LUDWIG-MAXIMILIANS-UNIVERSITÄT
TECHNISCHE UNIVERSITÄT MÜNCHEN

Masterarbeit
in Bioinformatik

Identification of Uni- and Multivariate Biomarkers in Metabolomics Data from Lymphoma Patients

Jonas Zierer



LUDWIG-MAXIMILIANS-UNIVERSITÄT
TECHNISCHE UNIVERSITÄT MÜNCHEN



Helmholtzzentrum München
Institute of Bioinformatics and Systems Biology

Masterarbeit
in Bioinformatik

Identification of Uni- and Multivariate Biomarkers in Metabolomics Data from Lymphoma Patients

Jonas Zierer

Aufgabensteller:	Prof. Mewes
Betreuer:	Gabi Kastenmüller, Jan Krumsiek
Abgabedatum:	15.09.2012

This document was created using LaTeX.

This is an adaption of the general template from
`http://www.matthiaspospiech.de`.

Ich versichere, dass ich diese Masterarbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

15.09.2012 _____

Jonas Zierer

Abstract

Lymphoma is a type of cancer that develops from malignant transformation of B- or T-lymphocytes, which are part of the immune system. The most common subtype of lymphoma is Diffuse Large B-Cell Lymphoma (DLBCL). It is usually treated with R-CHOP, a mixture of chemotherapeutic agents, but there is hardly any treatment for R-CHOP resistant cases. A new approach for treating these cases is the usage of histone deacetylase inhibitors, such as Panobinostat. Histone deacetylases regulate various cellular processes, including gene expression, by modification of histones and other proteins. However, so far only little is known about how Panobinostat prevents tumor growth.

In this thesis high throughput metabolomics data from a series of lymphoma studies were analyzed to reveal metabolic characteristics of DLBCL and Panobinostat treatment. These consist of 276 metabolites determined in the DLBCL cell line OCI-Ly1 and 1097 metabolites from blood samples of lymphoma patients in different states of disease and treatment. Besides univariate statistical methods, bi- and multivariate approaches were implemented to unveil relevant metabolic alterations.

The metabolic profiles of blood samples of patients with relapsed tumors significantly differ from primary tumors. Most changes are associated with oxidative stress and inflammatory pathways. Some changes, e.g. in the steroid hormone metabolism and the metabolism of eicosanoids, are reverted during the Panobinostat treatment while the concentration of other metabolites, e.g. long chain fatty acids, continuously increases from primary tumor to relapse and throughout the treatment. Samples from cell lines were analyzed to distinguish between direct, acute effects of Panobinostat on tumor cells and indirect, subchronic effects on the whole organism. These insights improve the understanding of lymphoma and might facilitate the further development of Panobinostat.

Zusammenfassung

Malignes Lymphom nennt sich eine Krebsart, die sich aus B- oder T-Lymphozyten entwickelt. Die am häufigsten auftretende Lymphom-Art besteht aus diffusen großzelligen B-Zell-Lymphomen (DLBCL). Diese werden in der Regel mit einer Mischung von Chemotherapeutika, R-CHOP, behandelt. Für Lymphome, die auf dieses Therapie nicht ansprechen, gibt es kaum andere Behandlungsmöglichkeiten. An einer dieser Alternativen, nämlich an Histondeacetlyase-Inhibitoren (HDACi), wird derzeit intensiv geforscht. Histondeacetylasen regulieren verschiedene zelluläre Prozesse, darunter auch die Genexpression, indem sie Histone und andere Proteine modifizieren. Allerdings ist wenig darüber bekannt, wie genau HDACi das Tumorwachstum aufhalten. In vorliegender Arbeit steht ein spezifischer HDACi im Fokus, nämlich Panobinostat.

Um Änderungen im Metabolismus von DLBCL und den Einfluss von Panobinostat auf diesen zu ergründen, wurden hier High-Throughput Metabolomics-Daten aus einer Reihe von Lymphom-Studien untersucht. Hierfür wurden 276 Metabolite in Proben der DLBCL Zelllinie OCI-Ly1 sowie 1097 Metabolite in Plasma- und Serumproben von Patienten in unterschiedlichen Stadien der Krankheit und der Behandlung gemessen. Neben univariaten wurden auch bi- und multivariate Methoden implementiert, um relevante Änderungen identifizieren zu können.

Die Metabolitenprofile in Blutproben von Patienten mit rezidivierendem Tumor unterscheiden sich deutlich von denen mit Primärtumoren. Viele der beobachteten Änderungen sind aus Entzündungsreaktionen oder von oxidativem Stress bekannt. Einige der Änderungen, z.B. die der Steroidhormone und Eicosanoide, werden durch Panobinostat rückgängig gemacht. Im Gegensatz dazu steigt die Konzentration anderer Metaboliten, z.b. die langkettiger Fettsäuren, kontinuierlich von der primären Erkrankung über den Rückfall und bis nach der Behandlung mit Panobinostat an. Zusätzlich zu den Blutproben wurden Zelldaten untersucht, um primäre, akute Effekte von Panobinostat auf die Tumorzellen von sekundären, mittelfristigen Effekten auf den gesamten Organismus zu unterscheiden. Diese Erkenntnisse tragen zum Verständnis von Lymphomen bei und können die Weiterentwicklung von Panobinostat fördern.

Contents

Abstract v

Contents ix

1 Introduction 1
1.1 Lymphoma . . . 1
1.1.1 Development of Lymphoma . . . 1
1.1.2 Hodgkin's Disease . . . 4
1.1.3 Non-Hodgkin-Lymphoma . . . 4
1.1.4 Diffuse Large B-Cell Lymphoma . . . 5
1.1.5 Symptoms . . . 8
1.1.6 Treatment . . . 8
1.2 Metabolomics . . . 11
1.3 Aim . . . 13

2 Data 15
2.1 Patient Inclusion Criteria . . . 16
2.2 Study Design . . . 16
2.3 Experimental Procedure . . . 18
2.4 Preprocessing . . . 20
2.4.1 Logarithmic transformation . . . 20
2.4.2 Outlier Removal . . . 21
2.4.3 Exclusion of Metabolites . . . 23
2.5 Assessment of Metabolomics Data . . . 23
2.6 Data Quality . . . 27
2.6.1 Metabolon CV Data . . . 27
2.6.2 Run Day and Experiment Comparison . . . 28

3 Methods 33
3.1 Correction for Confounding Factors . . . 33
3.2 Univariate Analyses . . . 33
3.2.1 Linear Regression . . . 34

3.2.2 T-statistics . . . 34
3.2.3 Test for different variances . . . 35
3.2.4 Missing Value Analysis . . . 36
3.3 Bivariate Analyses . . . 37
3.3.1 Metabolite Ratios . . . 37
3.3.2 Linear Regression with Metabolite Pairs . . . 38
3.3.3 Differential Correlation . . . 39
3.4 Multivariate Analyses . . . 40
3.4.1 Imputation . . . 40
3.4.2 Pathway Enrichment . . . 41
3.4.3 Principal Component Analysis . . . 42
3.4.4 Multivariate Regression . . . 42
3.4.5 Clustering . . . 43
3.5 Multiple Testing Correction . . . 44
3.6 Visualization . . . 44
3.7 Implementation and Availability . . . 44

4 Results and Discussion 47
4.1 Before vs. After Treatment with Panobinostat . . . 48
4.1.1 Results (In Vivo) . . . 48
4.1.2 Results (In Vitro) . . . 61
4.1.3 Discussion . . . 69
4.2 Primary vs. Relapsed/Refractory Lymphoma . . . 76
4.2.1 Results . . . 76
4.2.2 Discussion . . . 84
4.3 Resistant vs. Sensitive Lymphoma . . . 87
4.3.1 Results . . . 87
4.3.2 Discussion . . . 93
4.4 Plasma vs. Serum . . . 95
4.4.1 Results . . . 95
4.4.2 Discussion . . . 103

5 Conclusion and Perspectives 107
5.1 Summary . . . 107
5.2 Perspectives . . . 111

Bibliography 113

Acronyms 135

List of Figures 138

List of Tables 143

Supplements 145

A Content of Supplementary DVD 147

1 Introduction

In this thesis, a case-control metabolomics study investigating different aspects of lymphoma was analyzed. This introduction first gives an overview over hallmarks of lymphoma, its development, subtypes, treatment and prognosis and genetics. After that, the concept of metabolomics is described before the aim of this thesis is defined.

1.1 Lymphoma

Lymphoma is a type of cancer which develops from malignant transformed B- or T-lymphocytes, certain white blood cells which are part of the immune system. In the US, 79,000 new cases and 20,200 deaths caused by lymphoma were expected for 2012. This translates to an incidence of $\approx$ 25 new cases per 100,000 persons per year (Howlader et al., 1975). In total, nearly 700,000 people in the USA have or had lymphoma (Howlader et al., 1975; Leukemia and Lymphoma Society, 2013).

Furthermore, lymphoma has – compared to other tumors – a very high incidence in childhood: for males, aged 15 and younger, it is the 4. most frequent cancer leading to death for males in the age group from 15 to 34, it is even the 2. most frequent (Holland, 2000). In the age from 0 to 19 lymphoma accounts for nearly 15 % of all cancers (Leukemia and Lymphoma Society, 2013).

1.1.1 Development of Lymphoma

B-lymphocytes are a type of leukocytes (white blood cells) which are responsible for producing antibodies and thus form an important part of the body's adaptive immune system. All lymphocytes are produced in the primary lymphoid tissues: B-lymphocytes mostly in the bone marrow and T-lymphocytes usually in the thymus (even though there are B-cells which are produced in the thymus). When a mature B-lymphocyte comes into contact with an antigen, it absorbs and digests it and

presents peptides from the digested antigen – bound to Major Histocompatibility Complex (MHC) – at its cell surface. This allows armed T helper cells to bind and activate the B-cell (T-cell depended B-cell activation).

The activated B-cell migrates to secondary lymphoid tissues (e.g. lymph nodes) and forms a germinal center (GC) to adapt antibodies to the initial antigen. The GC is divided into two parts (see Figure 1.1): the light zone and the dark zone. The dark zone contains rapidly proliferating B-cells (about one cell cycle every 6 hours (Storb et al., 2001)) with non-cleaved nuclei – so called centroblasts. While centroblasts divide, they undergo Somatic Hypermutation (SHM). This is a mechanism which introduces mutations – mainly Single Nucleotide Polymorphisms (SNPs) – into the light chain of Immunoglobin (Ig) genes to increase the specificity of antibodies to the antigen (Storb et al., 2001; Klein and Dalla-Favera, 2008).

The light zone contains centrocytes which have a cleaved nucleus and do not divide (Schwartz, Lenz, and Staudt, 2010). B-cells are less dense within the light zone, instead Follicular Dentritic cells (FDCs) and T-cells are located there. Those FDCs interact with the centrocytes and T-cells to select beneficial mutations which enhance the affinity for the original antigen. Thereby, cells with disadvantageous mutations die by apoptosis. Furthermore, Class Switch Recombination (CSR) takes place in the light zone of the GC. Besides SHM, this is the second mechanism to adapt the antibody to the antigen. Induced by Activation-induced Cytidine Deaminase (AID) it changes Ig heavy chain classes from Ig-M to Ig-G, Ig-A or Ig-E while leaving the light chain untouched (Schwartz, Lenz, and Staudt, 2010; Klein and Dalla-Favera, 2008; Storb et al., 2001)

B-cells migrate back and forth between the dark and the light zone while undergoing alternating mutations and selection until highly specific antibodies can be produced. The majority of finally selected GC B-cells differentiate into plasmablasts which divide fast and further develop into plasma cells which in advance are specialized on antibody production. The other centrocytes mutate to memory B-cells. These are long-living, divide only slowly and do not produce antibodies. The whole process is called germinal center reaction (Janeway, 2001; LeBien and Tedder, 2008; Mauri and Bosma, 2012; Parker, 1993; Thomas et al., 2004; Klein and Dalla-Favera, 2008).

In most stages of the B-cell development and the germinal center reaction, malignant transformation of a B-cell can cause lymphoma. Possible causes – or at least supporting factors – for this are mutations which affect non-Ig genes due to imperfect selectivity of SHM or CSR. Depending on the stage of the B-cell, different lymphoma subtypes arise. Some lymphomas develop from mutated T-cells (T-cell lymphomas), but these are the minority and are not considered in this thesis.

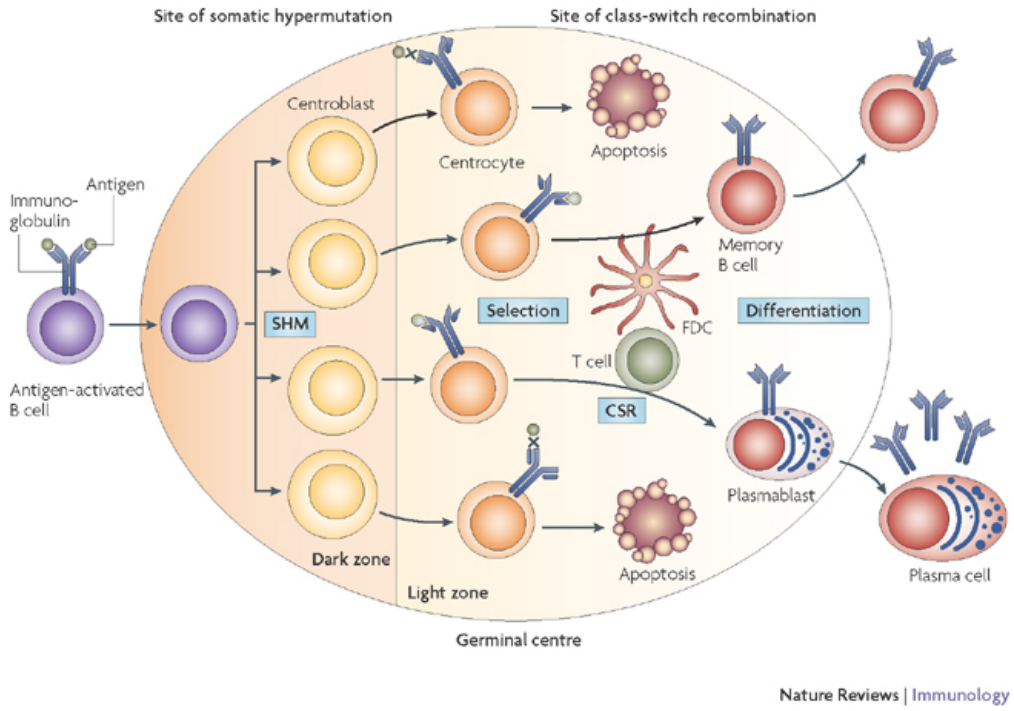


Figure 1.1: The germinal center reaction. An activated B-cell forms a GC. This is then divided in the dark zone where SHM changes the B-cell antibodies and in the light zone where Class Switch Recombination (CSR) switches Ig classes and beneficial mutations are selected by interaction with T-cells and Follicular Dentritic cell (FDC). B-cells migrate back and forth between the dark and the light zone while switching between mutation and selection. Finally, centrocytes differentiate into plasmablasts and then to plasma cells which produce antibodies. (figure from Klein and Dalla-Favera (2008), `http://www.nature.com/nri/journal/v8/n1/images/nri2217-f1.jpg`)

Usually lymphomas occur in the lymphatic system (nodal) and more specific the lymph nodes (about 75 %), but extranodal lymphomas, which arise from other organs, also occur. In those cases the stomach, bones, liver, spleen, renal and rarer heart, breast, salivary glands and thyroid can be affected. Extranodal lymphomas are usually aggressive forms. However, in many cases lymphoma invades other organs after developing in lymphic tissues. E.g., in nearly 25 % to 50 % of Non-Hodgkin-Lymphoma (NHL) patients, the tumor spreads into the liver (Holland, 2000).

1.1.2 Hodgkin's Disease

Lymphoma is classified into two different major types: The first is Hodgkin's disease (HD), which is characterized by the presence of Reed-Sternberg cells. Those cells presumably develop from mutated GC B-cells which were intended for apoptosis but for some reason (like deregulation of apoptosis genes such as p53) survived (Thomas et al., 2004). Reed-Sternberg cells are much larger than normal B-cells and often have multiple nuclei (Küppers, 2008).

HD has an incidence of about 2.80 per 100,000 persons per year and thus accounts for about 10 % of all lymphomas (Leukemia and Lymphoma Society, 2013). The proportion of HD is much higher in childhood: Until the age of 19, it accounts for more than 50 % of all lymphomas. The five years survival rate of HD is 86 % (Leukemia and Lymphoma Society, 2013).

1.1.3 Non-Hodgkin-Lymphoma

The second main type of lymphoma is the group of Non-Hodgkin-Lymphomas (NHLs). They form a diverse group of diseases which differ in their origin, histology, course and their treatment and prognosis.

With an incidence of 15 to 20 new cases per 100,000 per year (Howlader et al., 1975; Leukemia and Lymphoma Society, 2013; Fisher and Fisher, 2004), NHLs are much more frequent than HD and account for about 5 % of all new cancers (Holland, 2000). The incidence of NHLs highly depends on age and gender: For females between the ages of 20 and 24 the incidence (per 100,000 per year in the USA) is 1.9 while it is 2.9 for males of the same age. In the age group 60 to 64, the incidence for females is 39.0 and 54.70 for males (Leukemia and Lymphoma Society, 2013).

Risk factors apart from age and gender are various infections, especially with Epstein-Barr virus (EBV) (Gibson et al., 2011; Hoeller et al., 2010), Human Immodeficiency Virus (HIV) (Swerdlow, International Agency for Research on Cancer, and World Health Organization, 2008), Hepatitis-C Virus (HCV) (De Vita et al., 1997; Ascoli et al., 1998), and Helicobacter pylori. Furthermore, NHLs are associated with inherited and acquired disorders of the immune system (including autoimmune disorders). Even patients receiving immunosuppressive agents after e.g. an organ transplant share this increased risk for lymphoma. Besides those factors, NHLs are associated with various common cancer risk factors like radiation, prior chemotherapy and certain chemical substances like pesticides (Holland, 2000; Mey et al., 2012; Martelli et al., 2013).

The International Prognostic Index (IPI) is a classification scheme to assess the aggressiveness of NHLs. It summarizes certain risk factors and estimates the type based on the number of present risk factors. Those factors are amongst others age (over 60 or below), the spreading of the tumor and the level of lactate dehydrogenase in the serum. Tumors are classified into four stages: low-risk, low-intermediate-risk, high-intermediate-risk and high risk. Those differ strongly in their prognosis (73 % to 26 % overall survival (OS)) and the treatment is often based on the IPI.

Genetic Background

About 70 % to 90 % of all NHLs carry chromosomal abnormalities. Those are mostly translocations, but deletions and additions occur, as well. The modifications are not spread randomly over the genome but cluster at certain locations. The chromosomal region 14q is involved in 71 % of all translocations (particularly the translocation $t(14, 18)$ occurs in 85 % of all Follicular Lymphomas (FLs) – a subtype of NHL – cases) and 19 % of all NHLs have trisomy 12 (Le Beau, 1990; Bloomfield et al., 1983). In many cases Ig (or Ig regulatory regions) are fused with oncogenes causing their disregulation (Callén et al., 2007; Mey et al., 2012). A common example for this is the B-cell Lymphoma (BCL)-6 gene. It encodes the BCL-6 protein, a transcription factor with zinc finger structure, which regulates a group of genes including p53 and NF-κB which in turn repress apoptosis, differentiation and cell cycle progression (e.g. V-myc Avian Myelocytomatosis Viral Oncogene Homolog (MYC)) (Schwartz, Lenz, and Staudt, 2010; Martelli et al., 2013; Phan and Dalla-Favera, 2004). Another example is the activation of BCL-2 by amplification or translocation (20 %) which inhibits amongst others p53. The loss of these tumor suppressors allows malignant transformation. MYC itself is mutated or disregulated in 5 % to 10 % of all NHLs. Lymphomas carrying this mutation are associated with a bad diagnosis; 50 % of the affected patients suffer from progressive disease (PD) or partial response (PR) (Mey et al., 2012; Martelli et al., 2013).

Apart from those examples many other genes are mutated more often than expected by random. These specific mutations (and their combinations) determine – to a certain extent – the type of the lymphoma and hence influence histology and course of the disease (Le Beau, 1990; Bloomfield et al., 1983).

1.1.4 Diffuse Large B-Cell Lymphoma

As mentioned before, NHLs can be subdivided further into more homogenous types of lymphoma. This thesis will focus on the most common of those types – Diffuse

Large B-Cell Lymphoma (DLBCL).

Accounting for 30 % to 40 % of newly diagnosed NHLs (Schwartz, Lenz, and Staudt, 2010; Swerdlow, International Agency for Research on Cancer, and World Health Organization, 2008; The Non-Hodgkin's Lymphoma Classification Project, 1997), DLBCL is the most common subtype of NHL (Holland, 2000; American Cancer Society, 2013). Usually it occurs as primary tumor but it can also occur as secondary tumor after a disease with other, less aggressive types of lymphoma like FL, HD, lymphoplasmacytic lymphoma, or others. The median age of DLBCL diagnosis is between 60 and 70 years but it can occur at almost any age (Martelli et al., 2013; The Non-Hodgkin's Lymphoma Classification Project, 1997).

DLBCL is a histologically defined but nevertheless clinically and genetically very heterogeneous disease (Cigudosa et al., 1999). It is characterized by enlarged B-cells at least double the size of normal B-cells (Martelli et al., 2013) and either cleaved or non-cleaved nuclei, often with multiple nucleoli. When DLBCL grows, it does not show the usual pattern of round clusters like other lymphomas. Instead, it invades nearby tissue in a diffuse, disorganized pattern (Figure 1.2). The tumor cells usually origin from centroblasts or plasmablasts (Schwartz, Lenz, and Staudt, 2010) and thus express genes of mature B-cells and have a high mitotic rate (Holland, 2000; Martelli et al., 2013). In about half of the DLBCL neoplasms SHM is permanently continued (Martelli et al., 2013).

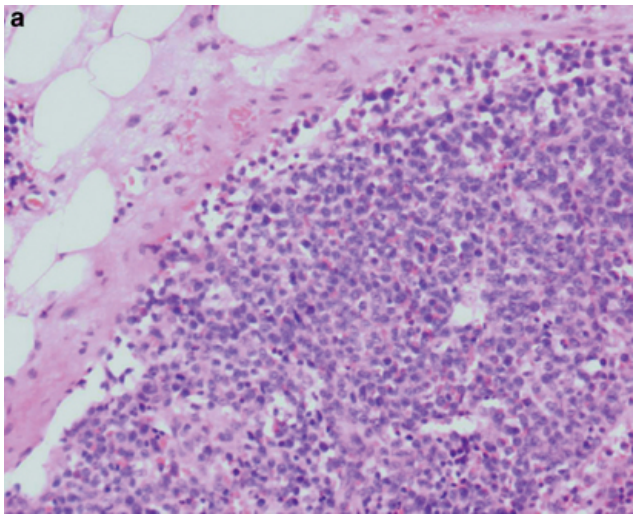


Figure 1.2: The histology of Diffuse Large B-Cell Lymphoma. The growth pattern of a DLBCL tumor can be seen. In the upper left corner the typical "diffuse" pattern of growth of lymphoma cells in the adjacent tissue can be seen. (from Morice et al. (2004))

1.1.4.1 Chromosomal/Genetic Changes in DLBCL

About 50 % of the DLBCL tumors have at least one mutation that leads to deregulation of oncogenes (Rao et al., 1998); many different genes are affected by those (Cigudosa et al., 1999). As for other NHLs, the genes BCL-6 (70 %) and BCL-2 are often deregulated by gene rearrangements (20 % to 30 %) or mutations in the regulatory regions (Rao et al., 1998; Cigudosa et al., 1999). The Reticuloendotheliosis Viral Oncogene Homolog (REL) gene which is part of the NF-κB pathway and controls various cellular processes like differentiation or apoptosis is amplified in 23 % of all cases (Rao et al., 1998)

1.1.4.2 Classification of DLBCL

According to the Cell-Of-Origin (COO) classification, DLBCL can be divided in different subtypes. The most common subtypes are the Activated B-cell Like (ABC) lymphoma and the Germinal-Center B-cell Like (GCB) lymphoma which together account for 85 % of all cases (Nathwani, 2012). Another less common subtype is the Primary Mediastinal B-Cell Lymphoma (PMBL). Even though those types do not differ in their histology, gene expression profiles vary significantly between the classes (Schwartz, Lenz, and Staudt, 2010; Cabanillas, 2011; Gupta et al., 2012). This classification is of clinical interest because of the different prognoses: GCB patients have 3- and 5-years survival rates of 87 % and 62 % respectively, while those rates are only 44 % and 26 % respectively for ABC patients (Wright, 2003; Choi et al., 2009).

The ABC type develops from post-GC B-cells which are arrested during differentiation to plasma cells (Gupta et al., 2012). The full differentiation is often blocked due to mutation of B-Lymphocyte-Induced Maturation Protein (BLIMP)-1. To suppress apoptosis, NF-κB is continuously active (Schwartz, Lenz, and Staudt, 2010; Cabanillas, 2011). There are efforts made to use this dependence on NF-κB for therapy e.g by inhibiting it with Bortezomib NF-κB (Cabanillas, 2011). ABC lymphomas remarkably express Ig-M which indicates that CSR has not been performed (Schwartz, Lenz, and Staudt, 2010). Furthermore, ABC tumors often have high levels of the ongogene AID (Schwartz, Lenz, and Staudt, 2010) and Signal Transducers and Activators of Transcription (STAT)-3 and are CD10 and BCL-6 negative. (Cabanillas, 2011)

In contrast to ABC, GCB tumors consist of mutated GC B cells (Gupta et al., 2012). They often have switched Ig classes due to CSR and SHM is continuously performed (Schwartz, Lenz, and Staudt, 2010; Martelli et al., 2013). In 39 % of the GCB cases, CREB Binding Protein (CREBBP) and p300 are inactivated by mutations. Both genes code for acetyltransferases which act as cofactors for the transcription of various

signaling genes. Amongst others the acetylation of BCL-6 is prevented by mutating these genes, leading to lacking suppression of BCL-6 (Martelli et al., 2013).

Similar to HD, PMBL is supposed to develop from thymic B-cells (Martelli et al., 2013). Moreover, it also has Reed-Sternberg like cells and its gene expression patterns are comparable to HD, as well (Rosenwald et al., 2003). In contrast to HD, PMBL tumors express genes of mature B-cells (which HD does not).

The validity of the COO classification is under discussion and other approaches exist. One example is the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (World Health Organization, 2004). It classifies lymphoma in the categories C81-C96 ("Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue") and more specifically DLBCL in the C83 category ("Non-follicular lymphoma") with the identifier C83.3. The DLBCL subtypes listed in the ICD-10 are "Anaplastic", "CD30-positive", "Centroblastic", "Plasmablastic", "Immunoblastic", "Subtype not specified" and "T-cell rich" lymphoma.

1.1.5 Symptoms

The most common symptoms of lymphomas are persistently swollen lymph nodes (lymphadenopathy). However, this can also be caused by infections or diseases affecting the immune system. In contrast to HD, where B-symptoms like fever, extreme weight loss and night sweats are very common, those occur in only 20 % of all NHL cases. Other symptoms for NHLs are pain in the bones, chest or the abdomen, cough and rashes (Holland, 2000; Leukemia and Lymphoma Society, 2013). There are no symptoms which are specific to NHLs and even asymptomatic courses occur.

1.1.6 Treatment

For more than 25 years the standard therapy for lymphoma-patients was a chemotherapy with a mixture of Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone (CHOP). The first two elements of it, cyclophosphamide and hydroxydaunorubicin damage Desoxyribonucleic Acid (DNA) following the idea that healthy cells can fix DNA damages faster than cancer cells can. Oncovin prevents cell division and the corticosteroid prednisone alters gene expression and has an anti-inflammatory effect.

About ten years ago, it was found that addition of Rituximab to the standard CHOP therapy (R-CHOP) significantly enhanced the prognosis. The percentage of patients older than 60 years with complete response (CR) was increased from 63 % to 75 % and five years' disease-free survival (DFS) from 45 % to 66 % (Feugier et al., 2005). For young patients with low-grade DLBCL it was shown that the addition of Rituximab enhanced the portion of patients with a three years' event-free survival (EFS) from 59 % to 79 % and the three years' OS from 84 % to 93 % (Pfreundschuh et al., 2006). The advantage of Rituximab was later shown for patients of any age in several studies (Sehn et al., 2005; Coiffier et al., 2010; Pfreundschuh et al., 2006). Dependent on the age and IPI, R-CHOP is usually given in 6 or 8 cycles every 14 or 21 days (Pfreundschuh et al., 2004). In total, 40 % to 60 % of all DLBCLs can be cured with R-CHOP therapy (Schwartz, Lenz, and Staudt, 2010; Cerchietti, 2013). Rituximab is a monoclonal antibody which targets the CD20 antigen which is expressed by all mature B-cells (Martelli et al., 2013). It selectively binds to CD20 and thereby enhances the efficiency of natural killer cells which leads to the depletion of B-cells (Rudnicka et al., 2013).

The question if radiotherapy should be applied after R-CHOP treatment for consolidation is still under debate (Pfreundschuh et al., 2006). Radiation can be applied, especially for extranodal sites (Mey et al., 2012; Leukemia and Lymphoma Society, 2013; Holland, Frei, and Kufe, 2003). However, recent studies failed to show a significant improvement or even suggested a worsened therapy success due to radiotherapy (Bonnet et al., 2007; Miller et al., 1998; Horning, 2004). The major problem of radiotherapy is the increased risk of new carcinogenesis.

Despite the enhanced prognosis due to Rituximab, 10 % still do not respond to the therapy and 10 % to 50 % (depending on the IPI) relapse after CR (Feugier et al., 2005; Martelli et al., 2013). In general one can say that prognosis is better the faster the CR is reached and the longer it holds. Patients with relapsed or refractory tumors have a bad prognosis (Weisdorf et al., 1992). Only 10 % of those patients reach a long DFS when treated again with standard chemotherapy (Martelli et al., 2013). Relapsed tumors are often treated with high dose chemotherapy (HDC) and Autologous Stem Cell Transplant (ASCT). Again, the addition of Rituximab enhances the cure rates (Martelli et al., 2013).

Various new drugs against lymphoma are in development (Martelli et al., 2013) for both separate therapies and improvements of R-CHOP. One family of such new drugs are Histone Deacetylase Inhibitors (HDACis).

Histone Deacetylase Inhibitors

Histones are proteins which together with the DNA form the chromatin. DNA is wrapped around the histones and thereby becomes condensed. Due to this condensation some parts of the DNA become unaccessible for other proteins and ca not be transcribed any more. As most processes in cells, this control of gene-expression by histones does not happen randomly but is influenced by various other processes. One of those is the modification of histone proteins. These (partially permanent) influence of histones on the DNA and gene expression is part of the epigenetic. One of the modifications, the (de-)acetylation of histones is controlled by Histone Deacetylases (HDACs).

But in contrast to what their names suggest, HDACs do not only control histones but also non-histone proteins like p35 and NF-κB (Bannister et al., 2000; Williams et al., 2006; Zhao et al., 2006; Cai et al., 2013; Lemoine and Younes, 2010). By regulating the degree of acetylation of various proteins, HDACs are involved in the regulation of cell survival, apoptosis, cell-cycle progression and angiogenesis (Lemoine and Younes, 2010). In humans, 18 different HDACs are known. Those are divided into three classes depending on their sequence homology with the according yeast homologs (Dokmanovic, Clarke, and Marks, 2007).

It was shown that Histone Acetyltransferase (HAT)-1 and HDAC-1 proteins are not expressed in normal B-cells but are highly abundant in DLBCL (and other) tumor cells (Gupta et al., 2012; Cai et al., 2013). Hence, they are thought to play an important role in the malignant transformation (Min et al., 2012). For that reason inhibitors of those proteins, Histone Deacetylase Inhibitors (HDACis), are developed as chemotherapeutic agents. An advantage of those agents is their low toxicity (Cabanillas, 2011).

One target of HDACs is the STAT-3 gene. This is a transcription factor which is expressed by DLBCL lymphomas of the ABC subtype (Cabanillas, 2011). It is a key regulator for development and various cellular processes like cell proliferation. STAT-3 has to be activated by phosphorylation and then forms a dimer before it can be imported into the nucleus. This dimerization is inhibited by an acetylation, which thus prevents its nuclear import. Hence, STAT-3 activity depends on HDACs. It was even shown that STAT-3 interacts physically with the HDAC p300 (Gupta et al., 2012; Klampfer et al., 2004). Elevated levels of HDAC in certain lymphoma cells lead to higher levels of deacetylated STAT-3 and thus to more active STAT-3 in the nucleus. The inhibition of HDAC with HDACi is counteracting this effect (Gupta et al., 2012; Cabanillas, 2011; Lee et al., 2012).

Another potential target of HDACis is BCL-6, which is crucial for the development of GCB subtypes. The activity of this transcription factor which regulates amongst others p53 and NF-κB is inhibited by acetylation. The treatment with HDACi increases the degree of acetylation of BCL-6 and leads to its reduced activity. This increases amongst others the expression of p53 and enhances apoptosis (Bereshchenko, Gu, and Dalla-Favera, 2002; Cabanillas, 2011; Lemoine and Younes, 2010; Phan and Dalla-Favera, 2004).

However, the exact mode of action of HDACis remains unknown and several other genes/proteins are supposed targets and epigenetic changes most likely play an important role as well(Lemoine and Younes, 2010; McLaughlin and La Thangue, 2004; Drummond et al., 2005).

One specific HDACi, which inhibits all three classes of HDACs, is Panobinostat (aka. LBH589)[1]. It was shown that Panobinostat increases the degree of acetylation of STAT-3 in lymphoma cell lines and decreases the STAT-3 concentration in nucleus (Gupta et al., 2012). Another study showed that Panobinostat promotes inhibition of cell growth, differentiation and proliferation and activates DNA damage repair genes as well as apoptosis genes in acute lymphoblastic leukemia cell lines (Scuto et al., 2008). In mice infected with adult T-cell leukemia/lymphoma, Panobinostat reduced tumor volume by up to 70 % by activating apoptosis pathways (Hasegawa et al., 2011).

A number of clinical trials are currently investigating the efficiency of Panobinostat in the treatment against HD (Dickinson et al., 2009), Cutaneous T-cell Lymphoma (CTCL), Multiple Myeloma (MM) and others (Stimson et al., 2009; Drummond et al., 2005). Most studies focus on refractory/relapsed lymphomas.

This thesis analyzes data from a clinical trial which investigates the treatment of relapsed/refractory DLBCL with Panobinostat[2]. For a deeper understanding of resistant DLBCL and the drug Panobinostat metabolite levels from serum and plasma samples are analyzed using an untargeted metabolomics approach.

1.2 Metabolomics

In each living organism thousands of chemical reactions are carried out. The complete set of all those reactions is referred to as metabolism. There are basically two types of molecules involved: “big” molecules (with a high molecular

[1] by Novartis Pharmaceuticals

[2] `http://clinicaltrials.gov/ct2/show/NCT01238692`

weight) like proteins and nucleic acids and "small" molecules (with a low molecular weight) like sugars, fatty acids, nucleotides which are summarized under the term metabolites (Stryer, 2012). Following the concept of genome/genomics and proteome/proteomics, the entirety of all metabolites is called metabolome and their systematic study is named metabolomics (Patti, Yanes, and Siuzdak, 2012; Kaddurah-Daouk, Kristal, and Weinshilboum, 2008). There are many different techniques for metabolomics, e.g. Nuclear Magnetic Resonance Spectrometry (NMR), Liquid Chromatography with Mass Spectrometry (LC/MS) or Gas Chromatography with Mass Spectrometry (GC/MS) in which mass spectrometry (MS) again is more a set of techniques than a single technique (Kaddurah-Daouk, Kristal, and Weinshilboum, 2008).

Two conceptually different approaches exist: targeted and untargeted metabolomics. While targeted metabolomics is defined as "measurement of defined groups of chemically characterized and biochemically annotated metabolites" (Roberts et al., 2012), untargeted metabolomics is the measurement of all (or as many as possible) metabolites within a sample (Roberts et al., 2012; Patti, Yanes, and Siuzdak, 2012). Both techniques have their advantages and disadvantages. Targeted analyses focus on a known, annotated set of metabolites, which are supposed to have an effect on the phenotype of interest (hypothesis-driven). This means that the result set is restricted to this predefined set. In contrast, one can uncover associations of phenotypes with any – even unexpected – metabolites when using untargeted metabolomics (hypothesis-generating). However, untargeted techniques usually have lower sensitivity, a lower dynamic range and are less robust than targeted metabolomics (Patti, Yanes, and Siuzdak, 2012; Baker, 2011).

One main characteristic of untargeted metabolomics is that the identification of metabolites is done as post-processing step after the actual experiment. When using LC/MS ions with unique retention times, mass/charge ratios and fragmentation spectra are identified. Each metabolite shows one or more of these ions and can thus be uniquely described. However, the chemical structure and other characteristic about the metabolite cannot be gathered from these information in a straightforward manner. Usually, known metabolites of interest have to be purified and then measured. The according retention time, mass/charge ratio and fragmentation spectrum are stored in a database which then can be used to compare the experiment results. However, many metabolites have not yet been identified. Their identity remains unknown as well as their function and involvement in pathways or reactions. This limits the "freedom" of untargeted metabolomics: All unexpected metabolites can be measured, but if there is no information available about them interpretation is limited (Patti, Yanes, and Siuzdak, 2012).

1.3 Aim

The main aim of this thesis is to investigate the effect of Panobinostat on the metabolism of DLBCL patients. Therefore, metabolites were measured before the treatment and again after the treatment with Panobinostat. The target of this comparison is to gain insights into the mode of action of Panobinostat, discover possible side effects and unveil possibilities for the improvement of the drug.

Furthermore, three more tasks are addressed in this thesis: The first one is to determine metabolic differences between the primary tumor (which is sensitive to R-CHOP) and the relapsed/refractory tumor. Understanding these differences is important to be able to design drugs specifically for R-CHOP resistant lymphomas.

The second challenge is to investigate if metabolic profiling of serum helps to distinguish patients who will be cured due to R-CHOP from those with resistant DLBCL before the treatment starts. As DLBCL is an aggressive form of cancer, time is crucial. If one could predict that R-CHOP treatment will fail high accuracy, the patient could receive another treatment – for example Panobinostat – without wasting time.

The last goal is to identify differences between serum and plasma of patients suffering from primary DLBCL (both primary, sensitive, untreated). The objective of this comparison is gaining new insights into lymphoma which might e.g. facilitate the design of new experiments.

In the following, a lymphoma case control study with untargeted metabolomics data is analyzed. The following chapter describes the analyzed data, how it was measured and its characteristics. Subsequently, the methods applied to the data are described before the results are presented and discussed.

2 Data

In this thesis, a case-control lymphoma study with altogether 105 patients was analyzed. The study was subdivided into five separate metabolomics experiments – two of them containing serum, two plasma samples and the last one samples from DLBCL cell lines. For each sample the metabolic profile was determined using an untargeted metabolomics approach (for details see sections 2.2 and 2.3). In total, 1097 metabolites were measured (c.f. Table 2.1). 527 of them are known chemical compounds with annotated pathway information while the remaining 570 are unknown.

The analyzed study is part of the clinical trial *A Randomized Phase II Study of Oral Panobinostat (LBH589) With or Without Rituximab to Treat Relapsed or Refractory Diffuse Large B Cell Lymphoma* (*NCT01238692*). The aim of this trial is amongst others to investigate the efficiency, safety and tolerability of Panobinostat.

In the following sections, the patient selection, study design and experimental settings are described. In the second part the preprocessing procedure is explained before an assessment of data characteristics and data quality is presented.

exp ID	*exp name*	*sample type*	*# samples*	*# metabolites*	*# days*
301	**primary-serum-1**	serum	26	765	1
302	**primary-serum-2**	serum	18	786	2
402	**pano-plasma-1**	plasma	45	605	2
602	**pano-plasma-2**	plasma	16	583	1
	pano-cells	cell lines	8	276	1
total			105	1097	

Table 2.1: Experiment Overview. The data was measured in four different experiments, each of them containing samples from independent patients. The number of patients and measured metabolites per run is shown. In the following, the experiments are referred to by their experiment name (second column).

2.1 Patient Inclusion Criteria

To participate in the study, patients had to meet certain criteria. The most important of them was the pathological confirmation of DLBCL in any stage. Furthermore, the patient must have had at least one prior treatment (usually with R-CHOP) and, wherever applicable, another treatment with ASCT and must have fully recovered from those treatments. The last treatment had to be finished at least three weeks before the registration for the trial. One of the previous treatments must have contained Rituximab; this treatment must have been finished at least six months before registration. Other inclusion criteria were the minimum age of 18 and an Eastern Cooperative Oncology Group (ECOG) status (Oken et al., 1982) of 0 or 1. This means that the patient must not be restricted by the disease at all or only in "physically strenuous activity", can be treated ambulatory and is able to do light work (e.g. office work) [1].

Patients, who received a prior cancer treatment with HDACi or comparable agents were excluded from the study as well as patients who showed severe concomitant diseases including cardiovascular disease, pulmonary disease, hypertension, unresolved diarrhea and impairment of gastrointestinal functions. Furthermore, patients with HIV or HCV infection were excluded. A complete list of inclusion and exclusion criteria can be found online[2].

2.2 Study Design

Serum and plasma samples in the four experiments were taken from independent patients at different stages of the disease (Figure 2.1). The first probes were taken after the initial DLBCL diagnosis, before the treatment with R-CHOP. These were later classified as primary, no relapse, untreated if the patient was cured and a EFS of two years. Otherwise the sample was classified as primary, relapse, untreated which includes relapsed as well as refractory lymphomas. To investigate the effect of Panobinostat on the metabolism, plasma samples were taken from relapsed patients before and after the treatment. Those are paired samples which were taken from the same patients.

The analyzed study was designed and conducted by Leandro Cerchietti from Weill Cornell Medical College, New York.

[1] `http://ecog.dfci.harvard.edu/general/perf_stat.html`
[2] *NCT01238692*, `http://clinicaltrials.gov/ct2/show/record/NCT01238692`

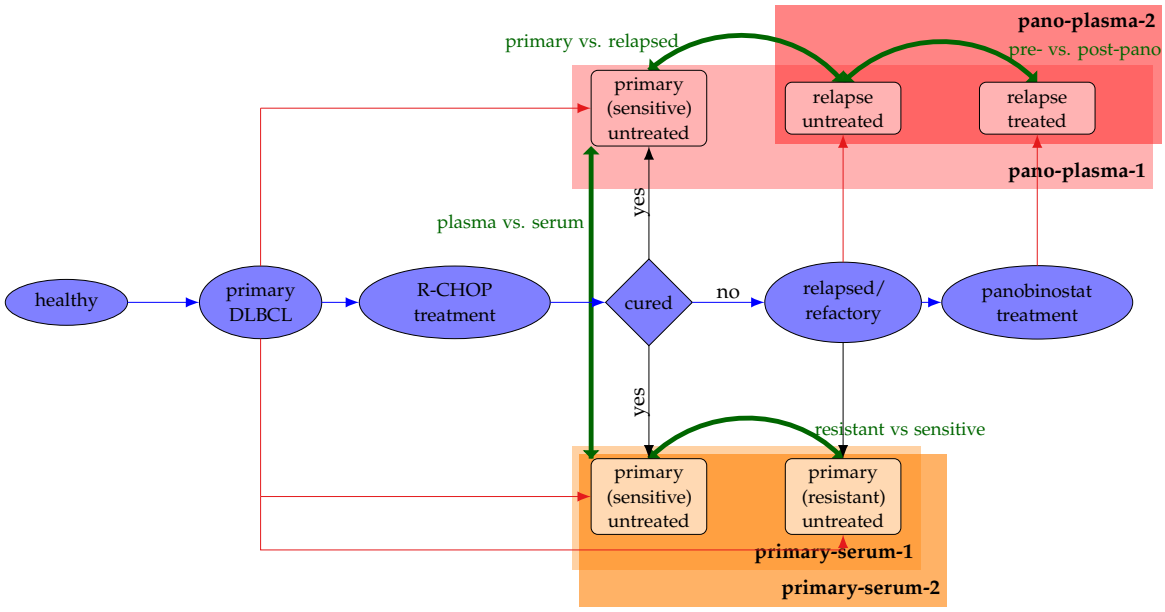


Figure 2.1: Overview over the study design. The central blue path indicates the progress of the disease, the upper red part illustrates the plasma samples and the lower orange rectangles the serum samples. Red arrows show the state of the disease and treatment at which the samples were taken. All serum samples were taken at the time of diagnosis of the primary tumor and classified later as sensitive or resistant. The transparent red and orange rectangles show which samples were measured in which experiment. In general, all samples are taken from different patients and therefore independent from each other. The only exception are the samples taken before and after the treatment with Panobinostat (plasma relapse untreated and plasma relapse treated); those are paired samples. The green arrows show which groups of samples were compared to each other.

