

Aus dem
CharitéCentrum für diagnostische und präventive Labormedizin (CC 5)
Institut für Pathologie
Direktor: Prof. Dr. Manfred Dietel

Habilitationsschrift

Metabolomanalyse solider Tumore: Beiträge zur Bioinformatik und translationalen Tumorforschung

zur Erlangung der Lehrbefähigung für das Fach Experimentelle Pathologie

vorgelegt dem Fakultätsrat der Medizinischen Fakultät der
Charité-Universitätsmedizin Berlin

von

**Dr. rer. nat. Jan Budczies
aus Berlin**

Eingereicht:	April 2015
Dekan:	Prof. Dr. Axel R. Pries
1. Gutachter:	Prof. Dr. Holger Moch
2. Gutachter:	Prof. Dr. Rainer Spang

Meinen Eltern
Gerda Budczies und Arnold Budczies

Inhaltsverzeichnis

Abkürzungen und Akronyme	7
1 Einleitung	9
1.1 Krebs als Stoffwechselerkrankung	9
1.1.1 Anaerobe Glykolyse (Warburg-Effekt)	9
1.1.2 Veränderungen im Glutamin-Stoffwechsel	10
1.1.3 Veränderungen im Fettstoffwechsel	11
1.2 Metabolomanalyse	13
1.3 Bioinformatik	14
1.3.1 Differentiell veränderte Metaboliten	15
1.3.2 Unüberwachte Mustererkennung	15
1.3.3 Überwachte Mustererkennung	16
1.3.4 Funktionelle Analyse	17
2 Zielsetzung	19
3 Ergebnisse	21
3.1 Metabolomanalyse des Ovarialkarzinoms	21
3.2 Metabolomanalyse des kolorektalen Karzinoms	22
3.3 METATarget: Detektion von Schlüsselstellen der Stoffwechselregulation aus Metabolomdaten	23
3.4 Metabolomanalyse des Mammakarzinoms: Tumor- versus Normalgewebe .	24
3.5 Metabolomanalyse des Mammakarzinoms: ER-negative versus ER-positive Tumore	26
3.6 Metabolomanalyse des Mammakarzinoms: Glutamat-Anreicherung als neu- er diagnostischer Parameter	27
4 Diskussion	29
4.1 Diagnostische, prognostische und prädiktive Marker	29
4.2 Bioinformatische Methoden	31
4.2.1 Differentiell veränderte Metaboliten	31
4.2.2 Unüberwachte Mustererkennung	33
4.2.3 Überwachte Mustererkennung	33
4.2.4 Funktionelle Analyse	34
4.3 GC-TOFMS-Analyse verschiedener Karzinome	35
4.4 2-Hydroxyglutarat (2-HG) als Onkometabolit	38
4.5 Ausblick	39
5 Zusammenfassung	41
Literatur	43
Danksagung	53
Zugrunde liegende Originalarbeiten	55
Erklärung	57

Abkürzungen und Akronyme

2-HG	2-Hydroxyglutarat
ABAT	4-Aminobutyrat-Aminotransferase, auch GABA-Transaminase genannt
ARROWS	Verbundprojekt Juli 2010 - Dezember 2013, Förderung im 7. EU-Forschungsrahmenprogramm (http://www.arrows-online.eu)
cancerclass	R-Paket zur Entwicklung und Validierung diagnostischer Tests (http://bioconductor.org/packages/release/bioc/html/cancerclass.html)
Cx:y	Notation für freie Fettsäuren: x = Zahl der Kohlenstoffatome, y = Zahl der Doppelbindungen
DASL	<i>cDNA-mediated annealing, selection, extension, and ligation assay</i> (Methode zur genomweiten Expressionsanalyse)
ER	Östrogenrezeptor
FDR	<i>False discovery rate</i>
FWER	<i>Familywise error rate</i>
GABA	Gamma-Aminobuttersäure
GEO	Gene Expression Omnibus = öffentliche Datenbank für Genexpressionsdaten (http://www.ncbi.nlm.nih.gov/geo)
GGR	<i>Glutamate/glutamine ratio</i>
GLS	Glutaminase 1
GLS2	Glutaminase 2
GLUL	Glutamat-Ammonium-Ligase
GC-TOFMS	<i>Gas chromatography combined with time-of-flight mass spectrometry</i> (Gaschromatographie kombiniert mit Flugzeit-Massenspektrometrie)
HER2	<i>Human epidermal growth factor receptor 2</i>
HMDB	<i>The Human Metabolome Database</i> (http://www.hmdb.ca)
HumanCyc	<i>Encyclopedia of Human Genes and Metabolism</i> (http://humancyc.org)
IDH1	Isocitrat-Dehydrogenase 1
IDH2	Isocitrat-Dehydrogenase 2
KEGG	<i>Kyoto encyclopedia of genes and genomes</i> (http://www.genome.jp/kegg)
LC-MS	<i>Liquid chromatography combined with mass spectrometry</i> (Flüssigkeitschromatographie kombiniert mit Massenspektrometrie)
METAcancer	Verbundprojekt April 2008 - März 2011, Förderung im 7. EU-Forschungsrahmenprogramm (http://www.metacancer-fp7.eu)

METAtarget	Methode zur Detektion signifikanter Änderungen des Substrat-Produkt-Gleichgewichts in Metabolomik-Datensätzen
NMR	<i>Nuclear magnetic resonance</i>
PROFILE-Clustering	<i>Projection from interaction lattice clustering</i> (Methode zur Analyse von Metabolitenänderungen im funktionellen Kontext)
R	Softwareumgebung für statistisches Rechnen und Visualisierung (http://www.r-project.org)
Reactome	<i>A curated pathway database</i> (http://www.reactome.org)
ROC	<i>Receiver operating characteristics</i>
TCA	<i>Tricarboxylic acid</i>

1 Einleitung

Das Metabolom ist die Gesamtheit der kleinen chemischen Moleküle in einer Zelle, in einem Gewebe oder in einem Organismus. Darunter fallen sehr unterschiedliche Moleküle mit einer großen Vielfalt an räumlichen chemischen Strukturen, wie z.B. Aminosäuren, Fettsäuren, Nukleotide, Zucker und viele andere. Metabolomik ist die Technologie und Wissenschaft zum Studium dieser Moleküle. Die hier vorgelegten sechs Studien beschäftigen sich mit der Analyse des Metaboloms in Krebsgeweben und der Entwicklung bioinformatischer Analysemethoden. Ziel der Analysen ist einerseits die Findung molekularer Zielstrukturen für eine therapeutische Intervention und andererseits die Entwicklung diagnostischer, prognostischer und prädiktiver Biomarker.

1.1 Krebs als Stoffwechselerkrankung

Veränderungen im Zellstoffwechsel gehören zu den typischen Veränderungen, die Krebszellen im Vergleich zu gesunden Normalzellen charakterisieren und die zum malignen Wachstum beitragen [1, 2, 3, 4]. In dem 2011er Update des viel rezipierten Artikels *hallmarks of cancer* wurden Veränderungen im Energiestoffwechsel als *emerging hallmark* zu den typischen Charakteristiken der Krebszelle hinzugefügt [5]. Im heute allgemein akzeptierten Modell der Krebsentstehung bilden Veränderungen im Erbgut der Körperzellen die Ursache für die maligne Transformation. Zu diesen Veränderungen gehören einerseits Alterationen der DNA-Sequenz, wie Punktmutationen, Deletionen, Insertationen und Kopienzahlvariationen (*copy number variations*), andererseits durch epigenetische Veränderungen wie zum Beispiel Hypo- oder Hypermethylierung der DNA. Während noch diskutiert wird, inwieweit die Veränderungen im Stoffwechsel eine Ursache oder Folge der malignen Transformation sind, ist klar, dass Veränderungen des Zellmetabolismus ein gemeinsames Charakteristikum vieler maligner Tumorzellen sind [5]. Die Veränderungen im Stoffwechsel der Krebszelle im Vergleich zur normalen Körperzelle umfassen u.a. Veränderungen im Glukose-Stoffwechsel (Warburg-Effekt), Veränderungen im Glutamine-Stoffwechsel, Veränderungen im Fettstoffwechsel.

1.1.1 Anaerobe Glykolyse (Warburg-Effekt)

Eukaryotische Zellen verfügen über zwei verschiedene Möglichkeiten, um aus aufgenommener Glukose Energie zu gewinnen [6]: Einerseits kann das durch Glykolyse aus Glukose entstandene Pyruvat im Zytosol weiter zu Lactat abgebaut werden. Andererseits kann das Pyruvat im Mitochondrium zu Acetyl-CoA umgewandelt und im TCA-Zyklus weiter metabolisiert werden. Der Energiegewinn erfolgt dann zum größten Teil durch die Wiederoxidation des bei diesem Prozess entstandenen NADH und FADH in der Atmungskette (oxidative Phosphorylierung). In der bahnbrechenden Arbeit „Über den Stoffwechsel der Karzinomzelle“ (erschienen 1924) stellte Otto Warburg eine Hochregulation der Glykolyse im Verhältnis zur oxidativen Phosphorylierung in malignen Tumorgeweben im Vergleich zu Normalgeweben und benignen Tumoren fest [7]. Lange vorher, von Luis Pasteur bereits 1857 beim Studien an Hefen entdeckt, war bereits bekannt, dass Zellen unter anaeroben Bedingungen mehr und schneller Glukose verbrauchen als unter aeroben Bedingungen und dass die Zufuhr von mehr Sauerstoff zu einer Hemmung der Glykolyse führt (Pasteur-Effekt). Ein wichtiger Aspekt von Otto Warburgs Entdeckung

ist eine starke Nutzung der Glykolyse in der Krebszelle auch unter aeroben Bedingungen (aerobe Glykolyse).

