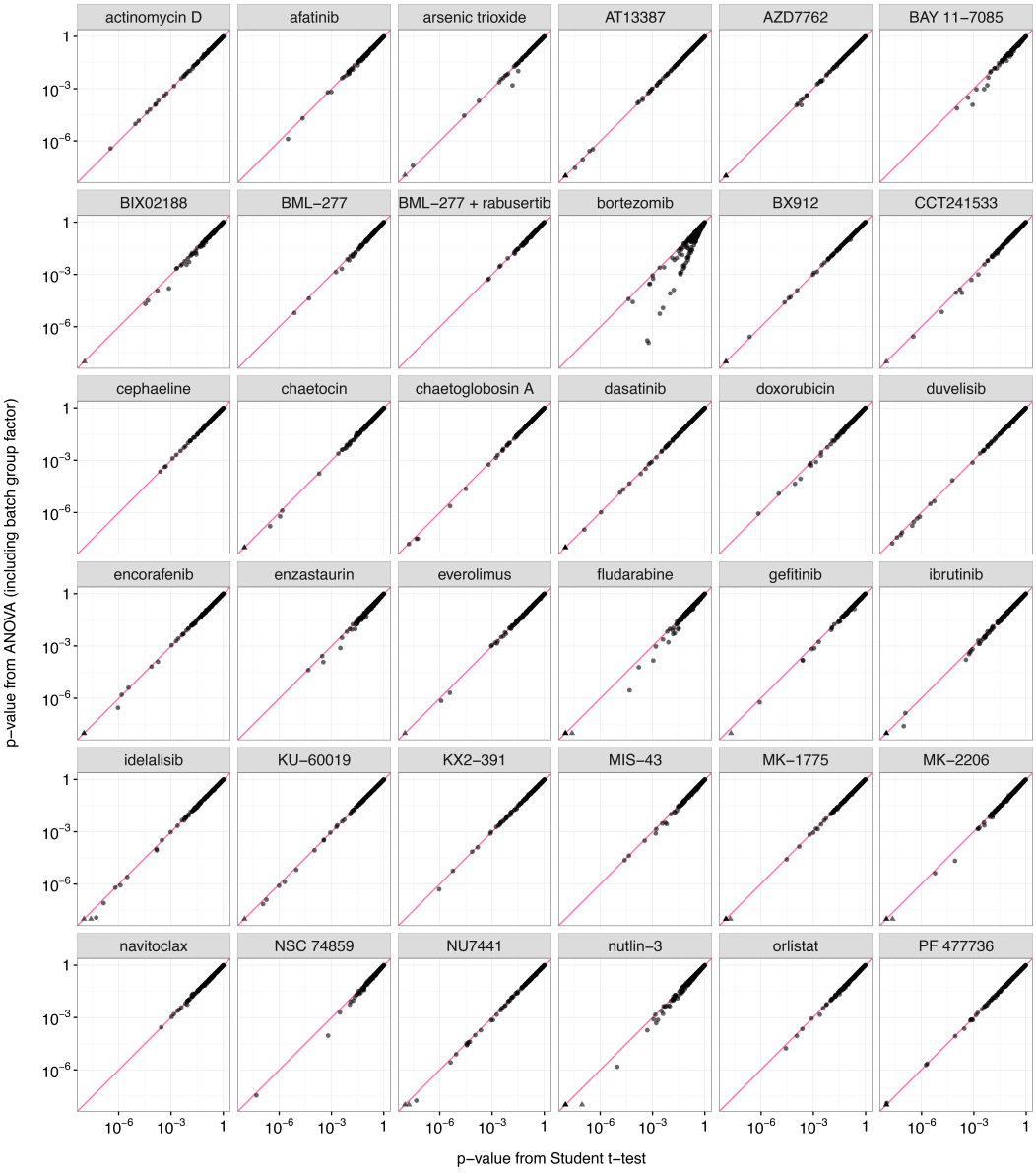
D

Drug ID	Pilot screen [μM]		Main screen [μM]				
	c1	c2	c1	c2	c3	c4	c5
D_001	0.1	0.05	1	0.25	0.063	0.016	0.004
D_002	10	1	40	10	2.5	0.625	0.156
D_003	10	1	40	10	2.5	0.625	0.156
D_004	0.2	0.1	4	1	0.25	0.063	0.016
D_005	10	-	-	-	-	-	-
D_006	10	1	40	10	2.5	0.625	0.156
D_007	5	1	20	5	1.25	0.313	0.078
D_008	5	2	20	5	1.25	0.313	0.078
D_009	5	1	-	-	-	-	-
D_010	10	1	40	10	2.5	0.625	0.156
D_011	5	2	40	10	2.5	0.625	0.156
D_012	10	1	40	10	2.5	0.625	0.156
D_013	4	2	15	5	1.667	0.556	0.185
D_014	10	1	-	-	-	-	-
D_015	10	1	40	10	2.5	0.625	0.156
D_016	10	-	-	-	-	-	-
D_017	1	0.1	10	2.5	0.625	0.156	0.039
D_018	10	-	-	-	-	-	-
D_019	10	-	-	-	-	-	-
D_020	10	1	40	10	2.5	0.625	0.156
D_021	5	2	40	10	2.5	0.625	0.156
D_022	10	-	-	-	-	-	-
D_023	10	-	40	10	2.5	0.625	0.156
D_024	5	2	40	10	2.5	0.625	0.156
D_025	10	-	40	10	2.5	0.625	0.156
D_026	10	-	-	-	-	-	-
D_027	10	-	-	-	-	-	-
D_028	10	-	-	-	-	-	-
D_029	5	2	40	10	2.5	0.625	0.156