As already mentioned in section 1.3, the main aim of this thesis is to investigate the influence of Panobinostat on the metabolism. This and three other comparisons are analyzed and discussed in the following chapters. For each of those comparisons different groups of samples were compared (as indicated in figure 2.1). They are listed in the following:

Pre- vs. Post-Pano The plasma samples "relapse untreated" were compared with "relapse treated" to unveil differences in the metabolism caused by Panobinostat treatment. This comparison contains samples from the experiment **pano-plasma-1** and all samples from **pano-plasma-2**. For this comparison, metabolomics data from a DLBCL cell line was analyzed in the same manner

as the in vivo data. This additional data is intended to facilitate the separation of primary effects due to the treatment from secondary effects e.g. caused by the liver. The goal of this comparison is to gain insights in the mode of action of Panobinostat and possible side effects.

Primary vs. Relapsed The comparison of the plasma samples "primary (sensitive) untreated" and "relapse untreated" was performed to investigate metabolic differences between the two different types of tumors, primary and relapsed. Therefore, samples from the **pano-plasma-1** and **pano-plasma-2** experiments were used. Information on the differences of these tumor types might facilitate the treatment of relapsed tumors.

Resistant vs. Sensitive For the classification of tumors as resistant or sensitive to R-CHOP treatment prior to the treatment, the serum sample groups "primary (sensitive) untreated" and "primary (resistant) untreated" were compared. This includes the experiments **primary-serum-1** and **primary-serum-2**. The distinction of these groups would allow to treat patients with R-CHOP-resistant DLBCL with an appropriate agent without spending time on unsuccessful treatments.

Plasma vs. Serum The samples "primary (sensitive) untreated" were compared between serum (measured in **primary-serum-1** and **primary-serum-2**) and plasma (experiment **pano-plasma-1**). Knowledge about their differences might facilitate the choice between serum and plasma for upcoming experiments. At the time of sample taking all patients in this comparison were diagnosed to have primary DLBCL and did not receive any treatment.

2.3 Experimental Procedure

The study design and sample acquisition was done by Leandro Cerchietti at the Weill Cornell Medical College. All data were measured by Metabolon Inc.[3], a commercial fee-for-service provider for metabolomics measurements, following their standard protocol. For the sample preparation the *MicroLab® STAR*[4] robotic system was used. For the pre- vs. post-pano comparison, OCI-Ly1 DLBCL cells were treated with Panobinostat for 24 h at 1 µM concentration. Cells were suspended in as little as possible water to obtain a homogenous solution. At least 150 µl were needed per sample, 100 µl for plating, 25 µl for protein determination and another

[3]Durham USA, `http://www.metabolon.com/`
[4]Hamilton Company, Salt Lake City, UT

25 µl for the c-matrix. 100 µl of each heparin-plasma, serum and cell sample was mixed with 400 µl extraction solvent (ethanol with recovery standards) and shaken for two minutes using the *Geno/Grinder 2000* [5]. Proteins were removed from the solution by centrifugation and the remainder was split in four aliquots: three for chromatography and one as reserve. The organic solvent was removed from each of them using a *Zymark TurboVap®*[6] before they were freeze-dried under vacuum.

A *Waters Acquity UPLC system*[7] with a directly attached *LTQ* mass spectrometer[8] was used to perform LC/MS measurements. It contains an electrospray ionization source and a linear ion-trap analyzer. Two aliquots were used for LC/MS, one optimized for acidic, positive ions and the other for basic, negative ions. Both were measured with separate injections in dedicated columns. For the first one, optimized for positive ions, 0.1 % formic acid in water and methanol respectively was used. The second one for basic ions was performed with 6.5 µM ammonium bicarbonate at pH 8.0 and water/methanol, as well.

The third aliquot was used for GC/MS on a *Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole MS* using electron impact ionization. The column contained 5 % phenyl and the temperature was ramped from 40 °C to 300 °C within 16 min (Evans et al., 2009; Boudonck et al., 2009).

Identification of metabolites was carried out by comparing LC/MS and GC/MS data to a proprietary database using a multi-parametric test. This database contains known metabolic entities as well as recurrent unknown metabolites. The metabolite identification as well as the peak identification were manually revised by Metabolon employees (Galagan et al., 2013).

Furthermore, Metabolon offers pathway information for each metabolite at two levels of detail: sub-pathways and super-pathways. 8 different super-pathways which contain 26 to 107 metabolites (mean 66) and 79 sub pathways which contain between 1 and 44 metabolites (mean 7) were defined. The list of all metabolites and their pathway assignments can be found on the supplemental DVD (cf. Section A).

[5] Glen Mills Inc., Clifton, NJ
[6] McKinley Scientific, Sparta, NJ
[7] Waters Corporation, Milford, MA
[8] Thermo Fisher Scientific Inc., Waltham, MA

2.4 Preprocessing

To gather comparable metabolite intensities across run days and experiments, all metabolites measured in serum and plasma samples were normalized by dividing their intensities by the median intensity of the metabolite within the run day (or the experiment if measured on one day).

Cell samples were scaled differently. For each of those the protein content was determined using the Bradford assay (Bradford, 1976). The metabolite intensities measured for cell samples were scaled by the amount of protein in the solution, determined by this assay. This was done to adjust the intensities for varying amount of cells in the samples. Apart from this, cell measurements were not processed or normalized in any other way.

For serum and plasma samples, the following preprocessing steps were performed.

2.4.1 Logarithmic transformation

Most statistical tests and methods like e.g. linear regression assume normally distributed data. To match this assumption metabolite intensities across all experiments were tested for normal distribution using the Shapiro-Wilk test for normality (Shapiro and Wilk, 1965; Royston, 1982). This tests the null hypothesis H_0 that the sample is normally distributed by comparing the variance of the sample with the variance of an estimated normal distribution. Thus high p-values suggest that the data is distributed approximately normally and lower p-values indicate deviations from normal distribution.

This test for normality was repeated after log-transformation and compared to the results before transformation, by calculating their ratio p_{fc} as defined in formula 2.1. Metabolites which were measured in less than 10 samples were excluded from this procedure.

$$p_{fc} = \frac{\text{p-value}_{\text{ after transformation}}}{\text{p-value}_{\text{ before transformation}}} \tag{2.1}$$

A $p_{fc} < 1$ means that logarithmic transformation worsens the normality of the metabolite. The distribution of p_{fc}'s is shown in Figure 2.2. 90 % of the 823 tested metabolites have $p_{fc} > 1$ which means that the log-intensities approximate normal distribution better than non-transformed intensities (examples are shown in figures

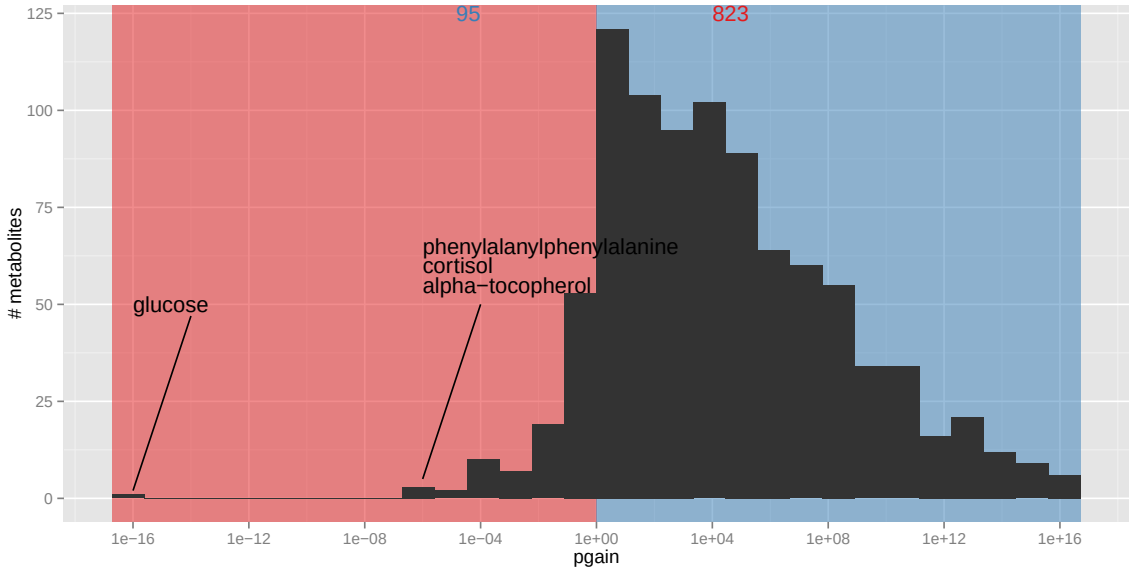


Figure 2.2: Logarithmically transformation of metabolite intensities. To achieve normally distributed metabolite intensities for serum and plasma samples, all values were logarithmized. To assess the effect of this transformation, a Shapiro-Wilk test for normality was performed before and after the transformation for each metabolite. The histogram shows the distribution of p_{fc} values which compare both tests.

2.3a and (b)). In 95 cases logarithmic transformation worsens the normal distribution (Figure 2.3c). Those were in average measured for less samples than the metabolites with $p_{fc} > 1$ (mean number of measurements 48 compared to 60). 179 metabolites were excluded from this tests because of too many missing values.

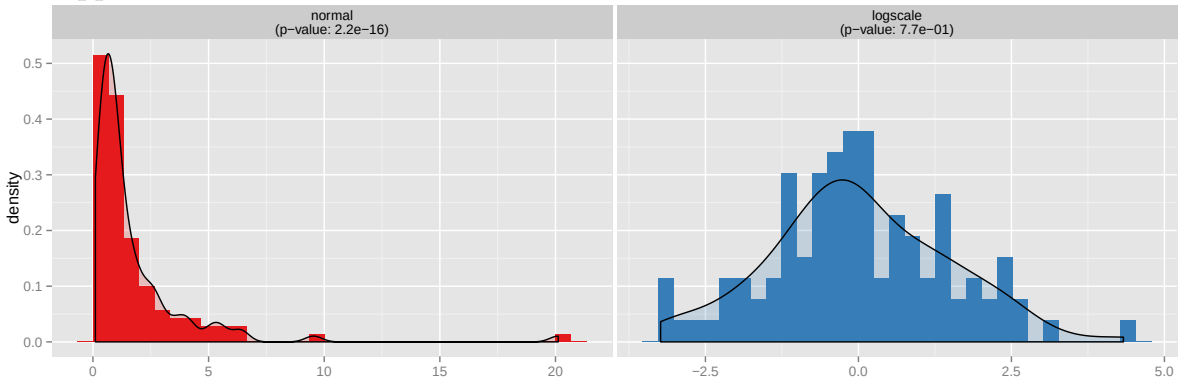
Consequently the measures intensities, including measurements from cell samples, were $\log 2$ transformed.

2.4.2 Outlier Removal

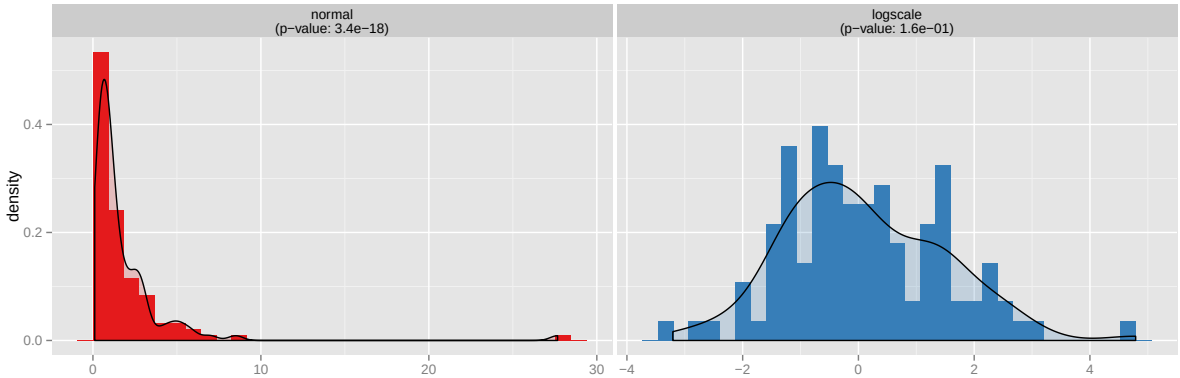
After the logarithmic transformation, values which differed more than three standard deviations from the mean intensity of each metabolite were excluded from all further analysis.

In average, 0.41 measurements were removed for each metabolite, 46.78 % of them due to extremely low values and 53.20 % due to extremely high values. The distribution of removed measurements per metabolite can be seen in Figure 2.4

(a) hippurate



(b) phenol sulfate



(c) 1,5-anhydroglucitol

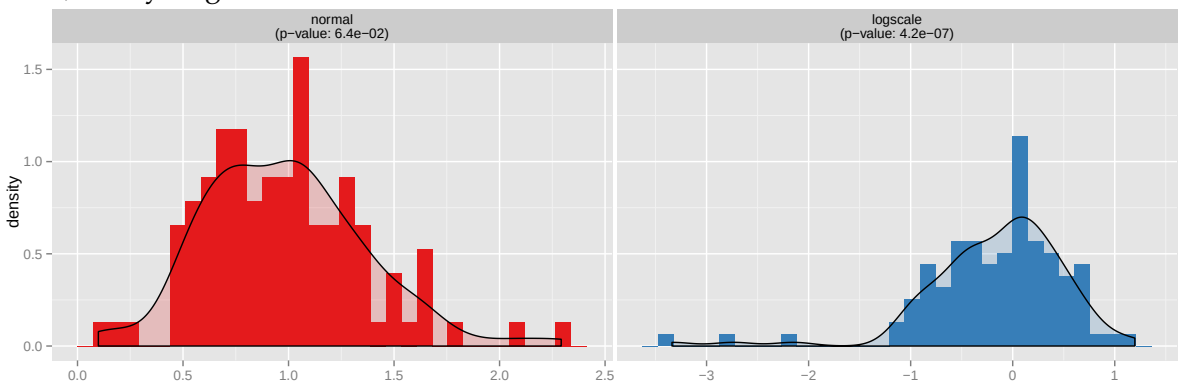


Figure 2.3: Examples for logarithmically transformed metabolite intensities of serum and plasma samples. Each figure (a)–(c) shows the density of metabolite intensities for all serum and plasma samples combined. Histograms in the left column (red) show the distribution of raw metabolite intensities while histograms in the right column (blue) show log-transformed values. Furthermore, the p-value of a Shapiro-Wilk-Test (testing for normal distribution) is shown for each distribution.

Figures (a) and (b) show metabolite intensities which approximate normal distribution better after logarithmic transformation. (c) is an example for one of the rare cases where this was not the case.

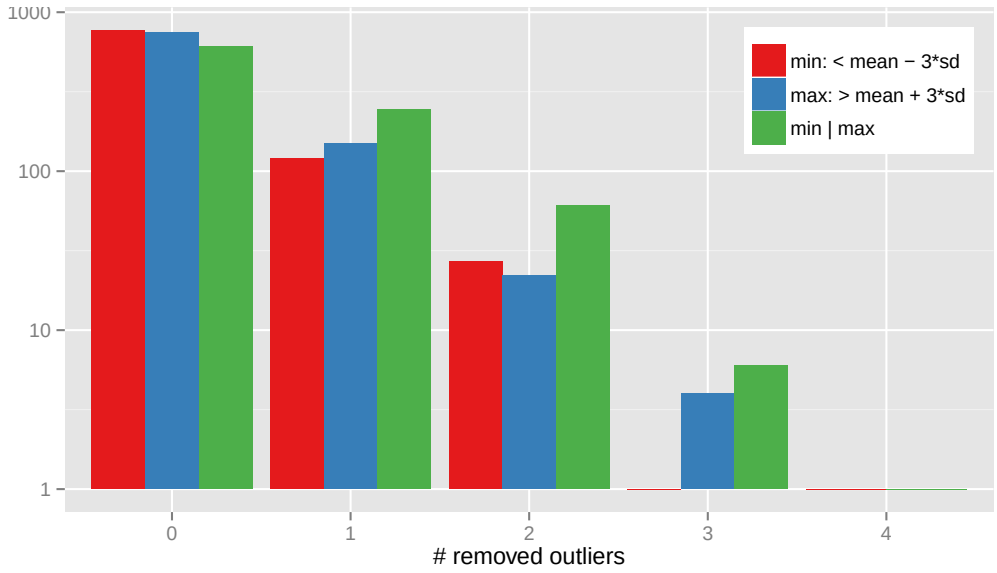


Figure 2.4: Removed outliers. The histogram shows the number of observations per metabolite which are more than $3 \cdot sd$ lower than the mean intensity of this metabolite (red), more than $3 \cdot sd$ higher than the mean intensity (blue) and the number of observations which differs more than $3 \cdot sd$ from the mean in any direction (green).

2.4.3 Exclusion of Metabolites

If a metabolite was measured in less than ten samples of any of the compared groups, it was excluded. For the *resistant vs. sensitive* and the *serum vs. plasma* comparisons this number was lowered to eight as the plasma, sensitive, untreated group, which is part of both comparisons, only contains nine samples (cf. Figure 2.5). For cell samples, a metabolite had to be measured in each of the four samples of both groups to be considered further.

2.5 Assessment of Metabolomics Data

The following section describes characteristics of the participating patients for each of the above defined problems. The number of patients per group varies from 9 to 33. In the *pre- vs. post-pano* comparison both groups contain the same number of samples because of the pairing. This is not the case for all other comparisons in which those numbers differ significantly (cf. Figure 2.5).

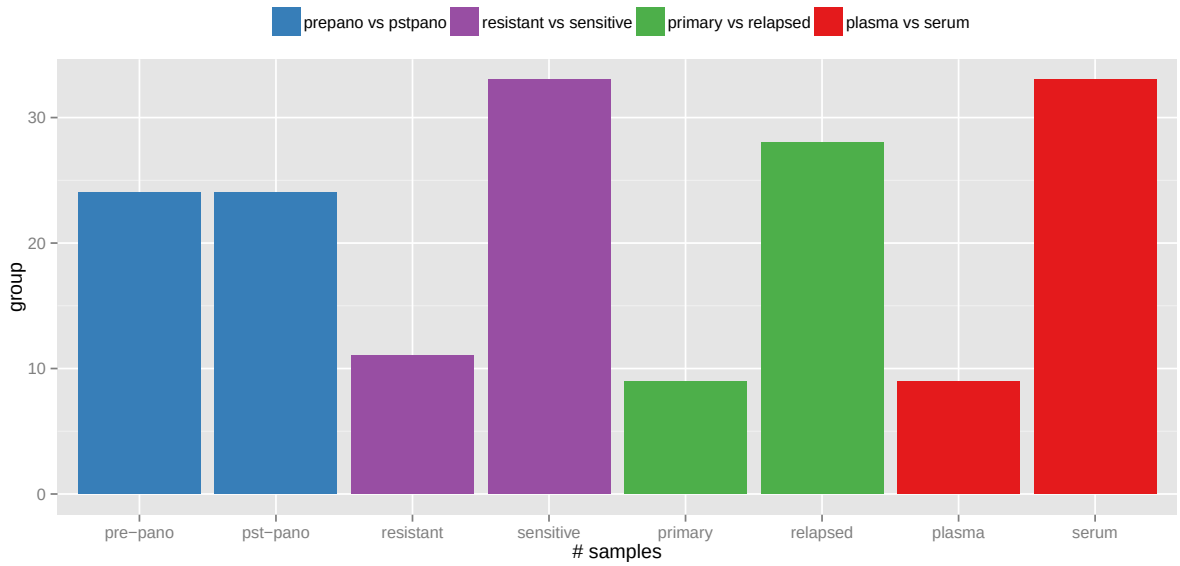


Figure 2.5: Number of samples per group. The histogram shows the number of samples per group; the colors indicate which groups are compared to each other. There are overlaps between the groups.

The overall age at time of diagnosis varied between 18 and 89 years. It is approximately normally distributed with median 59 and standard deviation 13.70. The distribution of the age of all patients and stratified by the four different problems is shown in Figure 2.6. 60 % of the participating patients were male. In average males in the study were 4 years older than females (Figure 2.8). It was previously shown that metabolite concentrations depend heavily on age and gender (Mittelstrass et al., 2011; Menni et al., 2013). As these vary significantly between patients, metabolite intensities were corrected for influences due to these factors in the following (see section 3.1).

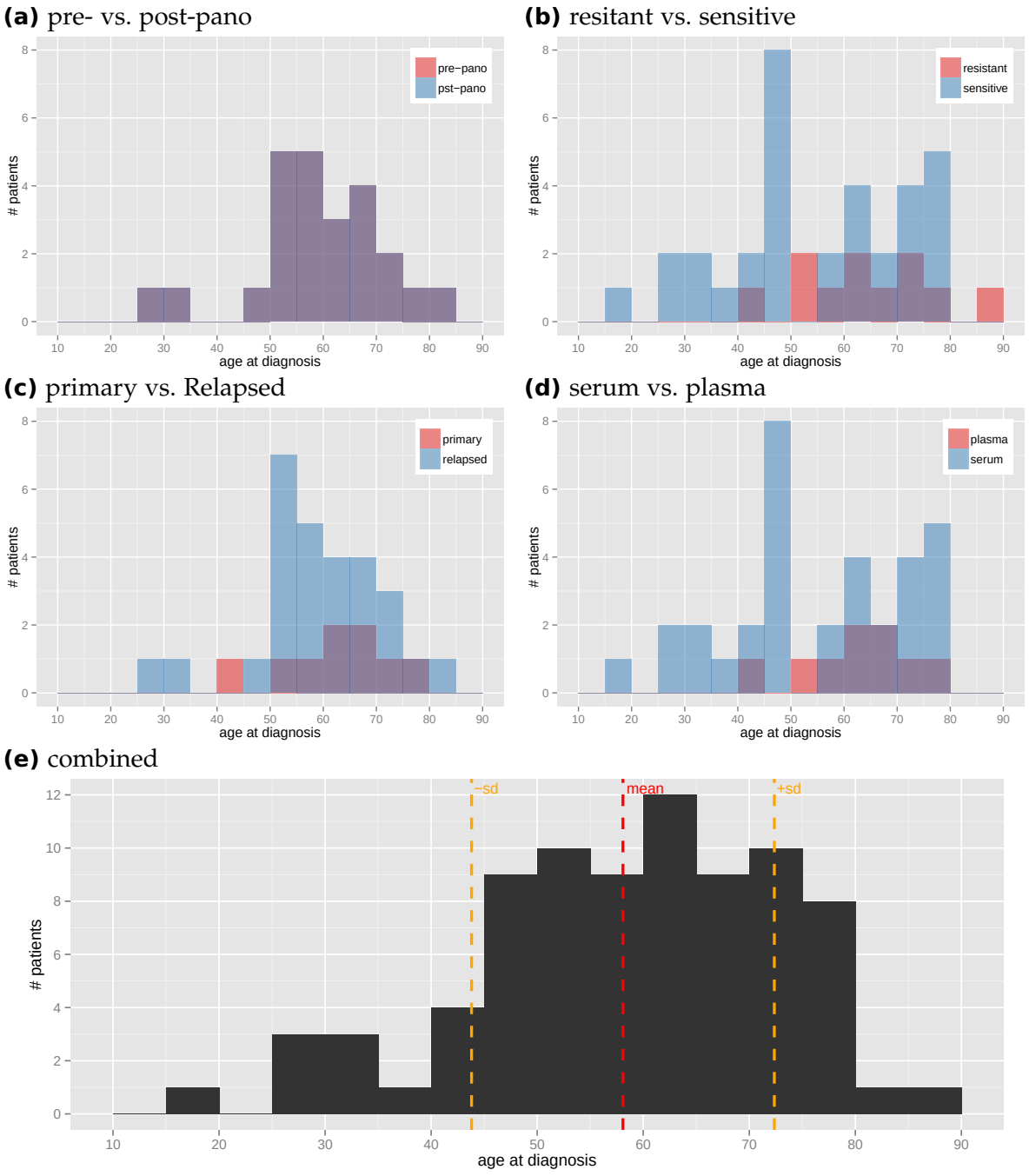


Figure 2.6: Age distribution. The figure shows the age distributions of samples compared between the two groups of all four comparisons in (a) - (d). (e) shows the age distribution of all patients combined. The red line indicates the mean age, orange lines its standard deviation.

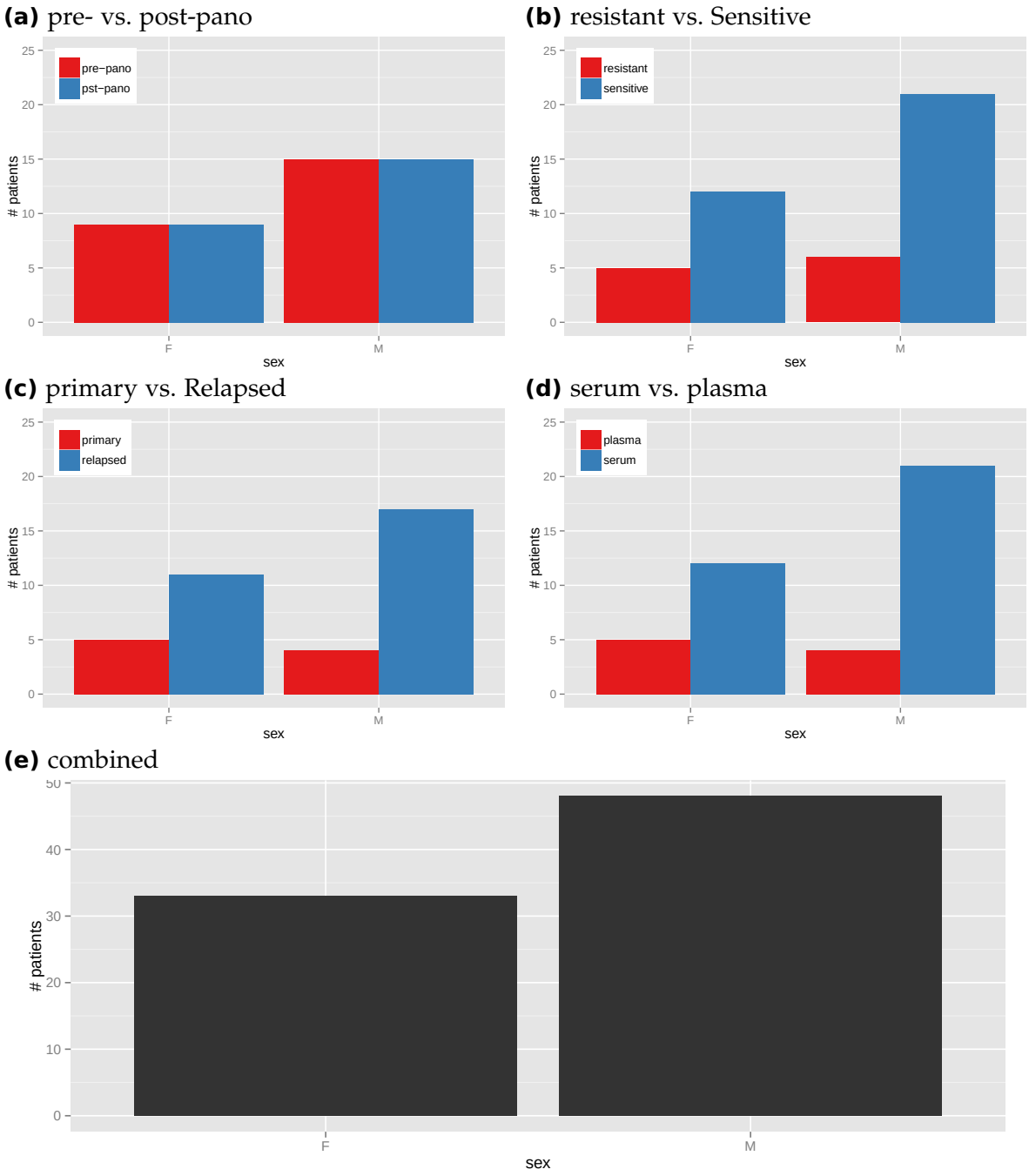


Figure 2.7: Sex distribution. The proportion of sexes between two groups for the four comparisons is shown in (a)-(d). The proportion of sexes for all samples combined is shown in (e).

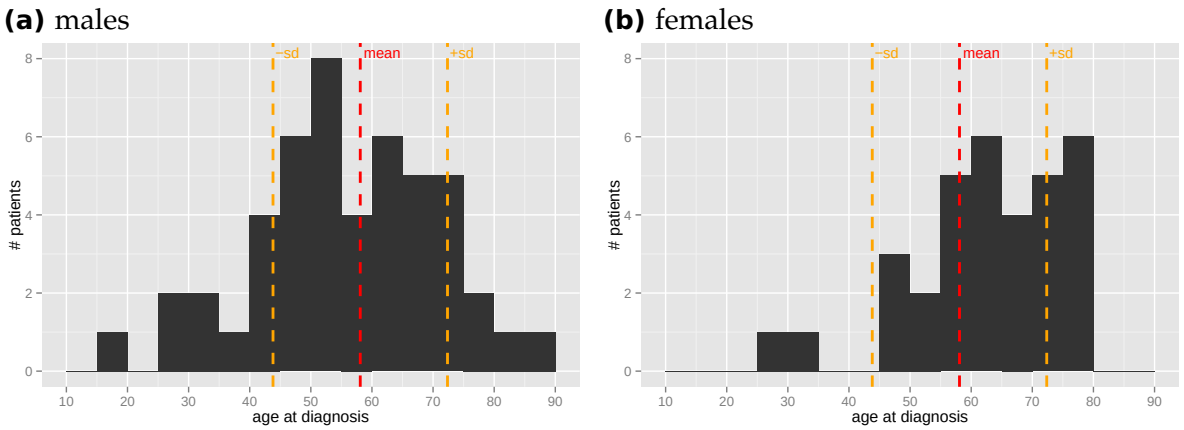


Figure 2.8: Age distribution stratified by gender. The histograms show the age distribution of all patients (as in Figure 2.7e) stratified by gender. The red lines indicate the mean age and the orange lines the standard deviation.

2.6 Data Quality

2.6.1 Metabolon CV Data

The robustness of metabolite measurements vary a lot between different techniques and even between different metabolites. As part of the KORA F4 study standard serum samples were measured for quality control and the Relative Standard Deviation (RSD) was determined for 267 metabolites from replicates (Suhre et al., 2011). It differs between 0 % (e.g. 4-acetamidophenol and ascorbate) and 204 % (levulinate). The latter is the most extreme exception, only 33 metabolites had a RSD $> 30\,\%$ and 8 had RSD $> 50\,\%$. In the other extreme 71 metabolites were observed to have RSD$< 10\,\%$.

Most of the highest RSDs can be observed in glycerophospholipids and glycerol-bound fatty acids. Amino acids in general have lower RSDs as well as hormones.

For the analyzed experiments there were no separate quality control samples available. Instead differences between the experiments and between run days were investigated. However those are not technical replicates and a considerable portion of variance within the data is expected.

2.6.2 Run Day and Experiment Comparison

Data often varies greatly between different experiments or run days (in case of experiment **primary-serum-2** and **pano-plasma-1**). In the following those differences between run days and experiments are shown. One common difference is that metabolites are only measured in one experiment but missing in (similar) experiments. Of course there are metabolites which were not measured because they are lower concentrated than the limit of detection of the technique, but there are other reasons. After the experiment Metabolon employees, so called data curators, manually analyze the measured data and exclude metabolites which do not meet their quality standards. As this assessment is subjective it differs between experiments. Another experimental bias is ion suppression. This is a matrix effect through which some ions – which might be highly concentrated in the sample – are not detected. The reasons for this effect are largely unknown, it might be caused by interactions of ions or e.g. coeluting of compounds (Annesley, 2003).

To investigate the strength of this influences the proportion of missing values NA_i of each metabolite M_i in run r was calculated as

$$NA_i = \frac{\sum_{s \in r} is.missing(M_i^{(s)})}{|s \in r|} \tag{2.2}$$

where $s \in r$ is the set of samples which are part of run r and $is.missing$ the indicator function

$$is.missing(M_i^{(s)}) = \begin{cases} 0 \mathrel{\widehat{=}} FALSE & \text{metabolite } i \text{ was measured for sample } s \\ 1 \mathrel{\widehat{=}} TRUE & \text{if } x \text{ otherwise} \end{cases} \tag{2.3}$$

The pairwise differences of these proportions between the runs are shown in Figure 2.10. It was distinguished between metabolites which were not measured at all in one run (meaning $NA_i = 1$, indicated by dark bars) and those which were measured at least once. The overlap of the measured metabolites between all runs is shown in Figure 2.11.

Besides the missing value patterns, metabolite intensities were tested for differences between the runs using Welch's t-tests (see Section 3.2.2 for details). This was done to test if there are metabolites whose concentration differed significantly between run days. As the groups (like pre- and post-pano) are randomly distributed between run days, significant differences between the concentrations might be a sign for biases. Causes for such biases are different people who carry out the experiments, different machines used on the days, contamination of probes, etc. Before testing

for differences, metabolite intensities were median scaled and logarithmized as described before.

The runs **primary-serum-2** and **pano-plasma-1** were measured at two days. The patterns of missing values (figures 2.9b and (d)) and the differences in metabolite concentrations between run days (figures 2.9a and (c)) were investigated as described above. In general the run days are very similar. None of the metabolites differed globally significant between the run days. The most diverging between the run days of **primary-serum-2** were 3-hydroxybutyrate and citrulline. Between the two run days of **pano-plasma-1** 5-oxoproline, glycerate and arginine differed the most (besides some unknown metabolites). Even if there are some metabolites which were measured on both days at different frequencies, the intersect of measured metabolites on the different days is very high. Only 29 and 41 metabolites respectively were measured in only one of the run days of **primary-serum-2** and **pano-plasma-1** but not the other. In those cases the metabolite was in average measured in less than 10 % of the samples of the other run. Consequently the run day separation was ignored in the following.

The overlap of measured metabolites between both serum experiments and both plasma experiments is very high. This can be seen in the comparison of the experiments **primary-serum-1** with **primary-serum-2** and **pano-plasma-1** with **pano-plasma-2** respectively (2.10a and (b)). There are much more metabolites which were only measured in one experiment than metabolites which were measured in both experiments at different frequencies. This suggests that different filtering, ion suppression or other experiment specific biases caused this differences.

When considering the comparison of plasma samples with serum samples (figures 2.10c-(f)) it is obvious that the differences are substantial. In each of the pairwise comparisons about one third of the metabolites was measured in only one of the experiments. This is of course at least partially caused by the differences between serum and plasma. However, these true differences can not be distinguished from differences due to biased. Thus samples from an experiment in which a metabolite was not measured at all were excluded when analyzing missing value structures of this metabolite (for further information see section 3.2.4). In general more metabolites were measured in serum samples. This can also be seen in Figure 2.11.

About one third of all metabolites (362) was measured in all runs (Figure 2.11).

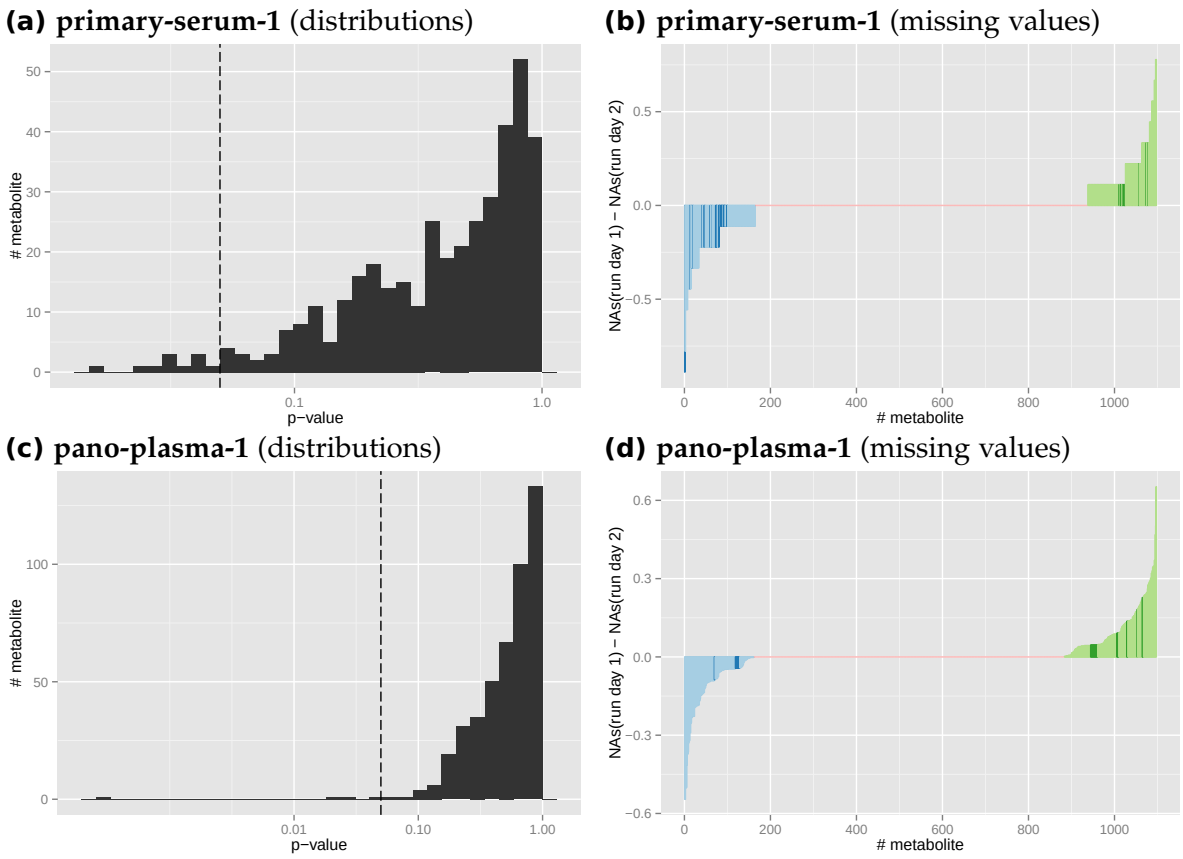


Figure 2.9: Run day comparison. For both experiments that were measured on two days the metabolite intensities were compared between the run days. Histograms in the right column ((b) and (d)) show the differences of proportions of missing values per metabolite between the run days. Dark blue and green bars represent metabolites which were measured on only one of both days but not the other, light green and blue bars metabolites which were measured on both days at different frequencies. The red part is the proportion of metabolites which were measured on both days at the same frequency. (a) and (c) show the uncorrected p-value distribution of Welch's t-test to compare the metabolite distributions between the run days. None of the metabolites is globally significant.

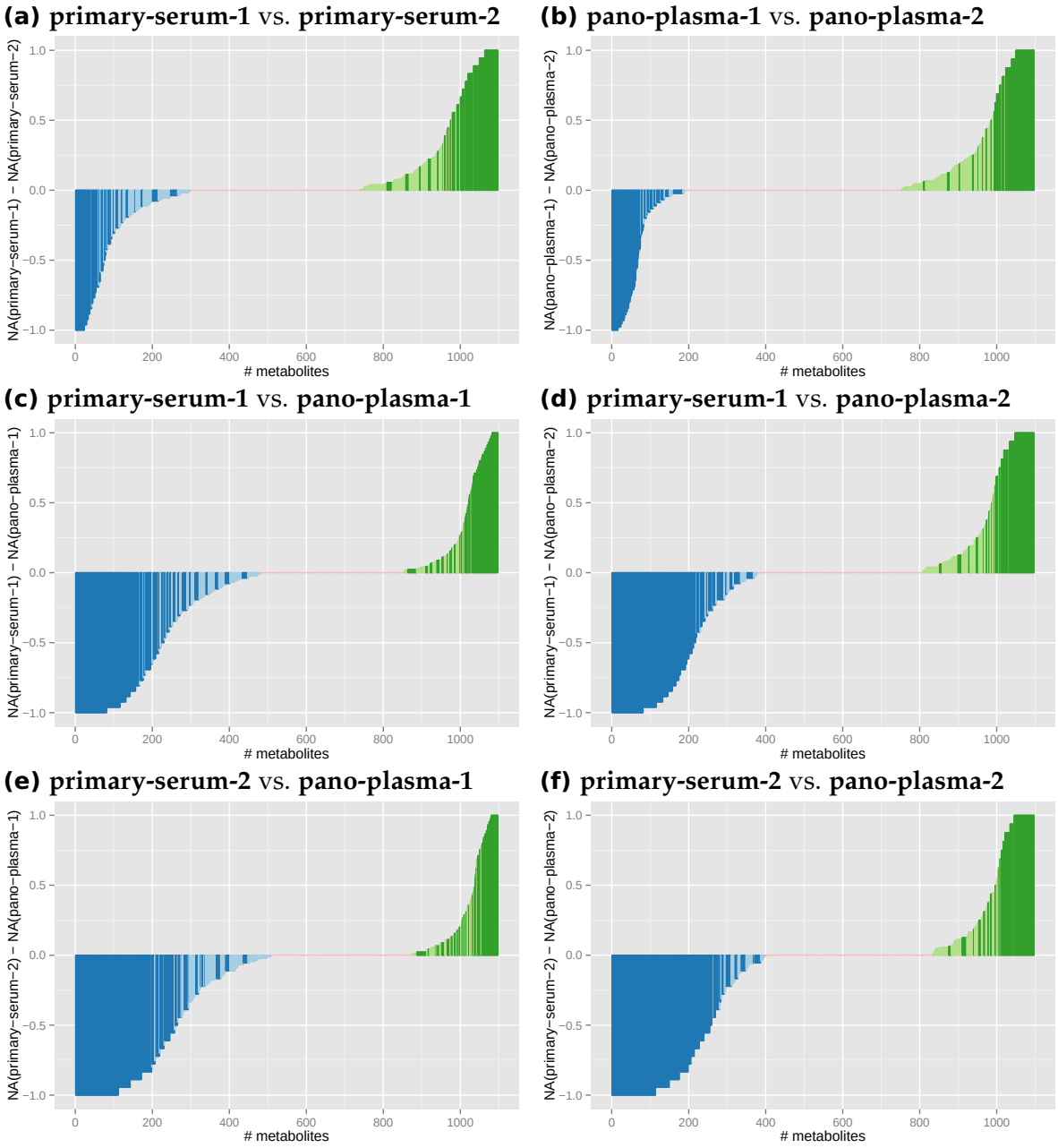


Figure 2.10: Missing value patterns. The proportion of missing values per metabolite (as defined in equation 2.2) was compared between each pair of runs. Each bar represents one metabolite with the difference of percentages of missing values in run 1 and run 2 plotted on the y-axis ($y = -1 = 0 - 1$ means measured for all samples of run 1 but none of 2).

Dark green and blue lines show metabolites which were not measured at all in one run while light green and blue lines represent metabolites which were measured in both runs but at different frequencies.