Die von Warburg aufgestellte Hypothese, dass dieser Effekt auf einer Beschädigung der Mitochondrien und einer nicht-funktionalen Atmungskette in den Krebszellen beruht, hat sich mittlerweile als jedenfalls nicht allgemein richtig herausgestellt [8]. Stattdessen liegt in den Krebszellen oft eine veränderte Regulation der Glykolyse vor. Eine Schlüsselrolle spielt dabei möglicherweise der Transkriptionsfaktor HIF1 α (*hypoxia-inducible factor 1 α*), der sowohl die Glukosetransporter als auch viele der an der Glykolyse beteiligten Enzyme reguliert [9]. HIF1 α kann einerseits durch Hypoxie, andererseits aber auch durch an der malignen Transformation beteiligte Onkogene aktiviert werden. Die anaerobe Glykolyse ist eine im Vergleich zu oxidativen Phosphorylierung ineffiziente Art und Weise der zellulären Energieproduktion (es entstehen nur 2 ATPs im Vergleich zu über 30 ATPs). Ein möglicher Vorteil für Krebszellen, dennoch diesen Pathway zu benutzen, liegt darin, dass er es erleichtert, Nahrungsstoffe (wie Glukose und Glutamin) aufzunehmen und als molekulare Bausteine für den Aufbau neuer Biomasse (zum Beispiel Nukleotide, Aminosäuren und Lipide) zu verwenden [3]. Zusätzlich ist spekuliert worden, dass die Erniedrigung des pH-Wert in der Microumgebung des Tumors immunsuppressiv wirkt und die Invasivität des Tumors begünstigt [1].

Zur Erreichung der hohen Glykolyseraten in malignen Tumoren ist eine starke Glukoseaufnahme durch die Krebszellen nötig. Insgesamt sind 14 Glukose-Transmembrantransporter (GLUT) beschrieben worden, von denen verschiedene eine besondere Rolle in Krebszellen spielen. Insbesondere ist der Transporter GLUT1 in vielen Krebsarten heraufreguliert [10]. Die vermehrte Aufnahme von Glukose durch Krebszellen wird in dem bildgebenden Verfahren der ^{18}F -*fluorodeoxyglucose position emission tomography* (FDG-PET) genutzt. Die FDG-PET beruht darauf, dass ein radioaktiv markiertes Glukose-Analog zwar von den Zellen über die Glukosetransporter aufgenommen wird und durch Phosphorylierung in diesen gefangen wird, aber wegen des Fehlens eines Sauerstoffatoms an der C2-Position nicht weiter metabolisiert werden kann [11]. Eine großen Metanalyse über die Verwendung der FDG-PET zur Detektion von soliden Tumoren ergab eine Sensitivität von 84% bei einer Spezifität von 86% [12].

1.1.2 Veränderungen im Glutamin-Stoffwechsel

Glutamin ist die am höchsten konzentrierte Aminosäure im menschlichen Blut. Zusätzlich zum Warburgeffekt findet man in vielen Krebszellen Veränderungen in der Aufnahme und Metabolisierung von Glutamin [13, 14, 15]. An Stelle von Glukose kann Glutamin verwendet werden, um den TCA-Zyklus wieder aufzufüllen und molekulare Bausteine (Kohlenstoff und Stickstoff) für die Synthese von Nukleotiden, Aminosäuren und Lipiden zu liefern. Dazu wird Glutamin zunächst in Glutamat und dann in Alpha-Ketoglutarat umgewandelt, das im TCA-Zyklus weiter metabolisiert werden kann. Dieser Prozess wird analog zur Glykolyse als Glutaminolyse bezeichnet. Eine weitere wichtige Funktion von Glutamin in proliferierenden Zellen ist die Produktion von Energie. In der Tat war in immortalisierten wie auch in malignen Zellen Glutaminolyse und nachfolgende oxidative Phosphorylierung die Hauptquelle von ATP [16].

Glutamin ist eine für den Menschen nicht-essentielle Aminosäure, die aus dem im TCA-Zyklus vorhandenen Alpha-Ketoglutarat über Glutamat synthetisiert werden kann. Im Gegensatz dazu ist schon lange bekannt, dass Glutamine für die Kultur einiger Krebs-

zellenlinien essential ist [17]. In Zellkulturmedien ist es üblicherweise die abundanteste Aminosäure mit einer mindestens zehnfach so hohen Konzentrationen wie jeder der anderen Aminosäuren. Neuere Studien zeigen, dass Glutamin zur Invasivität des Ovarialkarzinoms beitragen kann [18] und dass Glutamin das Wachstum des Pankreaskarzinoms über einen KRAS-regulierten Stoffwechselweg unterstützt [19].

Der erste Schritt der Verwendung von Glutamin besteht in seiner Umwandlung in Glutamat durch eine Glutaminase. Glutaminasen kommen in humanen Zellen in Form der beiden Isoenzyme Glutaminase 1 (GLS) und Glutaminase 2 (GLS2) vor. Durch die aktuelle Entdeckung von drei neuen Glutaminase-Inhibitoren (BPTES, 968 und CB-839), die das GLS-Enzym allosterisch blockieren, hat sich ein neues therapeutisches Fenster zur zielgerichteten Blockade des Glutamin-Stoffwechsels geöffnet [20, 21, 22]. CB-839 wird seit Anfang 2014 in Phase-I klinischen Studien zu soliden Tumoren und Leukämien getestet.

1.1.3 Veränderungen im Fettstoffwechsel

Die meisten Körperzellen decken ihren Bedarf an Lipiden durch Aufnahme aus dem Blutstrom, während die Neusynthese von Lipiden vor allem in der Leber, im Fettgewebe und der laktierenden Brust erfolgt [23]. Lipide werden u.a. benötigt zum Aufbau von Phosphoglyceriden, die den Hauptbestandteil der Zellmembranen bilden. Außerdem können Lipide als Triacylglyceride zum Speichern von Energie verwendet werden. Eine dritte Funktion besteht in der Verwendung von Lipiden als Signalmolekülen.

Wie normale Körperzellen können Krebszellen Lipide über die Blutversorgung aufnehmen. Seit über 50 Jahren ist jedoch bekannt, dass Krebszellen auch in der Lage sind Fettsäuren selbst zu synthetisieren [24]. In den letzten Jahren ist bei vielen Krebsarten eine Hochregulierung von an der Lipogenese beteiligten Proteinen gefunden worden [23]. Zur Bildung langkettiger Fettsäuren wird zunächst im Zytosol durch Aktivität von der *ATP citrate lyase* (ACLY) Citrat in Acetyl-CoA metabolisiert. Als zweiter Schritt wird Acetyl-CoA durch eine *CoA carboxylase* (ACACA) zu Malonyl-CoA carboxyliert. Der Schlüsselschritt zur Synthese von Lipiden besteht schließlich in der Kondensation von Acetyl-CoA und Malonyl-CoA zur 16-Kohlenstoff-Fettsäure Palmitinsäure durch das Protein *fatty acid synthase* (FASN). In vielen Karzinomen und Sarkomen ist eine hohe Expression der FASN gefunden worden, so z.B. im Prostatakarzinom [25], im kolorektalen Karzinom [26] und in vielen anderen. Sowohl Genexpression als auch Proteinexpression waren überexprimiert im invasiven Mammakarzinom sowie in nicht-invasiven Vorläuferläsionen [27]. In Zellkultursystemen und Tiermodellen wurde gezeigt, dass die Hemmung von ACLY das Tumorwachstum unterdrücken kann [28].

Endogene Lipogenese kann in folgender Weise zur Malignität proliferierender Zellen beitragen [23]: 1. Es werden genügend Phosphoglyceride zur Synthese neuer Membranen bereitgestellt. 2. Die neu synthetisierten Membranen haben einen hohen Anteil an gesättigten Fettsäuren und sind damit unempfindlicher gegen Oxidation. 3. Aus den Lipiden können Triacylglyceride synthetisiert werden, die als Energievorrat in Stresssituationen dienen können. 4. Auf- oder Abbau der langkettigen Fettsäuren können zur Regulierung der Redoxbalance der Zelle verwendet werden. 5. Durch Lipide vermittelte Signalwege können hochreguliert werden.