D_030	10	1	40	10	2.5	0.625	0.156
D_031	10	1	-	-	-	-	-
D_032	10	1	40	10	2.5	0.625	0.156
D_033	10	1	40	10	2.5	0.625	0.156
D_034	0.2	0.1	2	0.5	0.125	0.031	0.008
D_035	10	1	40	10	2.5	0.625	0.156
D_036	10	1	40	10	2.5	0.625	0.156
D_037	10	-	-	-	-	-	-
D_038	10	1	-	-	-	-	-
D_039	10	1	20	5	1.25	0.313	0.078
D_040	0.1	0.05	2	0.5	0.125	0.031	0.008
D_041	2	1	40	10	2.5	0.625	0.156
D_042	10	1	-	-	-	-	-
D_043	10	1	40	10	2.5	0.625	0.156
D_044	10	-	-	-	-	-	-
D_045	10	1	40	10	2.5	0.625	0.156
D_046	10	-	-	-	-	-	-
D_047	10	1	-	-	-	-	-
D_048	10	1	40	10	2.5	0.625	0.156
D_049	8	4	20	10	5	2.5	1.25
D_050	10	1	40	10	2.5	0.625	0.156
D_051	10	1	-	-	-	-	-
D_052	2	1	-	-	-	-	-
D_053	10	-	40	10	2.5	0.625	0.156
D_054	10	-	40	10	2.5	0.625	0.156
D_055	10	-	-	-	-	-	-
D_056	0.02	0.01	0.3	0.1	0.033	0.011	0.004
D_057	10	1	-	-	-	-	-
D_058	10	1	-	-	-	-	-
D_059	2	1	-	-	-	-	-
D_060	0.1	0.05	4	1	0.25	0.063	0.016
D_061	1	0.1	-	-	-	-	-
D_062	10	1	-	-	-	-	-
D_063	10	1	40	10	2.5	0.625	0.156
D_064	0.02	0.01	-	-	-	-	-
D_065	10	1	-	-	-	-	-
D_066	2	1	8	4	2	1	0.5
D_067	10	1	40	10	2.5	0.625	0.156
D_071	-	-	40	10	2.5	0.625	0.156
D_074	-	-	40	10	2.5	0.625	0.156
D_075	-	-	40	10	2.5	0.625	0.156
D_077	-	-	40	10	2.5	0.625	0.156

D_078	-	-	40	10	2.5	0.625	0.156
D_079	-	-	40	10	2.5	0.625	0.156
D_081	-	-	1	0.25	0.063	0.016	0.004
D_082	-	-	40	10	2.5	0.625	0.156
D_083	-	-	40	10	2.5	0.625	0.156
D_084	-	-	40	10	2.5	0.625	0.156
D_127	-	-	30	10	3.333	1.111	0.37
D_141	-	-	30	10	3.333	1.111	0.37
D_149	-	-	30	10	3.333	1.111	0.37
D_159	-	-	4	1	0.25	0.063	0.016
D_162	-	-	40	10	2.5	0.625	0.156
D_163	-	-	40	10	2.5	0.625	0.156
D_164	-	-	40	10	2.5	0.625	0.156
D_165	-	-	40	10	2.5	0.625	0.156
D_166	-	-	40	10	2.5	0.625	0.156
D_167	-	-	10	2.5	0.625	0.156	0.039
D_168	-	-	40	10	2.5	0.625	0.156
D_169	-	-	40	10	2.5	0.625	0.156
D_172	-	-	40	10	2.5	0.625	0.156
D_CHK	-	-	10	2.5	0.625	0.156	0.039

Influence of batch effect on drug-gene associations



Continued on the next page.