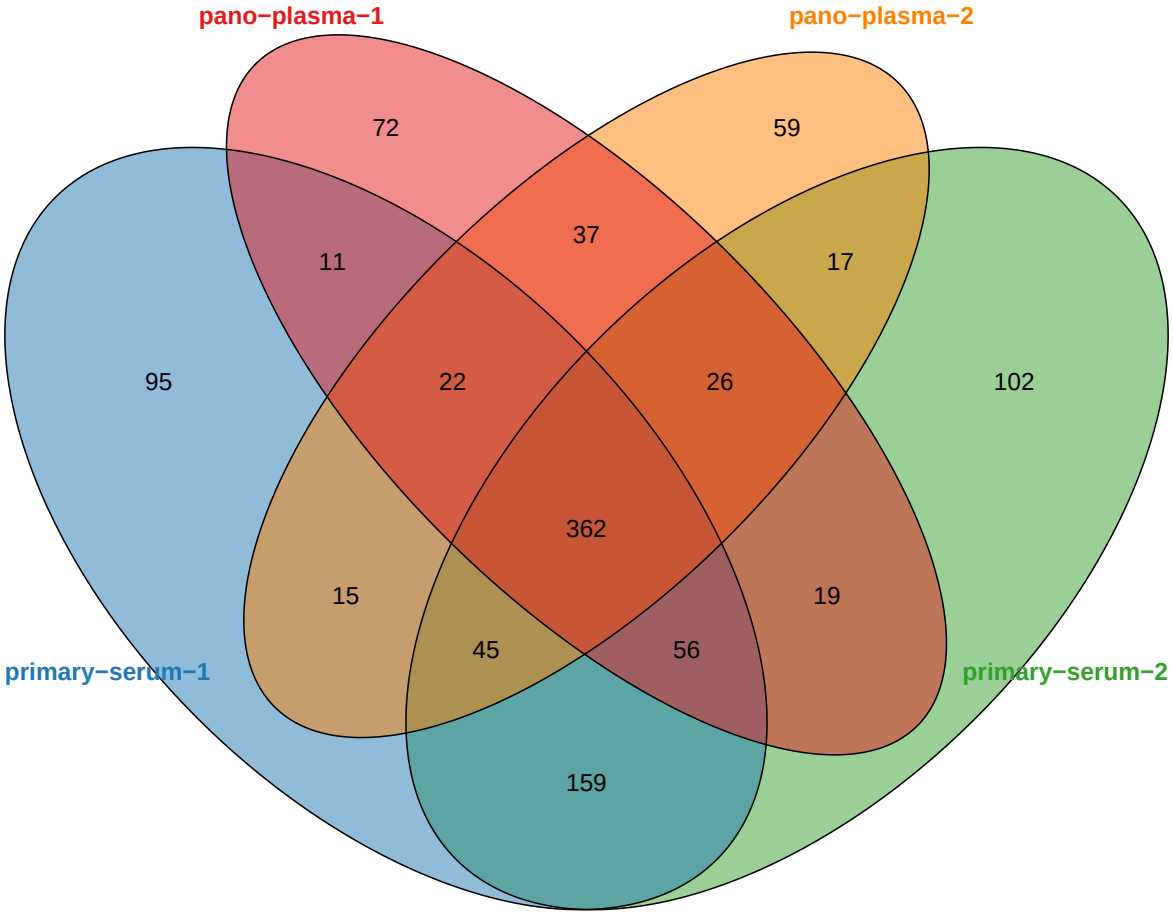


Figure 2.11: Metabolite overlap between studies. Each circle indicates the set of measured metabolites for each of the 4 runs. **primary-serum-1** (blue) and **primary-serum-2** (green) contain the serum samples while **pano-plasma-1** (red) and **pano-plasma-2** (orange) contain plasma samples. 362 metabolites are measured in all studies.

3 Methods

In the following the methods, which were applied to each of the above mentioned four comparisons, are described.

3.1 Correction for Confounding Factors

The metabolite intensities were corrected against variation due to the patient's gender and age. This was done by calculating linear models using the metabolite intensities as dependent variable and gender and age as independent variables (equation 3.1).

$$M^{(s)} \sim \beta_0 + \beta_1 \cdot gender + \beta_2 \cdot age + \epsilon^{(s)} \tag{3.1}$$

$M^{(s)}$ is the set of intensities for one metabolite M and all samples s. β_0 is the intercept of the model and $beta_{1,2}$ are the regression parameters for gender and age. Those are estimated for all samples using a least squares approximation. The $\epsilon^{(s)}$ is the error term for sample s. It describes the deviance between the observed metabolite intensity of sample s approximation of the linear model. It is individually calculated for each sample for each metabolite.

The set of all ϵs is called residuals. Those can be thought of the portion of variation within the metabolite intensities which can not be explained by gender and age. Those residuals were used for all further calculations instead of the metabolite intensities.

3.2 Univariate Analyses

The first step of data analysis was to investigate the influence of the phenotype on the concentration of single metabolites. Thereby three things were considered: First if a metabolite was more often measured in one groups than in the other. Second if

the mean intensity of the metabolite differs between the groups. And third if the variance of the metabolite intensity differs between the groups.

3.2.1 Linear Regression

To test for linear dependency of the metabolite intensity on the phenotype, linear regression models were fitted for each metabolite using the R method `glm`. Accordingly, the linear model was defined as

$$M \sim \beta_0 + \beta_1 \cdot group \tag{3.2}$$

with M being the metabolites intensity and $group$ a binary variable indicating the group (depending on the performed comparison). As metabolite intensities were corrected before, (cf. Section 3.1) no confounding factors were additionally included in the formula.

This model definition contains the metabolite intensity as response variable to investigate the influence of the phenotype on its concentration. The coefficient β_1 was reported together with its p-value and the p-value for the whole model. The latter was calculated using `anova` with an f-statistic.

3.2.2 T-statistics

For the *pre- vs. post-pano* comparison a t-test was performed instead of the linear regression to compare the central tendency of metabolite intensities between two groups. This was done to take pairing of samples into account. For each metabolite the differences between the pre- and post-pano samples of the patient were calculated. As metabolite intensities were logarithmized before, this conforms to log-foldchanges. The distribution of these differences were tested to have $mean = 0$. The according test statistic is

$$t = \frac{\overline{D}}{s_d / \sqrt{n}} \quad \text{with} \tag{3.3}$$

where

$$\overline{D} = \sum_{p \in pairs} \frac{(M_p^{post} - M_p^{pre})}{n} \tag{3.4}$$

is the mean difference between pre- and post-pano samples, s_d its standard deviation and $n = |pairs|$ the number of pairs. The test was performed using the R functions `t.test(M1, M2, paired=TRUE, var.equal=FALSE)`.

3.2.3 Test for different variances

The concentrations of many metabolites are tightly controlled in the body and stable metabolite levels (called homeostasis) are thought to be required to maintain health (Saude et al., 2007). A change in the variance of the metabolite concentration might be a sign for a dysfunction of the control mechanisms due to the influence of the disease or treatment.

Thus the variances of each metabolite in the compared groups were calculated and tested for significant differences. Two different tests were applied: Bartlett's (Bartlett, 1937) and Levene's (Levene, 1960) test for homogeneity of variances.

Bartlett's test relies on normally distributed samples and is sensitive to violation of this condition. For k different groups each consisting of N_i samples having the variance s_i^2, its test statistic is defined as

$$X^2 = \frac{(N-k)\ln(s_p^2) - \sum_{i=1}^{k}(N_i - 1)\ln(s_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}(\frac{1}{N_i - 1}) - \frac{1}{N-k}\right)} \tag{3.5}$$

Where $N = \sum N_i$ and $s_p^2 = \sum_i^k (N_i - 1)s_i^2/(N-k)$ is the weighted average of the variances of all groups. The test statistic is approximately χ^2 distributed with $k-1$ degrees of freedom.

Levene's test is more robust to deviations from normality, but this comes with the price of a lower sensitivity. The test statistic is

$$L = \frac{(N-k)\sum_{i=1}^{k} N_i(\overline{Z_i} - \overline{Z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Z_{ij} - \overline{Z})^2} \tag{3.6}$$

where N again is the total number of samples and k the number of compared groups. Z_{ij} is the intensity of sample j of group i, $\overline{Z_i}$ mean of all samples in group i and $\overline{Z}$

the overall mean.

For both tests median scaled, logarithmized and age- and gender-corrected metabolite intensities were used. P-values were adjusted for multiple testing as described in Section 3.5. The tests were performed using R with the `levene.test` function from `lawstat` package (Gastwirth et al., 2013) and the `bartlett.test` method from the `stats` package.

3.2.4 Missing Value Analysis

The above-named methods do not use information about missing values but omit samples for which the metabolite of interest was not measured. To investigate if phenotypes have an influence on the presence of a metabolite rather than its concentration Fisher's exact test (Fisher, 1970) was performed. It tests two categorical traits, in this case presence of metabolite M and the phenotype, for their independence. This can be represented by the rows and columns of a contingency table (example shown in Table 3.1).

	group 1	¬ group 1	$\sum$
M_i measured	A	B	$A+B$
¬ M_i measured	C	D	$C+D$
$\sum$	$A+C$	$B+D$	$A+B+C+D$

Table 3.1: Contingency Table. The table shows an exemplary contingency table as used for the missing value analysis.

Fisher's test was preferred over the χ^2-test because it is suitable for small sample numbers. Instead of calculating an estimated p-value from a sampling distribution as the χ^2-test does, the p-values are calculated "exact" using a hypergeometric distribution. The test was performed using the method `fisher.test` from the `stats` package in R.

$$p = \frac{\binom{A+B}{A}\binom{C+D}{C}}{\binom{A+B+C+D}{A+C}} = \frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{A!\,B!\,C!\,D!\,(A+B+C+D)!} \tag{3.7}$$

For missing value analysis all samples of a run were excluded if the analyzed metabolite was not measured at all in this run (see Section 2.6.2 for further explanations).

3.3 Bivariate Analyses

Metabolites are not separate units within the organism but react in chemical reactions – usually catalyzed by enzymes – with other compounds to form new metabolites. A series of metabolic transformations which is performed to fulfill a specified task (like transporting a signal/metabolite in a compartment or producing a certain compound) is called metabolic pathway.

Usually all reactions within a living cell are controlled through the abundance of needed cofactors, enzymes, localization, ... When the reaction between two metabolites is perturbated (e.g. due to mutations in the required enzyme) it occurs that this change is not visible when considering the involved metabolites separately, because of other reactions producing or consuming them. Instead, the relation of the metabolites to each other must be analyzed. Due to the pathway structure of the metabolic network this effect is often not visible in directly related metabolites but in those which are some reaction steps apart from each other in one pathway.

In the following three different methods for bivariate testing are described. For those methods median scaled, logarithmized and corrected intensities were used. As before metabolites which were measured in too little samples were excluded. Only samples for which both compared metabolites were measured were used.

The results of the tests which are described in the following were mapped on a metabolic reference network to analyze deregulated pathways and processes. Therefore, a metabolic network which was created by Krumsiek et al. (2012) using partial correlations and Gaussian graphical models was used.

3.3.1 Metabolite Ratios

The ratio between the product and the reactant of a reaction reflects the state of the chemical equilibrium of the reaction. This equilibrium constant can e.g. be changed by altered concentrations of required cofactors. Mutations in enzymes (or their regulation) which alter its efficiency or amount might also alter the ratio between two metabolites. As many metabolites are not just involved in one reaction but many different processes, alterations of one chemical equilibrium might not significantly

influence the concentration of the single metabolites but only their ratio. For this reason ratios of all pairs of metabolites were calculated and tested for different means between the group using the same linear model as defined in Section 3.2.1. Ratios of unrelated metabolites were calculated and tested as well. Those do not reflect chemical equilibrium constants but can be interpreted as patient-specific normalization of the metabolite levels Altmaier et al., 2008; Suhre et al., 2011.

To evaluate the benefit of testing metabolite ratios instead of separate metabolites p-gain statistics were calculated The p-gain of a metabolite ratio is defined as

$$\text{p-gain} = \frac{\min(p_{M1}, p_{M1})}{p_{M1/M2}} \tag{3.8}$$

where p_{M1} and p_{M1} are the p-values of linear regression of the separate metabolites and $p_{M1/M2}$ is the p-value of the linear regression using their ratio.

For the *pre- vs. post-pano* comparison, again a t-test was used to take pairing of samples into account. This was done in the same manner as for single metabolites (described in Section 3.2.2). In this case the p-gains were calculated using the separate p-values from t-tests and the ratio p-values from t-tests.

For metabolite ratios the tests for different variances (described in Section 3.2.3) were applied in the same way as for single metabolites. Thereby outliers, differing more than 3 standard deviations from the mean of the ratio, were excluded before the test. P-gains were calculated for the variance test in a similar manner as for the linear regression (Equation 3.8). For each metabolite two different p-gains were calculated: One ror Levene's test and the other for Barlett's test.

3.3.2 Linear Regression with Metabolite Pairs

Another way of investigating bivariate markers is to fit linear regression models with interaction terms. As this thesis investigates the influence of the Panobinostat treatment (and others) on the metabolite concentrations, the linear combination of metabolites was considered as outcome. This is not native supported by regression models, which is why the linear combinations of metabolites had to be pre-calculated. The weights w_i for the metabolites were determined by performing a Principal Component Analysis (PCA) with the two combined metabolites. The loadings of the first principal component were used as weights in the linear combination. Before calculating the principal components, metabolite intensities were transformed to z-scores $M^{(z)}$ to improve their comparability.

The linear regression model was defined as follows.

$$l_1 \cdot M_1^{(z)} + l_2 \cdot M_2^{(z)} \sim \beta_0 + \beta_1 \cdot group \tag{3.9}$$

where l_1 and l_2 are the loadings of the first principal component of the metabolites $M_1^{(z)}$ and $M_2^{(z)}$ respectively and $group$ is a binary variable describing the observed phenotype. For the paired samples of the *pre- vs. post-pano* comparison again a t-test was used instead. P-gains were calculated as for metabolite ratios.

3.3.3 Differential Correlation

Another way of representing a relationship between two metabolites is their correlation. Pearson correlation coefficients were calculated for all pairs of metabolites separately for both compared groups. The resulting two correlations were tested for significant difference. As correlations are not normally distributed, Fishers z'-transformation was applied (Fisher, 1915; Fisher, 1921; Anderson, 1996) to calculate z-scores z' from correlation coefficients r as

$$z' = 0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) \tag{3.10}$$

The difference of the z' scores between the both groups was defined as

$$z = \frac{z'_1 - z'_2}{\sigma_{z'_1 - z'_2}} \tag{3.11}$$

where the standard error of the difference σ is

$$\sigma_{z'_1 - z'_2} = \sqrt{\frac{1}{n_2 - 3} + \frac{1}{n_1 - 3}} \tag{3.12}$$

and n_i the number of samples used to calculate the correlation within the ith group.

The resulting z-scores are normally distributed and p-values $p = P[X > z]$ can be obtained by comparing it to a standard normally distribution. The test for differential correlation was only performed for pairs of metabolites, whose correlation was significant (uncorrected p-value < 0.05) in both compared groups. The final p-values were Benjamini-Hochberg (BH)-corrected as before.

3.4 Multivariate Analyses

The next step in the way from univariate to bivariate marker identification are multivariate analyses. The method for linear regression of metabolite pairs, described in Section 3.3.2, could easily be applied for any number of metabolites. The problem is that the amount of metabolite combinations increases exponentially with their size. That is on the one hand a computational problem and on the other a multiple testing problem. Therefore two different approaches can be pursued: To simplify the information by using external knowledge (Section 3.4.2) or use methods for dimension reduction (Section 3.4.3 and subsequent).

3.4.1 Imputation

The multivariate methods described below can not handle missing values internally. Therefore missing values had to be imputed for those methods. Two different approaches were implemented to handle "sparse" metabolites which have many missing values: One discarding them and the other using the information of those metabolites.

For both approaches metabolites were distinguished by the percentage of measured values. Missing values of metabolites with more than 80 % valid measurements were imputed to the mean intensity of the metabolite minus 5 times its standard deviation (formula 3.13). The first imputation approach was to discard sparse metabolites with more than 20 % missing values. The second approach was to generalize the intensities of those metabolites to 1 for measured values and 0 for missing values.

Thus the imputed value of metabolite i for sample j was calculated as

$$M_{ij} = \begin{cases} M_{ij} & \text{if } \frac{\sum is.missing(M_{ik})}{|M|} < 0.2 \wedge \neg is.missing(M_{ij}) \\ \overline{M_i} - 5 \cdot sd(M_i) & \text{if } \frac{\sum is.missing(M_{ik})}{|M|} < 0.2 \wedge \ is.missing(M_{ij}) \\ 1 & \text{if } \frac{\sum is.missing(M_{ik})}{|M|} > 0.2 \wedge \neg is.missing(M_{ij}) \\ 0 & \text{if } \frac{\sum is.missing(M_{ik})}{|M|} > 0.2 \wedge \ is.missing(M_{ij}) \end{cases} \tag{3.13}$$

with $\overline{M_i}$ being the mean intensity of metabolite i and $is.missing$ the indicator function defined in equation 2.3.

3.4.2 Pathway Enrichment

In the field of microarray analysis gene set enrichment analysis (GSEA) (Subramanian et al., 2005) is a well-established method. This algorithm tests if the observed differences in gene expression between two conditions accumulate in certain "gene sets". Those sets have to be defined before e.g. according to their function, sequence or protein structure. A very similar approach can also be applied to metabolomics data: Instead of sets of genes, sets of metabolites are over- or under-"expressed" respectively and instead of gene classifications metabolic pathways can be used.

As group annotation the pathway information provided by Metabolon was used (described in Section 2.3). This contains two different levels of detail, super-pathways and sub-pathways. All 570 unknown compounds were combined in an additional "unknown" pathway.

Three different publicly available methods were applied for the enrichment analysis: The first one is an adapted version of the standard GSEA approach of Subramanian et al. (2005) called Metabolite Set Enrichment Analysis (MSEA) (`R` package `mseapca`). It permutes metabolite set annotations and compares the observed enrichment to the sampled random background. As second method the Correlation Adjusted Mean Rank Gene Set Test (CAMERA) of Wu and Smyth (2012) (`limma` package) was used. CAMERA relies on a statistical test which compares the ranking of metabolites of one set to the others. The last approach was Generally Applicable Gene Set Enrichment for Pathway Analysis (GAGE) (`gage` package) of Luo et al. (2009). GAGE ranks the metabolites using a t-test to compare the intensity of each metabolite of the first group to the mean intensity of the second group. The resulting p-values are then combined to one p-value.

Another fourth method was additionally implemented. For that the intensities of all metabolites, belonging to one pathway, were combined to one value. The intensity value for the pathway p for sample s was calculated as mean z-score of the metabolites in the respective pathways:

$$I_p^s = \frac{\sum_{i=1}^{n_p} z_i^s}{n_p} \tag{3.14}$$

with n_p the number of metabolites belonging to pathway p and z_i^s the z-transformed intensity of metabolite i and sample s. This is defined as

$$z_i^s = \frac{m_i^s - \overline{m_i}}{sd_{m_i}} \tag{3.15}$$

where n_s is the number of samples, $\overline{m_i}$ the mean intensity of metabolite i across all samples and sd_{m_i} its standard deviation. The resulting pathway intensities were used to fit linear regression models in the same way as for single metabolites.

As the intensity values of pathways are defined as the mean z-scores of the contained metabolites, pathways can only be significant if all metabolites belonging to the pathway are either up- or down-regulated. If they are regulated in different directions the mean z-score will be ≈ 0.

3.4.3 Principal Component Analysis

Principal Component Analysis (PCA) is a standard technique for unsupervised reduction of dimensions. Correlated variables are merged to one principal component as a linear combination. The principal components are designed to be orthogonal to each other. Thus, the PCA is a projection of the data matrix into a lower dimensional space with each principal component representing one coordinate. The components are ordered: the first one accounts for the highest portion of variance of the data, all others contain a decreasing amount of variance.

As mentioned above, PCA is an unsupervised technique. That means that the phenotypes of samples (e.g. pre- or post-pano) do not influence the result directly. This means that the usefulness of PCA for the explanation of a phenotype is not known beforehand.

The principal components were calculated using the `R` methods `prcomp`, which uses singular value decomposition for the calculation.

3.4.4 Multivariate Regression

To assess the predictive power of multiple metabolites combined, one could in principal fit linear models with many metabolites and their combinations, but in practice this is usually not possible. The number linear models which have to be calculated increases exponentially with increasing number of metabolites contained in the models. Thus huge computational power and many samples would be needed to fit models for all combinations of metabolites. Even if that would be possible, multiple testing correction will most likely eliminate all significant results.

A good alternative for testing combinations of metabolites is penalized regression. It combines all variables in one linear model and adds a penalty term for the weights β_i, which ensures lower model complexity. One well established method, which

uses a penalty based on the L_1 norm, is the Least Absolute Shrinkage and Selection Operator (LASSO) regression. Its loss function, which depends on an arbitrary λ, is defined as

$$L_\lambda(\beta_1, \ldots, \beta_n) = \sum_{s=1}^{S} \left(y^{(s)} - \sum_{i=1}^{n} \beta_i \cdot x_i^{(s)} \right) + \lambda \sum_{i=1}^{n} abs(\beta_i) \tag{3.16}$$

where S is the number of samples, n the number of metabolites, $y^{(s)}$ the group annotation for sample s, $x_i^{(s)}$ the measurement of metabolite i of sample s and β_i the regression coefficient for metabolite i.

The first part of the term is a goodness-of-fit measure for the outcome y. The second part is the penalty term which rises with increasing sums of betas. The higher the weight λ is chosen, the higher is the influence of the penalty term $\sum_{i=1}^{n} abs(\beta_i)$ on the total loss function. As the sum of all β_i is penalized, weights of metabolites which do not improve the loss function are set to 0 to reduce the penalty and thus decrease the loss. The according metabolites are not considered in the model which means that the model complexity decreases.

To assess the quality of the model a leave-one-out (LOO) validation was performed. For each fold of the validation a nested cross validation (CV) was done to select the optimal λ (which minimizes the percentage of false classified samples). The (left out) test sample was predicted using the optimal λ_{min} and λ_{1se}, which is at most one standard error worse than the optimal lambda but uses the fewest possible metabolites. Thereby samples were weighted to assure that the sum of weights was the same for the compared groups.

3.4.5 Clustering

For each comparison samples and metabolites were clustered using hierarchical clustering with the manhattan distance function and average linkage. The manhattan distance of two samples is the absolute sum of all differences between each metabolite.

$$d(s_i, s_j) = \sum_{m=1}^{|M|} abs(M_m^{(i)} - M_m^{(j)}) \tag{3.17}$$

where $s_{i,j}$ are the samples, M the set of all metabolites and $M_m^{(i)}$ the intensity of the mth metabolite for the ith sample. When using average linkage the distance

between two clusters is the mean distance of all pairs of samples in the clusters. For metabolites clustering was performed in the same manner.

3.5 Multiple Testing Correction

All resulting p-values were corrected for multiple testing by controlling the false discovery rate (FDR). The FDR is the proportion of false positive amongst all positives. This was done by using the Benjamini-Hochberg (BH) method.

3.6 Visualization

The data and results were visualized using `R` and the `ggplot2` package (Wickham, 2009). For clustered heatmaps the `cim` method, which is part of the `mixOmics` package, was used (Dejean et al., 2013)

3.7 Implementation and Availability

All methods were implemented in `R` and were integrated in Konstanz Information Miner (KNIME) (Berthold et al., 2007). KNIME is an open source workbench program which integrates tools for data access, transformation, machine learning, visualization and other analytical tasks. Its user interface is based on Eclipse IDE[1] and allows users to build workflows consisting of nodes. Each node performs one tasks and provides its result to successive nodes.

All methods presented above – as well as some auxiliary tools – were implemented as KNIME nodes. This allows the user to configure the programs with a graphical user interface, integrate the scripts into own workflows or build new workflows using the presented methods together with standard KNIME nodes. The implementation of KNIME nodes and the KNIME workflow can be found on the supplemental DVD (cf. Section A)

[1] `http://www.eclipse.org/`

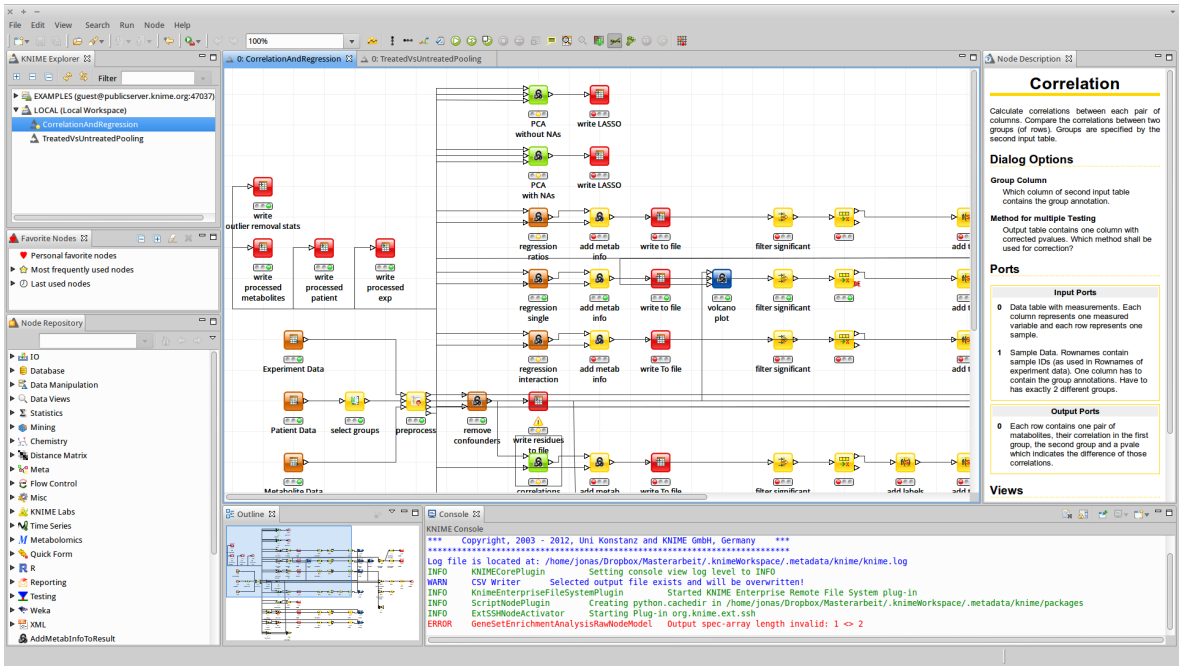


Figure 3.1: Screenshot of the KNIME interface. The main panel in the middle shows the workflow with nodes and connections between them. Different workflows can be chosen in the upper left panel. The available nodes are listed in the node repository below. On the left side the manual of the currently selected node is shown. It describes the node itself, input parameters and in- and output specifications. Each node can have different views to display e.g. log messages or results.

4 Results and Discussion

In this chapter the results of the four investigated questions are described and discussed. At first, the influence of Panobinostat on the metabolism was analyzed using paired plasma samples and samples from DLBCL cell lines. After that primary tumors were compared to relapsed tumors to figure out the metabolic changes between the two different types of tumors. To investigate if tumors which are resistant to R-CHOP differ from R-CHOP-sensitive tumors before the treatment starts, serum samples of these two groups were compared. The last aim was to compare blood plasma from serum of patients with untreated, primary DLBCL. This comparison might reveal advantages and disadvantages of the different probe types for the investigation of lymphoma (cf. Section 1.3).

Before analyzing the data it was scaled by the run-day median, log-transformed and corrected for influences due to gender and age (c.f. Sections 2.6.2, 2.4.1 and 3.1). The scaled and corrected data also was used for all plots of metabolite distributions as well. The presented p-values were adjusted for multiple testing using the BH method (Section 3.5). Unless otherwise specified results having (corrected) p-values ≤ 0.05 were considered to be significant. In case of bivariate testing a p-gain cutoff of 1000 was additionally applied. In most cases all significant results are shown, but in some cases the printed lists were restricted for the sake of briefness. In all cases complete lists including all metabolites and pairs of metabolites respectively can be found in the supplemental data on the attached DVD (cf. Supplements A).

4.1 Before vs. After Treatment with Panobinostat

To gain new insights in the mode of action of Panobinostat and figure out possible side effects, metabolite intensities were measured from patients with relapsed/refractory DLBCL before and after the treatment. This was done using paired plasma samples from the experiments **pano-plasma-1** and **pano-plasma-2**. The same comparison also was carried out for DLBCL cells from the cell line OCI-Ly1 before and after those were treated with Panobinostat. These were measured in experiment **pano-cells** and analyzed in the same manner as the in vivo data.

In the following the results of the in vivo, plasma samples are presented at first. After that the results of the cell line samples are shown in Section 4.1.2 before discussing both results in Section 4.1.3.

4.1.1 Results (In Vivo)

metabolite	***#***	***mean***^{1}	***mean***^{2}	***p-value***
4-androsten-3beta,17beta-diol disulfate 1*	24	−0.67	0.67	3.1×10^{-8}
pregnen-diol disulfate*	23	−0.57	0.57	1.5×10^{-5}
X -11440	23	−0.52	0.52	4.2×10^{-5}
X -12063	18	−0.30	0.27	1.8×10^{-4}
X -17299	24	−0.38	0.38	4.7×10^{-4}
5alpha-androstan-3beta,17beta-diol disulfate	19	−0.39	0.39	10.0×10^{-4}
betaine	24	0.23	−0.23	10.0×10^{-4}
uridine	24	0.24	−0.24	10.0×10^{-4}
X -17337	22	−0.34	0.31	1.1×10^{-3}
isoleucine	23	0.19	−0.20	1.3×10^{-3}
glycochenodeoxycholate	23	0.72	−0.75	3.6×10^{-3}
threonine	24	−0.34	0.34	5.8×10^{-3}
gamma-glutamylphenylalanine	24	−0.14	0.14	5.9×10^{-3}
2-hydroxyisobutyrate	23	−0.23	0.23	6.5×10^{-3}
X -18928	22	0.70	−0.76	6.5×10^{-3}
threonate	23	0.43	−0.41	6.5×10^{-3}
gamma-glutamylmethionine	21	−0.30	0.26	6.5×10^{-3}
3-methyl-2-oxovalerate	24	0.16	−0.16	10.0×10^{-3}

continued on next page

continued from previous page

metabolite	***#***	***mean***^{1}	***mean***^{2}	***p-value***
dehydroisoandrosterone sulfate (DHEA-S)	23	−0.07	0.07	1.0×10^{-2}
N-acetylalanine	24	−0.13	0.13	1.2×10^{-2}
gamma-glutamyltyrosine	24	−0.14	0.14	1.2×10^{-2}
urea	24	−0.19	0.19	1.3×10^{-2}
phenylalanylphenylalanine	11	0.20	−0.22	1.5×10^{-2}
glutamate	23	0.24	−0.25	1.6×10^{-2}
X -11727	15	−0.34	0.34	1.9×10^{-2}
2-hydroxystearate	24	0.12	−0.12	1.9×10^{-2}
phenylalanyltryptophan	23	−0.32	0.33	2.0×10^{-2}
X -11787	23	0.07	−0.07	2.1×10^{-2}
citrulline	24	0.22	−0.22	2.1×10^{-2}
gamma-glutamylisoleucine*	23	0.22	−0.23	2.2×10^{-2}
glycocholenate sulfate*	23	−0.20	0.19	2.3×10^{-2}
mannose	21	−0.27	0.27	2.3×10^{-2}
glycocholate	22	0.50	−0.54	2.4×10^{-2}
glycerol 3-phosphate (G3P)	24	0.18	−0.18	2.5×10^{-2}
2-hydroxypalmitate	24	0.11	−0.11	2.5×10^{-2}
1-stearoylglycerol (1-monostearin)	24	−0.22	0.22	3.1×10^{-2}
aspartate	24	0.13	−0.13	3.2×10^{-2}
N-acetylthreonine	16	−0.07	0.06	3.2×10^{-2}
C-glycosyltryptophan*	22	−0.11	0.11	3.4×10^{-2}
bilirubin (E,E)*	24	−0.35	0.35	3.4×10^{-2}
methionine	24	−0.17	0.17	3.4×10^{-2}
X -12846	16	0.15	−0.18	3.7×10^{-2}
glycerophosphorylcholine (GPC)	16	0.29	−0.29	3.9×10^{-2}
histidine	24	−0.10	0.10	4.1×10^{-2}

Table 4.2: Metabolites with different concentration between pre- and post-pano plasma samples. The table shows the results of the paired t-test (described in Section 3.2.2) comparing plasma samples before (^{1}) and after (^{2}) the treatment with Panobinostat. Apart from the metabolite names, the number of paired samples for which the metabolite was measured (#) is shown. The mean intensities for both groups were(using median-scaled, logarithmized and for age and gender corrected intensities). The p-values were BH-corrected.

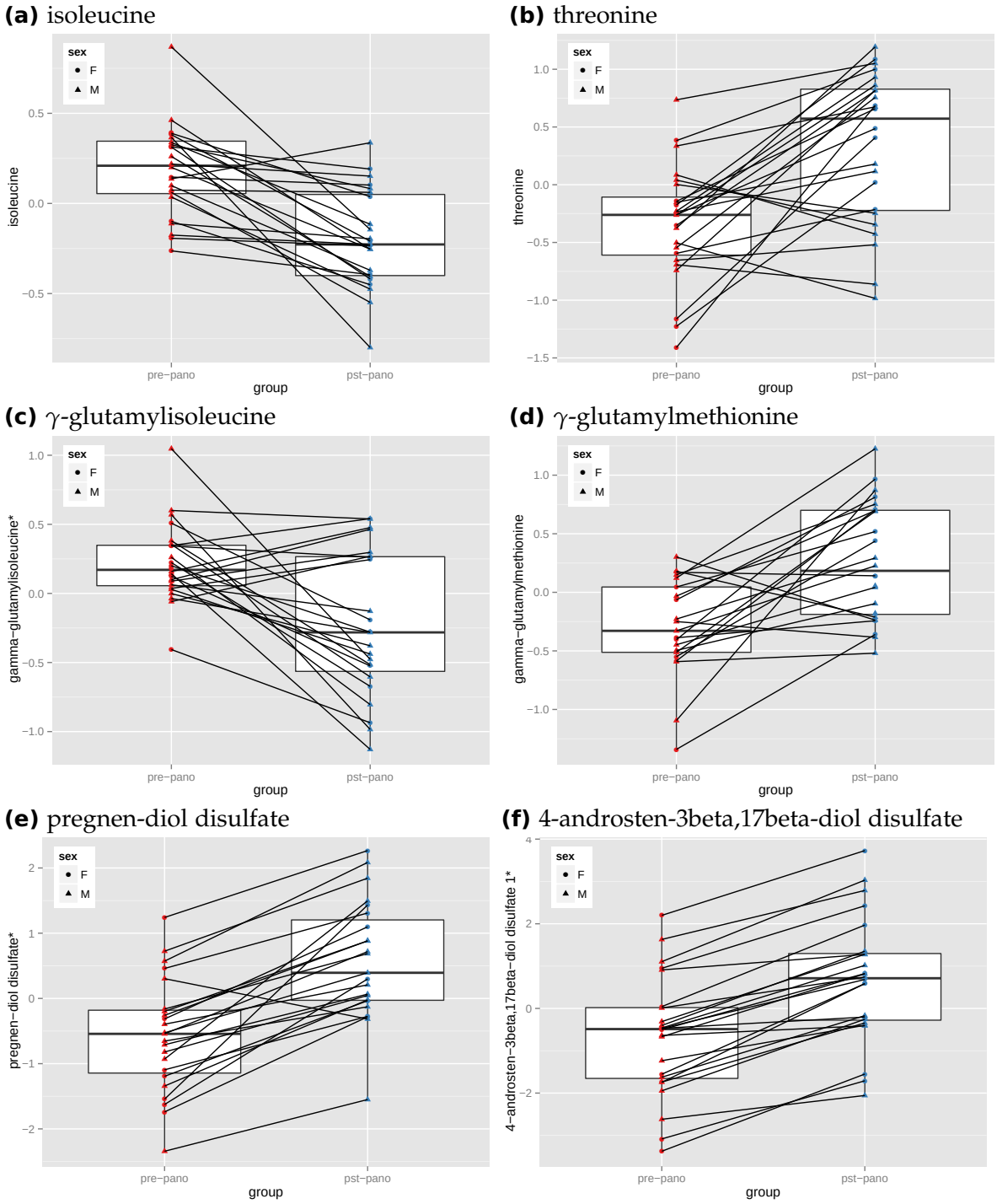


Figure 4.1: Examples for metabolites with different concentration between pre- and post-pano plasma samples. Each figure shows the distribution of median scaled, logarithmized and corrected (for sex and age) metabolite intensities of pre-pano (left) and post-pano (right) samples. The paired samples of the same patients are connected by lines. Boxes indicates the first and third quartile and the median of the distributions. The point shape represents the gender of the patient.

metabolite	*var*[1]	*var*[2]	*p-value (b)*	*p-value (l)*
gamma-glutamylisoleucine*	0.08	0.27	2.9×10^{-1}	2.0×10^{-1}
phenylacetylglutamine	0.51	1.15	6.5×10^{-1}	2.0×10^{-1}
inosine	0.36	1.35	2.7×10^{-1}	2.0×10^{-1}
X -18928	0.41	1.84	1.8×10^{-1}	2.0×10^{-1}
1-oleoylglycerophosphoethanolamine	0.26	0.81	3.9×10^{-1}	2.0×10^{-1}

Table 4.3: Metabolites with different variance between pre- and post-pano plasma samples. The table shows the results of the comparison of metabolite variances of pre-pano([1]) and post-pano([2]) samples. Corrected p-values of Barlett's test (b) and Levene's test (l) are shown in the two rightmost columns (cf. Section 3.2.3). The five metabolites with the best Levene p-value are presented which corresponds to the p-value cutoff ≤ 0.2.

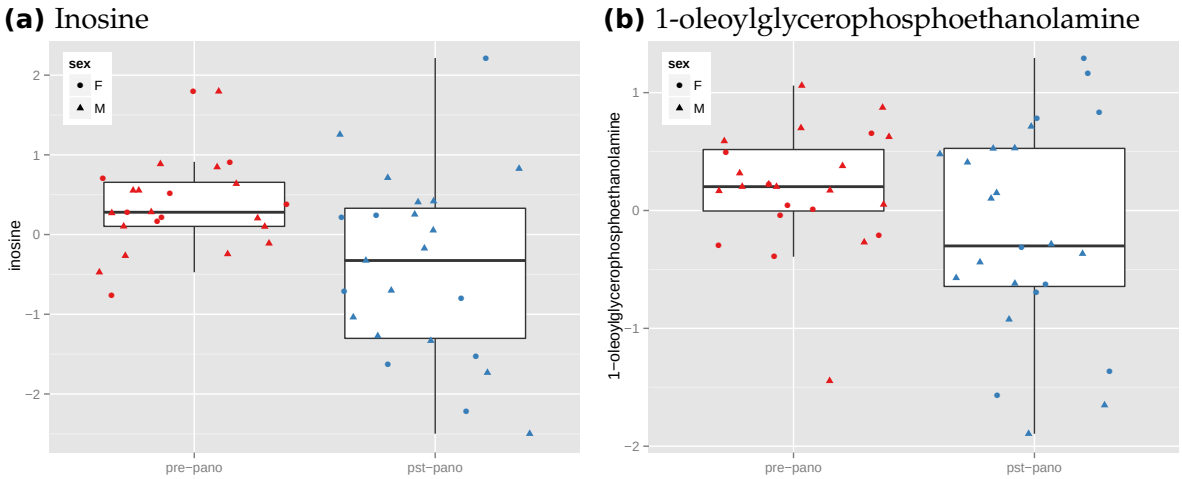


Figure 4.2: Metabolites with different variance between pre- and post-pano samples. The boxes represent distributions of scaled, logarithmized and age- and gender-adjusted metabolite intensities. Those are stratified by their group annotation (pre- or post-pano). Samples of the same patient are connected by lines. Both metabolites (a) and (b) have different variances between the groups with FDR≤ 0.2.

metabolite 1	*metabolite 2*	*slope*	*p-value*	*p-gain*
methionine	3- methyl- 2- oxovalerate	0.66	1.5×10^{-6}	4.3×10^{5}
3- methyl- 2- oxovalerate	4- methyl- 2- oxopentanoate	−0.29	2.7×10^{-6}	1.7×10^{5}
2- hydroxyisobutyrate	X - 11444	0.75	3.2×10^{-6}	7.7×10^{4}
isoleucine	methionine	−0.76	8.1×10^{-7}	7.6×10^{4}
cortisol	pregnenolone sulfate	0.67	8.1×10^{-5}	2.5×10^{4}
uridine	C- glycosyltryptophan*	−0.72	1.3×10^{-6}	2.4×10^{4}
isoleucine	gamma- glutamylphenylalanine	−0.68	2.1×10^{-6}	1.9×10^{4}
3- methyl- 2- oxovalerate	gamma- glutamylmethionine	−0.89	1.3×10^{-5}	1.1×10^{4}
pregnenolone sulfate	X - 19435	−0.60	1.8×10^{-4}	8.8×10^{3}
isoleucine	gamma- glutamylmethionine	−1.01	3.8×10^{-6}	7.4×10^{3}
phenylalanine	isoleucine	0.55	3.8×10^{-6}	7.2×10^{3}
leucine	phenylalanine	−0.33	4.4×10^{-4}	7.0×10^{3}
X - 11429	X - 12846	0.70	1.3×10^{-4}	6.4×10^{3}
X - 11429	X - 11470	0.56	2.0×10^{-4}	6.0×10^{3}
indolelactate	X - 11727	−0.97	6.0×10^{-5}	5.8×10^{3}
isoleucine	gamma- glutamyltyrosine	−0.68	4.8×10^{-6}	5.2×10^{3}
cortisol	X - 11470	0.58	2.7×10^{-4}	4.2×10^{3}
3- methyl- 2- oxovalerate	dehydroisoandrosterone sulfate (DHEA- S)	−0.46	3.9×10^{-5}	4.1×10^{3}
3- methyl- 2- oxovalerate	X - 11440	−1.38	3.4×10^{-7}	3.1×10^{3}
X - 11847	X - 11849	−1.31	8.7×10^{-4}	2.9×10^{3}

Table 4.4: Metabolite ratios differing between pre- and post-pano samples. The table shows the results of the paired t-test of metabolite ratios (as described in Section 3.3.1). The "slope" is the difference of the mean intensities of both groups. That corresponds to the beta of linear regression. Positive, green slope indicates higher ratios after the treatment. P-values were adjusted for multiple testing using BH corrections. The p-gains were calculates using the p-values of t-tests for separate metabolites as well as ratios. All ratios with p-value ≤ 0.05 and p-gain > 2000 are listed.