Zusätzlich zur Hochregulationen der Lipogenese findet sich in Krebsgeweben oft eine Hochregulation des Cholinmetabolismus [29]. Bereits in den 1980er-Jahren fiel in

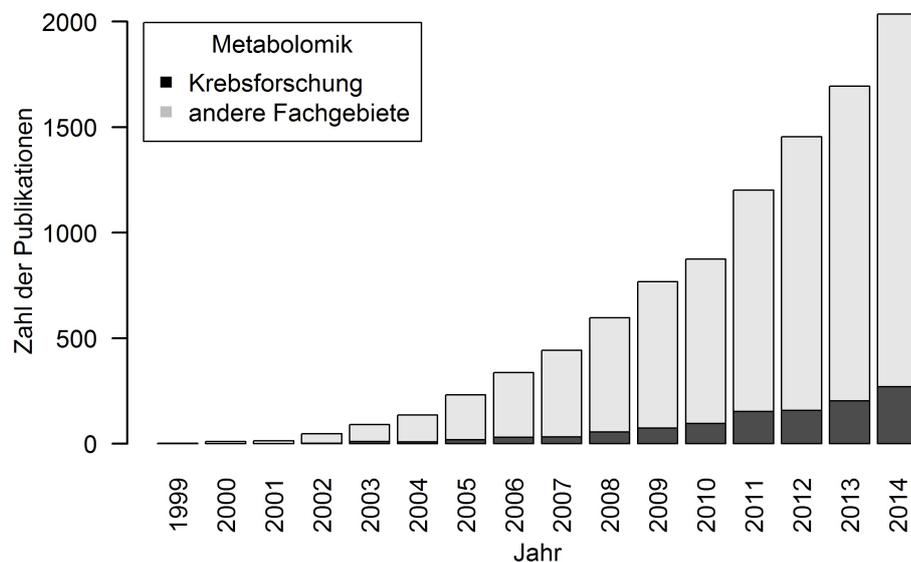


Abbildung 1: Publikationsaktivität in der Metabolomforschung von 1999 bis 2014. Anzahl der in PubMed gelisteten Publikationen, die im entsprechenden Jahr veröffentlicht wurden (Stand: 5.1.2015). Suchbegriffe (Titel oder Abstract): *metabolomic** OR *lipidomic** OR *metabonomic**, bzw. (*metabolomic** OR *lipidomic** OR *metabonomic**) AND (*cancer** OR *tumor** OR *carcinoma** OR *sarcoma** OR *leukemia**).

NMR-Studien an humanen Tumoren ein „Cholin-Peak“ auf und es stellte sich heraus, dass in vielen Tumoren die Gesamtkonzentration cholinhaltiger Metaboliten und insbesondere die Konzentration von Phosphocholin erhöht sind [30, 31, 32]. Später wurde gezeigt, dass die maligne Transformation und nicht alleine eine starke Proliferation Ursache des heraufregulierten Cholinmetabolismus ist [33]. Zur Bildung der Membranbestandteile Phosphatidylcholine und Phosphatidylethanolamine müssen Choline und Ethanolamine phosphoryliert und durch CDP aktiviert werden. Der erste dieser beiden Schritte wird durch die Cholinkinasen (CHKA und CHKB) katalysiert. Der zweite Schritt wird durch die CTP:Phosphocholin-Cytidylyltransferase (CCT) katalysiert. Sowohl CHKA als auch CCT in sind vielen in humanen Tumoren hochexprimiert oder aktiviert [29].

Aus epidemiologischen Studien ist bekannt, dass ein hoher Body-Mass-Index (BMI) mit einem höheren Risiko für verschiedene Krebsarten assoziiert ist. Es wird geschätzt, daß Fettleibigkeit ursächlich für etwa 20% der Mammakarzinome, etwa 20%-30% der kolorektalen Karzinome und etwa 50% der Endometriumkarzinome und Ösophaguskarzinome ist [34]. Ein interessantes Wechselspiel zwischen erhöhtem BMI und FASN-Expression ist im kolorektalem Karzinom gefunden worden: In der *nurses health study* war ein hoher BMI assoziiert mit erhöhtem Risiko für FASN-negative kolorektale Karzinome, aber unverändertem Risiko für FASN-positive Karzinome [35].

1.2 Metabolomanalyse

Metabolomik – in Analogie zu den anderen -omik-Wissenschaften Genomik, Transkriptomik und Proteomik – ist definiert als die Analyse der Gesamtheit der Metaboliten einer biologischen Probe. Unter Metaboliten werden dabei kleine Moleküle (molekulare Masse < 2000 Dalton) verstanden, die eine große Vielfalt an chemischen Strukturen und biologischen Funktionen aufweisen [36]. Die Begriffe Metabolom und Metabolomik (anfangs auch als Metabonomik bezeichnet) tauchen Ende der 1990er- bzw. Anfang der 2000er-Jahre erstmals in der biomedizinischen Literatur auf [37, 38]. Wie in Sumner et al. rezensiert [39], ist das Konzept der Metabolitenprofilierung (*metabolite profiling*) aber bereits in Arbeiten aus den 1970er-Jahren zu finden: Bis in die späten 1990er-Jahre gab es in diesem Fachgebiet eine niedrige weltweite Aktivität mit etwa 5-15 Publikationen jährlich. Danach setzte ein kontinuierliches Wachstum der Metabolomik-Forschung ein, das bis heute andauert und sich in mittlerweile mehr 2000 jährlich erscheinenden Publikationen manifestiert (Abbildung 1). Am Anfang dieser Entwicklung standen Anwendungen der Metabolomik zur funktionellen Charakterisierung von mutierten Hefezellen [40] und in der Pflanzenbiologie [38]. Später folgten erste Krebsstudien, zunächst in Modellsystemen [41, 42], dann an humanen Körperflüssigkeiten [43, 44] und an humanen Krebsgeweben, wie z.B. in unserer GC-TOFMS-Studie zum Ovarialkarzinom [45].

Im Jahr 2005 wurde in Kanada das *human metabolome project* (hmp) initiiert, das zum Ziel hat, alle Metaboliten, die in humanen Geweben und Körperflüssigkeiten vorkommen, zu identifizieren und zu katalogisieren. In der Projektdatenbank HMDB 3.0 sind bereits über 40.0000 Metaboliten erfasst, darunter 20.000 detektierte Metaboliten des humanen Metaboloms und über 19.000 weitere Metaboliten, von denen erwartet wird, dass sie Bestandteil des humanen Metaboloms sind [36]. Heute ist es in einer typischen Metabolomikanalyse möglich, 50 bis 1000 verschiedene Metaboliten in einer Probe zu messen und auf definierte chemische Strukturen abzubilden. Anders als bei Genexpressionsanalysen mittels RNAseq oder Microarrays, bei denen alle oder fast alle der exprimierten RNAs erfasst werden, bilden Metabolomikanalysen nur einen kleinen Teil der in der untersuchten Probe tatsächlich vorhandenen Metaboliten ab. Der Großteil der Komplexität der Metabolomik liegt in der riesigen Vielfalt der räumlichen chemischen Strukturen der Metaboliten begründet, im Gegensatz zur linearen Struktur von DNA-, RNA- und Aminosäuresequenzen.

Für Metabolomik-Analysen gibt es eine Vielzahl von verschiedenen Verfahren [46]. Die am häufigsten verwendeten Verfahren sind Kernspinresonanzspektroskopie (NMR) und Verfahren, die Massenspektroskopie (MS) mit chromatographischen Methoden kombinieren. NMR hat gegenüber massenspektrometrischen Verfahren den Vorteil die Proben unzerstört zu lassen und kann unter Umständen auch *in vivo* angewendet werden. Der Vorteil der massenspektrometrischen Verfahren liegt in einer höheren Sensitivität im Vergleich zur NMR. Die beiden am häufigsten mit der Massenspektrometrie kombinierten Verfahren sind die Gaschromatographie (GC) und die Flüssigkeitschromatographie (LC). Die größten kommerziellen und öffentlich verfügbaren Bibliotheken von Metabolitenspektren existieren derzeit für GC-MS-Daten. Bei der Analyse von biologischen Proben wie Geweben und Körperflüssigkeiten müssen die Metaboliten für die Untersuchung mit GC-MS zunächst durch Derivatisierung und/oder Temperaturerhöhung in die Gasphase gebracht werden. Daher können mit der GC-MS nur Substanzen mit einer relativ geringen Molekülmasse (< 1000 Dalton) und hinreichender thermischer Stabilität

untersucht werden. Der heutige Stand der Technik erlaubt mittels GC-MS typischerweise 100-200 Metaboliten in einer biologischen Probe zu detektieren und auf definierte chemische Strukturen abzubilden. Mit GC-MS detektierbar sind u.a.: Glukose und andere Zucker, fast alle proteinogenen Aminosäuren und weitere Aminosäuren, einige Nukleotide und einige ihrer Phosphate, zahlreiche freie Fettsäuren und einige Glycerophospholipide. Die LC-MS-Technik sowie die Bibliotheken für die entsprechenden Spektren haben sich in den letzten Jahren stark weiterentwickelt. Kürzlich wurde eine Plattform mit zwei separaten Injektionen vorgestellt, die die Detektion von 339 Metaboliten erlaubte [47]. Zur Abdeckung einer noch höheren Zahl von Metaboliten wurden in letzter Zeit öfter GC-MS und LC-MS miteinander kombiniert (getrennte Analyse zweier Aliquots und anschließende Kombination der Daten), so z.B. in [48] und [49].