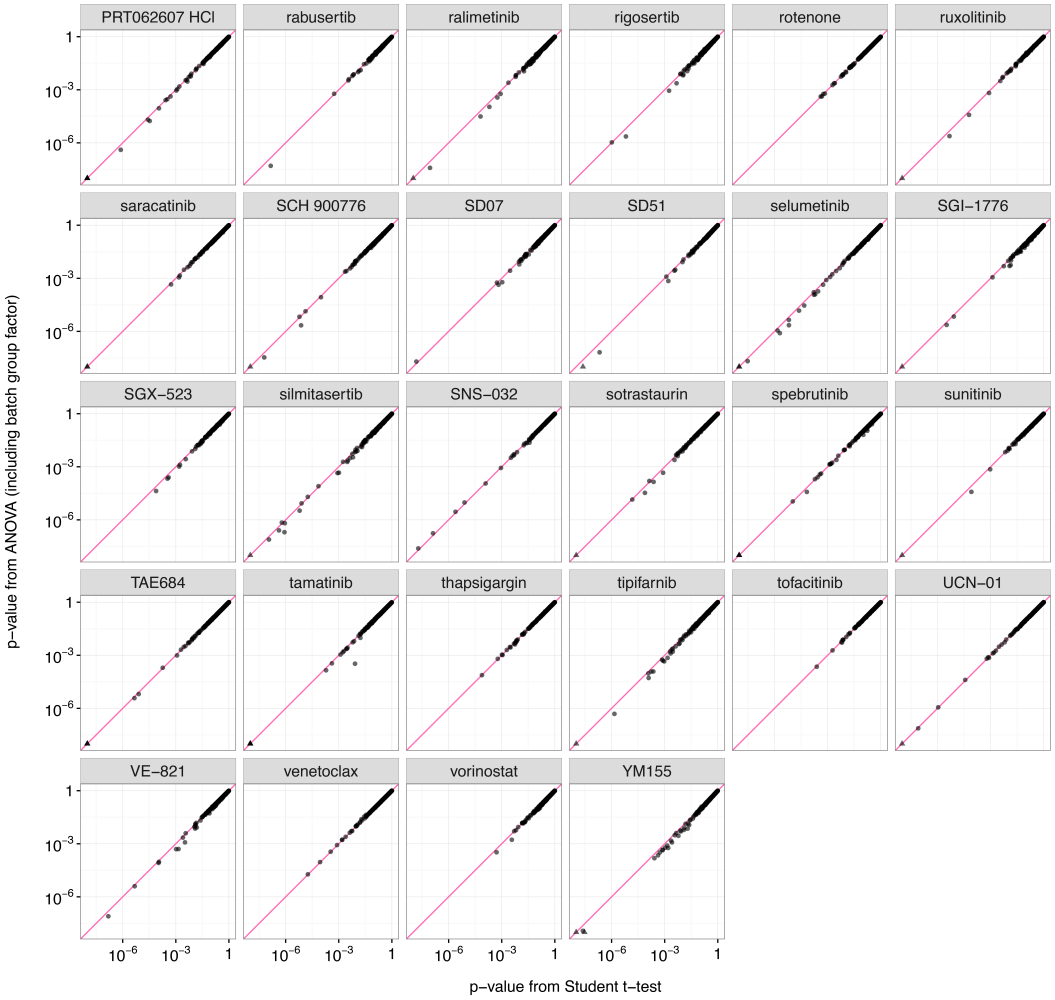


Figure E.1 Influence of batch effect on drug-gene associations.

To assess the potential influence of the three experimental batches of main screen on finding associations between drug responses and genetic features, we compared p-values computed in two different ways. We used two-way ANOVA with batch as a blocking factor (y-axis) and Student's t-test, which does not account for experimental batch group (x-axis). Triangles symbolize censored p-values ($< 10^{-8}$).

Dose-response curve fitting

F

A 4-parameter sigmoid curve given by the following equation (F.1) was fitted to the viability scores y in the function of log10 of drug concentration x (in μM).

$$\sigma(x, \xi, \kappa, \eta_1, \eta_0) = \frac{\eta_1 - \eta_0}{1 + e^{\kappa(x-\xi)}} + \eta_0, \tag{F.1}$$

where: ξ is the inflection point, κ is the curve slope in the inflection point, η_1 is the maximum asymptote, and η_0 is the minimum asymptote.

The fitting algorithm minimizes the value of the Γ function given by

$$\Gamma = (100 - \eta_1)^2 + \min(0, \eta_0)^2 + (\kappa - \kappa')^2 + \sum_{i=1}^{n} \left(y_i - \sigma\left(x_i, \xi, \kappa, \eta_1, \eta_0\right)\right)^2. \tag{F.2}$$

In the above equation, the weights of the different parts of the function Γ were identical and set to 1. The initial values of the parameters ξ and κ (ξ' and κ', respectively) were calculated for each drug and sample combination separately as follows. From the slope a and intercept b, the parameters obtained from fitting a linear equation to the two neighboring points between which there is maximum difference in viabilities, and by substituting y with 50, we obtain

$$\xi' = \frac{(50 - b)}{a}. \tag{F.3}$$

Then the first derivative of Eq. (F.1) for $x = \xi'$ is calculated, which gives the initial parameter for the slope.

$$\kappa' = -\frac{a}{25}. \tag{F.4}$$

Initial values for the parameters η_0 and η_1 are 0 and 100, respectively.

The fitting algorithm returns the parameters of the fitted sigmoid. To obtain the true value of IC_{50} we solve the sigmoid function equation for the 50% viability

$$IC_{50} = \xi + \frac{\ln\left(\frac{50-\eta 1}{\eta_0 - 50}\right)}{\kappa}. \tag{F.5}$$

The slope can be obtained by reversing Eq. (F.4) and calculating a.

Genes used in variant calling

G

Alphabetical list of genes for which variant calling was performed.

1. ABCC9
2. ACTB
3. ACTN2
4. ADAMTS7
5. ASXL1
6. ATM
7. BAZ2A
8. BCOR
9. BIRC3
10. BLK
11. BRAF
12. BRCC3
13. CARD11
14. CCND2
15. CDH2
16. CDKN1A
17. CDKN2A
18. CDKN2B
19. CDKN2B-AS1
20. CHD2
21. CHEK2
22. CREBBP
23. DCC
24. 3X
25. DENND4A
26. DYRK1A
27. EGFR
28. EGR2
29. ELF4
30. ERBB2IP
31. EWSR1
32. FAM50A
33. FAT1
34. FAT3
35. FBXW7
36. FLRT2
37. FUBP1
38. GNB1
39. HDAC2
40. HIST1H1E
41. HMCN1
42. IGLL5
43. IKZF3
44. IL26
45. IRF2BP2
46. IRF4
47. ITPKB
48. KLHL6
49. KRAS
50. LRP1B
51. MAG
52. MAP2K1
53. MAPK1
54. MED12
55. MED12L
56. MUC16
57. MYCBP2
58. MYD88
59. NFKBIE
60. NOTCH1
61. NRAS
62. NRXN1
63. NXF1
64. PCLO
65. PIM1
66. PLEKHG5
67. POT1
68. PRPF8
69. PTPN11
70. RFTN1
71. RIPK1
72. ROBO2
73. RPS15
74. RYR1
75. RYR2
76. RYR3
77. SAMHD1
78. SEMA4D
79. SF3B1
80. TET2
81. TGM7
82. TP53
83. TRAF2
84. TRAF3
85. UMODL1
86. XPO1
87. XPO4
88. ZMYM3