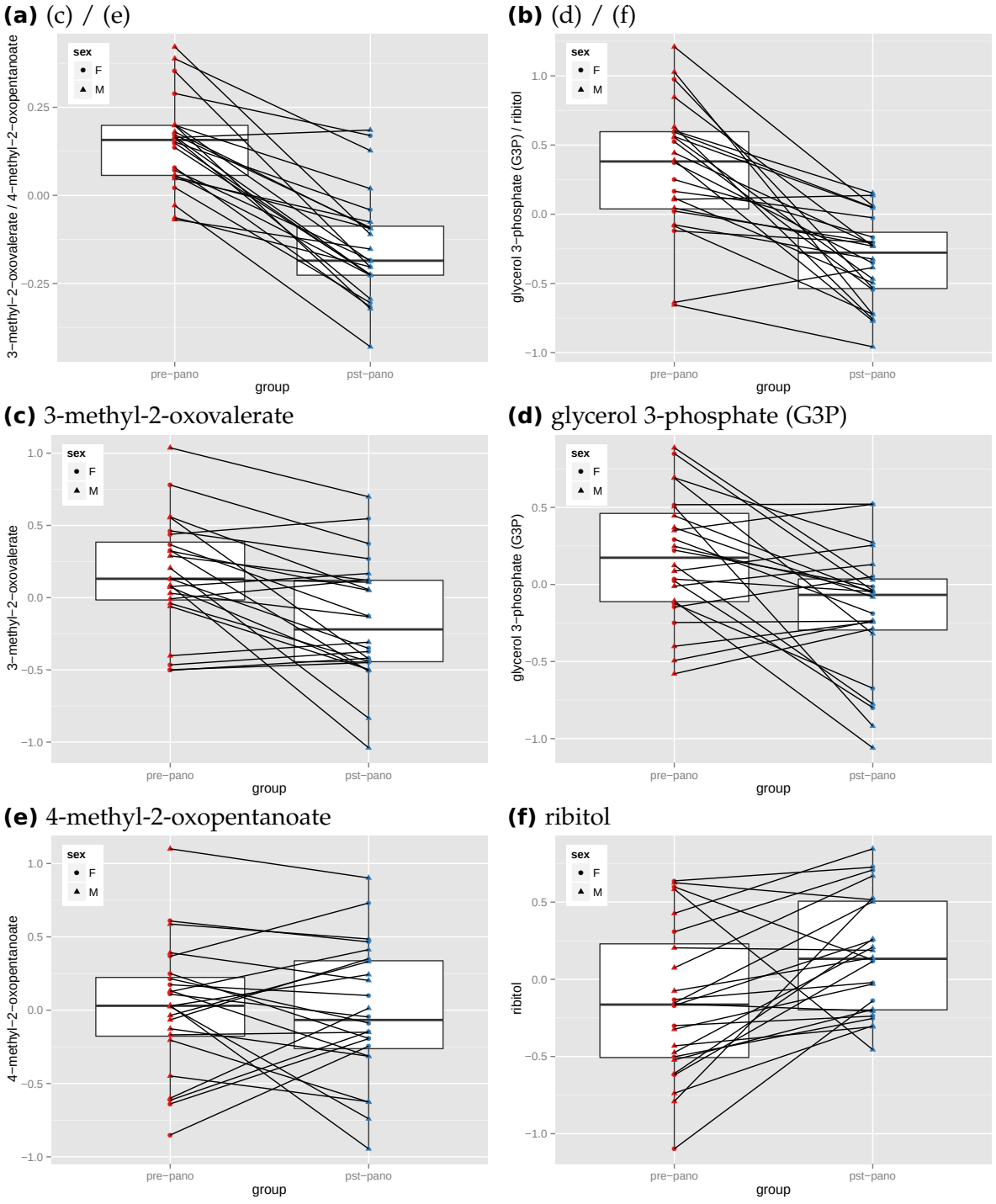


Figure 4.3: Metabolite ratios differing between pre- and post-pano samples. In the two uppermost plots (a) and (b) the distributions of the ratios of median scaled, logarithmized and corrected metabolite intensities are drawn. The lower plots, (c) - (f), show the distributions of the according metabolites separately. Boxes indicate the quartiles of the distributions, lines connect paired samples of the same patient

metabolite 1	*metabolite 2*	*l*	*slope*	*p-value*	*p-gain*
X -17138	X -19411	$+1$	0.90	3.5×10^{-5}	3.8×10^{5}
isoleucine	gamma-glutamyl-methionine	-1	-1.64	2.8×10^{-6}	3.5×10^{5}
isoleucine	methionine	-1	-1.51	2.8×10^{-6}	3.0×10^{5}
methionine	3-methyl-2-oxo-valerate	-1	1.15	1.4×10^{-5}	1.4×10^{5}
isoleucine	gamma-glutamylphenylalanine	-1	-1.50	1.0×10^{-5}	2.0×10^{4}
uridine	C-glycosyltryptophan*	-1	-1.16	1.0×10^{-5}	1.2×10^{4}
threonine	3-methyl-2-oxo-valerate	-1	1.25	3.5×10^{-5}	1.1×10^{4}
isoleucine	threonine	-1	-1.54	1.4×10^{-5}	1.1×10^{4}
X -11429	X -12846	-1	1.24	2.2×10^{-4}	9.0×10^{3}
threonine	X -17337	$+1$	1.67	1.4×10^{-5}	8.9×10^{3}
leucine	X -17453	-1	-1.19	5.3×10^{-4}	8.2×10^{3}
isoleucine	gamma-glutamyltyrosine	-1	-1.41	2.0×10^{-5}	5.5×10^{3}
X -11429	X -11470	-1	0.97	3.8×10^{-4}	5.5×10^{3}
docosapentaenoate (n6 DPA; 22:5n6)	X -17138	$+1$	0.99	4.2×10^{-4}	4.4×10^{3}
isoleucine	urea	-1	-1.44	2.1×10^{-5}	4.2×10^{3}
betaine	X -11470	$+1$	-1.25	1.7×10^{-5}	3.8×10^{3}
gamma-glutamyl-valine	X -17453	-1	-1.22	7.8×10^{-4}	3.7×10^{3}

Table 4.5: Linear regression results of combined metabolites comparing pre- and post-pano Samples. Pairs of metabolites were combined as linear combinations using their first principal component. Before the PCA was performed, metabolite ratios were scaled, logarithmized, corrected for gender and age, and standardized (divided by their standard deviation). The first principal component was used for the regression; it was defined as $l_1 \cdot M_1 + l_2 \cdot m_2$ where M_i are the metabolite intensities and l_i the loadings. Due to the scaling and standardization $|l_1| = |l_2|$. The column l in the table shows the ratio $l = l_1/l_2$ indicating the signs of the both loadings. The slope is defined as for ratios: It is the difference of the mean values of both groups, green translates to higher mean value of the linearly combined metabolites in post-pano samples. The p-values are FDR-corrected.

metabolite 1	*metabolite 2*	*cor*[1]	*cor*[2]	*p-value*
caprylate (8:0)	N-methyl proline	0.61	−0.56	2.2×10^{-2}
pyrophosphate (PPi)	phosphate	−0.51	0.65	2.2×10^{-2}
1-eicosadienoylglycerophosphocholine*	X -17453	−0.71	0.60	2.2×10^{-2}
7-beta-hydroxycholesterol	pro-hydroxy-pro	−0.60	0.73	3.3×10^{-2}
isovalerylcarnitine	X -15382	−0.66	0.44	3.3×10^{-2}
N-acetylaspartate (NAA)	pro-hydroxy-pro	−0.65	0.76	3.3×10^{-2}
caprate (10:0)	N-methyl proline	0.52	−0.56	3.3×10^{-2}
hippurate	paraxanthine	−0.59	0.60	3.8×10^{-2}
fumarate	maltotriose	−0.51	0.59	3.8×10^{-2}
glycolithocholate sulfate*	acisoga	0.57	−0.55	4.1×10^{-2}
2-hydroxydecanoic acid	X -19435	0.51	−0.65	3.3×10^{-2}
xanthine	N6-carbamoylthreonyladenosine	0.48	−0.62	3.8×10^{-2}
X -17269	2-hydroxydecanoic acid	−0.44	0.62	4.3×10^{-2}
beta-alanine	7-alpha-hydroxy-3-oxo-4-cholestenoate (7-Hoca)	−0.58	0.43	4.3×10^{-2}
pregnenolone sulfate	2-hydroxydecanoic acid	0.45	−0.64	3.8×10^{-2}

Table 4.6: Differentialy Correlated Metabolites in pre- and post-pano samples. Pairs of metabolites whose correlation differ significantly between pre-[1] and post-pano[2] samples are listed. All single correlation coefficients are significant ($p < 0.05$ without correction). For each pair of metabolites the correlations within the groups, pre- and post-pano, and the corrected p-values indicating the significance of their difference are reported. All pairs whose difference is globally significant (p-value < 0.05) are shown.

(a) caprylate vs N-methyl proline

(b) pyrophosphate (PPi) vs phosphate

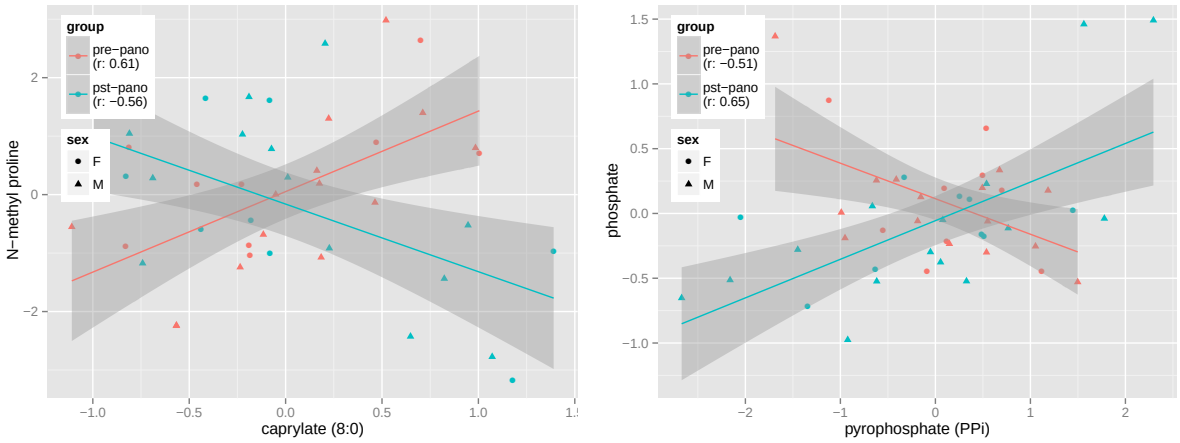


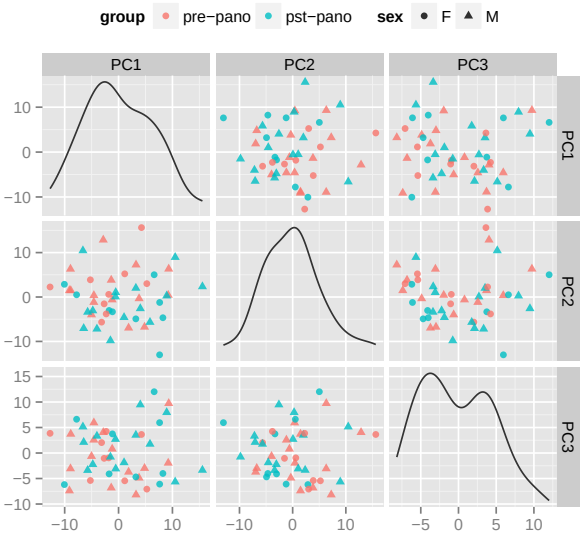
Figure 4.4: Differentialy Correlated Metabolites in pre- and post-pano samples. The plots show the correlation of two metabolites for pre- (blue) and post-pano samples (red). Males and female patients are distinguished by the point shapes. Regression lines and their confidence intervals are drawn separately for both groups. The metabolite levels were median-scaled, log-transformed and corrected for age and gender prior to the plotting.

At first changes of the central tendency of single metabolites were investigated. For 44 of the 447 tested compounds a significant influence of the Panobinostat treatment on their concentration could be observed by paired t-test. In 25 of these cases metabolites are higher concentrated in post-pano samples (Table 4.1).

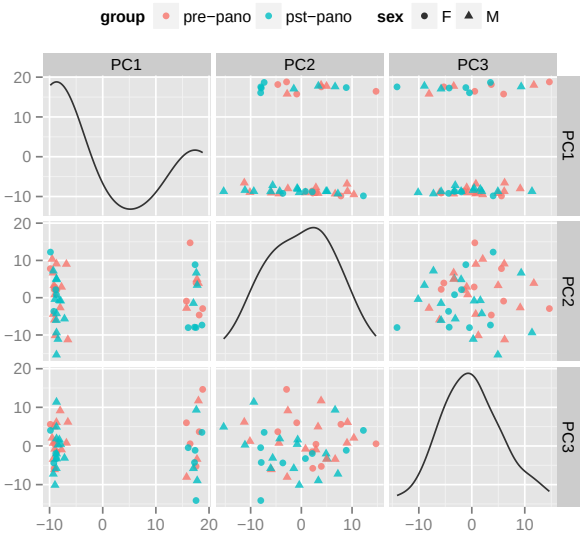
The test for different variances did not uncover any significant results but 5 metabolites were found to have higher variance in post-pano samples with FDR < 0.2 (Table 4.3). For metabolites with high fraction of missingness it was additionally tested if it was measured significantly more often in pre-pano samples than in samplepost-pano. By doing so, Fisher's exact test didn't unveil any significant results.

Ratios of metabolite intensities were investigated for two reasons: The first one is that the ratio of two related metabolites represents the state of their chemical equilibrium or if they are further apart in the same pathway the activity of the pathway. The second reason is that the ratio of two unrelated metabolites is a sort of intra-sample normalization and noise reduction (Altmaier et al., 2008; Suhre et al., 2011). As second possibility of contemplating interactions of metabolites, the first principal component of each pair of metabolites was tested for significant changes. This was done instead of linear regression with an interaction term to be able to use the metabolite interaction as outcome. By doing so 9026 metabolite ratios

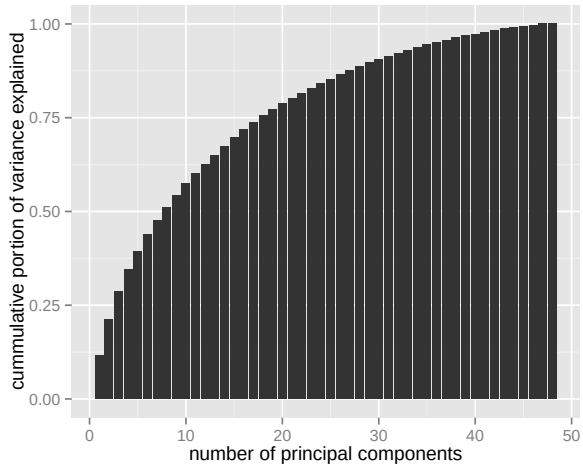
(a) PC scatter (w/o missing values)



(b) PC scatter (with missing values)



(c) Variance Explained (w/o missing values)



(d) Variance Explained (with missing values)

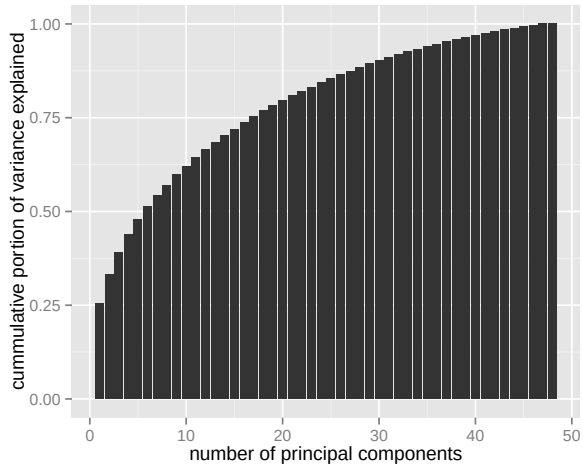
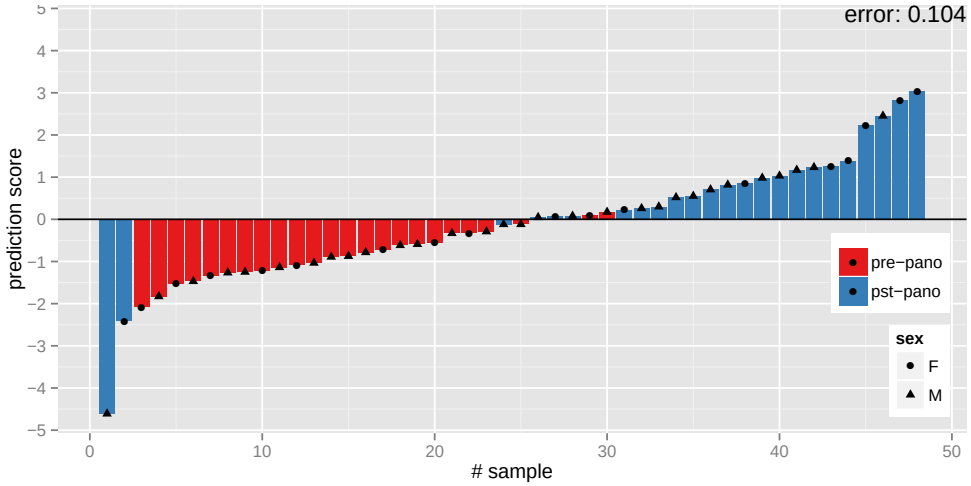


Figure 4.5: PCA of pre- and post-pano samples. The PCA was performed with two different types of imputed data: The first method was to omit sparse metabolites with more than 20 % missing values. The second was to use the sparse metabolites and generalize their intensities to 0 (missing) or 1 measured. (a) is the result of the first approach, (b) the second. The first three primary components were scattered against each other, their density is shown on the diagonal. The histograms (c) and (d) show the cumulative portion of explained variance for the two approaches. All data were corrected for influences due to gender and age, before.

(a) Performance



(b) Coefficients

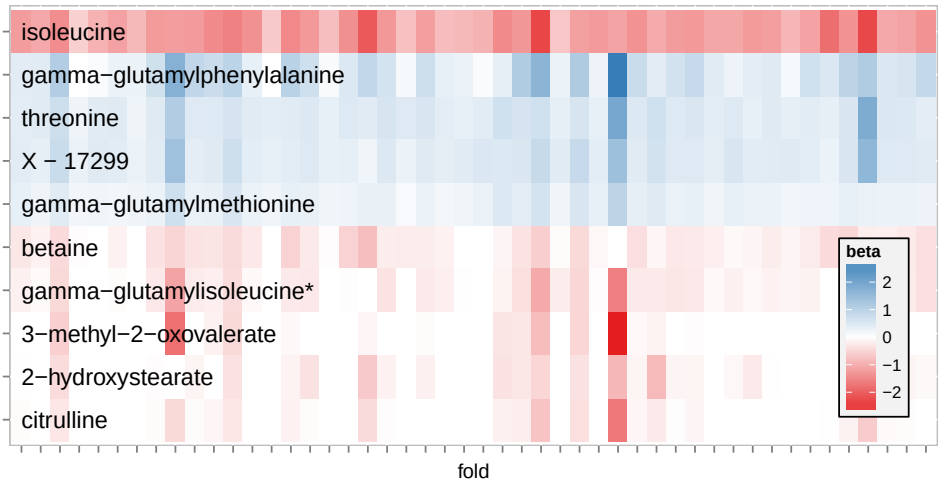


Figure 4.6: LASSO classification of pre-pano and post-pano samples. A LASSO regression was performed to asses the accuracy of class prediction. Therefore, a LOO CV was performed which means that for the prediction of each sample a separate LASSO model was trained. The prediction scores are shown in (a). Negative scores translate to "predicted as pre-pano" and positive scores to "predicted as post-pano". The real classes are indicated by the colors. Consequently, all blue samples with negative scores and red samples with positive scores are falsely classified. In (b) the beta coefficients for each of the trained models (rows) are shown. The ten metabolites with the highest average beta are shown.

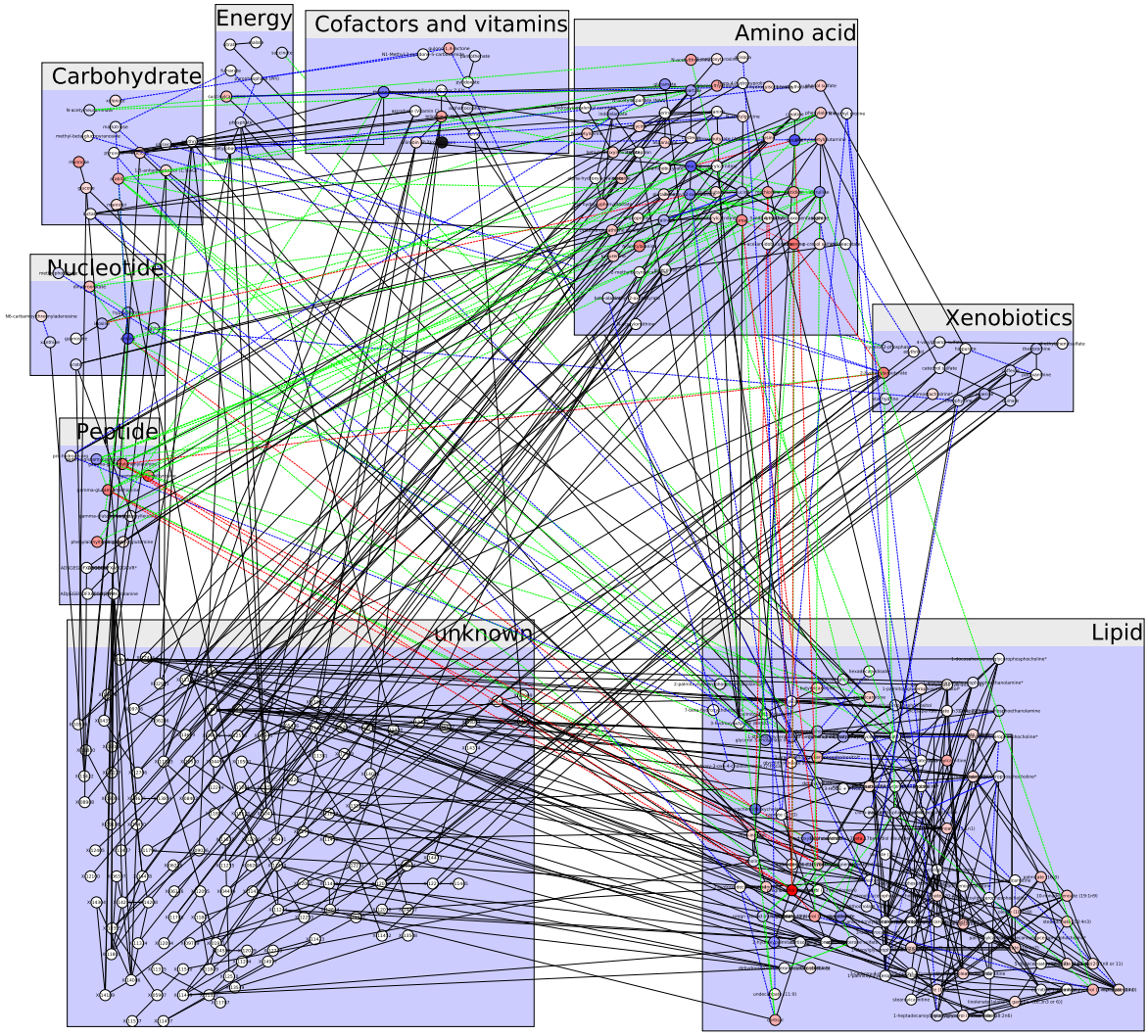


Figure 4.7: Mapping of results on a metabolic reference network. All results obtained from uni- and bivariate testing were mapped on a metabolic network. This network was inferred by Krumsiek et al. (2012) using partial correlations (black edges). Nodes, representing metabolites, were colored according to their p-values from the univariate t-test (red nodes are higher concentrated after treatment). Edges were colored based on the results of the bivariate tests: significantly differentially correlated metabolites were connected by green lines, significant ratios with blue and significant linear combinations with red lines (p-value 0.05, pgain > 100). If the edge was not present in the reference network it was added dashed (and colored, as well). Pathways were used for the grouping of nodes.

and 11,371 linear combinations of metabolite intensities were found to have altered concentrations. However, only 34 ratios and 31 linear combinations respectively have p-gains > 1000 and hence are likely to have a real combined effect instead of the large effect of the involved single metabolites (Tables 4.4 and Figure 4.3). According to Barlett's test, variances of metabolite ratios are significant in 14,050 cases, 6353 of them with p-gain > 1000, but only one of them – between pentadecanoate and serotonin (5HT) – is detected by Levene's test. Most of the significant p-values are rather caused by deviations from normality than variances differing between the groups. The variance is increased similar often before and after the treatment.

As third possibility of bivariate testing differential correlations were analyzed. If a reaction between two metabolites is compromised (e.g. due to mutations in the required enzyme), the correlation between them will change. In contrast the concentration of the individual metabolites might not change significantly due to other processes producing or degrading them. The test for differential correlation unveiled 20 significant pairs of metabolites. 8 of them were positively correlated in pre-pano and negatively in post-pano samples, for 12 pairs it was the other way round (Table 4.6, two examples are shown in 4.4).

The PCA unveils, that even though many metabolites differ between the pre- and post-pano samples, the greatest portion of variance of the data is not related to the treatment. The PCA was performed with two types of imputed data: For the first one sparse metabolites which had more than 20 % missing values were omitted. For the other one, the intensities of sparse metabolites were set to 0 (missing) or 1 (measured). The results of both approaches are shown. In the first case the first three principal components account for 27 % of the variance. The samples do not cluster in pre- and post-pano groups (figures 4.5a and (c)). In the latter case the greatest portion of variance is caused by the imputed sparse metabolites, which alone capture 25 % of the variance. However, it also does not separate the pre- from the post-pano samples and thus supports the results of the missing value analysis (Figure 4.5b and (d)).

The metabolites with the highest weights in the first principal component are – amongst others – fatty acids like palmitate, margarate and stearate. In the second principal component fatty acids from the eicosanoid metabolism like arachidonate, docosapentaenoate and lineoate have the highest weights while glycerophospholipids and alanine and glutamate are the main compounds of the third component.

By using the supervised LASSO regression for classification pre- and post-pano samples could be distinguished with an error rate of 12.5 %. This was assessed by performing a LOO CV. For each fold the optimal lambda and betas were calculated using all but one sample. The last sample was predicted using this model. The

trained models used in average 5 metabolites for classification, 3 of them, namely threonine, γ-glutamylmethionine and one unknown metabolite, were used in all folds. In total 18 metabolites were used together in all models, the 10 of them with the highest average weight (β) are shown in Figure 4.6b. The prediction for each sample is shown in Figure 4.6. Samples with negative score are predicted to be pre-pano, positive ones as post-pano.

The metabolites which were observed to be influenced by the panobinostat treatment were not significantly enriched in a significant pathway. The top ranked pathways were consistent among most of the different methods including the linear regression of combined metabolite z-scores. The most significant super-pathways were amino acids, nucleotides and peptides, the highest ranked sub-pathways γ-glutamyl, long-chain fatty acids and bile acids.

It was expected that most significant ratios and correlations would affect closely related metabolites and thus represent deregulated reactions or pathways. By mapping the results on a metabolic network it got obvious that this was not the case (Figure 4.7). The edges of the network were inferred using partial correlations by Krumsiek et al. (2012). The partial correlations are thought to reflect (direct) metabolic associations between metabolites. Most of the edges are within pathways. It can be seen that pairs of metabolites with significant correlations/ratios/linear combinations are usually not connected by an reference edge. There are much more colored edges between pathways than within pathways. Furthermore, it can be seen that many of the colored edges connect metabolites which were identified to be significantly altered by the univariate analysis. This suggests that testing for changes of metabolite ratios rather identified cases of “internal normalization” than real metabolic interactions.

4.1.2 Results (In Vitro)

As mentioned before, besides plasma samples, metabolites were also measured for DLBCL cell lines before and after those were treated with Panobinostat. The variance of metabolic profiles in cell samples is much lower than the variance of in vivo samples like blood plasma or serum. Hence only 4 cell samples were measured for both groups. Thus, changes in the variance of metabolites and metabolite ratios as well as missing value patterns could not be reliably detected. The results of the remaining tests are presented in the following.

The cell samples were not median scaled. Instead cell measurements were normalized by their protein content which was determined using the Bradford method.

After that data was logarithmized as all other measurements. Normalizing for sex or age was of course not necessary.

A first overview over the cell data is given by the heatmap in Figure 4.10. Samples (rows) and metabolites (columns) were clustered individually using hierarchical clustering with manhattan distance and average linkage. The differences between the metabolic profiles of treated and untreated cells are obvious, the groups form separated clusters. In contrast clustering of columns does not separate most of the super-pathways. At least lipids form a more or less isolated cluster (orange). Compared to the plasma samples the fold-changes of metabolite intensities are much higher but due to the pairing, p-values are nevertheless better for plasma samples.

The mean intensity of individual metabolites differed between cells before and after the treatment with panobinostat in 61 cases (Table 4.7). In total 40 metabolites were higher concentrated after Panobinostat treatment and 21 before (cf. Figure 4.8). Only urea, glycerol-3-phosphate and histidine were significant in both, cell and plasma samples, but there are several closely related metabolites which were significant in both comparisons.

Again metabolite ratios and PCA-linear combinations were tested for differences between the groups. 57 of the linear combinations and 104 ratios are significant in cells (Table 4.8). In about half of these cases either N-acetylmethionine, tyrosine, histidine or maltotetraose are contained. The first three are amino acids (derivatives), the latter is an oligosaccaride.

Just as the clustering, the first principal component perfectly separates between treated and untreated cells. This component accounts for $\approx 50\,\%$ of the observed variation (Figure 4.9). It is the linear combination of 1-oleoylglycerophosphoethanolamine, 2-palmitoylglycerol, guanosine, lysine, tyrosine and others.

(a) Plasma samples (paired t-test) **(b)** Cell samples (regression)

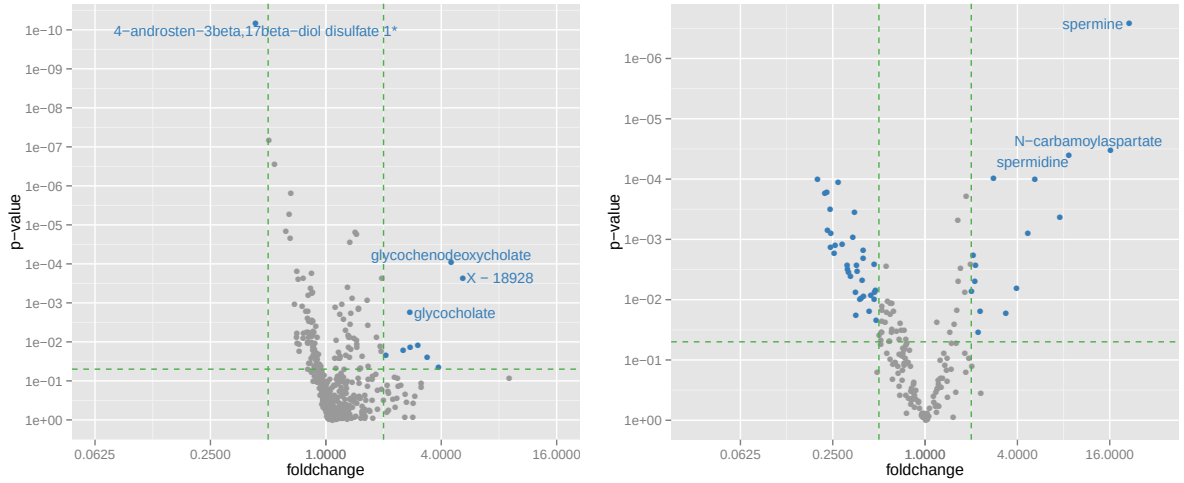


Figure 4.8: Volcano plots for pre- vs. post-pano comparison. The volcano plots show results of the univariate test for changes in the mean concentration of metabolites for plasma samples (a) and cell samples (b). P-values were not corrected for this plot. Due to the pairing of plasma samples several metabolites with relatively low fold-changes have high p-values.

(a) Primary components **(b)** Variance explained

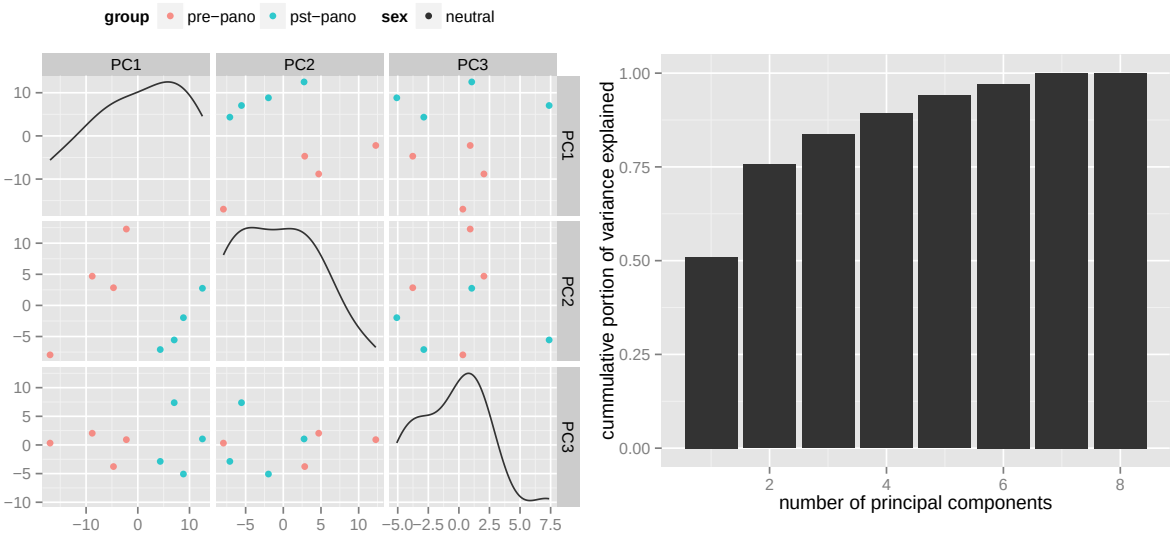


Figure 4.9: PCA of cell samples. (a) shows the scatterplot of the first three principal components against each other. Additionally the densities are drawn on the diagonal. (b) is the histogram of cumulative fraction of variance explained by the primary components.

metabolite	#	*lm p-value*	*beta*	*se*
spermine	8	4.7×10^{-5}	−4.40	0.18
N-carbamoylaspartate	8	2.4×10^{-3}	−3.98	0.36
spermidine	8	2.4×10^{-3}	−3.09	0.29
uridine 5′-diphosphate (UDP)	8	2.9×10^{-3}	−1.50	0.16
eicosenoate (20:1n9 or 11)	8	2.9×10^{-3}	2.35	0.26
maltopentaose	8	2.9×10^{-3}	−2.39	0.26
xanthosine	8	2.9×10^{-3}	1.88	0.21
dihomo-linoleate (20:2n6)	8	3.4×10^{-3}	2.14	0.26
docosadienoate (22:2n6)	8	3.4×10^{-3}	2.12	0.26
adenosine 5′-monophosphate (AMP)	8	3.4×10^{-3}	−0.90	0.11
myo-inositol	8	5.2×10^{-3}	2.10	0.29
1-stearoylglycerophosphoinositol	8	5.2×10^{-3}	1.58	0.22
orotate	8	5.9×10^{-3}	−2.86	0.41
isobutyrylcarnitine	8	6.1×10^{-3}	−0.71	0.10
dihomo-linolenate (20:3n3 or n6)	8	8.5×10^{-3}	2.19	0.35
glucose-6-phosphate (G6P)	8	8.5×10^{-3}	−2.19	0.35
erucate (22:1n9)	8	8.5×10^{-3}	2.09	0.34
1-palmitoylplasmenylethanolamine*	8	9.2×10^{-3}	1.61	0.27
10-nonadecenoate (19:1n9)	8	1.1×10^{-2}	1.85	0.32
mead acid (20:3n9)	8	1.1×10^{-2}	1.96	0.34
docosapentaenoate (n3 DPA; 22:5n3)	8	1.2×10^{-2}	2.17	0.39
1-stearoylglycerophosphoethanolamine	8	1.2×10^{-2}	1.38	0.25
docosahexaenoate (DHA; 22:6n3)	8	1.4×10^{-2}	2.11	0.39
cytidine 5′-monophosphate (5′-CMP)	8	1.4×10^{-2}	−1.07	0.20
Isobar: fructose 1,6-diphosphate, glucose 1,6-diphosphate, myo-inositol 1,4 or 1,3-diphosphate	8	1.5×10^{-2}	1.35	0.26

Table 4.7: Metabolites with different concentration between pre- and post-pano cell samples. All significantly differentially concentrated metabolites between cell samples before and after treatment with Panobinostat are shown (p-value ≤ 0.15). The linear model (lm) p-value, the regression coefficient beta and the standard error of the model (se) are shown. Positive beta indicates higher concentration after the treatment.

metabolite 1	*metabolite 2*	#	*p-value*	*p-gain*	*beta*
N- acetylmethionine	phosphoethanolamine	8	1.4×10^{-4}	2.4×10^{5}	−1.35
N- acetylmethionine	tyrosine	8	1.4×10^{-4}	1.8×10^{5}	−1.67
maltopentaose	X - 13005	8	1.0×10^{-5}	1.6×10^{5}	−3.48
choline phosphate	tyrosine	8	1.6×10^{-4}	8.2×10^{4}	−0.75
tyrosine	X - 13162	8	1.6×10^{-4}	5.2×10^{4}	1.23
phosphoethanolamine	serine	8	1.8×10^{-4}	5.2×10^{4}	0.77
proline	X - 13192	8	1.6×10^{-4}	4.2×10^{4}	0.88
phosphoethanolamine	X - 13162	8	1.9×10^{-4}	3.7×10^{4}	0.92
maltotetraose	X - 13005	8	1.6×10^{-4}	3.6×10^{4}	−2.09
glutathione, reduced (GSH)	histidine	8	1.6×10^{-4}	2.3×10^{4}	−1.54
2- methylbutyrylcarnitine (C5)	prolylglutamate	8	2.4×10^{-4}	2.0×10^{4}	−0.71
histidine	N- acetylmethionine	8	1.6×10^{-4}	1.9×10^{4}	1.81
N- acetylneuraminate	tyrosine	8	1.6×10^{-4}	1.7×10^{4}	−0.69
carnitine	X - 13162	8	2.9×10^{-4}	1.7×10^{4}	0.81
prolylglutamate	X - 13192	8	1.9×10^{-4}	1.6×10^{4}	1.05
glycine	pantothenate	8	1.6×10^{-4}	1.3×10^{4}	1.13
glutathione, reduced (GSH)	prolylglutamate	8	2.4×10^{-4}	1.2×10^{4}	−1.01

Table 4.8: Metabolite ratios differing between pre- and post-pano cell samples. The results of the linear regression comparing metabolite ratios between the groups are shown. The p-value were FDR-corrected. The p-gain is calculated as the ratio of the minimum p-value of the separate regression models of the involved metabolites divided by the p-value of the ratio regression mode. All ratios with p-gain $>$ 10,000 are reported.

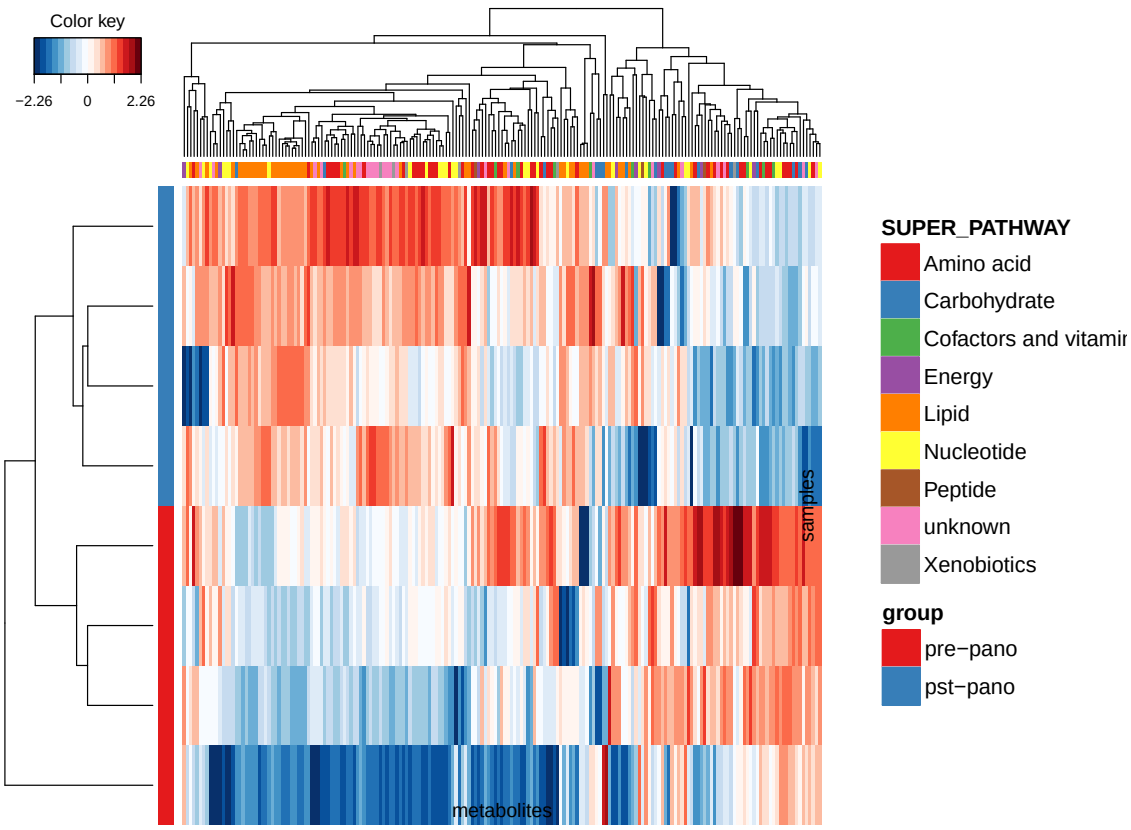


Figure 4.10: Heatmap of cell samples. The heatmap shows the scaled and logarithmized metabolite intensities for the cell samples. Rows represent samples, columns metabolites. The row colors indicate the group (pre- or post-pano), the column colours indicate pathways. Rows and columns were separately clustered using hierarchical clustering with Manhattan distance and average linkage.

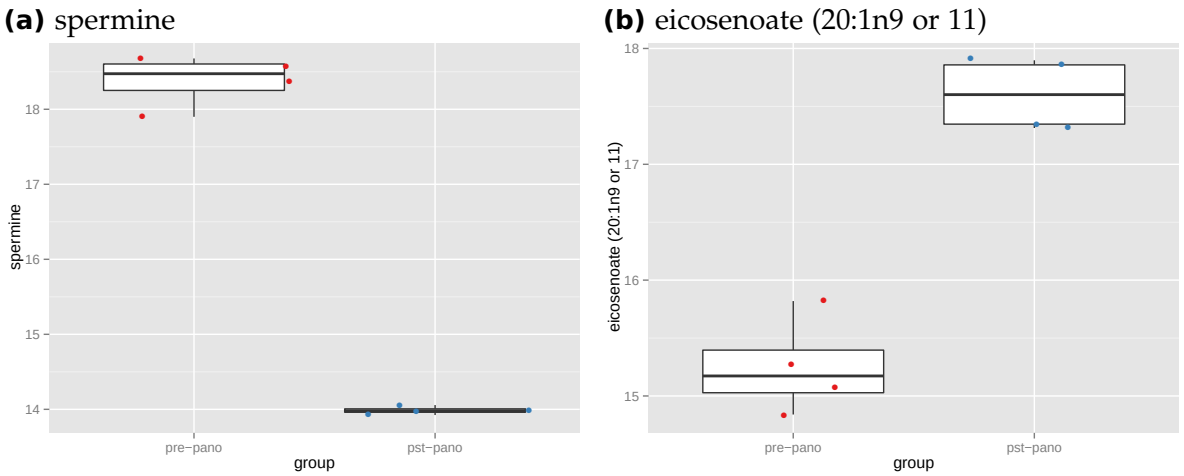


Figure 4.11: Examples for metabolites with different concentration between pre- and post-pano cell samples. The intensities of the cell samples before and after the Panobinostat treatment are shown. Left boxes with red points show pre-pano samples, blue points on the right post-pano samples.