Für die hier vorgelegten Studien wurde das Verfahren der Flugzeitmassenspektrometrie (time-of-flight mass spectrometry, TOFMS) kombiniert mit Gaschromatographie (gas chromatography, GC) verwendet. Es wurden Gefriergewebeproben verwendet, die an der Charité aufgearbeitet und durch einen Facharzt für Pathologie qualitätskontrolliert wurden. Zum besseren Aufschluss für die nachfolgenden Analysen wurden die Proben als $5\mu\text{m}$ - oder $10\mu\text{m}$ -Schnitte vorbereitet. Für alle Studien liegen positive Ethikvoten der Charité-Ethikkommission vor. Details dazu finden sich in den Originalpublikationen. Die GC-MSTOF-Analysen wurden unter Leitung von Prof. Oliver Fiehn am Max-Planck-Institut für Molekulare Pflanzenphysiologie in Potsdam Golm und am West Coast Metabolomics Center in Davis, CA, USA durchgeführt.

1.3 Bioinformatik

Die bioinformatische Analyse von Metabolomikdaten läßt sich in zwei voneinander unabhängige Schritte einteilen: Der erste Schritt besteht darin, aus den gemessenen NMR- oder Massenspektren quantitative Messungen von Metaboliten zu gewinnen. Der zweite Schritt besteht darin, den so entstandenen Metaboliten-Datensatz im Hinblick auf die biomedizinische Fragestellung zu untersuchen. Für die hier vorgestellten Studien wurde der erste bioinformatische Schritt direkt im Anschluss an die experimentelle Arbeit durch das Fiehn Lab durchgeführt. Die Detektion von Metaboliten in den GC-TOFMS-Spektren beruht dabei auf dem Abgleich mit Datenbanken, in denen Retentionszeiten und Massenspektren von Referenzproben der entsprechenden Metaboliten abgelegt sind.

Der zweite Schritt der bioinformatischen Analyse zielt auf die Aufdeckung neuer Zielstrukturen für eine therapeutische Intervention und die Detektion neuer diagnostischer, prognostischer und prädiktiver Biomarker. Dabei kommen folgende Methoden zum Einsatz: 1. Metabolitenweise Korrelation des Datensatzes mit klinisch-pathologischen Parametern oder anderen -omik-Daten, 2. Unüberwachtes Lernen zur Identifikation krankheitsrelevanter Metabolitenmuster, 3. Überwachtes Lernen zur Identifikation krankheitsrelevanter Metabolitenmuster und 4. Analyse der Metabolitendaten im Kontext der biochemischen Stoffwechselwege und im Kontext medizinischen Wissens. Während für die ersten drei Punkte dieselben Methoden wie bei der Analyse anderer -omik-Daten zum Einsatz kommen, benötigt man für den vierten Punkt auf Metabolikdaten spezialisierte Methoden.

1.3.1 Differentiell veränderte Metaboliten

Wie Genexpressionsdaten sind logarithmierte Metabolomikdaten oft in guter Näherung normalverteilt. Für die Korrelation mit klinisch-pathologischen Parametern können in diesem Fall parametrische Tests verwendet werden, z.B. der t-Test für den Vergleich zwischen zwei Gruppen oder der F-test für den Vergleich von mehr als zwei Gruppen. Nicht parametrische Tests (z.B. Wilcoxon-Ranksummen-Test oder Kruskal-Wallis-Test) sind robust gegen Abweichungen von der Normalverteilung, besitzen aber geringere Macht, falls die Normalverteilungshypothese erfüllt ist. Nachdem die Signifikanz jedes Metaboliten bewertet ist, ist es wichtig der Situation des multiplen Testens Rechnung zu tragen und die p-Werte entsprechend zu korrigieren. Die stringenteste Methode dazu ist die Bonferroni-Korrektur, die eine Kontrolle der *familywise error rate* (FWER) erlaubt. In genomweiten Genexpressionsdatensätzen, die Zehntausende von Genen umfassen, ist die Bonferroni-Methode fast immer zu streng und nicht erfolgversprechend. Die Situation ist etwas anders für Metabolomikdatensätze mit typischerweise einigen hundert Metaboliten, aber auch hier ist eine Bonferroni-Korrektur oft zu streng. Als Alternative greift man daher auf Methoden zur Kontrolle der *false discovery rate* (FDR) zurück. Hier können beispielsweise die Benjamini-Yekutieli oder die Benjamini-Hochberg-Prozedur zum Einsatz kommen [50, 51]. Eine weitere Möglichkeit besteht darin, eine p-Wert-Schwelle festzulegen und die entsprechende FDR mit einer Permutationsanalyse abzuschätzen. Dieses Verfahren haben wir für Genexpressionsdaten mehrfach erfolgreich angewendet [52, 53].

1.3.2 Unüberwachte Mustererkennung

Unüberwachtes Maschinenlernen dient der Entdeckung von Mustern in molekularen Daten, ohne dass ein Zielparameter (wie zum Beispiel das klinische Ergebnis oder die Zugehörigkeit zu einer diagnostischen Entität) in die Analyse miteinbezogen wird. Seit der ersten Anwendung von Heatmaps in der Molekularbiologie durch Eisen et al. (Eisen1998) haben sich diese zur wichtigsten und weitverbreitetsten Visualisierungs- und Clusteringmethode von -omik-Daten entwickelt. In einer Heatmap wird einerseits die Abundanz jedes Moleküls in jeder Probe durch die Farbe eines Pixels dargestellt, andererseits werden sowohl Moleküle als auch Proben bezüglich der Ähnlichkeit der molekularen Profile gruppiert. Die Gruppierung der Moleküle und der Proben erfolgt jeweils durch ein hierarchisches Clustering.

In wegweisenden Arbeiten von Sørlie, Perou und Koautoren wurden Heatmaps und hierarchisches Clustering zur Analyse der Genexpression des Mammakarzinoms verwendet [54, 55, 56]. Ergebnis war die Einteilung des Mammakarzinoms in molekulare fünf Subtypen: Luminal A, Luminal B, ERBB2+, Basal-like und Normal-Breast-like. Mit Hilfe von Überlebensanalysen wurde im Trainingsdatensatz und später in unabhängigen Validierungsdatensätzen gezeigt, dass diese fünf molekularen Subtypen ein unterschiedliches klinisches Outcome besitzen.

Die Methode der Heatmaps kann analog zur Analyse von Genexpressionsdaten zur Analyse von anderen -omik-Daten inklusive Proteomik- und Metabolomikdaten verwendet werden. Vor der Analyse werden die -omik-Daten üblicherweise auf logarithmische Skala gebracht und die Werte jedes Moleküls auf den Mittelwert Null skaliert. Für die Berechnung des Abstandes zwischen zwei molekularen Profilen wird oft die euklidische Distanz oder Eins minus Pearson-Korrelation verwendet. Zur Berechnung des Abstandes zwischen zwei Clustern ist die Verwendung der *average linkage* Methode vorteilhaft,

da diese im Vergleich zur *single linkage* oder *complete linkage* Methode eine größere Robustheit und Reproduzierbarkeit besitzt.

1.3.3 Überwachte Mustererkennung

Allgemein lässt sich die Problemstellung beim überwachten Lernen wie folgt formulieren: Für jeden Patienten wird eine Zahl von Inputvariablen gemessen, z.B. Genexpression-, Protein oder Metabolitenmessungen im Rahmen einer -omik-Studie. Ziel ist es nun, die Inputvariablen dazu zu verwenden, eine Outputvariable zu prädictieren. Bei der Outputvariable kann es sich um eine Diagnose (z.B. Malignität vs. Benignität), das klinische Outcome (z.B. Gesamtüberleben) oder das Ansprechen auf eine Therapie handeln. Synonym zur Bezeichnung „überwachtes Lernen“ wird auch der Terminus „Klassifikation“ benutzt.