Comparison of significant associations of drug response and mutation in the two drug screens

H

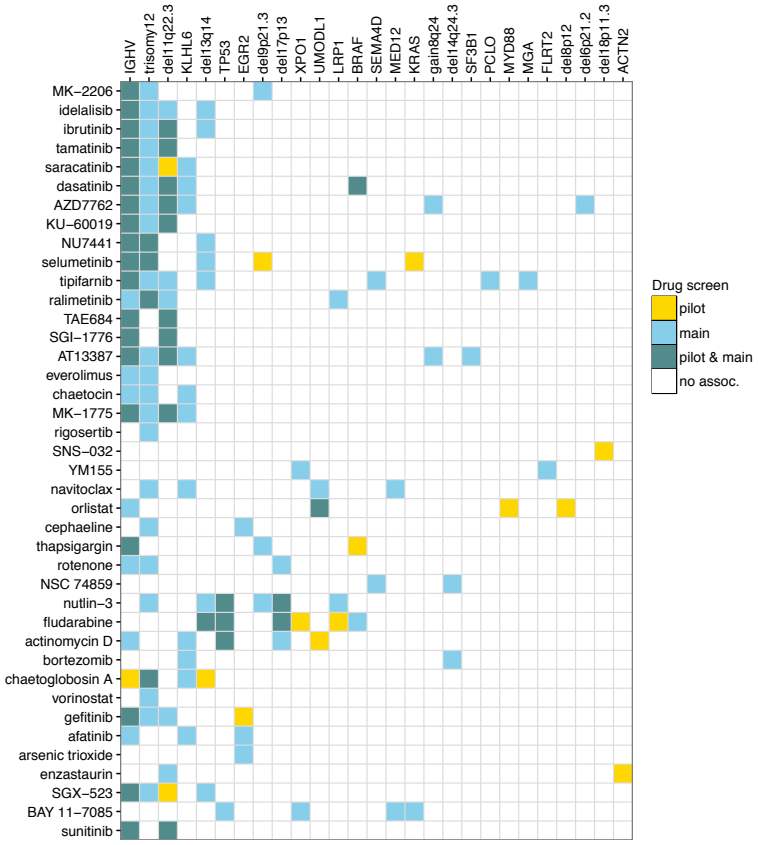


Figure H.1 Intersection of drug response modulators between drug screens in CLL.

The heat map shows associations of drug response (y-axis) and mutation (x-axis) which were statistically significant (FDR 10%) in the pilot and main screens. 39 mutations and 40 drugs were overlapping in the two analyses. Within this set 136 associations in total were found, from which 15 and 80 were uniquely identified in the pilot and in the main screen, respectively. 41 associations were in common. Association was deemed significant if any of the drug-concentration pairs showed significant result. Due to a lower number of both tested drug concentrations and patient samples in the pilot screen, there is less yellow squares than turquoise ones.

Results of multivariate Cox regression

We used the multivariate Cox model to assess which features from the well-established biomarkers and response to doxorubicin add to prediction of the overall survival in CLL ($n = 156$, including 24 deaths). Results presented in the table below show increased prediction accuracy by drug response profile of doxorubicin.

	p-value	HR	lower 95% CI	upper 95% CI
age (per 10 years)	0.12	1.4	0.92	2
pretreatment	1.8e-4	8.8	2.8	27
trisomy 12	0.01	6.1	1.5	25
del(11)(q22.3)	0.99	1	0.34	3
del(17)(p13)	0.92	1.1	0.3	3.8
TP53	0.72	0.81	0.26	2.5
U-CLL	0.04	2.9	1.1	8.1
doxorubicin	0.03	0.52	0.28	0.95

Analysis performed by Sascha Dietrich.

Bibliography

[1] Friedman, A. A., Letai, A., Fisher, D. E., and Flaherty, K. T. (2015) Precision medicine for cancer with next-generation functional diagnostics. *Nature Reviews Cancer*, **15**, 747–756.

[2] Weinberg, R. (2013) *The biology of cancer*. Garland science.

[3] Campo, E., Swerdlow, S. H., Harris, N. L., Pileri, S., Stein, H., and Jaffe, E. S. (2011) The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*, **117**, 5019–5032.

[4] Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M., and Vardiman, J. W. (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, **127**, 2391–2405.

[5] Gribben, J. G. (2010) How I treat CLL up front. *Blood*, **115**, 187–197.

[6] Hallek, M., et al. (2008) Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute–Working Group 1996 guidelines. *Blood*, **111**, 5446–5456.

[7] Damle, R. N., et al. (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*, **94**, 1840–1847.