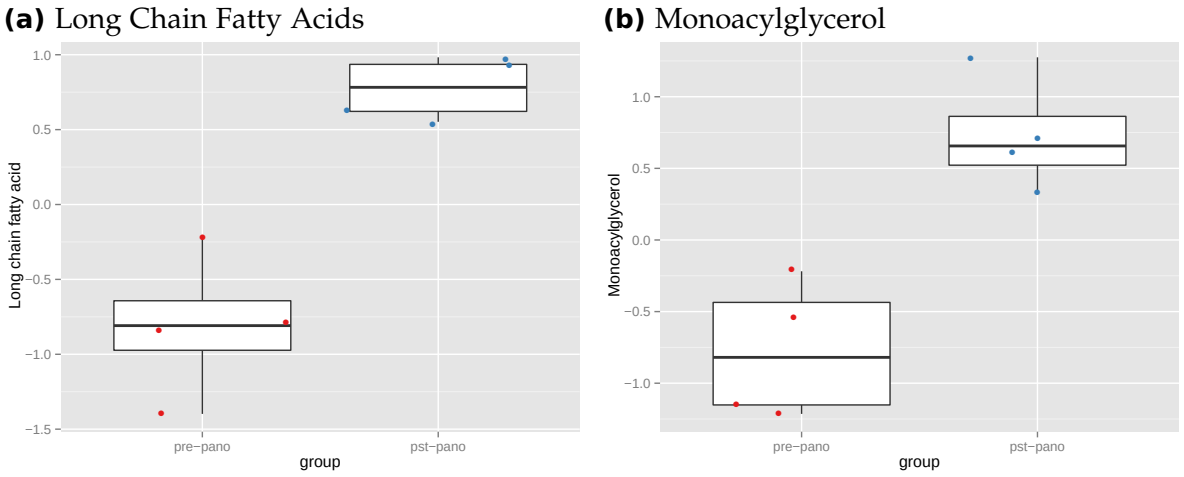


Figure 4.12: Pathway enrichment of metabolites which differ between pre- and post-pano cells. To analyze PWEA metabolite intensities were transformed to z-scores and pathway intensities were calculated as the mean of the z-scores of the involved metabolites. These were tested for significant changes as it was done for single metabolites. Two significant pathways are shown here. On the y-axis the mean metabolite z-scores are plotted.

metabolite 1	*metabolite 2*	*#*	*l*	*p-value*	*p-gain*	*beta*
maltotetraose	X -19378	8	-1	1.5×10^{-4}	3.7×10^{5}	-2.23
histidine	N-acetylmethionine	8	-1	1.5×10^{-4}	2.1×10^{5}	2.21
N-acetylmethionine	tyrosine	8	-1	1.5×10^{-4}	1.7×10^{5}	-2.22
creatine	N-acetylmethionine	8	-1	1.5×10^{-4}	5.6×10^{4}	2.31
maltotetraose	X -13005	8	-1	2.8×10^{-4}	5.2×10^{4}	-2.25
creatine	pantothenate	8	-1	1.8×10^{-4}	3.8×10^{4}	2.37
histidine	X -13192	8	-1	6.1×10^{-4}	2.2×10^{4}	2.10
isobutyrylcarnitin	mead acid (20:3n9)	8	-1	1.5×10^{-4}	1.9×10^{4}	-2.46
maltotetraose	X -13004	8	-1	6.1×10^{-4}	1.2×10^{4}	-2.16
cholesterol	glutathione, reduced (GSH)	8	-1	8.2×10^{-4}	1.0×10^{4}	1.85

Table 4.9: Metabolite combinations differing between pre- and post-pano cell samples. The linear models of PCA-combined metabolites are shown. The first primary component of each pair of metabolites was tested for difference between the groups using linear regression. The ratio of the loadings of the components (l) is shown. P-values were FDR-corrected before. Metabolite pairs with p-gain $>$ 10,000 are lsited.

4.1.3 Discussion

The mode of action of Panobinostat and other HDACis is largely unknown until today. One part of their effect is probably achieved by their influence on gene expression through the control of histone acetylation patterns. Some other proteins like STAT-3 and BCL-2 were proposed as potential targets of HDACs but real targets, the way how Panobinostat prevents tumor growth are not known, yet. One of these aspects id the influence of Panobinostat on the metabolism. Here plasma samples of relapsed/refractory DLBCL patients before and after the treatment were compared. The samples of DLBCL cells allow to distinguish between direct effects from the tumor cells and secondary effects due to the liver or other organs. Furthermore, cells are expected to exhibit acute effects after a relatively short Panobinostat treatment of 24 h compared to the subchronic effects of Panobinostat in plasma samples. Differences between these samples are obvious.

While PCA perfectly separates treated from untreated cells this is absolutely not the case for plasma samples. The much higher variance of the in vivo samples certainly covers several metabolic changes but there also might be metabolic changes in tumor cells which do not influence the blood or are immediately "corrected" by other organs. On the other hand effects which are seen in plasma samples but not tumor cells are likely to be off-target effects of panobinostat.

The metabolic differences which were observed to differ between pre- and post-pano samples can be divided in several groups. Those are summarized and discussed in the following.

4.1.3.1 Amino acids and Peptides

The first set of metabolites which showed substantial differences between the groups contains amino acids and peptides. Isoleucine, threonine, methionine and several γ-glutamyl-amino acids have different concentrations in the treated and untreated plasma samples. Specifically, γ-glutamyl-methionine, -isoleucine, -phenyalanine and -tyrosine are amongst the highest ranked metabolites of the plasma comparison. Furthermore, the peptides phenylalanylphenylalanine and phenylalanyltryptophan show significantly altered intensities. The amino acids and peptides are not consistently regulated but their directions alters. Some od them, like threonine or γ-glutamylphenylalanine are higher concentrated after the treatment, while others, like isoleucine and threonine have decreased concentration after the Panobinostat treatment.

Most of the amino acids and γ-glutamyl amino acids are higher concentrated in post-pano samples (see Figure 4.1b and (d)). The only exceptions are isoleucine and γ-glutamylisoleucine which are both higher concentrated before the treatment (Figure 4.1a and (c)). Furthermore, γ-glutamylisoleucine was also observed to have a higher variance in post-pano samples (Table 4.3). The ratios between γ-glutamylisoleucine and other γ-glutamyl-amino acids as well as other compounds are also significantly different in pre- and post-pano samples, but this might mainly reflect the reduced variance through the internal normalization.

For the plasma samples the group of amino acids and peptides is the most important for discrimination of the groups using multivariate regression. LASSO selects 6 metabolites from this group are amongst the 10 highest rated variables (Figure 4.6). However neither the group of amino acids nor peptides (which contains γ-glutamyl-amino acids) was found to be significant by any Pathway Enrichment Analysis (PWEA) method.

In the cell line samples the amino acids lysine, tyrosine and histidine were increased after treatment. Those and others were also found in many significant ratios and linear combinations. For example serine is higher concentrated in pre-pano samples but due to its high variance this difference is not significant. by normalizing its intensity by phosphoethanolamine, variance is drastically reduced and the difference detected as significant. This is similar for the dipeptide prolylglutamate which is significantly increased after treatment when normalized by the 2-methylbutyrylcarnitine intensity (and others).

The deregulation of N-acetylmethionine and N-acetylaspartate are detected by univariate tests but also much more significant when internally normalized. In plasma samples N-acetylalanine and N-acetylthreonine to be higher concentrated after the Panobinostat treatment. This could be a direct effect of Panobinostat which inhibits their deacetylation.

Several metabolites which are part of amino acid production or degradation show altered behavior after the panobinostat treatment in plasma samples. One of them is betaine which is a glycine precursor. It is higher concentrated in pre-pano samples. Another examples is 3-methyl-2-oxovalerate which is a degradation product of isoleucine. Additionally its ratios with 4-methyl-2-oxopentanoate and 3-methyl-2-oxobutyrate are highly significantly decreased after treatment. These are the degradation products of leucine and valine respectively indicating a shift in the metabolism of branched-chain amino acids.

The increased concentration of γ-glutamyl amino acids is presumably a sign for increased Gamma-Glutamyl Transferase (GGT) activity which in advance is a

biomarker for oxidative stress (Pandur et al., 2007). GGT is an enzyme that catalyzes the transfer of γ-glutamyl from gluthatione (GSH) to amino acids to form γ-glutamyl amino acids. Its expression was reported to be increased in many different neoplasms, including leukemia, before (Dominici et al., 1999). Lum and Gambino (1972) found increased GGT activity in lymphomas with liver involvement, which is the case in many DLBCL cases (Holland, 2000). A possible cause for the altered GGT activity between the groups is that Panobinostat alters its expression by changing DNA accessibility. Panobinostat might also influence transcription factors or GGT itself.

GGT is a main regulator of GSH which is a very effective antioxidant and involved in the detoxification of several metabolites mostly in the liver. It was suggested that specific levels of GSH are required for an optimal immune function and that changes of its level in leukocytes might have drastic effects (Dröge and Breitkreutz, 2000). GSH has a very ambiguous relationship to cancer. On the one hand it was suggested that increased levels enhance detoxification of Reactive Oxygen Species (ROS) and prevent tumor development. On the other hand, elevated GSH levels are associated with chemotherapy resistant tumors (Balendiran, Dabur, and Fraser, 2004). GSH is also part of the eicosanoid production (see below, Section 4.1.3.2).

In cell samples the increased levels of γ-glutamyl amino acids can not be observed but the oxidized form of gluthatione (GSSG) is significantly higher concentrated. Furthermore, the ratios between GSH and histidine and the dipeptide prolylglutamate respectively are significantly lower after panobinostat treatment. This indicates that higher levels of gluthatione are needed for detoxification of ROS.

Gulono-1,4-lactone is a precursor of vitamine-C and also a marker for oxidative stress. It is elevated in post-pano samples and therefore supports the hypothesis of increased oxidative stress after Panobinostat treatment. Its precursor myo-inositol had higher levels after the treatment in cell samples as well.

Summarizing, the increased levels of amino acids, γ-glutamyl amino acids and the related metabolites are thought to be a sign for increased activity of the γglutamylcycle (cf. Figure 5.1) and thus a sign for increased oxidative stress. This is supported by the presence of other stress-related metabolites and the increase of the oxidized from of GSH. The results of Ruefli (2001) suggest that the HDACi Suberoylanilide Hydroxamic Acid (SAHA), which is used as chemotherapeutic as well, induces apoptosis of tumor cells by activating the production of ROS. If this holds true the increased oxidative stress might be needed for HDACis to work and reduction of the stress by the increased activity of the γ-glutamyl cycle might be even counterproductive. This conforms with the observation that many resistant tumors show elevated GSH levels (Balendiran, Dabur, and Fraser, 2004).

However this is just a hypothesis and has to be further investigated in future. The complex function of GSH seem to be important for the treatment of cancer. The reason why isoleucine and its degradation products are altered contrary to the other amino acids also remains unknown and thus a subject to upcoming experiments.

4.1.3.2 Fatty Acids

When comparing DLBCL cells before and after panobinostat treatment it can Furthermore, be observed that many (long chain) fatty acids are higher concentrated after treatment. Those are enriched in the long chain fatty acid and the essential fatty acid subclasses. More specifically many eicosanoid precursors (e.g. arachidonate (20:4n6), eicosenoate (20:1n9 or 11), eicosapentaenoate (EPA; 20:5n3), dihomo-linoleate (20:2n6) and dihomo-linolenate (20:3n3 or n6) are up-regulated after panobinostat treatment.

Apart from the eicosanoid related fatty acids, other lysolipids namely 1-stearoylglycerophosphoinositol, 1-stearoylglycerophosphoethanolamine, 1-oleoylglycerophosphoethanolamine and 1-palmitoylplasmenylethanolamine are elevated after the treatment as well as the monoacylglycerols 1-stearoylglycerol and 2-palmitoylglycerol. The ratio of free phosphoethanolamine and N-acetylmethionine and serine respectively is higher as well after the treatment. cytidine 5'-diphosphocholine, which is an intermediate in the phospholipid synthesis, was increased after the treatment.

In plasma samples this trend was not that obviously but also visible. Glycerol-3-phosphate, the backbone of glycerophospholipids, was less abundant after panobinostat treatment. The Monoacylglycerol 1-stearoylglycerol is higher concentrated in post-pano samples and 1-oleoylglycerophosphoethanolamine has a higher variance after treatment. Instead, 2-hydroxystearate, which is an intermediate metabolite in the fatty acid oxidation, and isobutyrylcarnitine, a short chain fatty acid bound to the fatty acid transporter carnitine, are less abundant after the treatment. The medium chain fatty acids Caprylate and Caprate both are positively correlated with N-methyl proline in pre-pano but negatively in post-pano samples. Analogously, the correlation of 2-hydroxydecanoic acid with two unknown compounds changed after the treatment. 1-eicosadienoylglycerophosphocholine is the only fatty acid related to eicosanoids, which shows any changes between pre- and post-plasma samples. Its correlation with an unkown metabolite changed significantly. But 12-HETE which is a precursor of leukotrienes, a subclass of eicosanoids, is higher concentrated before the treatment. This difference did not reach global significance but the p-value of 0.076 (in the paired t-test) suggests that there might be a difference in its concentration.

Altered eicosanoid production can be observed DLBCL cell samples and to a certain extend in plasma samples after treatment with panobinostat. This suggests that they are actually produced by the tumor cells. The eicosanoid synthesis is closely linked to the GSH metabolism. GSH is incorporated to produce leukotriene-C4 from -A4 and γ-glutamyl amino acids are set free during the conversion of leuktriene-C4 to -D4 (Anderson, Allison, and Meister, 1982) (cf. Figure 5.1). The increased concentration of eicosanoid fatty acid metabolites allows for two possible explanation: The first is that eicosanoid production is increased and therefore its precursors are provided at higher concentrations. The second possible explanation is that eicosanoid production is hampered and therefore its precursors accumulate. Eicosanoids are signal lipids which are long known as pro-inflammatory agents (Khanapure et al., 2007), but it was also found that they support growth of different cancers like breast and colorectal cancer (Tronstad et al., 2003; Wang and DuBois, 2010). If this holds true for DLBCL their inhibition might be a positive effect of Panobinostat.

But there are many other fatty acids and related compounds which behaved different after treatment with Panobinostat. Long chain fatty acids accumulate after the treatment in both, cells and plasma samples, while their oxidation products and shorter fatty acids are less abundant. Fatty acids have various functions in the cell ranging from building membranes to signaling and energy storage. All those functions are important for malignant cells to proliferate as fast as possible. Hence, disregulation of lipid metabolism is a hallmark of cancer (Nomura and Cravatt, 2013; Baumann, Sevinsky, and Conklin, 2013). That is why fatty acids being higher concentrated after the treatment is counterintuitive. The reasons for this have to be figured out in coming experiments. Interestingly sugars like maltopentaose, maltotetraose, glucose-6-phosphate and other energy carrier like uridine 5'-diphosphate or were observed to be down regulated after the treatment instead.

4.1.3.3 Nucleotide metabolism

Another pathway which seams to be influenced by panobinostat is the metabolism of nucleotides. Inosine was observed to have lower mean concentration as well as lower variance in pre samples (Figure 4.2a). Inosine is a degradation product of purine bases. Further downstream in this pathway urate and then urea are produced. The latter is higher concentrated after the treatment in cells as well as in plasma samples. In the cell samples adenosine 5'-monophosphate and cytidine 5'-monophosphate were both lower concentrated after the treatment.

Furthermore, the purine metabolism uridine was observed at higher concentration in pre-pano samples. In cells one can observe its degradation product beta-

alanine at lower concentration after treating them with panobinostat as well. The urdine precursor N-carbamoylaspartate further reacts to orotate and then Uridine-5'-Monophosphate (UMP) and Uridine-5'-Diphosphate (UDP). Both UMP antecedents and UDP are higher concentrated before the treatment with HDACi in cell lines. Urea is part of the pyrimidine degradation as well.

For plasma samples the nucleotide pathway was the most significantly enriched pathway identified by MSEA, CAMERA and the pathway regression approach. Nevertheless it did not reach global significance.

In general the metabolites purine and pyrimidine pathways were observed to be down regulated after the treatment with HDACi in plasma samples as well as cell lines. The only exception is urea, but this has multiple functions and is, besides the nucleotide degradation, e.g also involved in amino acid metabolism. It was previously shown that cancer cells adapt expression of several genes to enhance production of nucleotides and prevent their degradation (Hatse, De Clercq, and Balzarini, 1999). This is done to meet the requirements of the accelerated proliferation of neoplasms. Thus decrease of nucleotide levels is a sign for reduced tumor growth. However it remains unknown if Panobinostat influences the nucleotide metabolism directly and thereby prevents tumor growth or if tumor growth is inhibited in any other way and decreased nucleotide levels are just a consequence of this. If the expression of genes which control the nucleotide levels are changed by the tumor cell HDACis could counteract this effect either through epigenetic effects of modification of transcription factors.

4.1.3.4 Steroid Hormones

Another group of metabolites which changed due to the treatment are the steroid hormones. Those are usually heavily depend on gender which is one of the reasons why data was corrected for gender (and age) before. The differences of the mean concentrations between pre- and post-pano samples are very low but the same changes can be seen for nearly all patients. Therefore, pairing of samples helped a lot in this case.

The corticosteroid, pregnen-diol disulfate, was found to be higher concentrated in post-pano samples. The ratios of it together with pregnenolone sulfate as well as the hydroxylated form hydroxypregnenolone disulfate and pregn steroid monosulfate respectively were shifted accordingly. Ratios of those and two unknown metabolites changed in the same way. 5-androsten-3-beta, 17-beta-diol disulphate and 5alpha-androstan-3beta,17beta-diol disulfate levels were elevated concordantly with the

pregnen-diol disulfate concentration in post-pano samples. Furthermore, their shared precursor dehydroisoandrosterone sulfate (DHEA-S) was higher concentrated after the treatment, as well. It was previously shown that androsten-17alpha-diol has the potential to inhibit tumor growth in different cancer cell lines (Loria and Graf, 2012; Graf, Jia, and Loria, 2007).

Steroids were not observed to behave different in any way in the cell samples. This suggests that the alteration of steroid balance is not a direct effect of panobinostat on the tumor but e.g. an off-target effect of panobinostat or a tumor suppressor produced by other organs. Pregnenolon is structurally very similar to the pharmacological, synthetic steroid prednisone. This is part of the standard R-CHOP treatment. Perhaps pregnenolone or its derivatives have a similar effect on DLBCL and are body-produced anti-cancer agents – and might be induced by Panobinostat. This should be further investigated in the ongoing trial.

Two bile acids, glycochenodeoxycholate and glycocholenate sulfate, which are – just like the discussed steroid hormones – metabolites of cholesterol had decreased levels after treatment in plasma samples. This could be an "off-target" effect which is caused by the higher rate of steroid production from cholesterol.

4.1.3.5 Spermine Biosynthesis

In DLBCL cells the polyamines spermidine and spermine both have lowered concentrations after the panobinostat treatment. Both are produced from S-adenosylmethionine (SAM) which also is less abundant after the treatment. As co-product of both, spermine and spermidine synthesis from SAM, 5-methylthioadenosine (MTA) is produced. MTA is also degradation product of methionine and there is evidence that it is involved in the regulation of various cellular processes like gene expression, proliferation, differentiation and apoptosis (Avila et al., 2004). MTA is also reduced due to panobinostat treatment. Spermin and spermidine are thought to act stabilizing on DNA (Casero and Marton, 2007).

SAM is synthesized from methionine and adenosine 5'-triphosphate. In cells N-acetylmethionine and adenosine 5'-monophosphate were also observed to be lower concentrated after the treatment. The first one seperated and the second one as metabolite ratio.

4.2 Primary vs. Relapsed/Refractory Lymphoma

The comparison of the plasma samples of primary and relapsed/refractory lymphoma patients is intended to unveil differences between the two types of tumors. In contrast to the *resistant vs. sensitive* comparison, it is not meant to predict the success of the R-CHOP-therapy beforehand. A major problem of DLBCL is that even though the R-CHOP treatment works good, patients who do not respond to the treatment or relapse have very bad prognosis. This comparison was intended to enhance the knowledge about the differences of primary tumors which respond to R-CHOP and relapsed tumors which did not respond. These differences might be crucial for the enhancement of drugs specifically designed for R-CHOP resistant tumors.

The comparison was performed using plasma samples measured in the experiments **pano-plasma-1** and **pano-plasma-2**. In total 9 samples of patients with primary DLBCL and 23 patients with relapsed DLBCL were analyzed. The primary DLBCL patients have not received any treatment at this time. The relapsed DLBCL patients were treated with R-CHOP before but this treatment was finished at least six month before the measurements. Thus, this is a comparison of two different types of lymphoma independent from influences of the treatment. The data was median scaled, logarithmized and corrected variation due to gender and age before as described in Chapters 2 and 3.

4.2.1 Results

The results of the comparison between plasma samples of patients with primary and relapsed/refractory DLBCL are presented in the following. All p-values were adjusted using BH correction. Metabolites which were measured in less than eight samples in any of the compared groups were omitted.

metabolite	#	*lm p-value*	*beta*	*se*
glycerol 3-phosphate (G3P)	37	1.5×10^{-4}	1.05	0.17
glycerophosphorylcholine (GPC)	29	4.2×10^{-4}	1.58	0.27
serine	37	4.4×10^{-4}	0.87	0.16
X -13478	28	3.2×10^{-3}	−0.93	0.19
histidine	37	3.3×10^{-3}	0.52	0.11
1-oleoylglycerophosphoethanolamine	36	3.3×10^{-3}	1.06	0.23
X -17629	29	7.8×10^{-3}	−2.12	0.48
1-stearoylglycerophosphoethanolamine	37	9.6×10^{-3}	0.95	0.23
acetylcarnitine	37	1.4×10^{-2}	−0.68	0.17
glycine	37	1.8×10^{-2}	0.61	0.16
X -11204	36	1.8×10^{-2}	−1.09	0.29
gamma-glutamylglutamine	29	1.8×10^{-2}	1.30	0.34
13-HODE + 9-HODE	35	2.4×10^{-2}	1.03	0.28
trans-4-hydroxyproline	37	2.4×10^{-2}	0.76	0.21
1-palmitoylglycerophosphoethanolamine	37	3.1×10^{-2}	0.74	0.21
uridine	37	3.4×10^{-2}	0.58	0.17
carnitine	35	3.4×10^{-2}	−0.34	0.10
hexanoylcarnitine	37	3.9×10^{-2}	−0.73	0.22
X -13230	28	3.9×10^{-2}	−0.44	0.13
2-arachidonoylglycerophosphoethanolamine*	37	3.9×10^{-2}	0.87	0.27
androsterone sulfate	37	4.4×10^{-2}	−1.49	0.46
pyrophosphate (PPi)	36	4.4×10^{-2}	−1.26	0.39
asparagine	37	4.4×10^{-2}	0.57	0.18
stearidonate (18:4n3)	35	4.4×10^{-2}	1.12	0.35
glutamate	37	4.4×10^{-2}	0.88	0.28
quinate	33	4.6×10^{-2}	2.15	0.68
1-oleoylglycerophosphocholine	37	4.6×10^{-2}	0.96	0.31

Table 4.10: Metabolites with different concentration between primary and relapse tumor samples. The table shows the results of the linear regression comparing median-scaled, log-transformed and corrected (for age and gender) metabolite intensities. All metabolites with significantly (p-value < 0.05) different concentrations between the groups are listed. The p-values of linear regression model (lm) together with its standard error (se) and regression coefficient (beta) are shown. Positive betas indicate that the metabolite is more abundant in relapsed tumor samples.

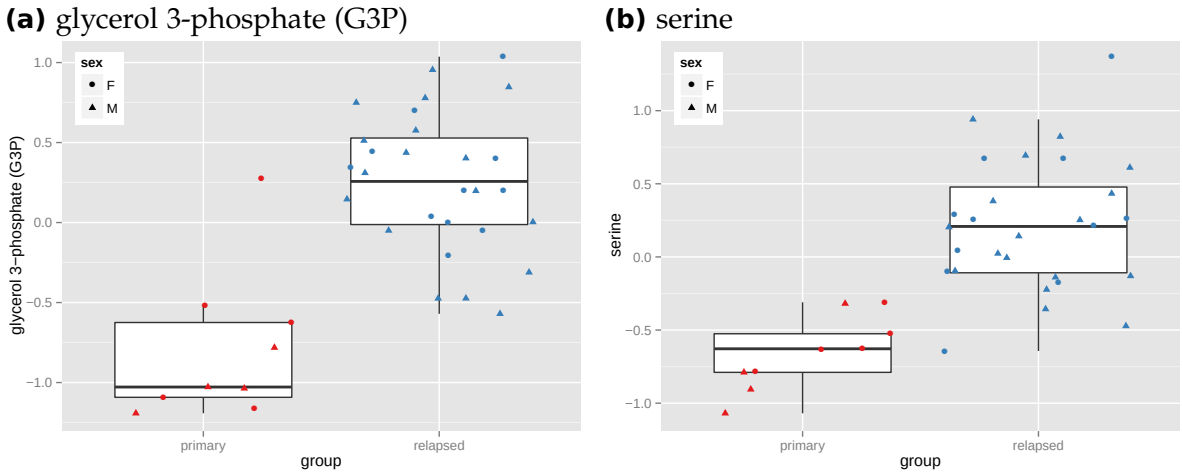


Figure 4.13: Examples for metabolites with different concentration between primary and relapse tumor samples. The distributions of median-scaled, logarithmized and age- and gender adjusted metabolite intensities are drawn stratified by the type of the tumor (colours). The boxes represent the quartiles of the metabolite distributions.

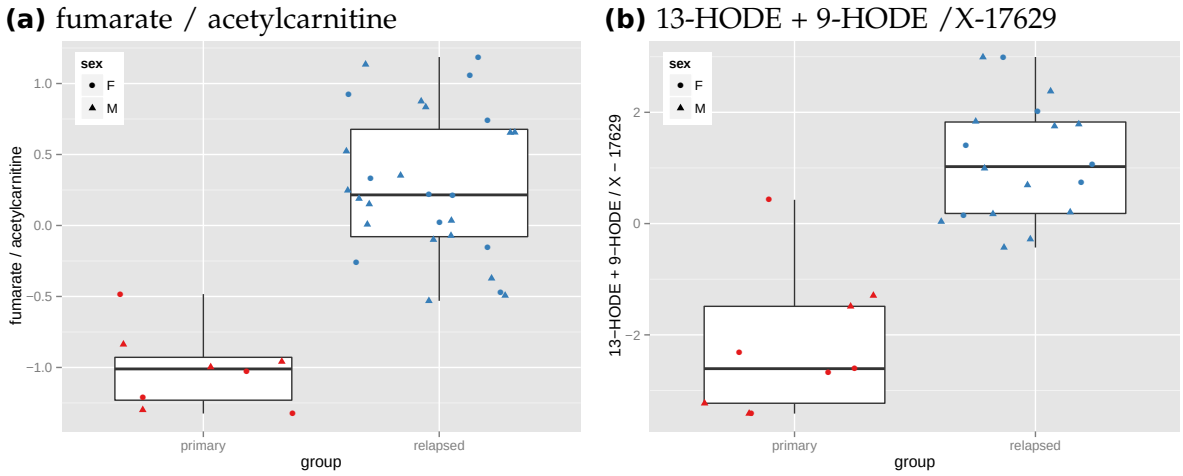


Figure 4.14: Examples of metabolite ratios differing between primary and relapse samples. The boxplots show distributions of metabolite ratios (of median-scaled, logarithmized and age- and gender corrected) metabolite intensities). The left boxes contain primary tumor samples, the right one relapsed.

metabolite 1	*metabolite 2*	*#*	*p-value*	*p-gain*	*beta*
fumarate	hexanoylcarnitine	36	8.4×10^{-4}	1.5×10^{4}	1.40
fumarate	X - 12100	36	1.7×10^{-3}	6.5×10^{3}	1.23
1- arachidonoyl-glycerophospho-choline*	2- arachidonoyl-glycerophospho-choline*	36	3.5×10^{-3}	6.3×10^{3}	0.85
fumarate	acetylcarnitine	36	6.9×10^{-4}	5.7×10^{3}	1.30
homocitrulline	X - 17299	37	2.6×10^{-3}	2.6×10^{3}	1.17
1- palmitoleoylglyc-erophosphocholine*	2- arachidonoyl-glycerophospho-choline*	36	3.2×10^{-3}	1.6×10^{3}	1.26
13- HODE + 9-HODE	X - 17629	27	8.4×10^{-4}	1.3×10^{3}	3.36
1- oleoylglyc-erophosphocholine	2- arachidonoyl-glycerophospho-choline*	36	2.2×10^{-3}	1.2×10^{3}	1.10
X - 11593	threitol	28	3.5×10^{-3}	1.0×10^{3}	-0.95

Table 4.11: Metabolite ratios differing between primary and relapse samples. The table shows the results of the linear regression of metabolite ratios (as described in Section 3.3.1). The number of samples for which both metabolites were measured, FDR-corrected p-values, the p-gain and the regression coefficient beta are shown. The p-gain is calculated as the ratio of the minimal p-value of the regression models of the separate metabolites and the p-value of the ratio. Positive beta indicates that the ratio is higher in relapsed tumor samples. All ratios with p-value < 0.05 and p-gain > 1000 are shown.

metabolite 1	***metabolite 2***	***cor***^{1}	***cor***^{2}	***p-value***
2-oleoylglycerophosphocholine*	X -17357	−0.69	0.91	2.4×10^{-3}
pantothenate	indoleacetate	−0.41	0.95	2.4×10^{-3}
1-palmitoylglycerophosphocholine	X -12776	−0.44	0.91	1.6×10^{-2}
erythritol	N6-carbamoylthreonyladenosine	0.49	0.98	1.9×10^{-2}
pelargonate (9:0)	pregn steroid monosulfate*	−0.60	0.79	1.9×10^{-2}
17-methylstearate	X -19380	−0.59	0.76	2.1×10^{-2}
pregn steroid monosulfate*	X -17269	−0.61	0.76	2.1×10^{-2}
2-ethylhexanoate	pregn steroid monosulfate*	−0.43	0.83	2.1×10^{-2}
caprylate (8:0)	pregn steroid monosulfate*	−0.54	0.77	2.2×10^{-2}
benzoate	pregn steroid monosulfate*	−0.48	0.86	1.6×10^{-2}
phenylalanine	X -17138	−0.56	0.73	2.4×10^{-2}
gamma-glutamylleucine	1-stearoylglycerophosphoinositol	−0.41	0.85	2.4×10^{-2}
docosapentaenoate (n6 DPA; 22:5n6)	X -19380	−0.53	0.74	2.4×10^{-2}
isovalerylcarnitine	pro-hydroxy-pro	−0.49	0.83	2.2×10^{-2}
5,6-dihydrouracil	glucose	−0.60	0.74	2.1×10^{-2}
cysteine	X -13435	−0.41	0.81	2.4×10^{-2}
choline	X -11550	−0.66	0.71	2.1×10^{-2}
dihydroorotate	5-oxoproline	0.57	−0.78	2.1×10^{-2}
xylitol	3-indoxyl sulfate	0.62	−0.72	2.4×10^{-2}
deoxycholate	2-linoleoylglycerophosphocholine*	0.68	−0.67	2.4×10^{-2}
benzoate	pregnenolone sulfate	−0.57	0.75	2.2×10^{-2}

Table 4.12: Metabolite pairs with significantly different correlation between primary and relapse samples. The correlations of each pair of metabolites were calculated for primary[1] and relapsed[2] samples. The pairs of metabolites which were significantly correlated within the groups were tested for difference of these correlation between the groups. The significant results (FDR-corrected p-value < 0.05) are shown in this table.

(a) pantothenate vs. indoleacetate

(b) pelargonate vs. pregn steroid monosulfate

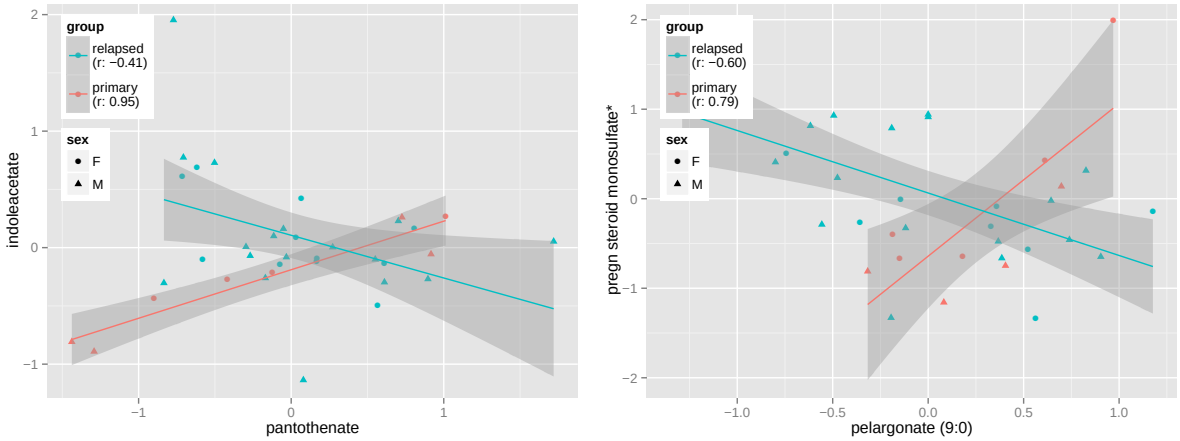


Figure 4.15: Examples for metabolite pairs with significantly different correlation between primary and relapse samples. These are scatter plots of pairs of metabolites. The tumor type, primary and relapsed, is indicated by the color of the points. Regression lines are calculated for the groups separately, the dark gray areas indicate the confidence of the regression line. All metabolite intensities were median-scaled, logarithmized and age- and gender-adjusted prior to plotting.

(a) Primary components

(b) Variance Explained

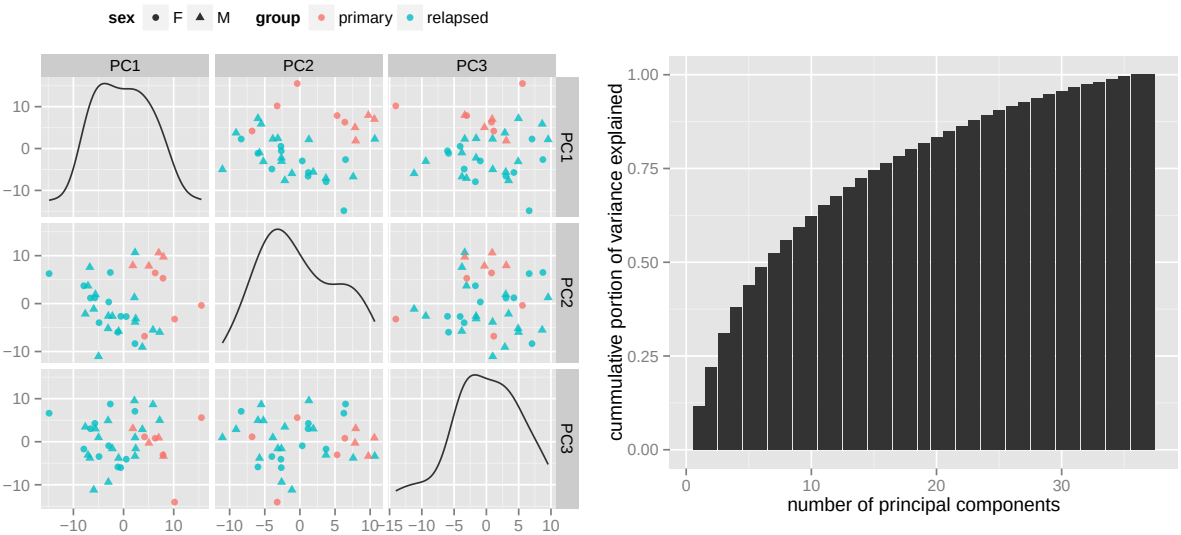
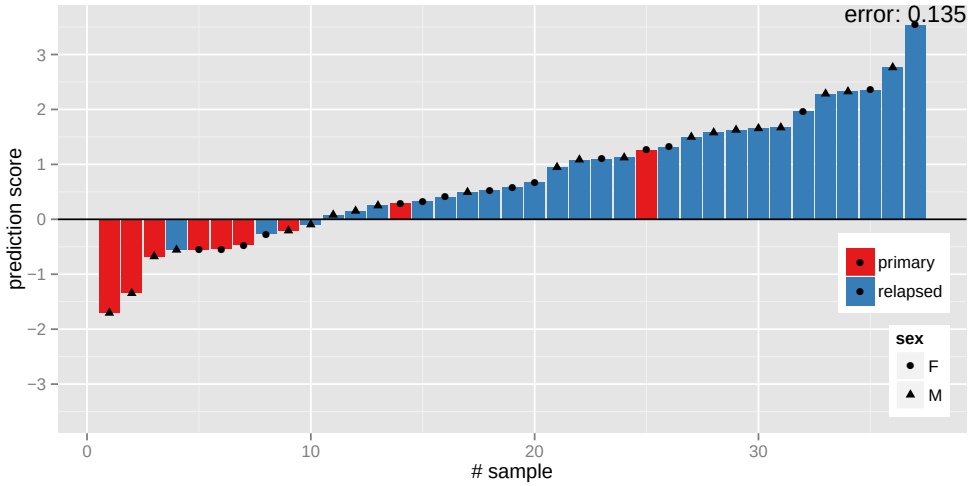


Figure 4.16: PCA of primary and relapsed plasma samples. The first three principal components of all samples of both groups were scattered against each other (a). The density of each component is shown on the diagonal. Before the PCA was performed, metabolite intensities were corrected for age and gender. Groups annotations are indicated by colour and the gender by the point shape. The histogram (b) shows the cumulative variance explained by the first x components.

(a) Performance



(b) Coefficients

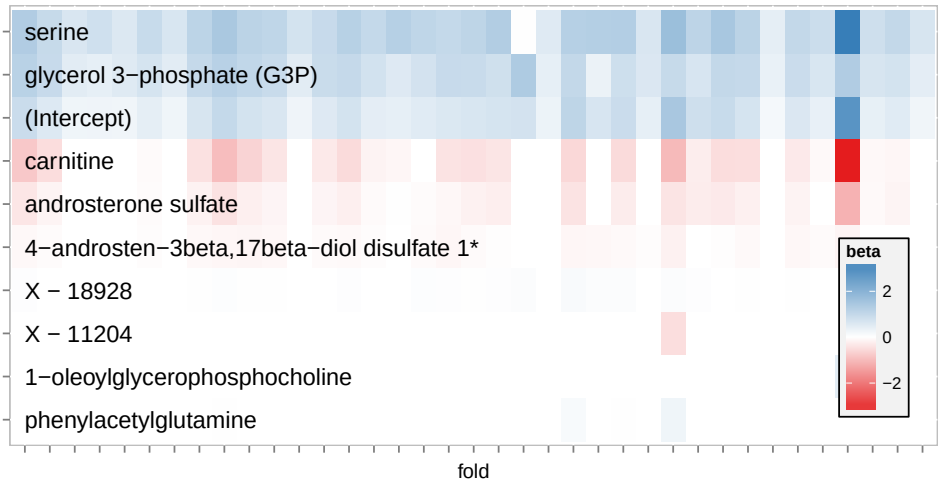


Figure 4.17: LASSO classification of primary and relapse samples. The performance of class prediction by a LASSO model is shown in (a). Each bar represents one sample, samples with positive prediction score are predicted to be relapsed and those with negative score are predicted as primary. The real classes are indicated by the colors. The predictions were made within a LOO CV in which the predicted samples was neither used for the selection of lambda nor for the training of the betas. Three relapsed samples were falsely predicted to be primary (blue bars with negative score) and two primaries were predicted as relapsed (red bars with positive score). Each row in (b) shows the regression parameters beta of the model which was used to predict the according bar (sample) in (a). The 10 metabolites with the highest average beta are shown.

To investigate if the concentrations of metabolites in plasma differ between patients with primary and relapsed tumors, linear regression models were fitted for each metabolite. According to those regression models 29 of the 342 tested metabolites had significantly different mean concentration, 65 % of them were more abundant in relapsed samples. The differentially concentrated metabolites do not cluster in any pathway. Only one metabolite, 2-hydroxyisobutyrate, was observed to have different variances between the groups. It had a variance of 1.45 in primary and 0.08 in relapsed samples (p-value 5.20×10^{-4}). The eicosanoid 12-HETE was measured in 23 of 28 relapsed probes but in none of the primary group. This is significant with a p-value of 1.9×10^{-3} (Fisher's test).

Metabolite ratios were calculated and tested for different means and variances. On the one hand this was shown to decrease variance significantly as shown by Altmaier et al. (2008). On the other hand metabolite ratios reflect equilibrium constants of chemical reactions which might be altered even though the separate metabolites did not change significantly. Linear regression unveiled Nine metabolite ratios which were globally significant and had p-gain > 1000 (Table 4.11). Fumarate divided by hexanoylcarnitine is the most significant and divided by acetylcarnitine the fourth significant. Apart from those, long chain fatty acids of different lengths bound to glycerophosphocholine are often changed. The octadecanoids 13-HODE + 9-HODE did not reach global significance separately but did so when normalized by the unknown metabolite X-17629 (Figure 4.14b). Variances of metabolite ratios differed in 87 cases with Levene's p-value < 0.05 and p-gain > 1000. In all those cases the variance is higher in the primary samples. Barlett's test revealed 1854 pairs at the same significance cutoffs.

Even though, there are 7067 linear combinations of metabolites significant, only seven of them have a p-gain > 500 and yet all of them only contain metabolites which are already significant when considered separately. Five of those seven pairs contain quinate, which already was identified to be increased in relapse samples by univariate analysis.

Another way of representing relationships of metabolites is their correlation. The correlations of 70 pairs of metabolites differ between the plasma samples of primary and relapsed tumor patients (Table 4.12). In more than the half of these cases either a lysolipid or steroid is part of the pair and in 40 % a metabolite from the amino acid metabolism is involved.

A PCA combines redundant variables as orthogonal linear combination, so called principal components, and orders those by the amount of captured variance. As it can be seen in Figure 4.16, the first two principal components essentially separate the relapse from the primary samples even though this separation is not perfect. These

components combined capture 20 % of the variance of the data. The supervised LASSO method which selects metabolites that enhance the separation of the groups can discriminate between the groups with less than 10 % error rate. The metabolites which were selected to gather this separation included serine, glycerol-3-phosphate, carnitine and 4-androsten-3beta,17beta-diol disulfate.

4.2.2 Discussion

There are substantial metabolic differences between the primary and the relapsed tumors prior to the treatment. The different types of tumors can be separated with reasonable accuracy using the two first principal components. Those account for nearly 25 % of the overall variance in the data. This makes clear that relapsed tumors are in fact different tumors compared to the primary tumors.

Many of the trends that have been observed in the comparison between pre- and post-pano samples can be seen between primary and relapsed samples in the same manner. For example long chain fatty glycerophospholipids of different lengths and saturation are higher concentrated in the plasma of patients with relapsed tumors. This tendency continues throughout the treatment with Panobinostat and those fatty acids were observed to be higher concentrated after the treatment (cf. type 1 in Figure 4.18). It was shown that isobutrylcarnitine is less abundant after the treatment and here it can be seen that acetylcarnitine, hexanoylcarnitine and carnitine itself are lower concentrated in relapsed than primary tumor samples. Other medium chain fatty acids show significantly altered correlations with other metabolites.

This increase of long chain fatty acids and simultaneous decrease of medium chain fatty acid seems to be a process from the primary tumor to the relapse tumor which is continued during the Panobinostat treatment. Therefore, it seems to be likely that this shift is not caused by Panobinostat but rather persists despite of the treatment.

The same is the case for amino acids. Many of them, including serine, histidine, glutamate, and also γ-glutamylglutamine are steadily increased from primary to relapsed tumor until after the treatment. This supports the theory that the GSH metabolism is related with the resistance of cancers to chemotherapy. An emerging question is if the increased activity of the GSH cycle enhances tumor growth or if this change is a consequence of DLBCL development. In any case it would be worth considering GSH as target of chemotherapy.