Aus dem Bereich des Maschinenlernens gibt es eine Vielzahl von Klassifikationsmethoden, von denen viele bereits auf molekulare Daten angewendet wurden. Einige Beispiele sind einfache Methoden wie *nearest-centroid classification*, lineare Diskriminanzanalyse, Klassifikationsbäume oder komplexere Methoden wie *support vector machines*, neuronale Netzwerke und *random forests*. Für eine Anwendung auf hoch-dimensionale Daten (z.B. auf Tausende von Genen oder Hunderte von Metaboliten) werden diese Methoden oft mit einer *feature selection* kombiniert, d.h. der Datensatz wird z.B. durch eine Filtration auf die wichtigsten informationstragenden Gene bzw. Metaboliten reduziert, bevor der eigentliche Klassifikationsalgorithmus angewendet wird. Bei der Analyse von Microarraydaten zeigte sich, dass oft einfache Methoden komplizierteren Methoden ebenbürtig oder sogar überlegen sind [57, 58]. Ein Ergebnis der MicroArray-Quality-Control (MAQC)-II-Studie war, dass für eine gute Modellperformanz die Erfahrung des Forschungsteams wichtiger war als die Wahl der Klassifikationsmethode [59]. Der kritische Punkt für eine korrekte Interpretation einer Klassifikationsstudie ist eine angemessene Methodik für Fehlerabschätzung und Validierung [60].

Als Folge der typischerweise großen Anzahl von Genen oder Metaboliten im Vergleich zu einer relativ kleinen Anzahl an Patienten, ergeben sich folgende methodischen Schwierigkeiten für überwachtes Lernen in molekularen Daten [61]: 1. Die Auswahl der Gene oder Metaboliten, die relevante Information zur Lösung des Klassifikationsproblems tragen (*feature selection*), 2. die Vermeidung von *overfitting* und 3. die Vermeidung einer fehlerhaften Überinterpretation der Klassifikationsergebnisse. Eine *feature selection* führt dazu, das Signal-zu-Rausch-Verhältnis in den Daten zu erhöhen, und kann helfen, Sensitivität und Spezifität des molekularen Tests zu erhöhen. Unter *overfitting* versteht man die Situation, dass ein molekularer Test auf dem Trainingsdatensatz, auf dem er entwickelt wurde, gute Ergebnisse liefert, die Prädiktionsfehler sich jedoch bei Generalisierung auf einen unabhängigen Validierungsdatensatz erheblich vergrößern. Die Gefahr eines *overfittings* steigt mit der Komplexität des mathematischen Modells [62]. Zur Überinterpretation von Klassifikationsergebnissen kann es durch ein fehlerhaftes Design der Trainings- und Validierungsverfahren kommen. Typische solche Fehler sind: 1. Die *feature selection* wird nicht mit in den Validierungsprozess einbezogen, was zu einem *selection bias* bei den Klassifikationsraten führt [63]. 2. Es werden verschiedene Klassifikationsverfahren ausprobiert, aber nur das Verfahren mit der besten Performanz auf dem Validierungsdatensatz wird publiziert (*publication bias*). 3. Der Validierungsdatensatz wird unter sehr homogenen Bedingungen produziert (z.B. innerhalb eines kurzen

Zeitraums), die nicht typisch für den beabsichtigten Anwendungsbereich des molekularen Tests sind. Dies birgt die Gefahr untypisch niedriger technischer Fehler im Validierungsdatensatz und einer daraus resultierenden Überschätzung der Klassifikationsergebnisse.

1.3.4 Funktionelle Analyse

Die Entschlüsselung der biochemischen Stoffwechselwege, die den humanen Metabolismus konstituieren, stellt eine der größten Leistungen der Biochemie der letzten 100 Jahre dar [64]. Nach der Sequenzierung des humanen Genoms wurde das Wissen über biochemische Pathways mit genomischer Information verknüpft und ist in Datenbanken wie KEGG, HumanCyc und Reactome verfügbar [65, 66, 67]. Diese Informationen stellen eine wertvolle Ressource zur Verbesserung der weltweiten Gesundheit und insbesondere für die Krebsforschung dar.

Aus Genexpressionstudien ist die Methode der Anreicherungsanalyse (*enrichment analysis*) bekannt, die hilfreich ist, um von einer Liste von differentiell exprimierten Genen zu einer funktionellen Interpretation zu gelangen. Aus folgenden zwei Gründen sind jedoch Anreicherungsanalysen in der Metabolomik weniger erfolgreich als in der Transkriptomik: 1. Die Abdeckung der gesamten Metaboloms durch die verschiedenen Metabolomikplattformen ist derzeit noch sehr lückenhaft. Daher kommt es zu einer sehr unvollständigen Abdeckung der meisten Stoffwechselwege und einer niedrigen Sensitivität der Anreicherungsanalyse. 2. Andererseits muss eine biologische wichtige Veränderung in einem Stoffwechselweg nicht zwangsläufig mit einer simultanen Ansammlung (oder Depletion) von vielen Metaboliten aus diesem Stoffwechselweg einhergehen. Wichtige Änderungen in Stoffwechselwegen können durchaus mit Ansammlung von nur einem oder nur wenigen Metaboliten einhergehen und würden bei einer Anreicherungsanalyse übersehen werden. Die Situation ist damit bei Metaboliten anders als bei Genen oder Proteinen, wo zur Ausübung einer biologischen Funktion oft Module aus mehreren oder vielen Genen reguliert werden müssen.

Da die Methoden zur funktionellen Analyse aus anderen -omik-Wissenschaften nur begrenzt übertragbar sind, besteht ein Bedarf an auf Metabolomik spezialisierten Methoden, die der Struktur und Funktion der biochemischen Stoffwechselwege Rechnung tragen. Ein Beispiel für einen solchen Ansatz ist MetaMapp, das die Visualisierung von Metabolomikdaten in einem biochemischen Netzwerk erlaubt [68]. Eine Besonderheit dieses Algorithmus ist es, dass er auch nicht auf bekannte Metaboliten abgebildete Metabolitenpeaks miteinbezieht. Dabei erfolgt die Einordnung ins Netzwerk mittels der Ähnlichkeit der entsprechenden Massenspektrometrieprofile. Im Rahmen der hier vorgelegten Studien entwickelten wir zwei neue bioinformatische Methoden zur Analyse und Interpretation von Metabolomikdaten im Kontext der Stoffwechselwege: PROFILE Clustering und METAtarget.

2 Zielsetzung

In den vorliegenden Studien wurde das Metabolom solider Tumore durch GC-TOFMS Analysen von Gewebeproben untersucht. Ziel war dabei die Identifikation therapeutischer Zielstrukturen und die Identifikation diagnostischer, prognostischer und prädiktiver Biomarker. Parallel dazu erfolgte die Neuentwicklung und Anwendung bioinformatischer Methoden.

1. Das Metabolom des Kolonkarzinoms und das Metabolom des Mammakarzinoms wurden mit dem Metabolom entsprechender Normalgewebe verglichen. Hierbei wurde einerseits eine Übersicht der metabolischen Änderungen im Tumorgewebe erstellt. Andererseits wurde untersucht, ob und wie eine Trennung von Tumorgewebe und Normalgewebe mittels Metabolitenmustern möglich ist. (Vgl. 3.2 und 3.4)
2. Das Metabolom des Ovarialkarzinoms wurde mit dem Metabolom von Borderline-Tumoren des Ovars verglichen. Diese Metabolomikuntersuchung sollte ein Beitrag zur Diagnostik von Tumoren des Ovars leisten vor dem Hintergrund, dass eine histopathologische Klassifikation nicht immer einfach ist, aber die Diagnose Borderline-Tumor mit einer guten Prognose einhergeht im Vergleich zu einer ungünstigen, oft infausten Prognose beim Ovarialkarzinom. (Vgl. 3.1)
3. Es besteht ein Mangel an Methoden zur Analyse und Interpretation von Metabolomikdaten in funktionellem Kontext. Um diese Lücke zu füllen, wurde PROFILE-Clustering als eine neue Methode zur Visualisierung und Interpretation von Metabolitenveränderungen im Kontext der Stoffwechselwege entwickelt. Dazu wurden Informationen über die Stoffwechselreaktionen aus der KEGG-Datenbank extrahiert und die Metaboliten bezüglich ihrer Distanz im Stoffwechsellnetzwerk geclustert. (Vgl. 3.2 sowie 3.4 und 3.6)
4. Ein von uns neu entwickelter Algorithmus METAtarget erlaubt es, Stoffwechselreaktionen zu identifizieren, bei denen das Gleichgewicht zwischen Substrat und Produkt signifikant verschoben ist. Zusätzlich liefert der Algorithmus eine Liste aller Enzyme, die eine Umwandlung des Substrats in das Produkt katalysieren. (Vgl. 3.3)
5. Es wurde untersucht, ob die molekularen Subtypen des Mammakarzinoms ein unterschiedliches Metabolom besitzen und welche Metabolitenkonzentrationen vom Hormonrezeptorstatus oder HER2-Status des Mammakarzinoms abhängen. Integrativ mit den Metabolomikdaten wurden Transkriptomikdaten derselben Tumore analysiert. (Vgl. 3.5)
6. Motiviert durch aktuelle Fortschritte bei der Entwicklung spezifischer Glutaminase-Inhibitoren wurde der Glutamin-Stoffwechsel im Mammakarzinom genauer untersucht. Dazu wurde das Konzentrationsverhältnis Glutamate/Glutamine mit den klinisch-pathologischen Daten und genomweiten Genexpressionsdaten des Mammakarzinoms korreliert. (Vgl. 3.6)

3 Ergebnisse

Die hier vorzustellenden Originalarbeiten umfassen die Metabolomanalyse von Ovarialtumoren (3.1), die Metabolomanalyse des kolorektalen Karzinoms (3.2), die Metabolomikanalyse des Mammakarzinoms (3.4, 3.5 und 3.6) sowie eine algorithmische Arbeit (3.3). In der Serie der experimentellen Arbeiten wurde die Metabolomikanalyse von frisch-gefrorenen Tumor- und Normalgeweben mittels kombinierter Gaschromatographie und Flugzeit-Massenspektrometrie (GC-TOFMS) etabliert und für verschiedene Tumorprobenkollektive durchgeführt. In den beiden letzten hier vorgestellten Studien (3.5 und 3.6) wurden Multiomik-Analysen durchgeführt, in denen die Metabolomikdaten integriert mit genomweiten Genexpressionsdaten und immunhistologischen Daten analysiert wurden.