[8] Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G., and Stevenson, F. K. (1999) Unmutated Ig VH genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*, **94**, 1848–1854.

[9] Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., Döhner, K., Bentz, M., and Lichter, P. (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *New England Journal of Medicine*, **343**, 1910–1916.

[10] Zenz, T., et al. (2010) *TP53* mutation and survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology*, **28**, 4473–4479.

[11] Skarbnik, A. P. and Goy, A. H. (2015) Mantle cell lymphoma: State of the art. *Clinical Advances in Hematology & Oncology*, **13**, 44–55.

[12] Tiacci, E., et al. (2015) Targeting mutant BRAF in relapsed or refractory hairy-cell leukemia. *New England Journal of Medicine*, **373**, 1733–1747.

[13] Tiacci, E., et al. (2011) *BRAF* mutations in hairy-cell leukemia. *New England Journal of Medicine*, **364**, 2305–2315.

[14] Dietrich, S., et al. (2015) Recurrent *CDKN1B* (p27) mutations in hairy cell leukemia. *Blood*, **126**, 1005–1008.

[15] Cawley, J., Burns, G., and Hayhoe, F. (1980) A chronic lymphoproliferative disorder with distinctive features: a distinct variant of hairy-cell leukaemia. *Leukemia Research*, **4**, 547–559.

[16] Dearden, C. (2012) How I treat prolymphocytic leukemia. *Blood*, **120**, 538–551.

[17] Moskowitz, A. J., Lunning, M. A., and Horwitz, S. M. (2014) How I treat the peripheral T-cell lymphomas. *Blood*, **123**, 2636–2644.

[18] Biggs, J. R. and Kraft, A. S. (2001) *Myeloid Cell Differentiation*. John Wiley & Sons, Ltd.

[19] Kumar, C. C. (2011) Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes & cancer*, **2**, 95–107.

[20] Papaemmanuil, E., et al. (2016) Genomic classification and prognosis in acute myeloid leukemia. *New England Journal of Medicine*, **374**, 2209–2221.

[21] MacDonald, D., Lachance, S., and Larratt, L. (2013) A Canadian perspective on the treatment of unfit patients with chronic lymphocytic leukemia. *New Evidence in Oncology*, pp. 118–135.

[22] Eichhorst, B., Goede, V., and Hallek, M. (2009) Treatment of elderly patients with chronic lymphocytic leukemia. *Leukemia & Lymphoma*, **50**, 171–178.

[23] Copelan, E. A. (2006) Hematopoietic stem-cell transplantation. *New England Journal of Medicine*, **354**, 1813–1826.

[24] Mukherjee, S. (2011) *The Emperor of All Maladies: A Biography of Cancer*. Scribner.

[25] Huang, M., Ye, Y.-c., Chen, S., Chai, J.-R., Lu, J.-X., Zhoa, L., Gu, L.-J., and Wang, Z.-Y. (1988) Use of All-*Trans* retinoic acid in the treatment of acute promyelocytic leukemia. *Blood*, **72**, 567–572.

[26] Druker, B. J., Sawyers, C. L., Kantarjian, H., Resta, D. J., Reese, S. F., Ford, J. M., Capdeville, R., and Talpaz, M. (2001) Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *New England Journal of Medicine*, **344**, 1038–1042.

[27] Flaherty, K. T., et al. (2010) Inhibition of mutated, activated BRAF in metastatic melanoma. *New England Journal of Medicine*, **363**, 809–819.

[28] Dietrich, S., Glimm, H., Andrulis, M., von Kalle, C., Ho, A. D., and Zenz, T. (2012) BRAF inhibition in refractory hairy-cell leukemia. *New England Journal of Medicine*, **366**, 2038–2040.

[29] Byrd, J. C., et al. (2001) Rituximab using a thrice weekly dosing schedule in B-cell chronic lymphocytic leukemia and small lymphocytic lymphoma demonstrates clinical activity and acceptable toxicity. *Journal of Clinical Oncology*, **19**, 2153–2164.

[30] Rossi, D., et al. (2009) The prognostic value of *TP53* mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness. *Clinical Cancer Research*, **15**, 995–1004.

[31] Gonzalez, D., Martinez, P., Wade, R., Hockley, S., Oscier, D., Matutes, E., Dearden, C. E., Richards, S. M., Catovsky, D., and Morgan, G. J. (2011) Mutational status of the *TP53* gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *Journal of Clinical Oncology*, **29**, 2223–2229.

[32] Longo, D. L. (2014) Cancer-drug discovery — let's get ready for the next period. *New England Journal of Medicine*, **371**, 2227–2228.

[33] Umar, A. and Rosenfeld, S. (2015) Role of network biology and network medicine in early detection of cancer. Thiagalingam, S. (ed.), *Systems Biology of Cancer*, chap. 29, pp. 457–463, Cambridge University Press.

[34] Malek, S. (2013) Molecular biomarkers in chronic lymphocytic leukemia. Malek, S. (ed.), *Advances in Chronic Lymphocytic Leukemia*, chap. 9, pp. 193–214, Springer New York.

[35] Wang, L., et al. (2011) *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *New England Journal of Medicine*, **365**, 2497–2506.

[36] Byrd, J. C., et al. (2013) Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *New England Journal of Medicine*, **369**, 32–42.

[37] Furman, R. R., et al. (2014) Idelalisib and Rituximab in relapsed chronic lymphocytic leukemia. *New England Journal of Medicine*, **370**, 997–1007.