The eicosanoids are a more complicated case. 2-arachidonoylglycerophosphoethanolamine was observed to be higher concentrated in relapsed samples just as other eicosanoid-related fatty acids between pre- and post-pano samples. 12-HETE was

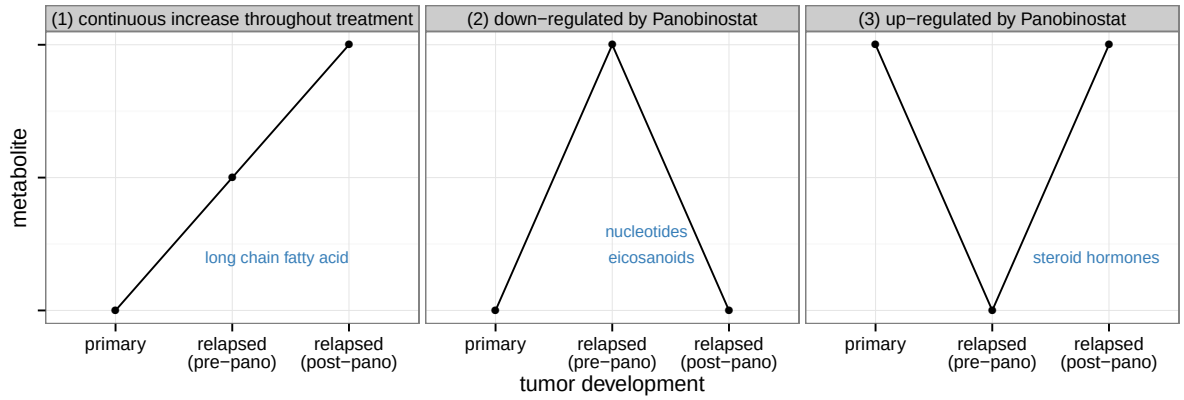


Figure 4.18: Schematic representation of the changes of metabolite concentrations from primary to relapsed tumor until after the treatment. Three different types of the progression of metabolite concentrations were observed in the tumor development from primary to relapsed tumor (before Panobinostat treatment) and continued throughout the treatment. The first type (leftmost) was that metabolite levels increased from primary to relapse tumor and the increase was continued in spite of the Panobinostat treatment. The second type (middle) was that metabolite levels were increased in relapsed DLBCL compared to primary DLBCL but decreased through the Panobinostat treatment. In the last case (rightmost) metabolite levels were lower in relapsed than primary tumors but elevated by the treatment.

measured significantly more often in relapsed than primary tumors. This indicates that 12-HETE is lower concentrated in primary samples as the technique can detect. This increase of the concentration of 12-HETE is reversed by the Panobinostat treatment, it is lower concentrated after the treatment than before. The same is the case for the related signaling lipids 13-HODE + 9-HODE, which are significantly increased (as ratio) in relapsed tumors but rather decreased – even if not significantly – after the treatment, as well (cf. type 2 in Figure 4.18). It was shown that both, 12-HETE and 13-HODE + 9-HODE, are markers for oxidative stress (Nazarewicz et al., 2007; Yoshida and Niki, 2004). Furthermore, it was suggested that prostagladins which are a subgroup of eicosanoids and also synthesized from arachidonic acid enhance growth of colorectal cancer and breast cancer (Tronstad et al., 2003; Wang and DuBois, 2010). This would explain increased eicosanoid levels in relapsed tumors.

In case of the steroid hormones, the tendency also seems to be changed by Panobinostat. After correction for gender and age, androsterone sulfate is detected to be lower concentrated in relapse tumors and its metabolite 4-androsten-3beta,17beta-diol disulfate higher after the treatment. 4-androsten-3beta,17beta-diol disulfate is also

one of the metabolites with the highest weight in LASSO regression of primary and relapsed samples. Pregnen-diol disulfate is another steroid hormone which is higher concentrated after the treatment than before. Its concentration is not significantly altered between primary and resistant samples but its correlation with several other compounds changed significantly. The same is the case for pregnenolone sulfate. This suggests that regulation of these steroid hormones differs between the tumor types. In this case Panobinostat might counteract the development of resistant tumor by activating the steroid hormone production (cf. type 3 in Figure 4.18). Apart from this, steroid levels highly depend on the gender. The association of steroids with tumor growth might be a reason for the different incidence of lymphoma for females and males. This might be of interest for subsequent analyses.

Another example for this effect are nucleotides. Uridine is increased in relapse tumors compared to the primary tumor samples and its concentration (or the concentration of its metabolites) is lowered again after Panobinostat treatment. However, uridine is the only noticeable metabolite from the nucleotide metabolism for the *primary vs. relapsed* comparison. In case of nucleotides it is more likely that their in-/decreased levels are rather a marker for tumor growth than the cause for tumors to grow or not.

Sugars were observed with decreased intensity after the Panobinostat treatment. This was not the case in any direction in this comparison. The only related metabolite is fumarate whose concentration – normalized by several other metabolites – is increased in relapsed DLBCL.

4.3 Resistant vs. Sensitive Lymphoma

The target of the *resistant vs. sensitive* comparison was to find differences in the metabolism of lymphomas, which allow for separation of R-CHOP-resistant from -sensitive lymphomas, prior to the treatment. The comparison was done using serum samples from 33 DLBCL patients, who were cured, and 11 patients with relapse or progressive disease. Those were measured in the **primary-serum-1** and **primary-serum-2** experiments. The results of this comparison are presented in the following section, their discussion follows in Section 4.3.2.

4.3.1 Results

Basically, there are no significantly differently concentrated metabolites or metabolite ratios or linear combinations of them which differ between sensitive and resistant tumors before the treatment. Nevertheless, there are some suggestive metabolites with FDR-corrected p-values ≤ 0.2. Those are shown in Figure 4.19. One of them, pyridoxate, is rather interesting due to the different variances within the groups: those are 0.27 (sensitive) and 2.34 (resistant). This difference has a Levene p-value of 1.0×10^{-4} and is the only significant deviation of variances.

in contrast, there is a long list of significant differences in variances of metabolite ratios (only the most significant ones are shown in Table 4.14). In total, 220 of them are rated with p-values < 0.05 by Levene's test and 150 of them have p-gains > 1000 (10,427 and 4325 respectively for Barlett's test). However, as it can be seen exemplary in Figure 4.20 the difference of variances are usually caused by some

metabolite	*#*	*lm p-value*	*beta*	*se*
xylose	39	1.8×10^{-1}	-1.43	0.37
caproate (6:0)	44	2.0×10^{-1}	0.48	0.14
pyridoxate	44	2.0×10^{-1}	-1.01	0.30
isobutyrylcarnitine	44	2.0×10^{-1}	-0.87	0.27
caprylate (8:0)	44	2.0×10^{-1}	0.51	0.16

Table 4.13: Metabolites with different concentration between R-CHOP resistant and sensitive DLBCL. All metabolites which differ between R-CHOP-resistant and sensitive lymphomas with FDR≤ 0.2 are shown. The regression coefficients beta are positive if the metabolite has elevated levels in sensitive samples.

metabolite 1	*metabolite 2*	#	*p-value (b)*	*p-value (l)*	*p-gain*
X -04357	X -16653	10	4.6×10^{-12}	8.6×10^{-4}	7.2×10^{5}
3-hydroxyisobutyrate	N-acetylneuraminate	11	9.0×10^{-8}	8.6×10^{-4}	3.4×10^{5}
X -11429	X -11533	10	6.3×10^{-10}	8.6×10^{-4}	3.2×10^{6}
X -03056	X -16653	10	8.0×10^{-17}	1.3×10^{-3}	2.9×10^{5}
X -04499	X -11533	11	2.0×10^{-5}	1.5×10^{-3}	9.8×10^{5}
1-stearoylglycerophosphoethanolamine	X -16653	10	7.8×10^{-9}	1.5×10^{-3}	1.5×10^{5}
3-(4-hydroxyphenyl)lactate	X -17269	10	7.6×10^{-7}	1.5×10^{-3}	8.1×10^{5}
X -11204	X -02249	10	3.8×10^{-14}	1.5×10^{-3}	1.2×10^{6}
X -10810	X -11378	10	2.3×10^{-9}	1.5×10^{-3}	1.2×10^{5}
3-dehydrocarnitine*	X -11818	10	1.4×10^{-4}	1.5×10^{-3}	2.7×10^{5}
X -11541	X -13429	10	2.0×10^{-9}	1.5×10^{-3}	1.7×10^{5}
X -06307	X -13457	11	8.5×10^{-10}	1.5×10^{-3}	6.1×10^{5}
dimethylglycine	X -11533	11	3.6×10^{-8}	1.6×10^{-3}	2.8×10^{5}
taurocholate	X -11550	10	9.1×10^{-9}	2.4×10^{-3}	1.1×10^{5}
eicosapentaenoate (EPA; 20:5n3)	X -11550	11	1.3×10^{-6}	2.4×10^{-3}	1.0×10^{5}
X -11538	X -13619	10	1.1×10^{-12}	3.0×10^{-3}	1.4×10^{5}

Table 4.14: Metabolite ratios with different variances between R-CHOP-resistant and sensitive DLBCL. All metabolite ratios were tested for different variances between the groups using Levene's (l) and Barlett's (b) tests. Those ratios with p-gain $> 10{,}000$ (using p-values from Levene's test) are listed.

metabolite 1	*metabolite 2*	*cor*[1]	*cor*[2]	*p-value*
isoleucine	ADSGEGDFXAEGGGVR*	−0.39	0.96	3.2×10^{-5}
X -03094	X -08402	0.79	−0.61	4.7×10^{-3}
X -11529	2-palmitoylglycerophosphocholine*	0.43	−0.89	4.7×10^{-3}
paraxanthine	2-palmitoylglycerophosphoethanolamine*	0.57	−0.80	5.5×10^{-3}
isoleucine	DSGEGDFXAEGGGVR*	−0.38	0.87	7.7×10^{-3}
theophylline	2-palmitoylglycerophosphoethanolamine*	0.53	−0.81	6.6×10^{-3}
3-(4-hydroxyphenyl)lactate	isovalerate	0.60	−0.66	2.1×10^{-2}
fucose	gamma-glutamylmethionine	−0.49	0.76	2.1×10^{-2}
X -08889	X -11541	0.35	−0.83	2.1×10^{-2}
X -11805	isovalerate	0.40	−0.83	2.1×10^{-2}
X -03003	X -11793	−0.38	0.83	2.1×10^{-2}
X -09789	N-acetyl-beta-alanine	0.45	−0.74	2.5×10^{-2}
alpha-ketoglutarate	1-stearoylglycerophosphocholine	0.37	−0.79	2.3×10^{-2}
X -05907	X -17337	−0.45	0.73	2.3×10^{-2}
isovalerate	X -16132	0.51	−0.81	1.2×10^{-2}
gamma-glutamylglutamine	X -16651	0.36	0.95	2.3×10^{-2}
xanthine	tryptophan betaine	0.41	−0.77	2.3×10^{-2}
tryptophan betaine	X -17441	−0.37	0.78	2.3×10^{-2}
1-palmitoylglycerol (1-monopalmitin)	tryptophan betaine	0.40	−0.81	2.1×10^{-2}
glycodeoxycholate	X -11847	−0.60	0.75	2.3×10^{-2}
carnitine	X -10429	0.44	−0.72	2.3×10^{-2}
glycodeoxycholate	X -12100	0.51	−0.76	2.3×10^{-2}

Table 4.15: Metabolite pairs with significantly different correlation between R-CHOP-resistant and sensitive samples. The metabolite pairs whose correlations differ significantly with FDR < 0.025 are listed. Correlation coefficients within the groups of resistant[1] and sensitive[2] samples are shown besides the FDR-corrected p-value. Only pairs which were significantly correlated within both groups were tested for differential correlation.

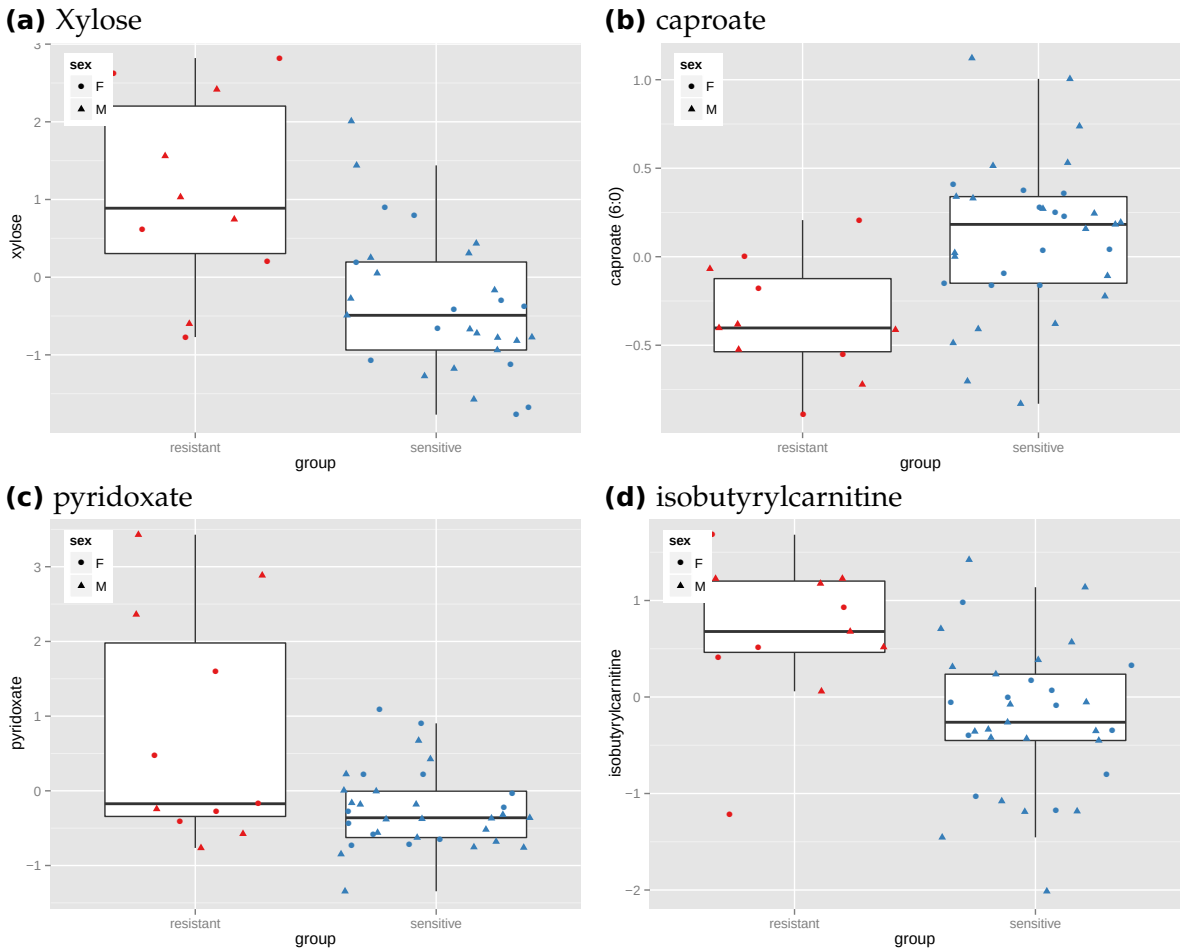
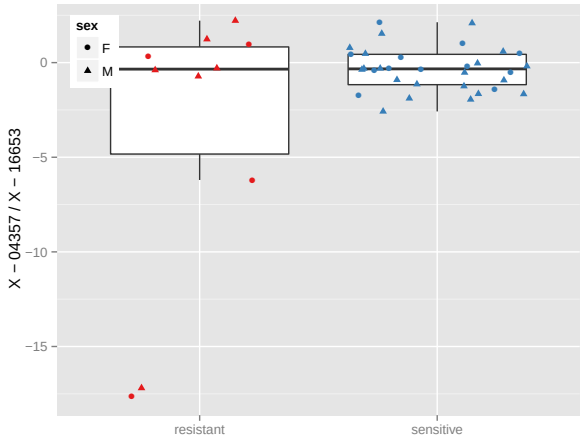


Figure 4.19: Examples for metabolites with different concentration between R-CHOP-resistant and sensitive DLBCL. The median-scaled, logarithmized and age- and gender-corrected metabolite intensities were plotted separately for R-CHOP-resistant and sensitive DLBCL.

extreme outliers. In all of the 220 significant cases the variance is higher in resistant lymphoma samples. This might be an artifact due to the much smaller number of resitant samples.

Testing for altered correlations of pairs of metabolites between the groups gave 137 globally significant results (Table 4.15). Isoleucine and two peptides from the fibrinogen alpha chain, ADSGEGDFXAEGGGVR and DSGEGDFXAEGGGVR, are amongst the highest rated hits (Figure 4.21). Amino acids are contained in about one third of the cases. Apart from, those, lipids are part of those correlations in half of all significant cases (exemplary shown in 4.21b). This is supported by the results

(a) X-04357 / X-16653



(b) 3-hydroxyisobutyrate / N-acetylneuraminate

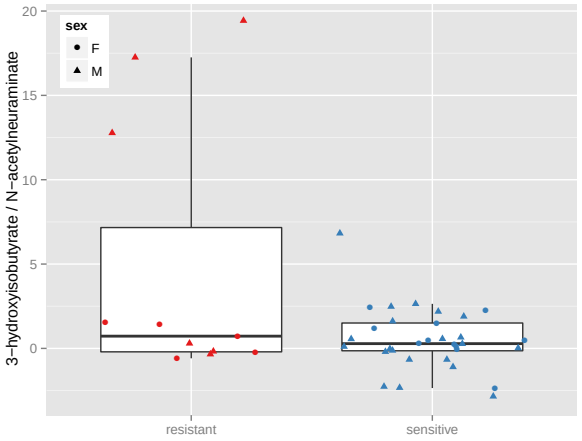
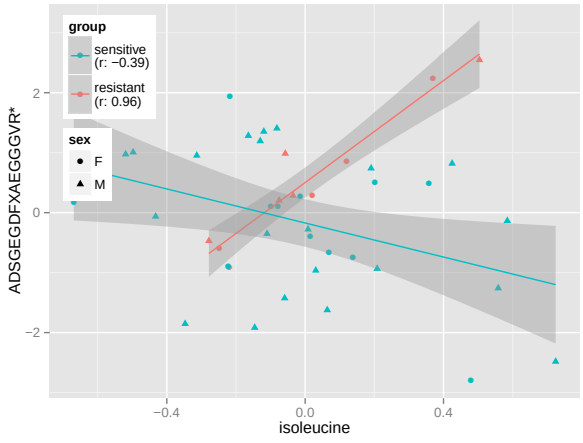


Figure 4.20: Examples for metabolite ratios with different variances between R-CHOP-resistant and sensitive DLBCL. Distributions of scaled, logarithmized and corrected metabolite intensity ratios of sensitive (left) and relapse/refractory (left) samples are shown.

(a) isoleucine vs ADSGEGDFXAEGGGVR



(b) paraxanthine vs 2-palmitoylglycerophosphoethanolamine

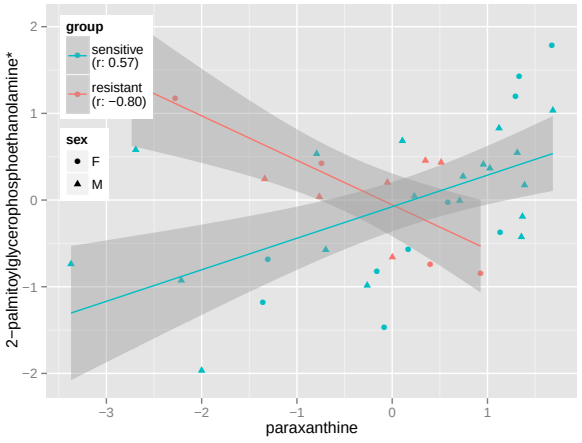


Figure 4.21: Examples for metabolite pairs with significantly different correlation between R-CHOP-resistant and sensitive samples. The scattered metabolites intensities were stratified by the resistance to R-CHOP treatment. For both groups separate regression lines and their confidence interval are drawn.

(a) Principal components

(b) Variance Explained

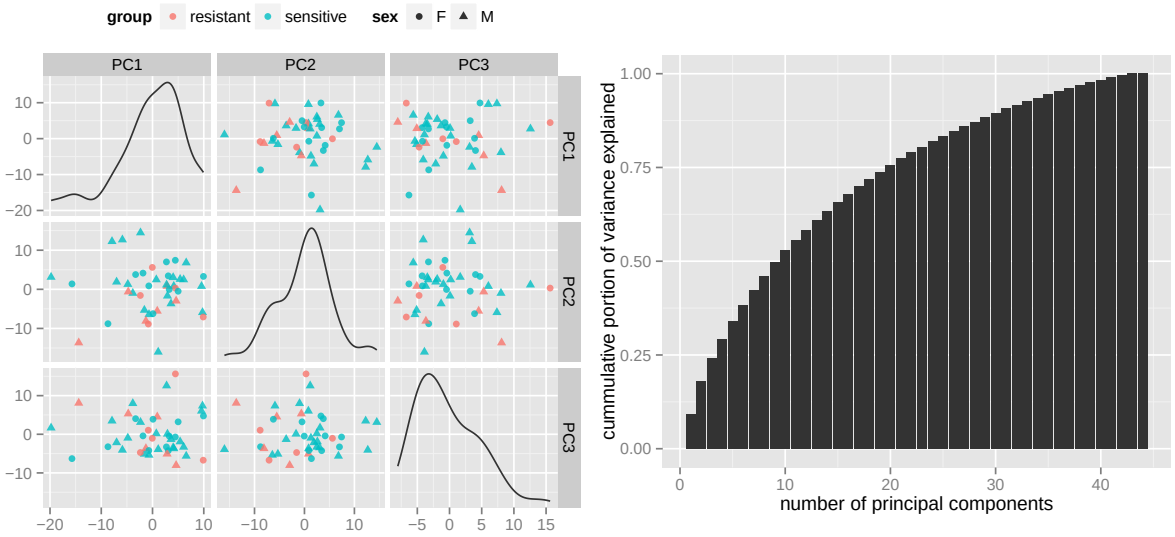


Figure 4.22: PCA of R-CHOP-resistant and sensitive DLBCL samples. (a) shows the first three principal components scattered against each other. The curves on the diagonal show the density of the according component. In (b) the cumulative proportion of explained variance is shown.

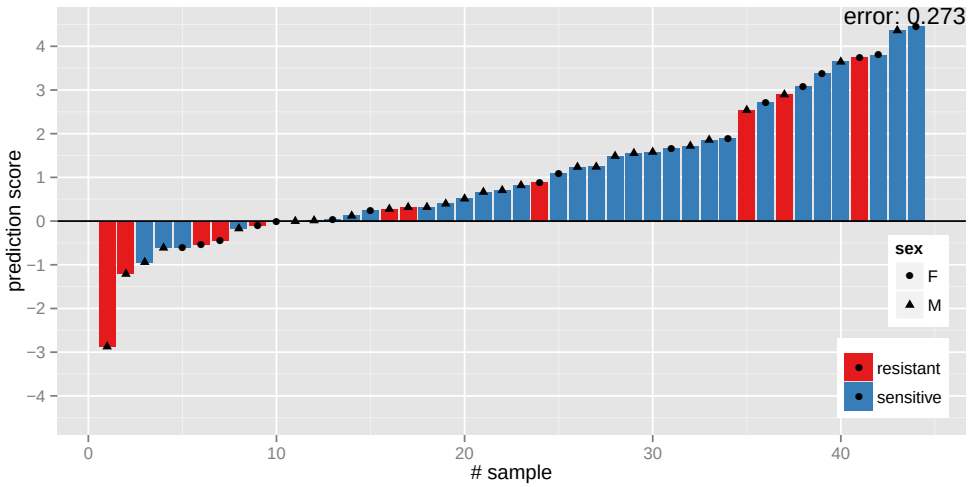


Figure 4.23: LASSO classification of R-CHOP-resistant and sensitive DLBCL. The performance of a predictive LASSO model to classify resistant and sensitive lymphomas was assessed using a LOO validation. The predicted outcome is plotted on the y-axis. Negative score means "predicted as resistant" and negative score "classified as sensitive". The real group annotation is indicated by the colors. The overall error-rate was 0.295.

of PWEA: Lipids, and more specific medium and long chain fatty acids are found to be the most enriched pathways having p-values of ≈ 0.02 by GAGE.

The results of PCA can be seen in Figure 4.22 and the performance of the LASSO regression in Figure 4.23. Neither the unsupervised approach nor the supervised managed to reliably separate the classes. The LASSO overall error rate was 0.295 while it is 0.364 for data with randomized group labels.

4.3.2 Discussion

The results show that R-CHOP sensitive and resistant lymphomas can not be easily distinguishes before the treatment. The results of PCA indicate that the greatest portion of variance within the data is not influenced by the resistance to R-CHOP and therefore unsupervised methods will probably fail to distinguish the groups (Figure 4.22). Even the supervised LASSO does not achieve group separation. The relatively good error rate of 0.295 comes just due to the imbalanced groups sizes (Figure 4.21).

However, applying a less stringent threshold for significance (FDR ≤ 0.2) reveals some differently concentrated metabolites. Differences in variances and correlations between metabolites can be detected as well. This suggests that there is a metabolic difference between resistant and non-resistant lymphomas. The most significant metabolite is xylose, a C_5 sugar. Three others are the medium chain fatty acids caproate (6:0) and caprylate (8:0) and isobutyrylcarnitine, which is a product in the degradation of fatty acids and branched-chain amino acids. pyridoxate which has particularly increased variance in resistant cases is part of the vitamin B6 catabolism.

Many of the cases of increased variances of metabolite ratios or altered correlations affect fatty acids. Paraxanthine and theophylline, which both are differentially correlated with 2-palmitoyl between the groups, are derivatives of xanthin, a degradation product of caffeine. These alterations might be caused by dietary habits which influence the result due to the small sample size. Nevertheless, lipid metabolism seems to be altered in resistant lymphomas. This is supported by PWEA, changes of variances and differential correlation results.

Additionally some sugars, xylose and fucose and the related metabolite alpha-ketoglutarate which is part of the citric acid cycle were observed to be disregulated.

Summarizing, it emerged that it is hard to distinguish between R-CHOP-resistant and -sensitive cases based on serum metabolic profiles. However, there is some evidence for distinct metabolic profiles. The main differences are observed in medium

chain and long chain fatty acid metabolism. Some changes are seen in amino acid metabolism and sugars. These are rather vague indications of disregulation than real differential concentrations. Nevertheless, these changes coincide with the results from the primary vs. relapsed comparison and might therefor be first signs of the subsequent relapse. However, lipids and amino acids are well known to be disregulated in cancers and the observed differences might be rather signs for aggressive lymphomas than specific R-CHOP resistance. To be sure about this, more specific studies with more patients in both groups must be carried out.

4.4 Plasma vs. Serum

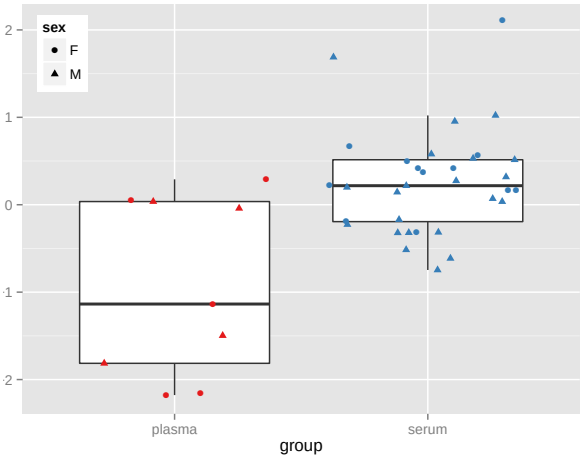
The last comparison was carried out between plasma and serum samples of patients with primary DLBCL before they received any treatment. This was not intended to be a "classic" plasma vs. serum comparison but more an analysis which sample type might be more suited to test a specific hypothesis in future experiments.

4.4.1 Results

In this section the results of the *plasma vs. serum* comparison are presented. Those were measured in distinct experiments. Data were median-scaled, logarithmized and corrected for influences due to gender and age as described before.

Testing single metabolites for altered central tendency between plasma and serum samples, yielded 21 significant results after correcting for multiple testing. 60 % of those were higher concentrated in serum, the other 40 % in plasma samples (Table 4.16). The variance of single metabolites differed in nine cases according to Levene's test. The variance was observed to be higher in plasma samples in all those cases.

(a) 5-oxoproline



(b) arachidonate

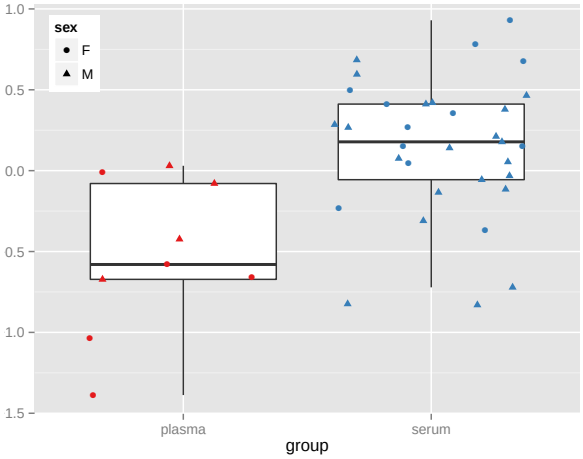


Figure 4.24: Examples for metabolites with different concentration between plasma and serum samples. The distribution of metabolite intensities os plotted for plasma and serum samples separately. Boxes indicate the quartiles of the distributions, points show the individual samples. Data were medium-scaled, logarithmized and corrected for age and gender before.

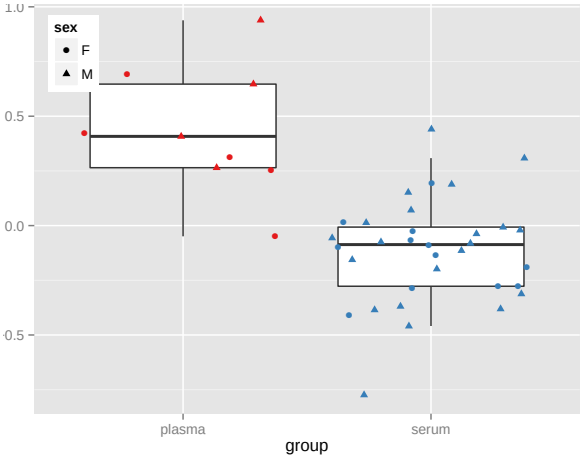
metabolite	*#*	*lm p-value*	*beta*	*se*
glycerophosphorylcholine (GPC)	41	6.6×10^{-4}	1.41	0.25
phosphate	42	5.7×10^{-3}	−0.73	0.16
5-oxoproline	42	5.7×10^{-3}	1.19	0.27
X -11204	41	5.7×10^{-3}	−1.05	0.24
serine	42	5.7×10^{-3}	0.75	0.17
X -13478	23	5.7×10^{-3}	−1.15	0.24
arachidonate (20:4n6)	42	8.4×10^{-3}	0.68	0.17
arginine	42	9.4×10^{-3}	−0.98	0.24
lactate	42	9.7×10^{-3}	0.81	0.21
gamma-glutamylglutamine	41	9.7×10^{-3}	1.08	0.27
glycerol 3-phosphate (G3P)	42	9.7×10^{-3}	0.90	0.23
glycerate	42	1.0×10^{-2}	0.84	0.22
dihomo-linolenate (20:3n3 or n6)	42	1.5×10^{-2}	0.64	0.17
glutamate	41	1.6×10^{-2}	0.85	0.23
fumarate	41	1.6×10^{-2}	0.65	0.18
X -13215	42	2.2×10^{-2}	0.80	0.23
benzoate	42	2.2×10^{-2}	−0.57	0.16
X -11261	41	2.2×10^{-2}	−1.12	0.32
X -11521	33	2.7×10^{-2}	−0.93	0.27
gamma-tocopherol	41	3.0×10^{-2}	−0.97	0.29
myristate (14:0)	42	4.4×10^{-2}	0.84	0.27

Table 4.16: Metabolites with different concentration between plasma and serum samples. The concentrations of the listed metabolites differ significantly (p-value < 0.05) between plasma and serum samples. The p-values of the according linear models (lm) were FDR-corrected. Positive beta indicates higher concentration of the metabolite in serum samples.

metabolite	*var*[1]	*var*[2]	*p-value (b)*	*p-value (l)*
X -12095	2.33	0.24	2.0×10^{-3}	4.3×10^{-3}
2-hydroxyisobutyrate	1.45	0.15	1.3×10^{-3}	4.3×10^{-3}
phosphate	0.57	0.09	1.9×10^{-2}	4.4×10^{-3}
N1-Methyl-2-pyridone-5-carboxamide	1.40	0.16	4.9×10^{-2}	9.1×10^{-3}
X -11497	1.14	0.17	1.1×10^{-2}	9.1×10^{-3}
pantothenate	0.96	0.21	1.2×10^{-1}	2.5×10^{-2}
X -13215	1.08	0.18	2.6×10^{-2}	2.5×10^{-2}
X -11317	0.78	0.19	1.6×10^{-1}	2.6×10^{-2}
ornithine	1.16	0.28	1.6×10^{-1}	4.4×10^{-2}

Table 4.17: Metabolites with different variances between plasma and serum samples. The table shows the results of the comparison of metabolite variances of plasma([2]) and serum([1]) samples. Corrected p-values of Barlett's test (b) and Levene's test (l) are shown (cf. Section 3.2.3).

(a) pelargonate / undecanoate



(b) pelargonate / docosahexaenoate

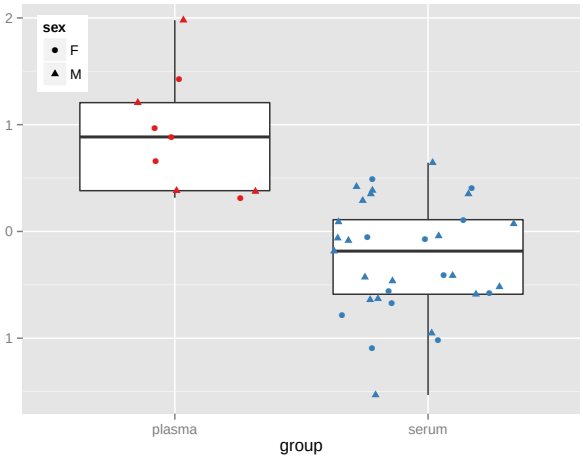


Figure 4.25: Examples for metabolite ratios differing between plasma and serum samples. The boxplots show distributions of metabolite ratios stratified by the type of sample.

metabolite 1	*metabolite 2*	*#*	*p-value*	*p-gain*	*beta*
arginine	serine	42	4.7×10^{-5}	5.1×10^{4}	−1.73
X - 13478	X - 14056	23	7.2×10^{-5}	2.4×10^{4}	−1.90
pelargonate (9:0)	undecanoate (11:0)	42	6.9×10^{-4}	2.3×10^{4}	−0.55
arginine	glutamate	41	1.6×10^{-4}	1.8×10^{4}	−1.85
arginine	glycerol 3- phosphate (G3P)	42	1.7×10^{-4}	9.8×10^{3}	−1.88
caprate (10:0)	pelargonate (9:0)	42	1.1×10^{-3}	8.9×10^{3}	0.79
pelargonate (9:0)	caproate (6:0)	42	1.2×10^{-3}	6.5×10^{3}	−0.50
benzoate	X - 13215	42	3.9×10^{-4}	5.2×10^{3}	−1.37
pelargonate (9:0)	docosahexaenoate (DHA; 22:6n3)	42	7.7×10^{-4}	4.1×10^{3}	−1.16
pelargonate (9:0)	X - 11497	42	1.4×10^{-3}	3.1×10^{3}	−1.07
glycerol 3- phosphate (G3P)	X - 13478	23	2.1×10^{-4}	2.8×10^{3}	2.39
caprate (10:0)	caprylate (8:0)	42	4.3×10^{-3}	2.6×10^{3}	0.46
lactate	arginine	42	3.6×10^{-4}	2.1×10^{3}	1.79
arachidonate (20:4n6)	arginine	42	3.6×10^{-4}	1.9×10^{3}	1.67
N6- carbamoylthreonyladenosine	X - 17138	22	2.1×10^{-3}	1.9×10^{3}	1.34
arginine	glycerophosphorylcho (GPC)	41	4.7×10^{-5}	1.8×10^{3}	−2.40
pelargonate (9:0)	X - 13215	42	6.6×10^{-4}	1.8×10^{3}	−1.28
benzoate	dihomo- linolenate (20:3n3 or n6)	42	4.7×10^{-4}	1.8×10^{3}	−1.21
laurate (12:0)	pelargonate (9:0)	42	2.4×10^{-3}	1.5×10^{3}	0.75
2- ethylhexanoate	X - 13215	42	6.6×10^{-4}	1.5×10^{3}	−1.21
serine	X - 13478	23	3.0×10^{-4}	1.4×10^{3}	2.20

Table 4.18: Metabolite ratios differing between plasma and serum samples. The table shows the results of the linear regression of metabolite ratios (as described in Section 3.3.1). The number of samples for which both metabolites were measured (#), the FDR-corrected p-value, the p-gain and the regression coefficient beta are reported. Positive beta stands for a higher ratio in the serum samples.

metabolite 1	*metabolite 2*	#	*p-value (b)*	*p-value (l)*	*p-gain*
threonine	glycerol	8	2.0×10^{-10}	2.3×10^{-4}	6.4×10^{7}
5-oxoproline	4-androsten-3beta,17beta-diol disulfate 1*	9	3.5×10^{-10}	2.9×10^{-4}	1.5×10^{5}
X -11787	N-acetyl-beta-alanine	9	1.3×10^{-6}	3.8×10^{-4}	7.7×10^{6}
N-acetylalanine	glycerol	8	2.8×10^{-8}	1.1×10^{-3}	9.9×10^{5}
3-hydroxyisobutyrate	octanoylcarnitine	8	1.9×10^{-8}	1.4×10^{-3}	7.2×10^{4}
arginine	choline	8	1.5×10^{-3}	2.1×10^{-3}	9.0×10^{5}
gamma-glutamyltyrosine	N-acetyl-beta-alanine	8	1.3×10^{-3}	2.1×10^{-3}	3.6×10^{5}
isobutyrylcarnitine	N-acetyl-beta-alanine	8	3.8×10^{-12}	2.1×10^{-3}	3.1×10^{5}
proline	indoleacetate	8	2.5×10^{-7}	2.1×10^{-3}	2.1×10^{5}
oleate (18:1n9)	3-methyl-2-oxovalerate	9	1.9×10^{-7}	2.1×10^{-3}	3.1×10^{5}
myo-inositol	21-hydroxypregnenolone disulfate	8	4.6×10^{-4}	2.4×10^{-3}	4.1×10^{4}
pyruvate	2-ethylhexanoate	9	1.2×10^{-8}	3.2×10^{-3}	5.2×10^{4}
1-linoleoyl-glycerophosphoethanolamine*	octanoylcarnitine	9	4.5×10^{-8}	3.2×10^{-3}	4.5×10^{5}
xylitol	glycerol	8	3.2×10^{-6}	3.4×10^{-3}	2.4×10^{5}
tyrosine	3-methyl-2-oxovalerate	9	5.3×10^{-4}	5.5×10^{-3}	1.1×10^{5}
gamma-glutamylglutamine	succinylcarnitine	9	3.6×10^{-3}	6.5×10^{-3}	2.4×10^{4}
X -11334	N-acetyl-beta-alanine	9	5.2×10^{-4}	8.3×10^{-3}	3.4×10^{4}

Table 4.19: Metabolite ratios with different variances between plasma and serum samples. The variances of metabolite ratios within plasma and serum samples were tested for significant differences by Levene's and Barlett's test. The p-gain was calculated using p-values from Levene's test for the ratio and the separate metabolites. Pairs with FDR-corrected p-value ≤ 0.05 and p-gain $> 20{,}000$ are reported. The full list can be found in the supplementary data (Section A).

metabolite 1	***metabolite 2***	***cor***^{1}	***cor***^{2}	***p-value***
beta-hydroxypyruvate	1-oleoylglycerophosphocholine	−0.52	0.95	1.2×10^{-3}
N-acetylalanine	erythronate*	0.47	0.99	1.8×10^{-3}
threonate	pregnenolone sulfate	−0.42	0.87	6.7×10^{-3}
threonate	pregnen-diol disulfate*	−0.54	0.82	7.1×10^{-3}
1-palmitoylglycerophosphocholine	ornithine	0.56	−0.88	4.5×10^{-3}
beta-hydroxypyruvate	2-oleoylglycerophosphocholine*	−0.50	0.86	1.3×10^{-2}
X -11299	stachydrine	0.50	−0.76	1.9×10^{-2}
1-palmitoylglycerophosphocholine	X -17337	−0.65	0.80	8.5×10^{-3}
1-stearoylglycerophosphocholine	ornithine	0.64	−0.77	1.2×10^{-2}
1-arachidonoylglycerophosphocholine*	X -17337	−0.43	0.87	1.2×10^{-2}
2-stearoylglycerophosphocholine*	ornithine	0.54	−0.79	1.9×10^{-2}
beta-hydroxypyruvate	1-eicosatrienoylglycerophosphocholine*	−0.58	0.78	1.9×10^{-2}
2-hydroxystearate	arabitol	−0.39	0.88	1.2×10^{-2}
1-palmitoylglycerophosphoethanolamine	cis-4-decenoyl carnitine	−0.36	0.87	9.7×10^{-3}
1-oleoylglycerophosphocholine	X -17337	−0.61	0.70	1.6×10^{-2}
3-indoxyl sulfate	erythronate*	0.58	−0.71	1.9×10^{-2}
phenylacetylglutamine	ornithine	−0.35	0.85	1.2×10^{-2}
1-linoleoylglycerol (1-monolinolein)	4-androsten-3beta,17beta-diol disulfate 2*	−0.41	0.89	1.0×10^{-2}
urea	glycoursodeoxycholate	−0.35	0.83	1.9×10^{-2}
X -12462	X -12644	−0.39	0.86	1.7×10^{-2}
citrate	carnitine	−0.41	0.86	8.6×10^{-3}

Table 4.20: Differentialy Correlated Metabolites in plasma and serum samples. All pairs of metabolites which are significantly differentially correlated (FDR-corrected p-value ≤ 0.02) between plasma (^{1}) and serum (^{2}) samples are listed. The separate correlations eithin both groups are reported together with the p-value.

(a) beta-hydroxypyruvate vs. 1-oleoylglycerophosphocholine

(b) threonate vs. pregnenolone sulfate

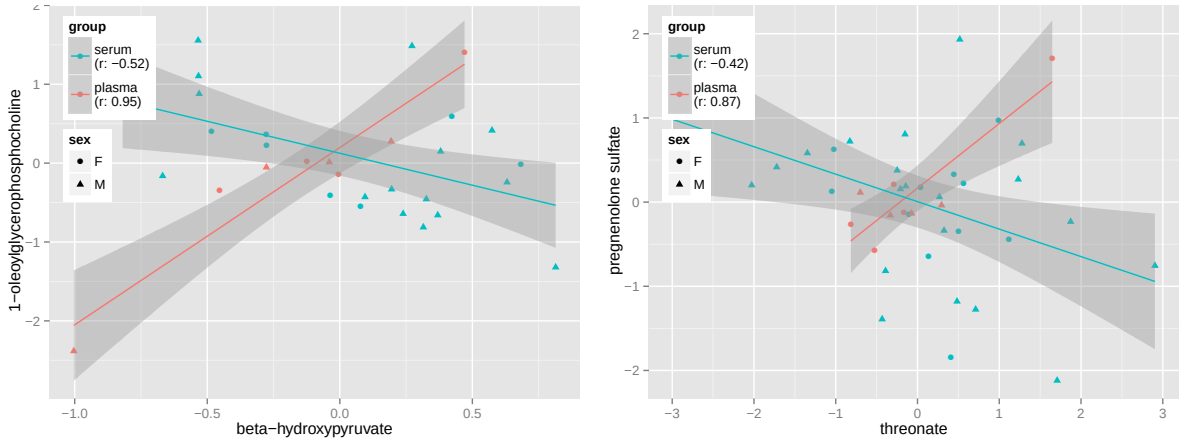


Figure 4.26: Examples for differentialy correlated metabolites in plasma and serum samples. Two pairs of metabolites are shown in the figures. The correlations of both differ significantly between the plasma and serum samples. Regression lines and their confidence (gray) are drawn separately for the two types os samples.