Es folgt ein kurzer Überblick über die wichtigsten selbst entwickelten und zur Anwendung gekommenen bioinformatischen Methoden: Für die Entwicklung von Metabolitenprädiktoren wurde das kürzlich vorgestellte multiple random cross-validation Protokoll [69] erweitert und auf die hier vorliegenden Datensätze angewendet (3.1, 3.2 und 3.4). Unsere Implementation und Erweiterung dieses Protokolls haben wir als Softwarepaket `cancerclass` für die statistische Sprache R publiziert [61]. PROFILE-Clustering ist eine neu entwickelte unüberwachte Methode zur Analyse der differentiellen Regulation von Metaboliten im Kontext des Netzwerkes der enzymatischen Reaktionen. PROFILE-Clustering wurde erstmals in der Studie zum kolorektalen Karzinom (3.2) vorgestellt und auch auf die Mammakarzinomdaten angewendet (3.4 und 3.6). METAtarget (3.3) ist ein neu entwickelter Algorithmus zur Extraktion der Schlüsselenzyme aus Metabolomikdaten durch die Integration mit in Datenbanken archiviertem Wissen über Stoffwechselwege.

3.1 Metabolomanalyse des Ovarialkarzinoms

Literatur: Carsten Denkert, Jan Budczies, Tobias Kind, Wilko Weichert, Peter Tabbak, Jalid Sehouli, Silvia Niesporek, Dominique Könsgen, Manfred Dietel, Oliver Fiehn. *Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. Cancer Research, 2006; 66(22): 10795-10804, <http://dx.doi.org/10.1158/0008-5472.CAN-06-0755>.*

Diese Studie ist weltweit eine der ersten Metabolomanalysen einer größeren Kohorte von humanen Gefriergeweben mittels GC-TOFMS. Untersucht wurden die Metabolitenprofile von verschiedenen Ovarialtumoren, invasiven Karzinomen (n=66) auf der einen Seite und Borderlinetumoren (n=9) auf der anderen Seite. Insgesamt wurden 291 Metabolitenpeaks identifiziert, von denen 114 auf Metaboliten mit wohldefinierter chemischer Struktur abgebildet werden konnten. 51 der Metabolitenpeaks (26 identifizierte Metaboliten) waren signifikant verschieden in den Karzinomen im Vergleich zu den Borderlinetumoren bei einer mittels Permutationsanalyse geschätzten FDR von 7,8%. Unter den in den Karzinomen heraufregulierten Metaboliten waren viele Aminosäuren, einschließlich Glutamat mit einer 3-fachen Hochregulation und Glutamin mit einer 2-fachen Hochregulation. Herunterreguliert war unter anderem eine Gruppe von gesättigten freien Fettsäuren (C17:0, C18:0, C19:0) und Benzoesäure.

Mittels überwachtem Lernen gelang die Trennung der Karzinome von den Borderlinetumoren mit einer Sensitivität von 89% und einer Spezifität von 88%. Dieses Klassifikationsergebnis konnte sowohl ohne *feature selection*, d.h. mit einem 291-Metaboliten-

Klassifikator, als auch mit Klassifikatoren, die aus einer sehr kleinen Zahl von Metaboliten (wie 5, 10 oder 20) aufgebaut wurden, erreicht werden. Ein Vergleich verschiedener Klassifikationsmethoden ergab die besten Ergebnisse für *nearest mean classification* (NMC) und *nearest centroid classification* (NCC) im Vergleich zu schlechteren Ergebnissen für lineare Diskriminanzanalyse (LDA) und für lineare Supportvektor-Maschinen (SVM). Schließlich wurde die Robustheit der Klassifikationsergebnisse gegen Wahl von verschiedenen Trainingsdatensätzen mittels des *multiple random cross-validation* Protokolls überprüft. Die Klassifikationsergebnisse lagen auch in diesem strengen Test für fast alle Trainingsdatensatzgrößen signifikant über der 50%-Basislinie.

Zusammengefasst zeigt die Studie, dass sowohl Einzelmetaboliten als auch Metabolitensignaturen in gefrorenem Tumorgewebe mittels GC-TOFMS verlässlich detektiert werden können. Trotz einer geringen Zahl von 9 Borderlinetumoren war das Verfahren präzise und reproduzierbar genug, um die Detektion von 2-fachen und auch 1,5-fachen Änderungen von Metabolitenkonzentrationen zwischen Karzinomen und Borderlinetumoren zu erlauben. Mögliche Schwankungen, die z.B. durch verschiedene Zeiten von der Gewebeentnahme bis zum Einfrieren während der Operation zustande kommen könnten, sind hinreichend klein und erlauben eine stabile Messung der Metabolitenbiomarker. Damit ergeben sich die folgende Perspektiven für die Metabolomanalyse in der Tumorforschung: 1. Der Vergleich von Tumor- und Normalgewebe sowie von aggressiveren Tumoren mit weniger aggressiven Tumoren kann bei der Identifikation neuer therapeutischer Zielstrukturen helfen. 2. Metaboliten (und Metabolitensignaturen), die als Biomarker für Diagnose, Prognose und das Ansprechen auf Therapien dienen, können durch Korrelation von Metabolomikdaten mit klinisch-pathologischen Patientendaten identifiziert werden.

3.2 Metabolomanalyse des kolorektalen Karzinoms

Literatur: Carsten Denkert*, Jan Budczies*, Wilko Weichert*, Gert Wohlgemuth, Martin Scholz, Tobias Kind, Silvia Niesporek, Aurelia Noske, Anna Buckendahl, Manfred Dietel, Oliver Fiehn. *Metabolite profiling of human colon carcinoma - deregulation of TCA cycle and amino acid turnover. Molecular Cancer, 2008; 7:72, <http://dx.doi.org/10.1186/1476-4598-7-72>. (Geteilte Erstautorschaft: C.D., J.B. und W.W.)*

In einer zweiten Metabolomiktudie wurden metabolischen Veränderungen im Kolonkarzinom (n=27) im Vergleich zu normalen Mucosa (n=18) analysiert. Mittels GC-TOFMS konnten insgesamt 206 Metabolitenpeaks identifiziert werden, von denen 82 (39,8%) signifikant ($p < 0,01$) verschieden zwischen den beiden Geweben waren. Eine überwachte Analyse zeigte, dass Metabolitenklassifikatoren sehr gut zwischen Krebs- und Normalgewebe unterscheiden konnten. Nearest-Centroid-Klassifikation mit einem 10-Metaboliten-Klassifikator ergab eine Sensitivität von 96% und eine Spezifität von 94%. Wieder wurde die Robustheit der Klassifikationsergebnisse gegen Wahl von verschiedenen Trainingsdatensätzen mittels des *multiple random cross-validation* Protokolls überprüft. Die Medianwerte von Sensitivität und Spezifität lagen typischerweise bei 90% und für viele Trainingsdatensatzgrößen signifikant über der Basislinie von 50%.

Im Rahmen dieser Studie wurde PROFILE-Clustering, eine Methode zu Visualisierung und Interpretation von metabolischen Veränderungen im Kontext metabolischer Pathways entwickelt und zum ersten Mal vorgestellt. PROFILE-Clustering integriert Ergebnisse aus der Analyse eines Metabolomikdatensatzes mit biochemischem Wissen.

Im ersten Schritt haben wir dazu Paare von Substraten und Produkten aus KEGG REACTION extrahiert und daraus ein Netzwerk aller in humanen Geweben möglichen Umwandlungen aufgebaut. Die Distanz zwischen zwei Knoten in diesem Netzwerk entspricht der Zahl der Reaktionen, die benötigt wird, um die entsprechenden Metaboliten ineinander umzuwandeln. Im zweiten Schritt haben wir diese Distanz zwischen allen Paaren von Metaboliten bestimmt und die Metaboliten mittels eines hierarchischen Clustering bezüglich dieser Distanz gruppiert. Der dritte Schritt war es, die Änderungen der Metaboliten (*fold changes*) im Vergleich zwischen Krebs- und Normalgewebe gegen die Anordnung des Clusterings als Barplot darzustellen.