[38] Roberts, A. W., et al. (2016) Targeting BCL2 with venetoclax in relapsed chronic lymphocytic leukemia. *New England Journal of Medicine*, **374**, 311–322.

[39] Landau, D. A., et al. (2015) Mutations driving CLL and their evolution in progression and relapse. *Nature*, **526**, 525–530.

[40] Puente, X. S., et al. (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, **526**, 519–524.

[41] Horwitz, R. I., Cullen, M. R., Abell, J., and Christian, J. B. (2013) (de)personalized medicine. *Science*, **339**, 1155–1156.

[42] Sawyers, C. L. (2008) The cancer biomarker problem. *Nature*, **452**, 548–552.

[43] de Gramont, A., Watson, S., Ellis, L. M., Rodón, J., Tabernero, J., de Gramont, A., and Hamilton, S. R. (2015) Pragmatic issues in biomarker evaluation for targeted therapies in cancer. *Nature Reviews Clinical Oncology*, **12**, 197–212.

[44] Garnett, M. J., et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, **483**, 570–575.

[45] Barretina, J., et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.

[46] Basu, A., et al. (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, **154**, 1151–1161.

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

[48] Consortium, C. C. L. E. and of Drug Sensitivity in Cancer Consortium, G. (2015) Pharmacogenomic agreement between two cancer cell line data sets. *Nature*, **528**, 84–87.

[49] Haverty, P. M., et al. (2016) Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature*, **533**, 333–337.

[50] Crystal, A. S., et al. (2014) Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science*, **346**, 1480–1486.

[51] Pemovska, T., et al. (2013) Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discovery*, **3**, 1416–1429.

[52] Tyner, J. W., et al. (2013) Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. *Cancer Research*, **73**, 285–296.

[53] Maxson, J. E., et al. (2013) Oncogenic *CSF3R* mutations in chronic neutrophilic leukemia and atypical CML. *New England Journal of Medicine*, **368**, 1781–1790.

[54] Pemovska, T., et al. (2015) Axitinib effectively inhibits BCR-ABL1 (T315I) with a distinct binding conformation. *Nature*, **519**, 102–105.

[55] Maxson, J. E., et al. (2016) Identification and characterization of tyrosine kinase nonreceptor 2 mutations in leukemia through integration of kinase inhibitor screening and genomic analysis. *Cancer Research*, **76**, 127–138.

[56] Boutros, M., Brás, L. P., and Huber, W. (2006) Analysis of cell-based RNAi screens. *Genome biology*, **7**, R66.

[57] 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. *Nature Chemical Biology*, **9**, 708–714.

[58] Yadav, B., et al. (2014) Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Scientific reports*, **4**.

[59] Jethwa, A., et al. (2013) Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. *British Journal of Haematology*, **163**, 496–500.

[60] Jones, D. T., et al. (2012) Dissecting the genomic complexity underlying medulloblastoma. *Nature*, **488**, 100–105.

[61] Jones, D. T., et al. (2013) Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nature genetics*, **45**, 927–932.

[62] Wang, K., Li, M., and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, **38**, e164–e164.

[63] Wu, T. D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

[64] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

[65] McKenna, A., et al. (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

[66] DePristo, M. A., et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491–498.

[67] Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, **29**, 308–311.

[68] Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., and Carey, V. J. (2013) Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, **9**, 1–10.

[69] Love, M., exomeCopy: Copy number variant detection from exome sequencing read depth.

[70] Arora, S., Morgan, M., Carlson, M., and Pagès, H., GenomeInfoDb: Utilities for manipulating chromosome and other 'seqname' identifiers.

[71] Fischer, B. and Pau, G., rhdf5: HDF5 interface to R.

[72] Pyl, P. T., Gehring, J., Fischer, B., and Huber, W. (2014) h5vc: scalable nucleotide tallies with HDF5. *Bioinformatics*, **30**, 1464–1466.

[73] TBD, T., BSgenome.Hsapiens.UCSC.hg19: Full genome sequences for Homo sapiens (UCSC version hg19).

[74] Marc, C. and BP, M., TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s).

[75] Bischl, B., Lang, M., Mersmann, O., Rahnenführer, J., and Weihs, C. (2015) BatchJobs and BatchExperiments: Abstraction mechanisms for using R in batch environments. *Journal of Statistical Software*, **64**, 1–25.

[76] Venkatraman, E. and Olshen, A. B. (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*, **23**, 657–663.

[77] Landry, J. J. M., et al. (2013) The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)*, **3**, 1213–1224.

[78] Olshen, A. B., Venkatraman, E., Lucito, R., and Wigler, M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **5**, 557–572.

[79] Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

[80] Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 1.

[81] Oakes, C. C., et al. (2016) DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nature Genetics*, **48**, 253–264.

[82] Ouillette, P. and Malek, S. (2013) Acquired genomic copy number aberrations in CLL. Malek, S. (ed.), *Advances in Chronic Lymphocytic Leukemia*, chap. 3, pp. 47–86, Springer New York.

[83] Martínez-Trillos, A., Quesada, V., Villamor, N., Puente, X. S., López-Otín, C., and Campo, E. (2013) Recurrent gene mutations in CLL. Malek, S. (ed.), *Advances in Chronic Lymphocytic Leukemia*, chap. 4, pp. 87–107, Springer New York.