(a) PC scatter (with missing values)

(b) PC scatter (without missing values)

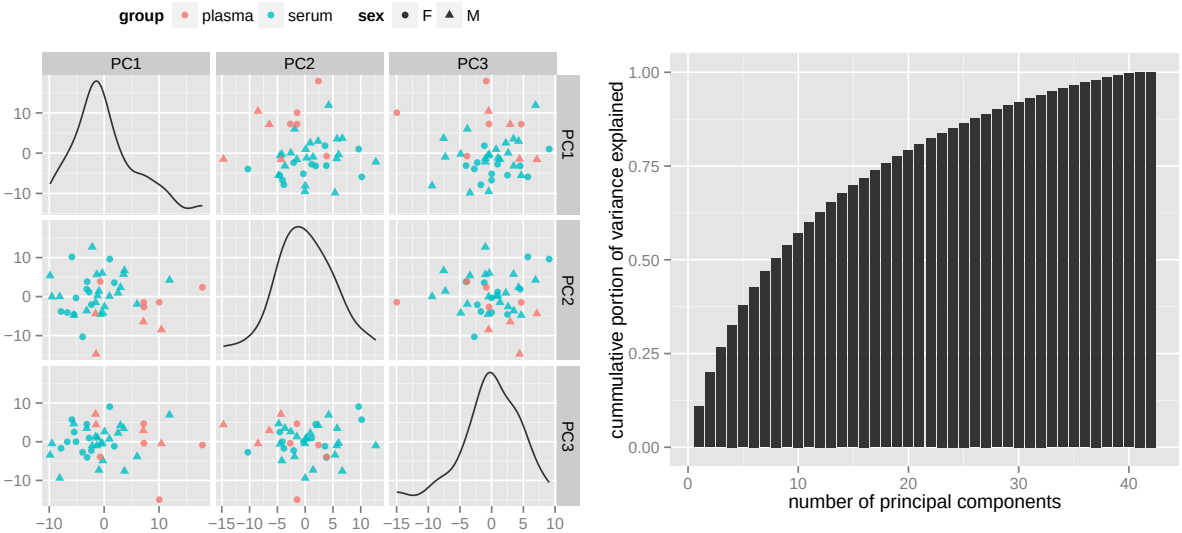
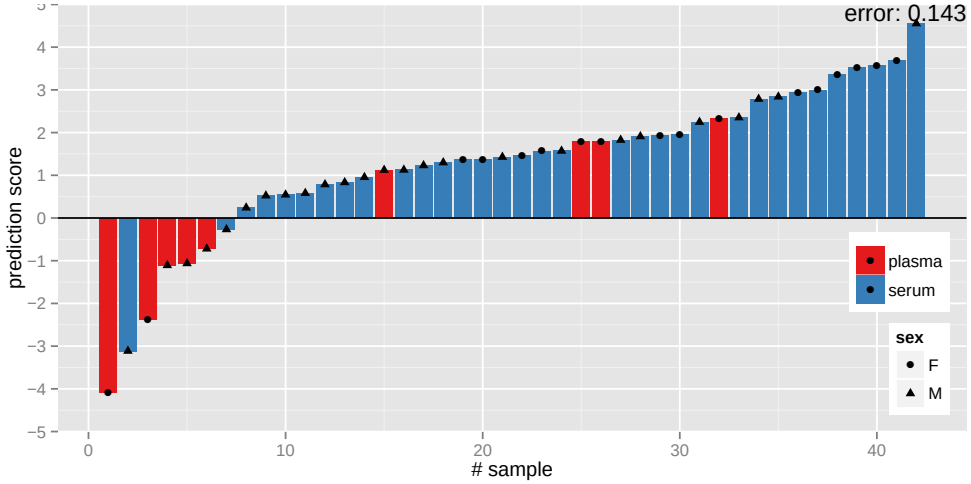


Figure 4.27: PCA of plasma and serum samples. In (a) the first three principal components of the plasma and serum samples combined are scattered against each other. Additionally, the density functions of the three components are drawn. (b) is the histogram of the cumulative fraction of explained variance of the primary components.

(a) Performance



(b) Coefficients

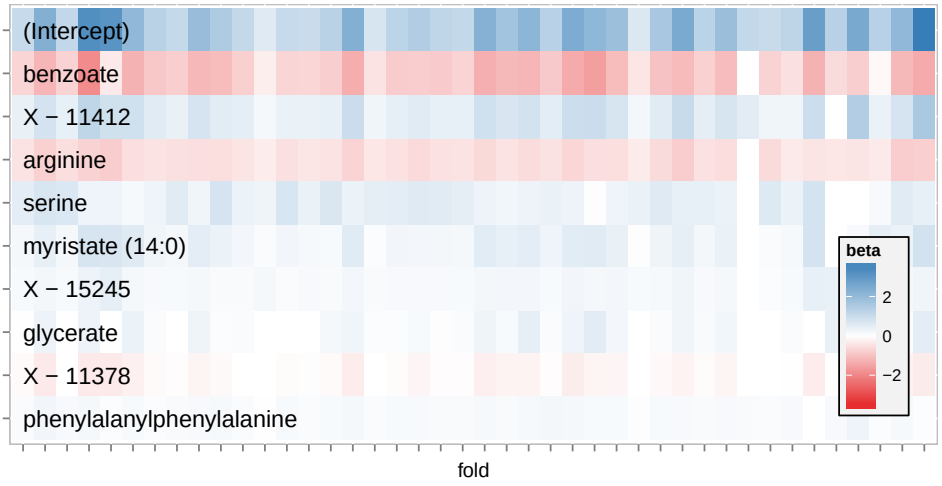


Figure 4.28: LASSO classification of plasma and serum samples. The upper plot shows the prediction performance of LASSO. Each bar indicates one serum (blue) or plasma (red) sample. Samples with negative score are predicted as plasma, positive score translates to serum. The performance was assessed in a LOO CV. The beta coefficients which were used for the samples are shown in (b). Each column represents the beta coefficients of a models which was trained for one fold of the CV. The ten variables with the highest average beta are shown.

Ratios of metabolite intensities are thought to reflect either chemical equilibrium states in case of closely related metabolites or are a type of "intra-sample" normalization. Therefore, ratios of all pairs of metabolites were tested for altered concentration or variance between the groups. 21 of these pairs had significantly different concentration with a corrected p-value < 0.05 and p-gain > 1000. In nearly the half of these cases pelagonate is involved, mostly together with another medium chain fatty acid (Figure 4.25). Ratios containing arginine show altered intensity in one third of the significant cases (Table 4.18). Levene's test yielded 259 significantly different metabolite ratios where 254 show higher variances of plasma samples (Table 4.19).

Investigating differences of linear combinations of metabolites between the groups is a second possibility of bivariate testing. 11 of them were detected to be significantly different but all of them only consist of metabolites which have significantly altered concentration alone. This is different for differential correlations. 24 pairs of metabolites are positively correlated in plasma samples and negatively in serum, 25 metabolites the other way round (Table 4.20, Figure 4.26). In 37 of these pairs none of the two metabolites was observed to behave differently between the groups separately.

By using PCA, some of the plasma samples somehow separate from the serum samples but a strict distinction of the samples based on the first three primary components is not possible (Figure 4.27a). Those account for about 26 % of the variance in the data (Figure 4.27b). The LASSO regression does not achieve proper separation, as well. The LOO CV shows that the groups can be distinguished with an error rate of 14.3 %. However, this is at least better than the random error rate of 33 %. Thereby, less than the half of the plasma samples are correctly classified. The most important metabolite for that is benzoate, followed by an unknown compound and the amino acids arginine and serine (Figure 4.28).

4.4.2 Discussion

The first thing to mention is that the expected changes in ratios and correlations were not visible. The difference between plasma and serum is that while blood clotting is prevented in plasma, it is frozen immediately after blood taking. In contrast serum probes incubate for some time during the clotting process. In this time some enzymes are still active and some reactions are carried out. These reactions are thought to lead to a decoupling of the involved metabolites. But as mentioned before, this effect could not be observed.

Neither PCA nor LASSO properly separate between the groups. Nevertheless there are significant differences. Lactate had a higher concentration in serum samples. Mey et al. (2012) showed that the level of lactate dehydrogenase in serum is associated with the prognosis. Thus, an altered level of lactate is not surprising. Apart from this, many amino acids have different concentrations between the groups, most of them are higher concentrated in serum. This was previously reported to be a common difference between serum and plasma of healthy individuals (Barri and Dragsted, 2013; Liu et al., 2010).

Nevertheless, it attracts attention that amino acids also were observed to be differentially concentrated between primary and relapse tumors. 5-oxoproline which is a major metabolite of the GSH cycle is elevated in serum, as well. This is also the case for long chain fatty acids (dihomo-linolenate and arachidonate). Medium chain fatty acids did not show altered concentrations but their ratios were significant in many cases. In all those cases pelargonate (9:0) divided by any other fatty acid – no matter if longer or shorter – is higher concentrated in plasma samples. There is no direction recognizable, that would suggest that longer fatty acids are higher concentrated in one sample and shorter in the other.

Many correlations of long chain glycerophospholipids with e.g. ornithine, a metabolite of arginine, or Hydroxypyruvic acid, which is a metabolite of serine changed significantly between serum and plasma. Furthermore, there are steroid hormones involved in those correlations. Pregnenolone sulfate, pregnen-diol disulfate and 4-androsten-3beta,17beta-diol disul-fate have altered correlations with threonate, a metabolite of vitamin C or fatty acids. The glycerophospholipids as well as the steroid hormones also showed altered correlations between primary and relapsed tumors.

A possible explanation of those parallelisms between the *primary vs. relapsed* comparison and the *serum vs. plasma* comparison is that certain metabolic processes are carried out during the progression of lymphoma. Those processes influence fatty acid, amino acid and steroid hormone metabolism, which is why those groups of metabolites are observed to differ between primary and relapse tumor. If those processes are active when the blood sample is taken they might be continued in serum samples during the clotting process while being interrupted in plasma samples almost immediately. However, even if this is not the case, these results do not affect the prior results as all other comparisons were performed with only plasma or only serum samples, but not a mixture of both.

The real reason for the observed differences and the conformance of the results of the different comparisons can not be figured out based on the available data. When comparing both, plasma and serum samples, having both types of samples from the

same patients would help a lot. The pairing would eliminate many biases through age, gender or different stages or subtypes of DLBCL from the very beginning. A comprehensive comparison of serum and plasma of healthy individuals would allow to exclude common differences between the two types of blood samples and concentrate on uncommon differences which are specific to lymphoma patients. This would facilitate the interpretation.

All significant variances were increased in plasma samples. Almost certainly this is an artifact caused by the lower number of samples in the plasma group. For this reason no conclusion can be made concerning the suitability of serum and plasma for lymphoma studies. Nevertheless, it can at least be concluded that mixing both will produce spurious results.

5 Conclusion and Perspectives

5.1 Summary

In this thesis it was shown that even though lymphoma is usually not considered as metabolic disease, it has a strong influence on the metabolism. Several metabolic changes were reported to enhance tumor growth, like alterations of the GSH cycle or eicosanoids. Furthermore it was shown that Panobinostat significantly influences the metabolism of tumor cells themselves and blood samples of patients, as well. Many of to observed changes reoccur in both comparisons and can be seen between plasma and serum samples of primary lymphoma patients, as well. For the alterations of the GSH cycle, the amino acid metabolism and the steroid hormone pathway the results of these comparisons were summarized on Figures 5.1, 5.2 and 5.3 respectively.

Each of the Figures shows the relevant parts of the according pathways as it is annotated in KEGG (Kanehisa et al., 2012; Kanehisa and Goto, 2000). For each metabolite it was highlighted if it was observed at different concentration in the *pre- vs. post-pano* (plasma and cells), the *primary vs. relapsed* and the *serum vs. plasma* comparison. The *sensitive vs. resistant* comparison was not further considered as there were no significant results.

Metabolites which are part of the γ-glutamyl cycle are disregulated in each of the comparisons (Figure 5.1). Amino acids did not change consistently, some are up- and others down-regulated in all comparisons. In contrast fatty acids and specifically the eicosanoid and octadecanoid precursor were up-regulated consistently in relapsed tumors, after the treatment (in cells and plasma) and in serum compared to plasma. In contrast eicosanoids and octadecanoids themselves are up-regulated in relapsed tumors compared to primaries. This is reverted by Panobinostat. This suggests that Panobinostat has an influence on the reaction of the fatty acids to the signaling peptides. Most amino acids have changed concentrations due to the Panobinostat treatment (Figure 5.2). But there is no consistent pattern visible. This makes an interpretation difficult. It can just be concluded that amino acid metabolism is altered which might depend on the alterations of the GSH metabolism.

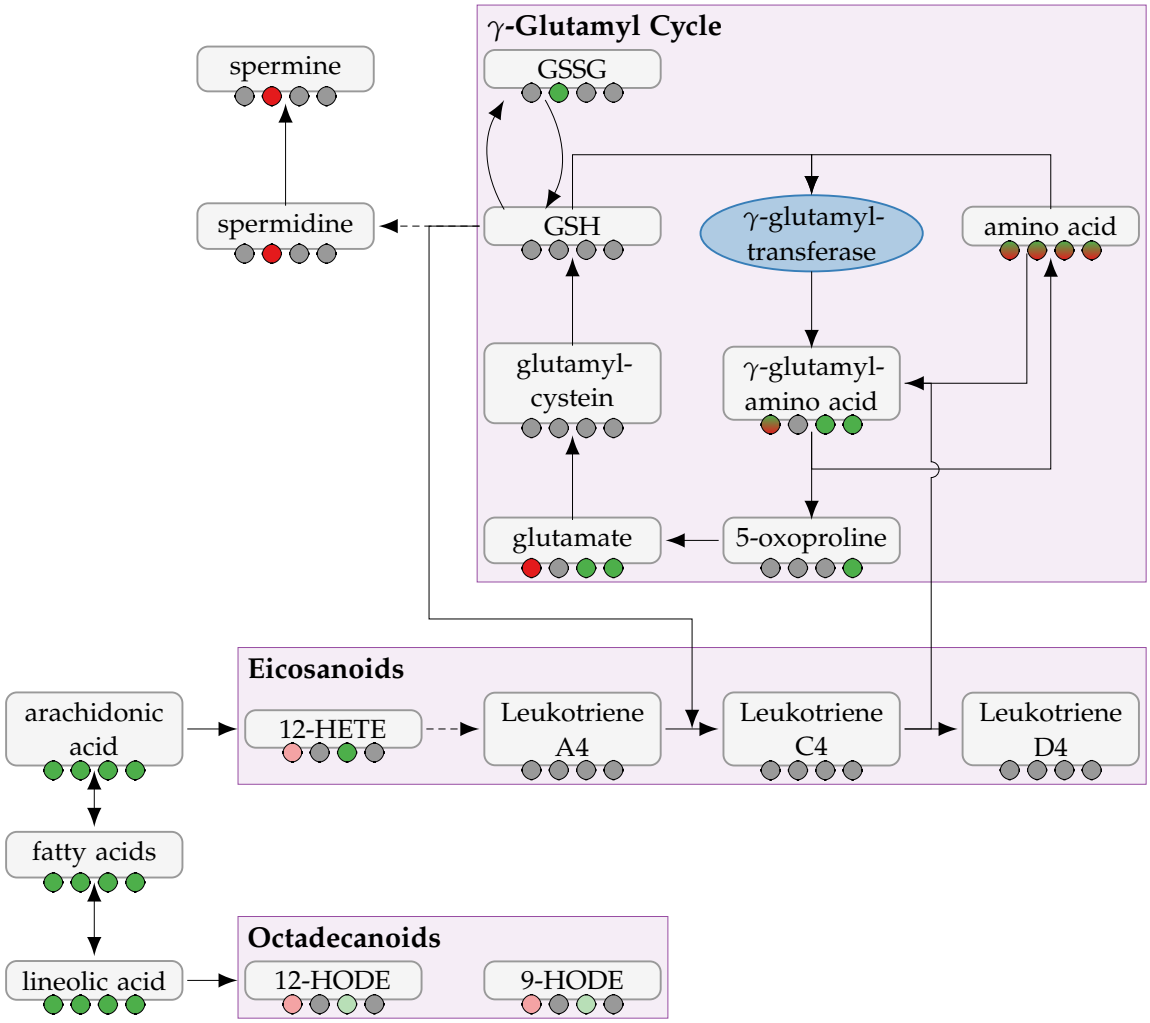


Figure 5.1: Gamma-glutamyl-cycle and signal lipids. Excerpts from the GSH metabolism as annotated in KEGG (Kanehisa et al., 2012; Kanehisa and Goto, 2000) are shown here. The four points below each metabolite represent the four comparisons *pre- vs. post-pano plasma*, *pre- vs. post-pano cells*, *primary vs. relapsed* and *serum vs. plasma* respectively. Red indicates that the metabolite is higher concentrated in the first group, green in the second. Light green and red metabolites were either observed as ratios or did not reach global significance. Nodes which are red and green were observed to be inconsistently altered in the according comparison. Solid edges were derived from KEGG, dashed edges represent indirect connections with some pathway.steps between the metabolites. The blue ellipse shows an enzyme.

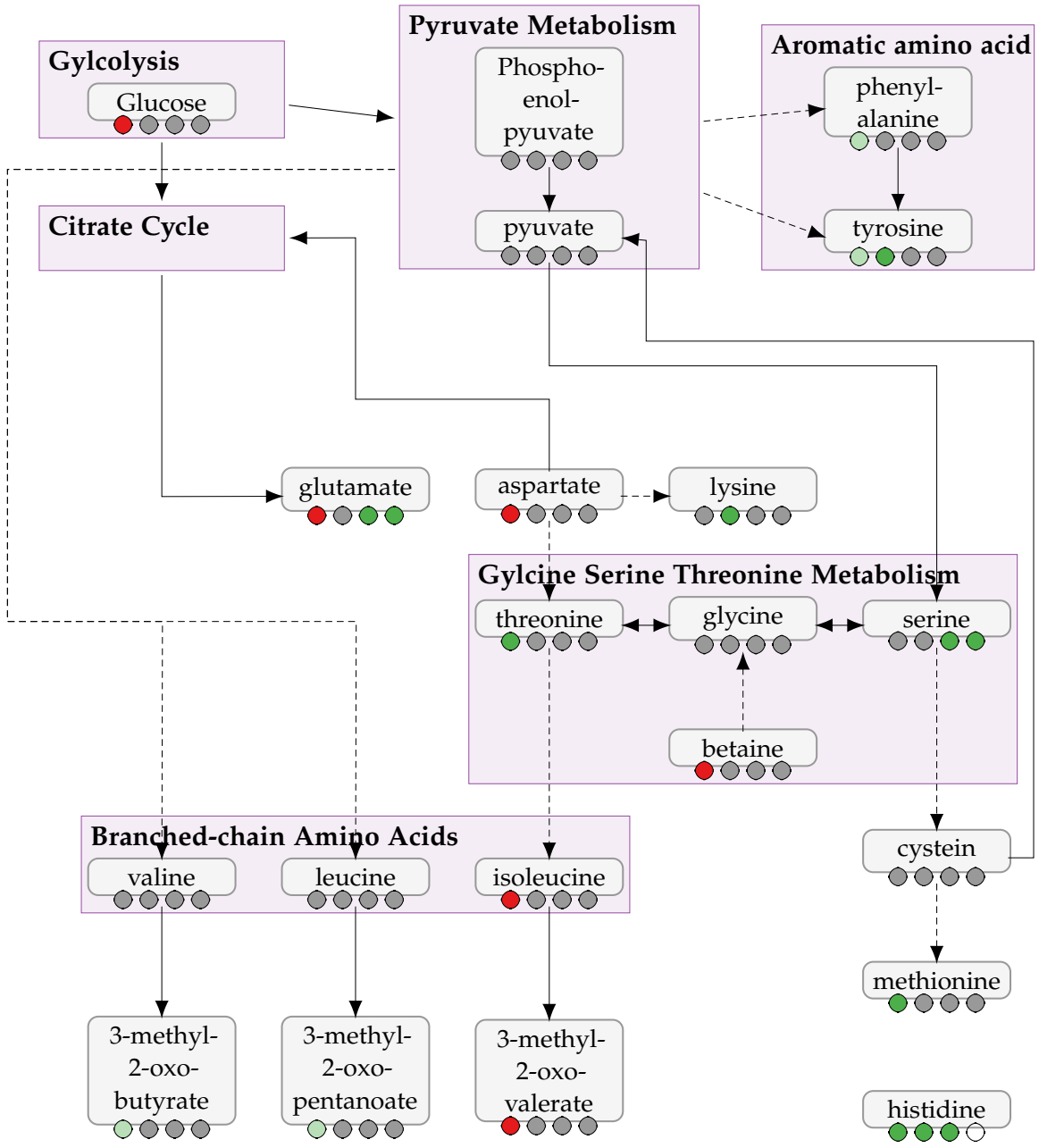


Figure 5.2: Amino acid pathway. The metabolism of amino acids is shown. As in Figure 5.1, edges were derived from KEGG (Kanehisa et al., 2012; Kanehisa and Goto, 2000). Again the circles below the metabolites represent the changes of the metabolite in the four comparisons *pre- vs. post-pano plasma*, *pre- vs. post-pano cells*, *primary vs. relapsed* and *serum vs. plasma* respectively. Dashed edges represent indirect interactions of metabolites.

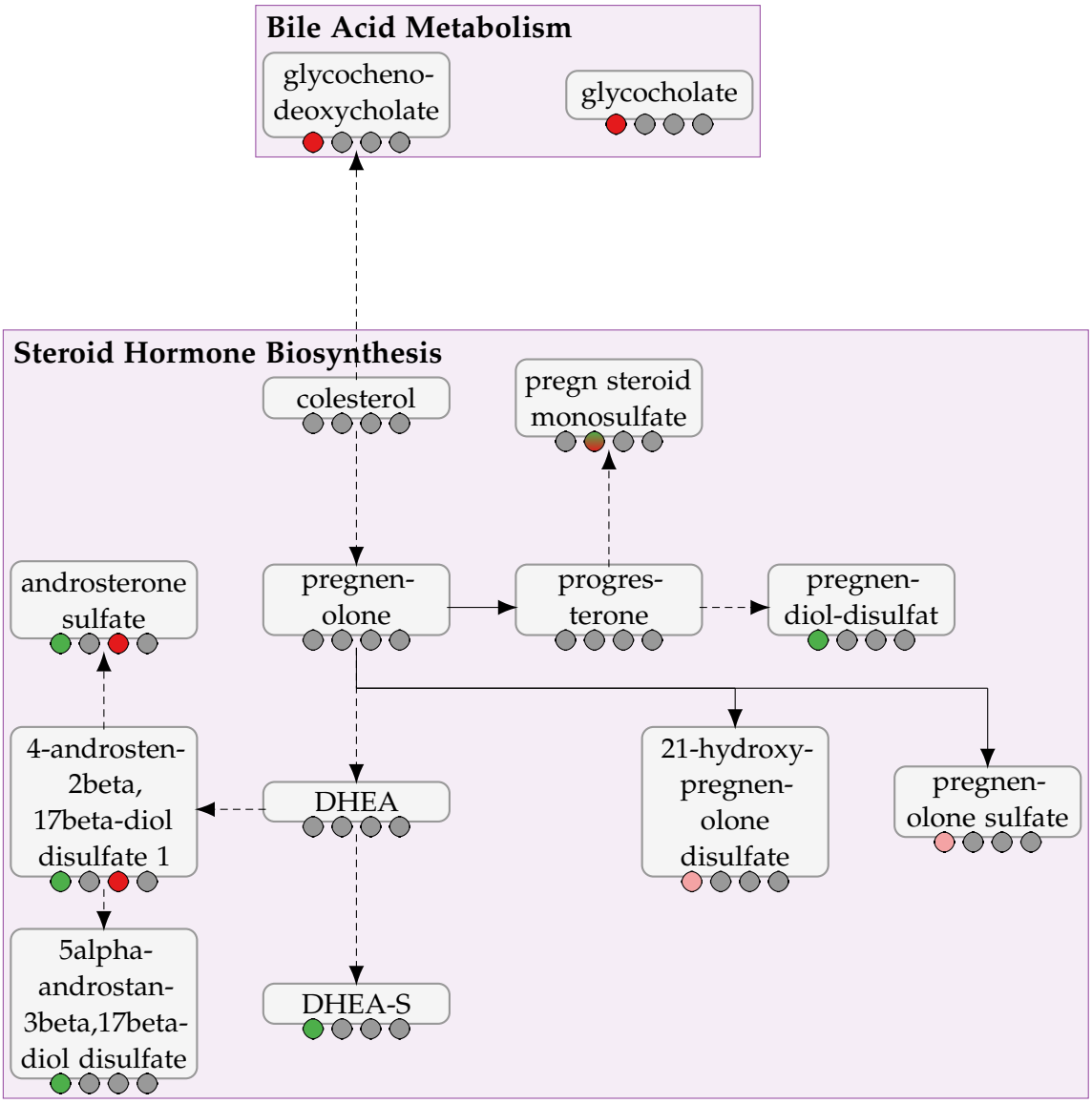


Figure 5.3: Steroid hormone pathway. Excerpts from the hormone steroid metabolism can bee seen in this figure. It was designed in the same way as Figures 5.1 and 5.2 The colored points indicate that a metabolite is higher (red) or lower (green) concentrated in the first group of the according comparison (*pre- vs. post-pano plasma*, *pre- vs. post-pano cells*, *primary vs. relapsed*, *serum vs. plasma*). Edges are derived from KEGG and dashed edges represent metabolites which are not directly connected but some metabolic steps apart from each other.

Steroid hormones are lower concentrated in relapsed than in plasma samples. This shift is reverted by Panobinostat which raises there concentration. This tendency can't be seen in cell lines, wherefore this are assumed to be an secondary effect on the organism instead of directly affecting the tumor cells. Nevertheless this seems to be one part of the anticancer effect of Panobinostat.

Apart from those, the nucleotide metabolism, fatty acid metabolism and the spermine biosynthesis were observed to be influenced by Panobinostat. Thereby, the increased fatty acid levels are thought to be caused by the progression of cancer, but the decrease of nucleotide synthesis is likely to be counteracting the tumor development. An open questions is if it is caused by Panobinostat rather is a sign that Panobinostat works.

5.2 Perspectives

Even though some new characteristics of DLBCL and the Panobinostat treatment were observed, many questions remain unanswered. The disregulation of the GSH-cycle is for example an objective for further investigations. Most likely it is involved in R-CHOP-resistance as well as the Panobinostat treatment. It should be analyzed if Panobinostat enhances the regulation of GSH, and if this is not the case it might be a target for other chemotherapeutics.

Integration with other data like genomics and proteomics experiments would help to distinguish between effects of Panobinostat due to gene regulation and effects caused by protein modifications. Specifically measurements of the γ-glutamyl transpeptidase would facilitate the understanding of the GSH metabolism. As in most cases larger sample sizes would have helped to increase the statistical power. This might e.g. allow for an distinction between R-CHOP-resistant and R-CHOP-sensitive tumors prior to the treatment.

This thesis is just a small step in the development of new chemotherapeutics, but after decades of CHOP and R-CHOP treatment the development of new drugs – especially against aggressive types of lymphoma – continues.

Bibliography

The following literature was used for this thesis.

[1] A A Alizadeh, M B Eisen, R E Davis, C Ma, I S Lossos, A Rosenwald, J C Boldrick, H Sabet, T Tran, X Yu, et al. "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling". In: *Nature* 403.6769 (Feb. 2000), pp. 503–511. DOI: 10.1038/35000501.

[2] E. Altmaier, S. L. Ramsay, A. Graber, H.-W. Mewes, K. M. Weinberger, and K. Suhre. "Bioinformatics Analysis of Targeted Metabolomics–Uncovering Old and New Tales of Diabetic Mice under Medication". In: *Endocrinology* 149.7 (Mar. 2008), pp. 3478–3489. DOI: 10.1210/en.2007-1747. URL: http://endo.endojournals.org/cgi/doi/10.1210/en.2007-1747 (visited on 09/10/2013).

[3] Maria R. Ambrosio, Pier P. Piccaluga, Maurilio Ponzoni, Bruno J. Rocca, Valeria Malagnino, Monica Onorati, Giulia De Falco, Valeria Calbi, Martin Ogwang, Kikkeri N. Naresh, et al. "The Alteration of Lipid Metabolism in Burkitt Lymphoma Identifies a Novel Marker: Adipophilin". In: *PLoS ONE* 7.8 (Aug. 2012). Ed. by Ann M. Moormann, e44315. DOI: 10.1371/journal.pone.0044315. URL: http://dx.plos.org/10.1371/journal.pone.0044315 (visited on 09/10/2013).

[4] American Cancer Society. *Non-Hodgkin Lymphoma Overview*. 2013. URL: http://www.cancer.org/cancer/non-hodgkinlymphoma/overviewguide/index (visited on 04/22/2013).

[5] M E Anderson, R D Allison, and A Meister. "Interconversion of leukotrienes catalyzed by purified gamma-glutamyl transpeptidase: concomitant formation of leukotriene D4 and gamma-glutamyl amino acids". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 79.4 (Feb. 1982), pp. 1088–1091.

[6] T. M. Annesley. "Ion Suppression in Mass Spectrometry". In: *Clinical Chemistry* 49.7 (July 2003), pp. 1041–1044. DOI: 10.1373/49.7.1041. URL: http://www.clinchem.org/cgi/doi/10.1373/49.7.1041 (visited on 09/11/2013).

[7] V Ascoli, F Lo Coco, M Artini, M Levrero, M Martelli, and F Negro. "Extranodal lymphomas associated with hepatitis C virus infection". eng. In: *American journal of clinical pathology* 109.5 (May 1998), pp. 600–609.

[8] Matías A Avila, Elena R García-Trevijano, Shelly C Lu, Fernando J Corrales, and José M Mato. "Methylthioadenosine". eng. In: *The international journal of biochemistry & cell biology* 36.11 (Nov. 2004), pp. 2125–2130. DOI: 10.1016/j.biocel.2003.11.016.

[9] Monya Baker. "Metabolomics: from small molecules to big ideas". In: *Nature Methods* 8.2 (Feb. 2011), pp. 117–121. DOI: 10.1038/nmeth0211-117. URL: http://www.nature.com/doifinder/10.1038/nmeth0211-117 (visited on 08/09/2013).

[10] Ganesaratnam K. Balendiran, Rajesh Dabur, and Deborah Fraser. "The role of glutathione in cancer". In: *Cell Biochemistry and Function* 22.6 (Nov. 2004), pp. 343–352. DOI: 10.1002/cbf.1149. URL: http://doi.wiley.com/10.1002/cbf.1149 (visited on 09/10/2013).

[11] A J Bannister, E A Miska, D Görlich, and T Kouzarides. "Acetylation of importin-alpha nuclear import factors by CBP/p300". eng. In: *Current biology: CB* 10.8 (Apr. 2000), pp. 467–470.

[12] Thaer Barri and Lars Ove Dragsted. "UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: Effect of experimental artefacts and anticoagulant". In: *Analytica Chimica Acta* 768 (Mar. 2013), pp. 118–128. DOI: 10.1016/j.aca.2013.01.015. URL: http://linkinghub.elsevier.com/retrieve/pii/S0003267013001244 (visited on 09/09/2013).

[13] Jan Baumann, Christopher Sevinsky, and Douglas S. Conklin. "Lipid biology of breast cancer". In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1831.10 (Oct. 2013), pp. 1509–1517. DOI: 10.1016/j.bbalip.2013.03.011. URL: http://linkinghub.elsevier.com/retrieve/pii/S1388198113000814 (visited on 09/10/2013).

[14] Oksana R Bereshchenko, Wei Gu, and Riccardo Dalla-Favera. "Acetylation inactivates the transcriptional repressor BCL6". eng. In: *Nature genetics* 32.4 (Dec. 2002), pp. 606–613. DOI: 10.1038/ng1018.

[15] C D Bloomfield, D C Arthur, G Frizzera, E G Levine, B A Peterson, and K J Gajl-Peczalska. "Nonrandom chromosome abnormalities in lymphoma". eng. In: *Cancer research* 43.6 (June 1983), pp. 2975–2984.

[16] Nathan Blow. "Metabolomics: Biochemistry's new look". In: *Nature* 455.7213 (Oct. 2008), pp. 697–700. DOI: 10.1038/455697a. URL: http://www.nature.com/doifinder/10.1038/455697a (visited on 08/09/2013).

[17] Christophe Bonnet, Georges Fillet, Nicolas Mounier, Gérard Ganem, Thierry Jo Molina, Catherine Thiéblemont, Christophe Fermé, Bruno Quesnel, Claude Martin, Christian Gisselbrecht, et al. "CHOP alone compared with CHOP plus radiotherapy for localized aggressive lymphoma in elderly patients: a study by the Groupe d'Etude des Lymphomes de l'Adulte". eng. In: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 25.7 (Mar. 2007), pp. 787–792. DOI: 10.1200/JCO.2006.07.0722.

[18] Kurt J. Boudonck, Matthew W. Mitchell, Jacob Wulff, and John A. Ryals. "Characterization of the biochemical variability of bovine milk using metabolomics". In: *Metabolomics* 5.4 (Apr. 2009), pp. 375–386. DOI: 10.1007/s11306-009-0160-8. URL: http://link.springer.com/10.1007/s11306-009-0160-8 (visited on 09/11/2013).

[19] M M Bradford. "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding". eng. In: *Analytical biochemistry* 72 (May 1976), pp. 248–254.

[20] Fernando Cabanillas. "Non-Hodgkin's Lymphoma: The Old and the New". In: *Clinical Lymphoma Myeloma and Leukemia* 11 (June 2011), S87–S90. DOI: 10.1016/j.clml.2011.03.029. URL: http://linkinghub.elsevier.com/retrieve/pii/S2152265011000437 (visited on 07/31/2013).

[21] Ying Cai, Wenli Cui, Weixiang Chen, Ping Wei, Yayun Chi, Ping Zhang, Rui Bi, and Xiaoyan Zhou. "The effects of a histone deacetylase inhibitor on biological behavior of diffuse large B-cell lymphoma cell lines and insights into the underlying mechanisms". eng. In: *Cancer cell international* 13.1 (2013), p. 57. DOI: 10.1186/1475-2867-13-57.

[22] Elsa Callén, Mila Jankovic, Simone Difilippantonio, Jeremy A Daniel, Hua-Tang Chen, Arkady Celeste, Manuela Pellegrini, Kevin McBride, Danny Wangsa, Andrea L Bredemeyer, et al. "ATM prevents the persistence

and propagation of chromosome breaks in lymphocytes". eng. In: *Cell* 130.1 (July 2007), pp. 63–75. DOI: 10.1016/j.cell.2007.06.016.

[23] Robert A. Casero and Laurence J. Marton. "Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases". In: *Nature Reviews Drug Discovery* 6.5 (May 2007), pp. 373–390. DOI: 10.1038/nrd2243. URL: http://www.nature.com/doifinder/10.1038/nrd2243 (visited on 09/10/2013).

[24] Leandro Cerchietti. "Cellular metabolism and serum metabolomic profiling to identify prognostic markers and novel therapeutic targets in non-Hodgkin lymphomas." In: *Weill Cornell Medical College of Cornell University* (2013).

[25] Leandro Cerchietti and Maria Nieves Calvo-Vidal. "The metabolism of lymphomas". In: *Weill Cornell Medical College of Cornell University* (2013).

[26] W. W.L. Choi, D. D. Weisenburger, T. C. Greiner, M. A. Piris, A. H. Banham, J. Delabie, R. M. Braziel, H. Geng, J. Iqbal, G. Lenz, et al. "A New Immunostain Algorithm Classifies Diffuse Large B-Cell Lymphoma into Molecular Subtypes with High Accuracy". In: *Clinical Cancer Research* 15.17 (Aug. 2009), pp. 5494–5502. DOI: 10.1158/1078-0432.CCR-09-0113. URL: http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-09-0113 (visited on 08/08/2013).

[27] J C Cigudosa, N Z Parsa, D C Louie, D A Filippa, S C Jhanwar, B Johansson, F Mitelman, and R S Chaganti. "Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas". eng. In: *Genes, chromosomes & cancer* 25.2 (June 1999), pp. 123–133.

[28] Bertrand Coiffier, Catherine Thieblemont, Eric Van Den Neste, Gérard Lepeu, Isabelle Plantier, Sylvie Castaigne, Sophie Lefort, Gérald Marit, Margaret Macro, Catherine Sebban, et al. "Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte". eng. In: *Blood* 116.12 (Sept. 2010), pp. 2040–2045. DOI: 10.1182/blood-2010-03-276246.

[29] Max D. Cooper and Matthew N. Alder. "The Evolution of Adaptive Immune Systems". In: *Cell* 124.4 (Feb. 2006), pp. 815–822. DOI: 10.1016/j.cell.2006.02.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867406001528 (visited on 08/01/2013).

[30] *Current protocols in molecular biology*. Brooklyn, N.Y. : Media, Pa: Greene Pub. Associates ; J. Wiley, order fulfillment, 1987.

[31] S De Vita, C Sacco, D Sansonno, A Gloghini, F Dammacco, M Crovatto, G Santini, R Dolcetti, M Boiocchi, A Carbone, et al. “Characterization of overt B-cell lymphomas in patients with hepatitis C virus infection”. eng. In: *Blood* 90.2 (July 1997), pp. 776–782.

[32] Michael Dickinson, David Ritchie, Daniel J DeAngelo, Andrew Spencer, Oliver G Ottmann, Thomas Fischer, Kapil N Bhalla, Angela Liu, Katie Parker, Jeffrey W Scott, et al. “Preliminary evidence of disease response to the pan deacetylase inhibitor panobinostat (LBH589) in refractory Hodgkin Lymphoma”. eng. In: *British journal of haematology* 147.1 (Oct. 2009), pp. 97–101. DOI: 10.1111/j.1365-2141.2009.07837.x.

[33] Milos Dokmanovic, Cathy Clarke, and Paul A Marks. “Histone deacetylase inhibitors: overview and perspectives”. eng. In: *Molecular cancer research: MCR* 5.10 (Oct. 2007), pp. 981–989. DOI: 10.1158/1541-7786.MCR-07-0324.

[34] S Dominici, M Valentini, E Maellaro, B Del Bello, A Paolicchi, E Lorenzini, R Tongiani, M Comporti, and A Pompella. “Redox modulation of cell surface protein thiols in U937 lymphoma cells: the role of gamma-glutamyl transpeptidase-dependent H2O2 production and S-thiolation”. eng. In: *Free radical biology & medicine* 27.5-6 (Sept. 1999), pp. 623–635.

[35] W Dröge and R Breitkreutz. “Glutathione and immune function”. eng. In: *The Proceedings of the Nutrition Society* 59.4 (Nov. 2000), pp. 595–600.

[36] Daryl C Drummond, Charles O Noble, Dmitri B Kirpotin, Zexiong Guo, Gary K Scott, and Christopher C Benz. “Clinical development of histone deacetylase inhibitors as anticancer agents”. eng. In: *Annual review of pharmacology and toxicology* 45 (2005), pp. 495–528. DOI: 10.1146/annurev.pharmtox.45.120403.095825.

[37] Hong Duan, Caroline A Heckman, and Linda M Boxer. “Histone deacetylase inhibitors down-regulate bcl-2 expression and induce apoptosis in t(14;18) lymphomas”. eng. In: *Molecular and cellular biology* 25.5 (Mar. 2005), pp. 1608–1619. DOI: 10.1128/MCB.25.5.1608-1619.2005.

[38] Anne M. Evans, Corey D. DeHaven, Tom Barrett, Matt Mitchell, and Eric Milgram. “Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems”. In: *Analytical Chemistry* 81.16 (Aug. 2009), pp. 6656–6667. DOI: 10.1021/ac901536h. URL: http://pubs.acs.org/doi/abs/10.1021/ac901536h (visited on 01/27/2013).

[39] I S Fentiman. “Gamma-glutamyl transferase: risk and prognosis of cancer”. In: *British Journal of Cancer* 106.9 (Apr. 2012), pp. 1467–1468. DOI: 10.1038/bjc.2012.128. URL: http://www.nature.com/doifinder/10.1038/bjc.2012.128 (visited on 08/28/2013).

[40] P Feugier, A Van Hoof, C Sebban, P Solal-Celigny, R Bouabdallah, C Fermé, B Christian, E Lepage, H Tilly, F Morschhauser, et al. “Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d’Etude des Lymphomes de l’Adulte”. eng. In: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 23.18 (June 2005), pp. 4117–4126. DOI: 10.1200/JCO.2005.09.131.

[41] Richard I Fisher, Thomas P Miller, and Owen A O’Connor. “Diffuse aggressive lymphoma”. eng. In: *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* (2004), pp. 221–236. DOI: 10.1182/asheducation-2004.1.221.

[42] Susan G Fisher and Richard I Fisher. “The epidemiology of non-Hodgkin’s lymphoma”. eng. In: *Oncogene* 23.38 (Aug. 2004), pp. 6524–6534. DOI: 10.1038/sj.onc.1207843.

[43] James E. Galagan, Kyle Minch, Matthew Peterson, Anna Lyubetskaya, Elham Azizi, Linsday Sweet, Antonio Gomes, Tige Rustad, Gregory Dolganov, Irina Glotova, et al. “The Mycobacterium tuberculosis regulatory network and hypoxia”. In: *Nature* 499.7457 (July 2013), pp. 178–183. DOI: 10.1038/nature12337. URL: http://www.nature.com/doifinder/10.1038/nature12337 (visited on 09/11/2013).

[44] Sarah E Gibson, Steven H Swerdlow, Fiona E Craig, Urvashi Surti, James R Cook, Michael A Nalesnik, Chris Lowe, Katrina M Wood, and Chris M Bacon. “EBV-positive extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue in the posttransplant setting: a distinct type of posttransplant lymphoproliferative disorder?” eng. In: *The American journal of surgical pathology* 35.6 (June 2011), pp. 807–815. DOI: 10.1097/PAS.0b013e3182190999.

[45] *Goldman’s Cecil medicine*. 24th ed. Philadelphia: Elsevier/Saunders, 2012.

[46] M R Graf, W Jia, and R M Loria. “The neuro-steroid, 3beta androstene 17alpha diol exhibits potent cytotoxic effects on human malignant glioma and lymphoma cells through different programmed cell death pathways”.

eng. In: *British journal of cancer* 97.5 (Sept. 2007), pp. 619–627. DOI: 10.1038/sj.bjc.6603894.

[47] M Gupta, J J Han, M Stenson, L Wellik, and T E Witzig. "Regulation of STAT3 by histone deacetylase-3 in diffuse large B-cell lymphoma: implications for therapy". eng. In: *Leukemia* 26.6 (June 2012), pp. 1356–1364. DOI: 10.1038/leu.2011.340.

[48] N L Harris, E S Jaffe, H Stein, P M Banks, J K Chan, M L Cleary, G Delsol, C De Wolf-Peeters, B Falini, and K C Gatter. "A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group". eng. In: *Blood* 84.5 (Sept. 1994), pp. 1361–1392.

[49] H Hasegawa, Y Yamada, K Tsukasaki, N Mori, K Tsuruda, D Sasaki, T Usui, A Osaka, S Atogami, C Ishikawa, et al. "LBH589, a deacetylase inhibitor, induces apoptosis in adult T-cell leukemia/lymphoma cells via activation of a novel RAIDD-caspase-2 pathway". eng. In: *Leukemia* 25.4 (Apr. 2011), pp. 575–587. DOI: 10.1038/leu.2010.315.

[50] Sigrid Hatse, Erik De Clercq, and Jan Balzarini. "Role of antimetabolites of purine and pyrimidine nucleotide metabolism in tumor cell differentiation". In: *Biochemical Pharmacology* 58.4 (Aug. 1999), pp. 539–555. DOI: 10.1016/S0006-2952(99)00035-0. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006295299000350 (visited on 09/10/2013).

[51] Sylvia Hoeller, Alexandar Tzankov, Stefano A Pileri, Philip Went, and Stephan Dirnhofer. "Epstein-Barr virus-positive diffuse large B-cell lymphoma in elderly patients is rare in Western populations". eng. In: *Human pathology* 41.3 (Mar. 2010), pp. 352–357. DOI: 10.1016/j.humpath.2009.07.024.

[52] James F Holland. *Cancer medicine*. English. Hamilton, Ontario; New York: B.C. Decker, 2000.

[53] James F. Holland, Emil Frei, and Donald W. Kufe. *Cancer medicine 6*. 6th ed. Hamilton, Ont. ; Lewiston, NY: BC Decker, 2003.

[54] S. J. Horning. "Chemotherapy With or Without Radiotherapy in Limited-Stage Diffuse Aggressive Non-Hodgkin's Lymphoma: Eastern Cooperative Oncology Group Study 1484". In: *Journal of Clinical Oncology* 22.15 (Aug. 2004), pp. 3032–3038. DOI: 10.1200/JCO.2004.06.088. URL: http://www.jco.org/cgi/doi/10.1200/JCO.2004.06.088 (visited on 08/08/2013).

[55] N Howlader, AM Noone, M Krapcho, J Garshell, N Neyman, SF Altekruse, CL Kosary, M Yu, J Ruhl, Z Tatalovich, et al. "SEER Cancer Statistics Review". eng. In: *National Cancer Institute* (1975).