Für den Kolonkarzinomdatensatz ergab der Vergleich von Krebs- zu Normalgewebe das folgende Regulationsmuster: Harnsäurezyklus und viele andere Aminosäuren waren hochreguliert, wobei Beta-Alanin (4,9-fach) und Methionin (4,3-fach) die im Gesamtdatensatz am stärksten regulierten Metaboliten waren. Glutamat zeigt eine 1,6-fache Hochregulation, während Glutamin eine der wenigen herunterregulierten (2,1-fach) Aminosäuren war. Im Fettstoffwechsel fanden wir eine Herunterregulation vieler Metaboliten einschließlich der freien Fettsäuren C12:0, C18:0, C18:1 und C20:4. Im Kohlenhydratmetabolismus zeigte sich eine Herunterregulation von Glucose und Fructose. Im Purin- und Pyrimidinmetabolismus fanden starke Veränderungen mit Hoch- und Herunterregulation verschiedener Metaboliten statt.

3.3 METAtarget: Detektion von Schlüsselstellen der Stoffwechselregulation aus Metabolomdaten

Literatur: Jan Budczies, Carsten Denkert, Berit M. Müller, Scarlet F. Brockmüller, Manfred Dietel, Jules L. Griffin, Matej Oresic and Oliver Fiehn. METAtarget - extracting key enzymes of metabolic regulation from high-throughput metabolomics data using KEGG REACTION information. German Conference on Bioinformatics 2010 (Vortrag und Manuskript, peer-reviewed). Lecture Notes in Informatics, 2010; P-173:103-112, <http://subs.emis.de/LNI/Proceedings/Proceedings173/article5798.html>.

Bei dieser algorithmischen Arbeit handelt es sich um einen Ansatz des *reverse engineering*, der darauf zielt, aus Metabolitemessungen Hypothesen über die zugrunde liegenden Regulationsmechanismen zu generieren. Änderungen des Metabolitenprofils sind die finale Antwort einer Zelle auf genetische Veränderungen oder Veränderungen ihrer Umgebung. Sie liegen damit am Ende einer Kausalkette von Regulationen auf DNA-Ebene, auf RNA-Ebene auf Proteinebene und auf Ebene der Enzymaktivitäten.

Der neu entwickelte Algorithmus METAtarget verwendet als Input Daten eines Metabolomik-Experiments, das zwei Krankheitszustände miteinander vergleicht. Es kann sich dabei z.B. um den Vergleich von Krebs- und Normalgewebe oder den Vergleich von Tumorsubtypen handeln. Im ersten Schritt extrahiert METAtarget mit Hilfe der Datenbank KEGG RPAIR alle Paare von Metaboliten, die Substrat und Produkt in einer metabolischen Reaktion sind. Im zweiten Schritt wird mit einer Statistik für jedes dieser Paare die Stärke der Regulation zwischen den beiden Krankheitszuständen bewertet. Die Statistik ähnelt der Welchs t-Statistik und besitzt zwei unterschiedliche Interpretationen: Erstens misst sie, wie sehr sich der Einfluss, den der Krankheitszustand auf das Produkt hat von dem unterscheidet, den der Krankheitszustand auf das Substrat hat. Zweitens kann sie interpretiert werden als Einfluss des Krankheitszustandes auf das Verhältnis der Konzentration von Produkt und Substrat. Eine signifikante Abweichung

der Statistik von der Nullverteilung generiert die Hypothese einer verschiedenen Regulation der Umwandlung in den beiden Krankheitszuständen. Im dritten Schritt extrahiert METAtarget alle zu den Reaktionspartnern gehörenden Reaktionen und Enzyme. Die extrahierten Enzyme können mittels zusätzlicher -omik-Datensätze auf Mutationen oder Veränderungen in Gen- und Proteinexpression untersucht werden.

In der vorliegenden Studie haben wir mittels METAtarget die Metabolomikprofile von Hormonrezeptor-negativem (ER-) Brustkrebs (n=58) mit Hormonrezeptor-positivem (ER+) Brustkrebs (n=188) verglichen. Insgesamt konnten in diesem Datensatz 91 Substrat-Produkt-Paare analysiert werden, von denen 13 eine hochsignifikante (nach Bonferroni-Korrektur) Veränderung aufwiesen. Darunter waren sowohl das Verhältnis Glukose-6-Phosphat/Glukose mit einer 2,4-fachen Hochregulation im (ER-)-Mammakarzinom sowie das Verhältnis Glutamat/Glutamin mit einer 2,9-fachen Hochregulation (ER-)-Mammakarzinom. Es wurden sechs Enzyme extrahiert, die eine Umwandlung zwischen Glukose und Glukose-6-Phosphat katalysieren: vier Hexokinasen (HK1, HK2, HK3 und GCK) und zwei Glukose-6-Phosphatasen (G6PC, G6PC2). Des Weiteren wurden 13 Enzyme extrahiert, die eine Umwandlung zwischen Glutamin und Glutamat katalysieren. Darunter waren zwei Glutaminasen (GLS und GLS2), eine Glutamat-Ammonium-Ligase (GLUL) und zehn weitere Enzyme, die komplexere Reaktionen katalysieren, bei denen Glutamin als Aminodonor dient.

Ein externer Genexpressionsdatensatz wurde verwendet, um die von METAtarget detektierten Enzyme auf Änderungen der Genexpression zu untersuchen. Dieser wurde durch die Fusion von drei über das Repository GEO verfügbaren Datensätzen generiert und umfasste 508 (ER+)-Mammakarzinome und 176 (ER-)-Mammakarzinome. Unter den sechs Enzymen, die die Phosphorylierung von Glukose zu Glukose-6-Phosphat katalysieren, zeigte nur die Hexokinase 3 (HK3) eine Veränderung und zwar eine Hochregulation in den (ER-)- im Vergleich zu den (ER+)-Tumoren. Für die Umwandlung von Glutamin in Glutamate zeigte sich ein komplexeres Expressionsmuster, das die Hochregulation von GLS, sowie die Herunterregulation GLS2 und GLUL beinhaltet.

Glykolyse und Glutaminolyse sind zwei Stoffwechselwege, die von humanen Zellen zur Energiegewinnung verwendet werden können und in vielen malignen Tumoren hochreguliert sind. Zusammengefasst wurde in dieser Studie gezeigt, dass die Anfangsreaktionen sowohl der Glykolyse als auch die Anfangsreaktion der Glutaminolyse im (ER-)-Mammakarzinom im Vergleich zum (ER+)-Mammakarzinom heraufreguliert sind. Diese Regulationen führen zu einer Verschiebung des Gleichgewichts der beiden Reaktionen vom Substrat zum Produkt. Dies passt gut zur klinisch-pathologischen Situation, aus der das (ER-)-Mammakarzinom als aggressiverer Tumor mit höheren Tumorgaden und schlechterer Prognose bekannt ist.

3.4 Metabolomanalyse des Mammakarzinoms: Tumor- versus Normalgewebe

Literatur: Jan Budczies, Carsten Denkert, Berit M. Müller, Scarlet F. Brockmüller, Frederick Klauschen, Balazs Györfy¹, Manfred Dietel, Christiane Richter-Ehrenstein, Ulrike Marten, Reza M. Salek, Julian L. Griffin, Mika Hilvo, Matej Orešič, Gert Wohlgemuth, Oliver Fiehn. *BMC Genomik*, 2012; 13:334, <http://dx.doi.org/10.1186/1471-2164-13-334>.

Die drei folgenden Studien (3.4, 3.5 und 3.6) haben die Metabolomikanalyse des

Mammakarzinoms zum Thema und wurden innerhalb des Verbundprojekts METAcancer durchgeführt, das im 7. Rahmenprogramm der Europäischen Kommission gefördert wurde. In METAcancer wurde das Mammakarzinom mittels drei Metabolomikmethoden (GC-MS, LC-MS und NMR) untersucht mit dem Ziel, neue diagnostische, prognostische und prädiktive Biomarker zu identifizieren. Dazu wurde an der Charité eine Biobank aus frischgefrorenen und paraffineingebetteten Proben des primären Mammakarzinoms aufgebaut. Für die Metabolomanalysen standen gefrorene Proben von 271 Tumorgeweben und 98 Normalgeweben der Mamma zu Verfügung. Für immunhistologische Studien waren darüber hinaus formalinfixierte und paraffineingebettete Proben in einer Teilkohorte vorhanden. In METAcancer wurde prospektiv eine Aufteilung der Gewebe in einen Trainingsdatensatz (2/3) und einen Validierungsdatensatz (1/3) vorgenommen. Durch Minimierung der Kullback-Leibler-Divergenz wurde die Aufspaltung so optimiert, dass die beiden Datensätze bezüglich der wichtigsten klinisch-pathologischen Parameter gut übereinstimmten. Die drei hier vorzustellenden Studien behandeln die Analyse der in METAcancer generierten GC-TOFMS-Daten unter Einbeziehung von genomweiten Genexpressionsdaten und von immunohistologischen Daten.