[84] Chiaretti, S. and Foà, R. (2013) Novel molecular acquisitions in leukemias. Pfeffer, U. (ed.), *Cancer Genomics: Molecular Classification, Prognosis and Response Prediction*, pp. 453–493, Springer Netherlands.

[85] Oza, V., et al. (2012) Discovery of checkpoint kinase inhibitor (s)-5-(3-fluorophenyl)-n-(piperidin-3-yl)-3-ureidothiophene-2-carboxamide (AZD7762) by structure-based design and optimization of thiophenecarboxamide ureas. *Journal of Medicinal Chemistry*, **55**, 5130–5142.

[86] Blasina, A., et al. (2008) Breaching the DNA damage checkpoint via PF-00477736, a novel small-molecule inhibitor of checkpoint kinase 1. *American Association for Cancer Research*, **7**, 2394–2404.

[87] Zenz, T., Mertens, D., Küppers, R., Döhner, H., and Stilgenbauer, S. (2010) From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature Reviews Cancer*, **10**, 37–50.

[88] Guo, A., Lu, P., Galanina, N., Nabhan, C., Smith, S. M., Coleman, M., and Wang, Y. L. (2015) Higher BTK-dependent cell proliferation in unmutated chronic lymphocytic leukemia confers increased sensitivity to ibrutinib. *Oncotarget*, **7**, 4598–4610.

[89] Wang, M. L., et al. (2013) Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. *New England Journal of Medicine*, **369**, 507–516.

[90] Rahal, R., et al. (2014) Pharmacological and genomic profiling identifies NF-κB-targeted treatment strategies for mantle cell lymphoma. *Nature medicine*, **20**, 87–92.

[91] Hahn, C. K., et al. (2009) Proteomic and genetic approaches identify Syk as an AML target. *Cancer Cell*, **16**, 281–294.

[92] Thomas, X. and Elhamri, M. (2007) Tipifarnib in the treatment of acute myeloid leukemia. *Biologics*, **1**, 415–424.

[93] Fiedler, W., et al. (2015) A phase I/II study of sunitinib and intensive chemotherapy in patients over 60 years of age with acute myeloid leukaemia and activating *FLT3* mutations. *British Journal of Haematology*, **169**, 694–700.

[94] DeAngelo, D. J., Neuberg, D., Amrein, P. C., Berchuck, J. E., Wadleigh, M., Sirulnik, L. A., Galinsky, I., Golub, T., Stegmaier, K., and Stone, R. M. (2014) A phase II study of the EGFR inhibitor gefitinib in patients with acute myeloid leukemia. *Leukemia Research*, **38**, 430–434.

[95] Garg, R., Wierda, W., Ferrajoli, A., Abruzzo, L., Pierce, S., Lerner, S., Keating, M., and O'Brien, S. (2012) The prognostic difference of monoallelic versus biallelic deletion of 13q in chronic lymphocytic leukemia. *Cancer*, **118**, 3531–3537.

[96] Stevenson, F. K., Sahota, S. S., Ottensmeier, C. H., Zhu, D., Forconi, F., and Hamblin, T. J. (2001) The occurrence and significance of V gene mutations in B cell-derived human malignancy. *Advances in cancer research*, **83**, 81–116.

[97] Stevenson, F. K., Krysov, S., Davies, A. J., Steele, A. J., and Packham, G. (2011) B-cell receptor signaling in chronic lymphocytic leukemia. *Blood*, **118**, 4313–4320.

[98] Borche, L., Lim, A., Binet, J.-L., and Dighiero, G. (1990) Evidence that chronic lymphocytic leukemia B lymphocytes are frequently committed to production of natural autoantibodies. *Blood*, **76**, 562–569.

[99] Veldurthy, A., Patz, M., Hagist, S., Pallasch, C. P., Wendtner, C.-M., Hallek, M., and Krause, G. (2008) The kinase inhibitor dasatinib induces apoptosis in chronic lymphocytic leukemia cells in vitro with preference for a subgroup of patients with unmutated IgVH genes. *Blood*, **112**, 1443–1452.

[100] McCaig, A. M., Cosimo, E., Leach, M. T., and Michie, A. M. (2011) Dasatinib inhibits B cell receptor signalling in chronic lymphocytic leukaemia but novel combination approaches are required to overcome additional pro-survival microenvironmental signals. *British Journal of Haematology*, **153**, 199–211.

[101] Goldstein, R. L., et al. (2015) Pharmacoproteomics identifies combinatorial therapy targets for diffuse large B cell lymphoma. *The Journal of Clinical Investigation*, **125**, 4559–4571.

[102] Bourgon, R., Gentleman, R., and Huber, W. (2010) Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences*, **107**, 9546–9551.

[103] Buchner, M., Baer, C., Prinz, G., Dierks, C., Burger, M., Zenz, T., Stilgenbauer, S., Jumaa, H., Veelken, H., and Zirlik, K. (2010) Spleen tyrosine kinase inhibition prevents chemokine- and integrin-mediated stromal protective effects in chronic lymphocytic leukemia. *Blood*, **115**, 4497–4506.

[104] Burger, J. A. (2013) The CLL cell microenvironment. Malek, S. (ed.), *Advances in Chronic Lymphocytic Leukemia*, chap. 2, pp. 25–45, Springer New York.

[105] Riley, T., Sontag, E., Chen, P., and Levine, A. (2008) Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol*, **9**, 402–412.