[56] Charles Janeway. *Immunobiology: the immune system in health and disease*. English. New York: Garland Pub., 2001.

[57] Ricky W Johnstone and Jonathan D Licht. "Histone deacetylase inhibitors in cancer therapy: is transcription the primary target?" eng. In: *Cancer cell* 4.1 (July 2003), pp. 13–18.

[58] Rima Kaddurah-Daouk, Bruce S. Kristal, and Richard M. Weinshilboum. "Metabolomics: A Global Biochemical Approach to Drug Response and Disease". In: *Annual Review of Pharmacology and Toxicology* 48.1 (Feb. 2008), pp. 653–683. DOI: 10.1146/annurev.pharmtox.48.113006.094715. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.pharmtox.48.113006.094715 (visited on 08/01/2013).

[59] William Kaelin and Steven McKnight. "Influence of Metabolism on Epigenetics and Disease". In: *Cell* 153.1 (Mar. 2013), pp. 56–69. DOI: 10.1016/j.cell.2013.03.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867413002900 (visited on 04/09/2013).

[60] M Kanehisa and S Goto. "KEGG: kyoto encyclopedia of genes and genomes". eng. In: *Nucleic acids research* 28.1 (Jan. 2000), pp. 27–30.

[61] Minoru Kanehisa, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. "KEGG for integration and interpretation of large-scale molecular data sets". eng. In: *Nucleic acids research* 40.Database issue (Jan. 2012), pp. D109–114. DOI: 10.1093/nar/gkr988.

[62] Ralf Küppers. "The biology of Hodgkin's lymphoma". In: *Nature Reviews Cancer* 9.1 (Dec. 2008), pp. 15–27. DOI: 10.1038/nrc2542. URL: http://www.nature.com/doifinder/10.1038/nrc2542 (visited on 08/01/2013).

[63] Subhash P Khanapure, David S Garvey, David R Janero, and L Gordon Letts. "Eicosanoids in inflammation: biosynthesis, pharmacology, and therapeutic frontiers". eng. In: *Current topics in medicinal chemistry* 7.3 (2007), pp. 311–340.

[64] Amit Khot, Michael Dickinson, and H Miles Prince. "Panobinostat in lymphoid and myeloid malignancies". ENG. In: *Expert opinion on investigational drugs* (July 2013). DOI: 10.1517/13543784.2013.815165.

[65] Lidija Klampfer, Jie Huang, Laurie-Anne Swaby, and Leonard Augenlicht. "Requirement of histone deacetylase activity for signaling by STAT1". eng. In: *The Journal of biological chemistry* 279.29 (July 2004), pp. 30358–30368. DOI: 10.1074/jbc.M401359200.

[66] Ulf Klein and Riccardo Dalla-Favera. "Germinal centres: role in B-cell physiology and malignancy". In: *Nature Reviews Immunology* 8.1 (Jan. 2008), pp. 22–33. DOI: 10.1038/nri2217. URL: http://www.nature.com/doifinder/10.1038/nri2217 (visited on 08/02/2013).

[67] Jan Krumsiek, Karsten Suhre, Anne M. Evans, Matthew W. Mitchell, Robert P. Mohney, Michael V. Milburn, Brigitte Wägele, Werner Römisch-Margl, Thomas Illig, Jerzy Adamski, et al. "Mining the Unknown: A Systems Approach to Metabolite Identification Combining Genetic and Metabolic Information". In: *PLoS Genetics* 8.10 (Oct. 2012). Ed. by Mark I. McCarthy, e1003005. DOI: 10.1371/journal.pgen.1003005. URL: http://dx.plos.org/10.1371/journal.pgen.1003005 (visited on 01/27/2013).

[68] M M Le Beau. "Chromosomal abnormalities in non-Hodgkin's lymphomas". eng. In: *Seminars in oncology* 17.1 (Feb. 1990), pp. 20–29.

[69] T. W. LeBien and T. F. Tedder. "B lymphocytes: how they develop and function". In: *Blood* 112.5 (Aug. 2008), pp. 1570–1580. DOI: 10.1182/blood-2008-02-078071. URL: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2008-02-078071 (visited on 08/01/2013).

[70] H. Lee, P. Zhang, A. Herrmann, C. Yang, H. Xin, Z. Wang, D. S. B. Hoon, S. J. Forman, R. Jove, A. D. Riggs, et al. "Acetylated STAT3 is crucial for methylation of tumor-suppressor gene promoters and inhibition by resveratrol results in demethylation". In: *Proceedings of the National Academy of Sciences* 109.20 (Apr. 2012), pp. 7765–7769. DOI: 10.1073/pnas.1205132109. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1205132109 (visited on 08/09/2013).

[71] Manuela Lemoine and Anas Younes. "Histone deacetylase inhibitors in the treatment of lymphoma". eng. In: *Discovery medicine* 10.54 (Nov. 2010), pp. 462–470.

[72] Leukemia and Lymphoma Society. *Facts and Statistics*. 2013. URL: http://www.lls.org/diseaseinformation/getinformationsupport/factsstatistics/ (visited on 04/22/2013).

[73] Linsheng Liu, Jiye Aa, Guangji Wang, Bei Yan, Ying Zhang, Xinwen Wang, Chunyan Zhao, Bei Cao, Jian Shi, Mengjie Li, et al. "Differences in metabolite profile between blood plasma and serum". In: *Analytical Biochemistry* 406.2 (Nov. 2010), pp. 105–112. DOI: 10.1016/j.ab.2010.07.015. URL: http://linkinghub.elsevier.com/retrieve/pii/S0003269710004616 (visited on 09/09/2013).

[74] Roger M Loria and Martin R Graf. "17alpha-androstenediol-mediated oncophagy of tumor cells by different mechanisms is determined by the target tumor". eng. In: *Annals of the New York Academy of Sciences* 1262 (July 2012), pp. 127–133. DOI: 10.1111/j.1749-6632.2012.06602.x.

[75] G Lum and S R Gambino. "Serum gamma-glutamyl transpeptidase activity as an indicator of disease of liver, pancreas, or bone". eng. In: *Clinical chemistry* 18.4 (Apr. 1972), pp. 358–362.

[76] Maurizio Martelli, Andrés J M Ferreri, Claudio Agostinelli, Alice Di Rocco, Michael Pfreundschuh, and Stefano A Pileri. "Diffuse large B-cell lymphoma". eng. In: *Critical reviews in oncology/hematology* 87.2 (Aug. 2013), pp. 146–171. DOI: 10.1016/j.critrevonc.2012.12.009.

[77] Claudia Mauri and Anneleen Bosma. "Immune Regulatory Function of B Cells". In: *Annual Review of Immunology* 30.1 (Apr. 2012), pp. 221–241. DOI: 10.1146/annurev-immunol-020711-074934. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev-immunol-020711-074934 (visited on 08/01/2013).

[78] McGraw-Hill. *Dictionary of chemistry*. 2nd ed. New York: McGraw-Hill, 2003.

[79] Fiona McLaughlin and Nicholas B La Thangue. "Histone deacetylase inhibitors open new doors in cancer therapy". eng. In: *Biochemical pharmacology* 68.6 (Sept. 2004), pp. 1139–1144. DOI: 10.1016/j.bcp.2004.05.034.

[80] A Meister. "Selective modification of glutathione metabolism". eng. In: *Science (New York, N.Y.)* 220.4596 (Apr. 1983), pp. 472–477.

[81] C. Menni, G. Kastenmuller, A. K. Petersen, J. T. Bell, M. Psatha, P.-C. Tsai, C. Gieger, H. Schulz, I. Erte, S. John, et al. "Metabolomic markers reveal novel pathways of ageing and early development in human populations". In: *International Journal of Epidemiology* (July 2013). DOI: 10.1093/ije/dyt094. URL: http://www.ije.oxfordjournals.org/cgi/doi/10.1093/ije/dyt094 (visited on 09/11/2013).

[82] Ujm Mey, F Hitz, A Lohri, S Pederiva, C Taverna, A Tzankov, O Meier, K Yeow, and C Renner. "Diagnosis and treatment of diffuse large B-cell lymphoma". In: *Swiss Medical Weekly* (Jan. 2012). DOI: 10.4414/smw.2012.13511. URL: http://doi.emh.ch/smw.2012.13511 (visited on 07/31/2013).

[83] T P Miller, S Dahlberg, J R Cassady, D J Adelstein, C M Spier, T M Grogan, M LeBlanc, S Carlin, E Chase, and R I Fisher. "Chemotherapy alone compared with chemotherapy plus radiotherapy for localized intermediate- and high-grade non-Hodgkin's lymphoma". eng. In: *The New England journal of medicine* 339.1 (July 1998), pp. 21–26. DOI: 10.1056/NEJM199807023390104.

[84] Soo Kee Min, Young Ho Koh, Yunwoong Park, Hyo Jung Kim, Jinwon Seo, Hye-Rim Park, Seong Jin Cho, and In Sun Kim. "Expression of HAT1 and HDAC1, 2, 3 in Diffuse Large B-Cell Lymphomas, Peripheral T-Cell Lymphomas, and NK/T-Cell Lymphomas". eng. In: *Korean journal of pathology* 46.2 (Apr. 2012), pp. 142–150. DOI: 10.4132/KoreanJPathol.2012.46.2.142.

[85] Kirstin Mittelstrass, Janina S. Ried, Zhonghao Yu, Jan Krumsiek, Christian Gieger, Cornelia Prehn, Werner Roemisch-Margl, Alexey Polonikov, Annette Peters, Fabian J. Theis, et al. "Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers". In: *PLoS Genetics* 7.8 (Aug. 2011). Ed. by Mark I. McCarthy, e1002215. DOI: 10.1371/journal.pgen.1002215. URL: http://dx.plos.org/10.1371/journal.pgen.1002215 (visited on 09/11/2013).

[86] *Molecular cell biology*. 7th ed. New York: W.H. Freeman and Co, 2013.

[87] William G Morice, Fausto J Rodriguez, James D Hoyer, and Paul J Kurtin. "Diffuse large B-cell lymphoma with distinctive patterns of splenic and bone marrow involvement: clinicopathologic features of two cases". In: *Modern Pathology* 18.4 (Oct. 2004), pp. 495–502. DOI: 10.1038/modpathol.3800297. URL: http://www.nature.com/doifinder/10.1038/modpathol.3800297 (visited on 09/10/2013).

[88] L. M. Morton. "Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001". In: *Blood* 107.1 (Jan. 2006), pp. 265–276. DOI: 10.1182/blood-2005-06-2508. URL: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2005-06-2508 (visited on 04/22/2013).

[89] B. N. Nathwani. “A promising new biologic prognostic model in diffuse large B-cell lymphoma”. In: *Blood* 120.11 (Sept. 2012), pp. 2161–2162. DOI: 10.1182/blood-2012-07-441253. URL: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2012-07-441253 (visited on 08/08/2013).

[90] Rafal R Nazarewicz, Woineshet J Zenebe, Arti Parihar, Mordhwaj S Parihar, Michael Vaccaro, Cameron Rink, Chandan K Sen, and Pedram Ghafourifar. “12(S)-hydroperoxyeicosatetraenoic acid (12-HETE) increases mitochondrial nitric oxide by increasing intramitochondrial calcium”. eng. In: *Archives of biochemistry and biophysics* 468.1 (Dec. 2007), pp. 114–120. DOI: 10.1016/j.abb.2007.09.018.

[91] Jeremy K. Nicholson and John C. Lindon. “Systems biology: Metabonomics”. In: *Nature* 455.7216 (Oct. 2008), pp. 1054–1056. DOI: 10.1038/4551054a. URL: http://www.nature.com/doifinder/10.1038/4551054a (visited on 01/27/2013).

[92] Daniel K. Nomura and Benjamin F. Cravatt. “Lipid metabolism in cancer”. In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* (Aug. 2013). DOI: 10.1016/j.bbalip.2013.08.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S1388198113001583 (visited on 09/10/2013).

[93] M M Oken, R H Creech, D C Tormey, J Horton, T E Davis, E T McFadden, and P P Carbone. “Toxicity and response criteria of the Eastern Cooperative Oncology Group”. eng. In: *American journal of clinical oncology* 5.6 (Dec. 1982), pp. 649–655.

[94] Online Mendelian Inheritance in Man, OMIM. *OMIM Entry - # 236000 - LYMPHOMA, HODGKIN*. Johns Hopkins University, Baltimore. 2013. URL: http://www.omim.org/entry/236000 (visited on 12/10/2012).

[95] Ia M Ozhga. “[Electrocardiographic examinations of children in the Pamir mountains]”. rus. In: *Zdravookhranenie Kirgizii* 3 (June 1975), pp. 27–32.

[96] Seila Pandur, Serhiy Pankiv, Mona Johannessen, Ugo Moens, and Nils-Erik Huseby. “Gamma-glutamyltransferase is upregulated after oxidative stress through the Ras signal transduction pathway in rat colon carcinoma cells”. eng. In: *Free radical research* 41.12 (Dec. 2007), pp. 1376–1384. DOI: 10.1080/10715760701739488.

[97] F.Nina Papavasiliou and David G Schatz. “Somatic Hypermutation of Immunoglobulin Genes: Merging Mechanisms for Genetic Diversity”. In: *Cell* 109.2 (Apr. 2002), S35–S44. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867402007067.

[98] D C Parker. “T Cell-Dependent B Cell Activation”. In: *Annual Review of Immunology* 11.1 (Apr. 1993), pp. 331–360. DOI: 10.1146/annurev.iy.11.040193.001555. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.iy.11.040193.001555 (visited on 08/01/2013).

[99] Gary J. Patti, Oscar Yanes, and Gary Siuzdak. “Innovation: Metabolomics: the apogee of the omics trilogy”. In: *Nature Reviews Molecular Cell Biology* 13.4 (Mar. 2012), pp. 263–269. DOI: 10.1038/nrm3314. URL: http://www.nature.com/doifinder/10.1038/nrm3314 (visited on 08/01/2013).

[100] Helen Pearson. “Meet the human metabolome”. In: *Nature* 446.7131 (Mar. 2007), pp. 8–8. DOI: 10.1038/446008a. URL: http://www.nature.com/doifinder/10.1038/446008a (visited on 08/12/2013).

[101] Michael Pfreundschuh, Lorenz Trümper, Anders Osterborg, Ruth Pettengell, Marek Trneny, Kevin Imrie, David Ma, Devinder Gill, Jan Walewski, Pier-Luigi Zinzani, et al. “CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group”. eng. In: *The lancet oncology* 7.5 (May 2006), pp. 379–391. DOI: 10.1016/S1470-2045(06)70664-7.

[102] Michael Pfreundschuh, Lorenz Trümper, Marita Kloess, Rudolf Schmits, Alfred C Feller, Christian Rübe, Christian Rudolph, Marcel Reiser, Dieter K Hossfeld, Hartmut Eimermacher, et al. “Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of elderly patients with aggressive lymphomas: results of the NHL-B2 trial of the DSHNHL”. eng. In: *Blood* 104.3 (Aug. 2004), pp. 634–641. DOI: 10.1182/blood-2003-06-2095.

[103] Ryan T Phan and Riccardo Dalla-Favera. “The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells”. eng. In: *Nature* 432.7017 (Dec. 2004), pp. 635–639. DOI: 10.1038/nature03147.

[104] A. M. Puzio-Kuter. "The Role of p53 in Metabolic Regulation". In: *Genes & Cancer* 2.4 (May 2011), pp. 385–391. DOI: 10.1177/1947601911409738. URL: http://gan.sagepub.com/lookup/doi/10.1177/1947601911409738 (visited on 07/30/2013).

[105] P H Rao, J Houldsworth, K Dyomina, N Z Parsa, J C Cigudosa, D C Louie, L Popplewell, K Offit, S C Jhanwar, and R S Chaganti. "Chromosomal and gene amplification in diffuse large B-cell lymphoma". eng. In: *Blood* 92.1 (July 1998), pp. 234–240.

[106] Lee D Roberts, Amanda L Souza, Robert E Gerszten, and Clary B Clish. "Targeted metabolomics". eng. In: *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 30 (Apr. 2012), Unit 30.2.1–24. DOI: 10.1002/0471142727.mb3002s98.

[107] Simone Rochfort. "Metabolomics Reviewed: A New "Omics" Platform Technology for Systems Biology and Implications for Natural Products Research". In: *Journal of Natural Products* 68.12 (Dec. 2005), pp. 1813–1820. DOI: 10.1021/np050255w. URL: http://pubs.acs.org/doi/abs/10.1021/np050255w (visited on 08/01/2013).

[108] Andreas Rosenwald, George Wright, Karen Leroy, Xin Yu, Philippe Gaulard, Randy D Gascoyne, Wing C Chan, Tong Zhao, Corinne Haioun, Timothy C Greiner, et al. "Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma". eng. In: *The Journal of experimental medicine* 198.6 (Sept. 2003), pp. 851–862. DOI: 10.1084/jem.20031074.

[109] D. Rudnicka, A. Oszmiana, D. K. Finch, I. Strickland, D. J. Schofield, D. C. Lowe, M. A. Sleeman, and D. M. Davis. "Rituximab causes a polarization of B cells that augments its therapeutic function in NK-cell-mediated antibody-dependent cellular cytotoxicity". In: *Blood* 121.23 (Apr. 2013), pp. 4694–4702. DOI: 10.1182/blood-2013-02-482570. URL: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2013-02-482570 (visited on 09/11/2013).

[110] A. A. Ruefli. "The histone deacetylase inhibitor and chemotherapeutic agent suberoylanilide hydroxamic acid (SAHA) induces a cell-death pathway characterized by cleavage of Bid and production of reactive oxygen species". In: *Proceedings of the National Academy of Sciences* 98.19 (Sept. 2001), pp. 10833–10838. DOI: 10.1073/pnas.191208598. URL:

http://www.pnas.org/cgi/doi/10.1073/pnas.191208598 (visited on 09/07/2013).

[111] Erik J. Saude, Darryl Adamko, Brian H. Rowe, Tom Marrie, and Brian D. Sykes. "Variation of metabolites in normal human urine". In: *Metabolomics* 3.4 (Sept. 2007), pp. 439–451. DOI: 10.1007/s11306-007-0091-1. URL: http://link.springer.com/10.1007/s11306-007-0091-1 (visited on 09/02/2013).

[112] Robert S. Schwartz, Georg Lenz, and Louis M. Staudt. "Aggressive Lymphomas". In: *New England Journal of Medicine* 362.15 (Apr. 2010), pp. 1417–1429. DOI: 10.1056/NEJMra0807082. URL: http://www.nejm.org/doi/abs/10.1056/NEJMra0807082 (visited on 02/27/2013).

[113] Anna Scuto, Mark Kirschbaum, Claudia Kowolik, Leo Kretzner, Agnes Juhasz, Peter Atadja, Vinod Pullarkat, Ravi Bhatia, Stephen Forman, Yun Yen, et al. "The novel histone deacetylase inhibitor, LBH589, induces expression of DNA damage response genes and apoptosis in Ph- acute lymphoblastic leukemia cells". eng. In: *Blood* 111.10 (May 2008), pp. 5093–5100. DOI: 10.1182/blood-2007-10-117762.

[114] Laurie H Sehn, Jane Donaldson, Mukesh Chhanabhai, Catherine Fitzgerald, Karamjit Gill, Richard Klasa, Nicol MacPherson, Susan O'Reilly, John J Spinelli, Judy Sutherland, et al. "Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia". eng. In: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 23.22 (Aug. 2005), pp. 5027–5033. DOI: 10.1200/JCO.2005.09.137.

[115] Janet Stavnezer, Jeroen E J Guikema, and Carol E Schrader. "Mechanism and regulation of class switch recombination". eng. In: *Annual review of immunology* 26 (2008), pp. 261–292. DOI: 10.1146/annurev.immunol.26.021607.090248.

[116] L Stimson, V Wood, O Khan, S Fotheringham, and N B La Thangue. "HDAC inhibitor-based therapies and haematological malignancy". eng. In: *Annals of oncology: official journal of the European Society for Medical Oncology / ESMO* 20.8 (Aug. 2009), pp. 1293–1302. DOI: 10.1093/annonc/mdn792.

[117] U. Storb, H. Ming Shen, N. Michael, and N. Kim. "Somatic hypermutation of immunoglobulin and non-immunoglobulin genes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 356.1405 (Jan. 2001),

pp. 13–19. DOI: 10.1098/rstb.2000.0743. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2000.0743 (visited on 08/08/2013).

[118] Lubert Stryer. *Biochemistry*. 7th ed. New York: W.H. Freeman, 2012.

[119] Karsten Suhre. *Genetics meets metabolomics: from experiment to systems biology*. New York: Springer, 2012.

[120] Karsten Suhre, So-Youn Shin, Ann-Kristin Petersen, Robert P. Mohney, David Meredith, Brigitte Wägele, Elisabeth Altmaier, CARDIoGRAM, Panos Deloukas, Jeanette Erdmann, et al. "Human metabolic individuality in biomedical and pharmaceutical research". In: *Nature* 477.7362 (Aug. 2011), pp. 54–60. DOI: 10.1038/nature10354. URL: http://www.nature.com/doifinder/10.1038/nature10354 (visited on 01/27/2013).

[121] Surveillance, Epidemiology and End Results (SEER). *B-cell Non-Hodgkin Lymphoma - SEER Stat Fact Sheets*. URL: http://seer.cancer.gov/statfacts/html/bcellnhl.html (visited on 08/01/2013).

[122] Steven H Swerdlow, International Agency for Research on Cancer, and World Health Organization. *WHO classification of tumours of haematopoietic and lymphoid tissues*. English. Lyon, France: International Agency for Research on Cancer, 2008.

[123] The International Non-Hodgkin's Lymphoma Prognostic Factors Project. "A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project". eng. In: *The New England journal of medicine* 329.14 (Sept. 1993), pp. 987–994. DOI: 10.1056/NEJM199309303291402.

[124] The Non-Hodgkin's Lymphoma Classification Project. "A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma". eng. In: *Blood* 89.11 (June 1997), pp. 3909–3918.

[125] Roman K Thomas, Daniel Re, Jürgen Wolf, and Volker Diehl. "Part I: Hodgkin's lymphoma–molecular biology of Hodgkin and Reed-Sternberg cells". eng. In: *The lancet oncology* 5.1 (Jan. 2004), pp. 11–18.

[126] G J Thorbecke, A R Amin, and V K Tsiagbe. "Biology of germinal centers in lymphoid tissue". eng. In: *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 8.11 (Aug. 1994), pp. 832–840.

[127] Karl J Tronstad, Kjetil Berge, Rolf Kristian Berge, and Øystein Bruserud. "Modified fatty acids and their possible therapeutic targets in malignant diseases". eng. In: *Expert opinion on therapeutic targets* 7.5 (Oct. 2003), pp. 663–677. DOI: 10.1517/14728222.7.5.663.

[128] Nawaporn Vinayavekhin and Alan Saghatelian. "Untargeted metabolomics". eng. In: *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 30 (Apr. 2010), Unit 30.1.1–24. DOI: 10.1002/0471142727.mb3001s90.

[129] Dingzhi Wang and Raymond N. DuBois. "Eicosanoids and cancer". In: *Nature Reviews Cancer* 10.3 (Feb. 2010), pp. 181–193. DOI: 10.1038/nrc2809. URL: http://www.nature.com/doifinder/10.1038/nrc2809 (visited on 09/09/2013).

[130] D J Weisdorf, J W Andersen, J H Glick, and M M Oken. "Survival after relapse of low-grade non-Hodgkin's lymphoma: implications for marrow transplantation". eng. In: *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 10.6 (June 1992), pp. 942–947.

[131] J B Whitfield. "Gamma glutamyl transferase". eng. In: *Critical reviews in clinical laboratory sciences* 38.4 (Aug. 2001), pp. 263–355. DOI: 10.1080/20014091084227.

[132] Samuel A Williams, Lin-Feng Chen, Hakju Kwon, Carmen M Ruiz-Jarabo, Eric Verdin, and Warner C Greene. "NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation". eng. In: *The EMBO journal* 25.1 (Jan. 2006), pp. 139–149. DOI: 10.1038/sj.emboj.7600900.

[133] World Health Organization. *International statistical classification of diseases and related health problems*. 10th revision, 2nd edition. Geneva: World Health Organization, 2004.

[134] G. Wright. "A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma". In: *Proceedings of the National Academy of Sciences* 100.17 (July 2003), pp. 9991–9996. DOI: 10.1073/pnas.1732008100. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1732008100 (visited on 08/08/2013).

[135] Yasukazu Yoshida and Etsuo Niki. "Detection of lipid peroxidation in vivo: total hydroxyoctadecadienoic acid and 7-hydroxycholesterol as oxidative stress marker". eng. In: *Free radical research* 38.8 (Aug. 2004), pp. 787–794.

[136] Ying Zhao, Shaoli Lu, Lipeng Wu, Guolin Chai, Haiying Wang, Yingqi Chen, Jia Sun, Yu Yu, Wen Zhou, Quanhui Zheng, et al. "Acetylation of p53 at lysine 373/382 by the histone deacetylase inhibitor depsipeptide induces expression of p21(Waf1/Cip1)". eng. In: *Molecular and cellular biology* 26.7 (Apr. 2006), pp. 2782–2790. DOI: 10.1128/MCB.26.7.2782-2790.2006.

[137] E. Zucca and F. Cavalli. "Extranodal lymphomas". In: *Annals of Oncology* 11.suppl 3 (Jan. 2000), pp. 219–222. DOI: 10.1093/annonc/11.suppl_3.219. URL: http://annonc.oxfordjournals.org/cgi/doi/10.1093/annonc/11.suppl_3.219 (visited on 08/08/2013).

Software

The following methods, software and software package were used for this thesis.

[138] T. W. Anderson. "R. A. Fisher and multivariate analysis". In: *Statistical Science* 11.1 (Jan. 1996), pp. 20–34. DOI: 10.1214/ss/1032209662. URL: http://projecteuclid.org:80/Dienst/getRecord?id=euclid.ss/1032209662/ (visited on 07/03/2013).

[139] Baptiste Auguie. *gridExtra: functions in Grid graphics*. 2012. URL: http://CRAN.R-project.org/package=gridExtra.

[140] M. S. Bartlett. "Properties of Sufficiency and Statistical Tests". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 160.901 (May 1937), pp. 268–282. DOI: 10.1098/rspa.1937.0109. URL: http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1937.0109 (visited on 08/27/2013).

[141] Henrik Bengtsson. *R.matlab: Read and write of MAT files together with R-to-Matlab connectivity*. 2013. URL: http://CRAN.R-project.org/package=R.matlab.

[142] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995), pp. 289–300. URL: http://www.jstor.org/stable/2346101.

[143] Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Christoph Sieb, Kilian Thiel, and Bernd Wiswedel. "KNIME: The Konstanz Information Miner". In: *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007.

[144] Hanbo Chen. *VennDiagram: Generate high-resolution Venn and Euler plots*. 2013. URL: http://CRAN.R-project.org/package=VennDiagram.

[145] Gabor Csardi and Tamas Nepusz. "The igraph software package for complex network research". In: *InterJournal* Complex Systems (2006), p. 1695. URL: http://igraph.sf.net.

[146] Sebastien Dejean, Ignacio Gonzalez, Kim-Anh Le Cao with contributions from Pierre Monget, Jeff Coquery, FangZou Yao, and Benoit Liquet. *mixOmics: Omics Data Integration Project*. 2013. URL: http://CRAN.R-project.org/package=mixOmics.

[147] G Fasola, R Fanin, F Gherlinzoni, P Galieni, D Taruscio, G Frezza, P Mazza, S Pileri, and M Baccarani. "Serum LDH concentration in non-Hodgkin's lymphomas. Relationship to histologic type, tumor mass, and presentation features". eng. In: *Acta haematologica* 72.4 (1984), pp. 231–238.

[148] R. A. Fisher. "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population". In: *Biometrika* 10.4 (May 1915), p. 507. DOI: 10.2307/2331838. URL: http://www.jstor.org/stable/2331838?origin=crossref (visited on 07/03/2013).

[149] Ronald Aylmer Fisher. "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample". In: *Metron* 1.3-32 (1921), pp. 205–235. URL: http://digital.library.adelaide.edu.au/dspace/handle/2440/15169 (visited on 07/03/2013).

[150] Ronald Aylmer Fisher. *Statistical methods for research workers*. 14th ed., revised and enlarged. Edinburgh: Oliver and Boyd, 1970.

[151] Jerome H. Friedman, Trevor Hastie, and Rob Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (Feb. 2010), 1–22. URL: http://www.jstatsoft.org/v33/i01.

[152] Joseph L. Gastwirth, Yulia R. Gel, W. L. Wallace Hui, Vyacheslav Lyubchich, Weiwen Miao, and Kimihiro Noguchi. *lawstat: An R package for biostatistics, public policy, and law*. 2013. URL: http://CRAN.R-project.org/package=lawstat.

[153] Frank E. Harrell. *Hmisc: Harrell Miscellaneous*. 2013. URL: http://CRAN.R-project.org/package=Hmisc.

[154] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. "Unbiased Recursive Partitioning: A Conditional Inference Framework". In: *Journal of Computational and Graphical Statistics* 15.3 (Sept. 2006), pp. 651–674. DOI: 10.1198/106186006X133933. URL: http://www.tandfonline.com/doi/abs/10.1198/106186006X133933 (visited on 08/14/2013).

[155] H. Levene. "Robust Tests for Equality of Variances". In: *Contributions to Probability and Statistics*. Ed. by I. Olkin. Palo Alto, CA: Stanford University Press, June 1960, pp. 278–292.

[156] Weijun Luo, Michael S Friedman, Kerby Shedden, Kurt D Hankenson, and Peter J Woolf. "GAGE: generally applicable gene set enrichment for pathway analysis". In: *BMC Bioinformatics* 10.1 (2009), p. 161. DOI: 10.1186/1471-2105-10-161. URL: http://www.biomedcentral.com/1471-2105/10/161 (visited on 07/03/2013).

[157] R. Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2012. URL: http://www.R-project.org/.

[158] Revolution Analytics. *doMC: Foreach parallel adaptor for the multicore package*. 2013. URL: http://CRAN.R-project.org/package=doMC.

[159] J. P. Royston. "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples". In: *Applied Statistics* 31.2 (1982), p. 115. DOI: 10.2307/2347973. URL: http://www.jstor.org/stable/10.2307/2347973?origin=crossref (visited on 06/25/2013).

[160] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3-4 (Dec. 1965), pp. 591–611. DOI: 10.1093/biomet/52.3-4.591. URL: http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/52.3-4.591 (visited on 06/25/2013).

[161] Charlie Sharpsteen and Cameron Bracken. *tikzDevice: R Graphics Output in LaTeX Format*. 2012. URL: http://R-Forge.R-project.org/projects/tikzdevice/.

[162] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43 (Oct. 2005), pp. 15545–15550. DOI: 10.1073/pnas.0506580102.

[163] B. L. Welch. "The generalization of Student's problem when several different population variances are involved". In: *Biometrika* 34.1-2 (1947), pp. 28–35. DOI: 10.1093/biomet/34.1-2.28. URL: http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/34.1-2.28 (visited on 06/28/2013).

[164] Wickham and Hadley. "Reshaping data with the reshape package". In: *Journal of Statistical Software* 21.12 (2007). URL: http://www.jstatsoft.org/v21/i12/paper.

[165] Hadley Wickham. *Ggplot2: elegant graphics for data analysis*. Use R! New York: Springer, 2009.

[166] Hadley Wickham. "The Split-Apply-Combine Strategy for Data Analysis". In: *Journal of Statistical Software* 40.1 (2011), 1–29. URL: http://www.jstatsoft.org/v40/i01/.

[167] D. Wu and G. K. Smyth. "Camera: a competitive gene set test accounting for inter-gene correlation". In: *Nucleic Acids Research* 40.17 (May 2012), e133–e133. DOI: 10.1093/nar/gks461. URL: http://www.nar.oxfordjournals.org/cgi/doi/10.1093/nar/gks461 (visited on 07/03/2013).

[168] J. Xia and D. S. Wishart. "MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data". In: *Nucleic Acids Research* 38 (May 2010), W71–W77. DOI: 10.1093/nar/gkq329. URL: http://www.nar.oxfordjournals.org/cgi/doi/10.1093/nar/gkq329 (visited on 07/03/2013).

Acronyms

ABC	Activated B-cell Like (DLBCL subtype)
AID	Activation-induced Cytidine Deaminase (protein)
ASCT	Autologous Stem Cell Transplant (lymphoma treatment)
BCL	B-cell Lymphoma (gene)
BH	Benjamini-Hochberg (method for multiple testing correction)
BLIMP	B-Lymphocyte-Induced Maturation Protein (gene)
CAMERA	Correlation Adjusted Mean Rank Gene Set Test
CHOP	Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone (DLBCL drug)
COO	Cell-Of-Origin
CR	complete response
CREB	cAMP Response Element-Binding Protein (gene)
CREBBP	CREB Binding Protein (gene)
CSR	Class Switch Recombination (aka. isotype switching)
CTCL	Cutaneous T-cell Lymphoma (lymphoma subtype)
CV	cross validation
DFS	disease-free survival
DLBCL	Diffuse Large B-Cell Lymphoma (Lyphoma subtype)
DNA	Desoxyribonucleic Acid
EBV	Epstein-Barr virus
ECOG	Eastern Cooperative Oncology Group
EFS	event-free survival
FDC	Follicular Dentritic cell
FDR	false discovery rate

FL	Follicular Lymphoma (lymphoma subtype)
GAGE	Generally Applicable Gene Set Enrichment for Pathway Analysis
GC	germinal center
GCB	Germinal-Center B-cell Like (DLBCL subtype)
GC/MS	Gas Chromatography with Mass Spectrometry
GGT	Gamma-Glutamyl Transferase (protein)
GSEA	gene set enrichment analysis
GSH	gluthatione
HAT	Histone Acetyltransferase (enzyme)
HCV	Hepatitis-C Virus
HD	Hodgkin's disease (aka Hodkin's lymphoma, lymphoma subtype)
HDAC	Histone Deacetylase (enzyme)
HDACi	Histone Deacetylase Inhibitor
HDC	high dose chemotherapy
HIV	Human Immodeficiency Virus
ICD-10	International Statistical Classification of Diseases and Related Health Problems
Ig	Immunoglobin
IPI	International Prognostic Index
KNIME	Konstanz Information Miner
LASSO	Least Absolute Shrinkage and Selection Operator
LC/MS	Liquid Chromatography with Mass Spectrometry
LOO	leave-one-out
MHC	Major Histocompatibility Complex
MM	Multiple Myeloma (lymphoma subtype)
MS	mass spectrometry
MSEA	Metabolite Set Enrichment Analysis
MTA	5-methylthioadenosine

MYC	V-myc Avian Myelocytomatosis Viral Oncogene Homolog (gene)
NF-κB	Nuclear Factor Kappa B (gene)
NHL	Non-Hodgkin-Lymphoma
NMR	Nuclear Magnetic Resonance Spectrometry
OS	overall survival
PCA	Principal Component Analysis
PD	progressive disease
PMBL	Primary Mediastinal B-Cell Lymphoma (DLBCL subtype)
PR	partial response
PWEA	Pathway Enrichment Analysis
R-CHOP	Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone with Rituximab
REL	Reticuloendotheliosis Viral Oncogene Homolog (gene)
ROS	Reactive Oxygen Species
RSD	Relative Standard Deviation
SAHA	Suberoylanilide Hydroxamic Acid
SAM	S-adenosylmethionine
SHM	Somatic Hypermutation
SNP	Single Nucleotide Polymorphism
STAT	Signal Transducers and Activators of Transcription (group of genes/proteins)
UDP	Uridine-5′-Diphosphate
UMP	Uridine-5′-Monophosphate

List of Figures

1 Introduction 1
1.1 The germinal center reaction 3
1.2 The histology of Diffuse Large B-Cell Lymphoma 6

2 Data 15
2.1 Overview over the study design 17
2.2 Logarithmically transformation of metabolite intensities 21
2.3 Examples for logarithmically transformed metabolite intensities of serum and plasma samples 22
(a) hippurate
(b) phenol sulfate
(c) 1,5-anhydroglucitol
2.4 Removed outliers 23
2.5 Number of samples per group 24
2.6 Age distribution 25
(a) pre- vs. post-pano
(b) resitant vs. sensitive
(c) primary vs. Relapsed
(d) serum vs. plasma
(e) combined
2.7 Sex distribution 26
(a) pre- vs. post-pano
(b) resistant vs. Sensitive
(c) primary vs. Relapsed
(d) serum vs. plasma
(e) combined
2.8 Age distribution stratified by gender 27
(a) males
(b) females
2.9 Run day comparison 30
(a) primary-serum-1 (distributions)

(b) primary-serum-1 (missing values)
(c) pano-plasma-1 (distributions)
(d) pano-plasma-1 (missing values)
2.10 Missing value patterns 31
(a) primary-serum-1 vs. primary-serum-2
(b) pano-plasma-1 vs. pano-plasma-2
(c) primary-serum-1 vs. pano-plasma-1
(d) primary-serum-1 vs. pano-plasma-2
(e) primary-serum-2 vs. pano-plasma-1
(f) primary-serum-2 vs. pano-plasma-2
2.11 Metabolite overlap between studies 32

3 Methods 33
3.1 Screenshot of the KNIME interface 45

4 Results and Discussion 47
4.1 Examples for metabolites with different concentration between pre- and post-pano plasma samples 50
(a) isoleucine
(b) threonine
(c) γ-glutamylisoleucine
(d) γ-glutamylmethionine
(e) pregnen-diol disulfate
(f) 4-androsten-3beta,17beta-diol disulfate
4.2 Metabolites with different variance between pre- and post-pano samples 51
(a) Inosine
(b) 1-oleoylglycerophosphoethanolamine
4.3 Metabolite ratios differing between pre- and post-pano samples 53
(a) (c) / (e)
(b) (d) / (f)
(c) 3-methyl-2-oxovalerate
(d) glycerol 3-phosphate (G3P)
(e) 4-methyl-2-oxopentanoate
(f) ribitol
4.4 Differentialy Correlated Metabolites in pre- and post-pano samples . . 56
(a) caprylate vs N-methyl proline
(b) pyrophosphate (PPi) vs phosphate
4.5 PCA of pre- and post-pano samples 57
(a) PC scatter (w/o missing values)
(b) PC scatter (with missing values)

(c) Variance Explained (w/o missing values)
(d) Variance Explained (with missing values)
4.6 LASSO classification of pre-pano and post-pano samples 58
(a) Performance
(b) Coefficients
4.7 Mapping of results on a metabolic reference network 59
4.8 Volcano plots for pre- vs. post-pano comparison 63
(a) Plasma samples (paired t-test)
(b) Cell samples (regression)
4.9 PCA of cell samples 63
(a) Primary components
(b) Variance explained
4.10 Heatmap of cell samples 66
4.11 Examples for metabolites with different concentration between pre- and post-pano cell samples 67
(a) spermine
(b) eicosenoate (20:1n9 or 11)
4.12 Pathway enrichment of metabolites which differ between pre- and post-pano cells 67
(a) Long Chain Fatty Acids
(b) Monoacylglycerol
4.13 Examples for metabolites with different concentration between primary and relapse tumor samples 78
(a) glycerol 3-phosphate (G3P)
(b) serine
4.14 Examples of metabolite ratios differing between primary and relapse samples 78
(a) fumarate / acetylcarnitine
(b) 13-HODE + 9-HODE /X-17629
4.15 Examples for metabolite pairs with significantly different correlation between primary and relapse samples 81
(a) pantothenate vs. indoleacetate
(b) pelargonate vs. pregn steroid monosulfate
4.16 PCA of primary and relapsed plasma samples 81
(a) Primary components
(b) Variance Explained
4.17 LASSO classification of primary and relapse samples 82
(a) Performance
(b) Coefficients

4.18 Schematic representation of the changes of metabolite concentrations from primary to relapsed tumor until after the treatment 85
4.19 Examples for metabolites with different concentration between R-CHOP-resistant and sensitive DLBCL 90
(a) Xylose
(b) caproate
(c) pyridoxate
(d) isobutyrylcarnitine
4.20 Examples for metabolite ratios with different variances between R-CHOP-resistant and sensitive DLBCL 91
(a) X-04357 / X-16653
(b) 3-hydroxyisobutyrate / N-acetylneuraminate
4.21 Examples for metabolite pairs with significantly different correlation between R-CHOP-resistant and sensitive samples 91
(a) isoleucine vs ADSGEGDFXAEGGGVR
(b) paraxanthine vs 2-palmitoylglycerophosphoethanolamine
4.22 PCA of R-CHOP-resistant and sensitive DLBCL samples 92
(a) Principal components
(b) Variance Explained
4.23 LASSO classification of R-CHOP-resistant and sensitive DLBCL 92
4.24 Examples for metabolites with different concentration between plasma and serum samples 95
(a) 5-oxoproline
(b) arachidonate
4.25 Examples for metabolite ratios differing between plasma and serum samples 97
(a) pelargonate / undecanoate
(b) pelargonate / docosahexaenoate
4.26 Examples for differentialy correlated metabolites in plasma and serum samples 101
(a) beta-hydroxypyruvate vs. 1-oleoylglycerophosphocholine
(b) threonate vs. pregnenolone sulfate
4.27 PCA of plasma and serum samples 101
(a) PC scatter (with missing values)
(b) PC scatter (without missing values)
4.28 LASSO classification of plasma and serum samples 102
(a) Performance
(b) Coefficients

5 Conclusion and Perspectives 107

5.1 Gamma-glutamyl-cycle and signal lipids 108
5.2 Amino acid pathway 109
5.3 Steroid hormone pathway 110

Bibliography 113

A Content of Supplementary DVD 147

List of Tables

1 Introduction 1

2 Data 15
2.1 Experiment Overview 15

3 Methods 33
3.1 Contingency Table 36

4 Results and Discussion 47
4.2 Metabolites with different concentration between pre- and post-pano plasma samples 49
4.3 Metabolites with different variance between pre- and post-pano plasma samples 51
4.4 Metabolite ratios differing between pre- and post-pano samples 52
4.5 Linear regression results of combined metabolites comparing pre- and post-pano Samples 54
4.6 Differentialy Correlated Metabolites in pre- and post-pano samples . . 55
4.7 Metabolites with different concentration between pre- and post-pano cell samples 64
4.8 Metabolite ratios differing between pre- and post-pano cell samples . . 65
4.9 Metabolite combinations differing between pre- and post-pano cell samples 68
4.10 Metabolites with different concentration between primary and relapse tumor samples 77
4.11 Metabolite ratios differing between primary and relapse samples . . . 79
4.12 Metabolite pairs with significantly different correlation between primary and relapse samples 80
4.13 Metabolites with different concentration between R-CHOP resistant and sensitive DLBCL 87
4.14 Metabolite ratios with different variances between R-CHOP-resistant and sensitive DLBCL 88

4.15 Metabolite pairs with significantly different correlation between R-CHOP-resistant and sensitive samples 89
4.16 Metabolites with different concentration between plasma and serum samples 96
4.17 Metabolites with different variances between plasma and serum samples 97
4.18 Metabolite ratios differing between plasma and serum samples 98
4.19 Metabolite ratios with different variances between plasma and serum samples 99
4.20 Differentialy Correlated Metabolites in plasma and serum samples . . 100

5 Conclusion and Perspectives 107

Bibliography 113

A Content of Supplementary DVD 147

Supplements

A Content of Supplementary DVD

Results All produced results with complete p-value lists and plots are contained in the `results` folder. This contains one sub-folder for each of the above described comparisons. The executed KNIME workflow with the according data loaded is contained as `workflow.zip` in each of the results folders. Additionally the results of the assessment of the data described in chapter 2.

KNIME Nodes All methods were implemented as `R` scripts and incorporated as KNIME nodes. The implemented scripts as well as the nodes can be found on the DVD in the `KNIME nodes (source)` folder. Additionally the nodes are available as `jar`-package for the use it with KNIME (contained in the `KNIME nodes (jar)` folder).

KNIME Workflow The workflow which was designed and used for this thesis is contained on the supplemental DVD in the folder `KNIME workflow`

Metabolites and Pathways The list of measure metabolites and their pathway annotation can be found in the file `metabolites.csv`.