[106] Rosenwald, A., et al. (2004) Fludarabine treatment of patients with chronic lymphocytic leukemia induces a p53-dependent gene expression response. *Blood*, **104**, 1428–1434.

[107] Trbusek, M. and Malcikova, J. (2013) *TP53* aberrations in chronic lymphocytic leukemia. Malek, S. (ed.), *Advances in Chronic Lymphocytic Leukemia*, chap. 5, pp. 109–131, Springer New York.

[108] Zenz, T., et al. (2009) Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, *TP53* mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial. *Blood*, **114**, 2589–2597.

[109] Robertson, L., Chubb, S., Meyn, R., Story, M., Ford, R., Hittelman, W., and Plunkett, W. (1993) Induction of apoptotic cell death in chronic lymphocytic leukemia by 2- chloro-2'-deoxyadenosine and 9-beta-D-arabinosyl-2-fluoroadenine. *Blood*, **81**, 143–150.

[110] Tacar, O., Sriamornsak, P., and Dass, C. R. (2013) Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *Journal of Pharmacy and Pharmacology*, **65**, 157–170.

[111] Greiner, T., Moynihan, M., Chan, W., Lytle, D., Pedersen, A., Anderson, J., and Weisenburger, D. (1996) p53 mutations in mantle cell lymphoma are associated with variant cytology and predict a poor prognosis. *Blood*, **87**, 4302–4310.

[112] Thompson, P. A., Ferrajoli, A., O'Brien, S., Wierda, W. G., Keating, M. J., and Burger, J. A. (2015) Trisomy 12 is associated with an abbreviated redistribution lymphocytosis during treatment with the BTK inhibitor ibrutinib in patients with chronic lymphocytic leukaemia. *British Journal of Haematology*, **170**, 125–128.

[113] Decker, S., et al. (2012) Trisomy 12 and elevated GLI1 and PTCH1 transcript levels are biomarkers for Hedgehog-inhibitor responsiveness in CLL. *Blood*, **119**, 997–1007.

[114] Delbridge, A. and Strasser, A. (2015) The BCL-2 protein family, BH3-mimetics and cancer therapy. *Cell Death & Differentiation*, **22**, 1071–1080.

[115] Anderson, M. A., et al. (2016) The BCL2 selective inhibitor venetoclax induces rapid onset apoptosis of CLL cells in patients via a *TP53*-independent mechanism. *Blood*, **127**, 3215–3224.

[116] Davies, H., et al. (2002) Mutations of the *BRAF* gene in human cancer. *Nature*, **417**, 949–954.

[117] Jebaraj, B. M. C., Kienle, D., Bühler, A., Winkler, D., Döhner, H., Stilgenbauer, S., and Zenz, T. (2013) *BRAF* mutations in chronic lymphocytic leukemia. *Leukemia & lymphoma*, **54**, 1177–1182.

[118] Hall-Jackson, C. A., Eyers, P. A., Cohen, P., Goedert, M., Boyle, F. T., Hewitt, N., Plant, H., and Hedge, P. (1999) Paradoxical activation of Raf by a novel Raf inhibitor. *Chemistry & biology*, **6**, 559–568.

[119] Callahan, M. K., et al. (2012) Progression of RAS-mutant leukemia during RAf inhibitor treatment. *New England Journal of Medicine*, **367**, 2316–2321.

[120] Goodman, R. H. and Smolik, S. (2000) CBP/p300 in cell growth, transformation, and development. *Genes & Development*, **14**, 1553–1577.

[121] Puente, X. S., et al. (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*, **475**, 101–105.

[122] Lapalombella, R., et al. (2012) Selective inhibitors of nuclear export show that CRM1/XPO1 is a target in chronic lymphocytic leukemia. *Blood*, **120**, 4621–4634.

[123] Brown, J. R., et al. (2012) Integrative genomic analysis implicates gain of *PIK3CA* at 3q26 and *MYC* at 8q24 in chronic lymphocytic leukemia. *Clinical Cancer Research*, **18**, 3791–3802.

[124] Kämpjärvi, K., et al. (2015) Somatic MED12 mutations are associated with poor prognosis markers in chronic lymphocytic leukemia. *Oncotarget*, **6**, 1884–1888.

[125] Norberg, M., Lindhagen, E., Kanduri, M., Rickardson, L., Sundström, C., Stamatopoulos, K., Rosenquist, R., and ÅLESKOG, A. (2012) Screening for cytotoxic compounds in poor-prognostic chronic lymphocytic leukemia. *Anticancer Research*, **32**, 3125–3136.

[126] Boulesteix, A.-L., Bin, R. D., Jiang, X., and Fuchs, M. (2015), IPF-LASSO: integrative L1-penalized regression with penalty factors for prediction based on multi-omics data.

[127] Friedman, J., Hastie, T., and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.

[128] Ferreira, P. G., et al. (2014) Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Research*, **24**, 212–226.

[129] Hothorn, T. and Lausen, B. (2002) Maximally selected rank statistics in R. *R News*, **2/1**.

[130] Nakahara, T., et al. (2007) YM155, a novel small-molecule survivin suppressant, induces regression of established human hormone-refractory prostate tumor xenografts. *Cancer research*, **67**, 8014–8021.

[131] Na, Y.-S., et al. (2012) YM155 induces EGFR suppression in pancreatic cancer cells. *PLoS One*, **7**, e38625.