In der ersten der drei Studien wurden die Metabolitenveränderungen im Krebsgewebe im Vergleich zum Normalgewebe analysiert. Insgesamt konnten in den Geweben 468 Metabolitenpeaks identifiziert werden, von denen die Mehrzahl von 368 (79%) signifikant verschieden zwischen Krebs- und Normalgeweben war ($p < 0.05$ in Trainings- und Validierungsdatensatz). Zur Interpretation dieser Veränderungen wurden diese als PROFILE- Clustering und in Form eines metabolischen Netzwerks visualisiert. Im Glukosestoffwechsel fanden wir eine Herunterregulation verschiedener Zucker (Glukose, Fruktose und Sucrose) bei einer Hochregulation von Glukose-6-Phosphat. Dieses Ergebnis kann als Folge des Warburgeffekts, nämlich einer stärkeren Aufnahme der Zucker durch die Krebszellen und einer höheren Aktivität der Glykolyse, interpretiert werden. Im Glutaminstoffwechsel fanden wir eine Hochregulierung der meisten Aminosäuren inklusive Glutamat, während Glutamin keine signifikante Änderung zeigte. In Purin- und Pyrimidinmetabolismus zeigt sich eine Hochregulation vieler Metaboliten, kompatibel mit einem stärkeren Bedarf an diesen Molekülen für den für den Aufbau von DNA und RNA in stark proliferierenden Zellen.

Viele freie Fettsäuren (C14:0, C15:0, C16:0, C17:0, C18:0 und C20:0) waren herunterreguliert in den Krebsgeweben. Dieses scheint auf dem ersten Blick in Widerspruch zu stehen zu dem Paradigma einer verstärkten Lipogenese in Krebszellen und einer Hochregulation der fatty acid synthase (FASN) in vielen Krebszellen. Möglicherweise ist die Lipogenese tatsächlich erhöht im Mammakarzinom, aber die freien Fettsäuren werden schnell weiter zu komplexeren Glycerophospholipiden metabolisiert. Tatsächlich wurde in einer Lipidomikstudie im METAcancer-Projekt eine Hochregulation vieler Membranlipide in Krebsgeweben im Vergleich zu Normalgeweben gefunden [70]. Ebenfalls zu dieser Hypothese passt die im GC-TOFMS-Datensatz detektierte Hochregulation von Phospho-Ethanolamine bei Herunterregulation von Ethanolamine. Die Phosphorylierung von Ethanolamine bildet den ersten Schritt des Kennedy-Pathways, der zur Synthese von Membranlipiden führt.

Schließlich haben wir die Fragestellung behandelt, wie gut sich Krebs- von Normalgeweben durch Metabolitenklassifikatoren trennen lassen und wie viele Metaboliten für einen solchen Klassifikator benötigt werden. Die Mamma besteht aus verschiedenen Gewebestrukturen, einerseits den Loboli, den Milchgängen und Muskelgewebe (im Folgen-

den als normales Mammagewebe bezeichnet), andererseits enthält sie eine große Menge an Fettgewebe. Die Gewebehomogenität spiegelte sich in den Gewebeproben wider, für die getrennt der Anteil an Tumorzellen, der Anteil an normalem Mammagewebe und der Anteil an Fettgewebe bestimmt wurde. An einen Klassifikator wurde die Forderung gestellt, sowohl Krebszellen von normalem Mammagewebe als auch Krebszellen von Fettgewebe trennen zu können. Mit der Forderung, dass ein Marker beide Trennungen mit mindestens 80% Sensitivität und 80% Spezifität erreichen soll, wurden die möglichen Biomarker auf eine Liste von 20 Metaboliten reduziert. 13 dieser Marker waren hochreguliert in den Krebsgeweben („Tumormarker“), 7 dieser Marker waren herunterreguliert in den Krebsgeweben („Normalgewebemarker“).

Im nächsten Schritt wurden 2-Metabolitenmarker, die aus Quotienten von Tumormarkern und Normalgewebemarkern gebildet wurden, auf ihre Klassifikationsgenauigkeit untersucht. Die beste Trennung wurde durch das Verhältnis $R = \text{Cytidine-5-Monophosphat} / \text{Pentadecansäure}$ erreicht, mit einer Fläche unter der ROC-Kurve von 0,97, einer Sensitivität von 95% und einer Spezifität von 94%. Andererseits zeigt sich eine fast ebenso gute Performanz von vielen anderen 2-Metabolitenklassifikatoren, die sich statistisch nicht vom Marker R unterschied. Abschließend untersuchten wir, ob sich die Ergebnisse für die Trennung durch eine höhere Zahl von Metaboliten im Klassifikator und andere Klassifikationsmethoden verbessern lassen. Weder ein mittels *nearest centroid classification* optimierter 2-Metaboliten-Klassifikator, noch ein 20-Metaboliten-Klassifikator, noch ein alle Metaboliten mit bekannter chemischer Struktur einbeziehender 162-Metaboliten-Klassifikator erzielten ein signifikant besseres Ergebnis als der Marker R.

3.5 Metabolomanalyse des Mammakarzinoms: ER-negative versus ER-positive Tumore

Literatur: Jan Budczies, Scarlet F. Brockmüller, Berit M. Müller, Dinesh K. Barupal, Christiane Richter-Ehrenstein, Anke Kleine-Tebbe, Julian L. Griffin, Matej Orešič, Manfred Dietel, Carsten Denkert, Oliver Fiehn. *Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. Journal of Proteomik, 2013; 94:279-288, <http://dx.doi.org/10.1016/j.jprot.2013.10.002>.*

In der zweiten Metabolomikstudie zum Mammakarzinom wurde der Östrogenrezeptor-negative (ER-) mit dem Östrogenrezeptor-positiven (ER+) molekularen Subtyp verglichen. 40 Metabolitenpeaks, von denen 19 auf eine bekannte chemische Struktur abgebildet werden konnten, waren signifikant verschieden zwischen den beiden Subtypen ($p < 0.05$ in Trainings- und Validierungsdatensatz). Unter diesen Metaboliten zeigte Beta-Alanin den stärksten und signifikantesten Unterschied zwischen (ER-) und (ER+)-Tumoren (fold change = 2,4; $p=1,5E-20$). Weiter fanden sich Unterschiede im Glutamin-Stoffwechsel mit einer Hochregulation von Glutamate und 2-Hydroxyglutarat sowie eine Herunterregulation von Glutamine im (ER-)-Mammakarzinom.

In einer Subkohorte von 154 Tumoren wurde eine genomweite Genexpressionsanalyse mittels des DASL-Assays durchgeführt. Aus diesen Daten konnte die Expression von 1368 in der KEGG-Datenbank annotierten Enzymen extrahiert werden. In einer unüberwachten Analyse wurden die 162 identifizierten Metaboliten mit den Expressionslevels der 1368 Enzyme korreliert. Von diesen Metaboliten-Transkript-Paaren zeigten fünf Paare eine starke und hochsignifikante Spearman-Korrelation von $|R| > 0,5$

(alle p-Werte $< 1,0E - 11$). Unten den beiden allerstärksten Korrelationen war eine hochsignifikante negative Korrelation der Aminosäure Beta-Alanin mit dem Enzym 4-Aminobutyrate-Aminotransferase ABAT ($R=-0,62$). Interessanterweise lieferte die unüberwachte Korrelationsanalyse hier einen biochemisch sehr gut interpretierbaren Zusammenhang: Beta-Alanin ist ein Substrat von ABAT durch das es zu Malonate-Semialdehyde transaminiert wird. ABAT ist besser bekannt als GABA-Transaminase, da ABAT neben Beta-Alanine auch die Gamma-Aminobuttersäure (GABA) als Substrat hat. GABA spielt als Neurotransmitter eine wichtige Rolle im menschlichen Gehirn. Obwohl der GC-TOFMS-Analyse grundsätzlich zugänglich, konnte in den Brustkrebsgeweben keine signifikante Konzentration von GABA detektiert werden.

Zusätzlich zum Expressionslevel von ABAT analysierten wir mittels Immunhistochemie die Expression des ABAT-Proteins. Wir fanden eine starke und hochsignifikante positive Korrelation von ABAT-Protein und ABAT-RNA ($p=2,4E-07$) sowie eine starke und hochsignifikante negative Korrelation von ABAT-Protein und Beta-Alanin ($p=4,8E-07$). Von den (ER⁻)-Tumoren zeigten nur 2,1%, von den (ER⁺)-Tumoren hingegen 41,0% eine starke zytoplasmatische Färbung für ABAT. Schließlich analysierten wir den Einfluß der ABAT-Expression auf das rezidivfreie Überleben beim Mammakarzinom. In einem großen Datensatz von 2896 Patienten, der mittels des Tools Kaplan-Meier-Plotter [71] untersucht wurde, war ABAT ein starker und hochsignifikanter positiver Prognosemarker ($HR=0,60$; $p=3,2E-15$). Ebenfalls ein signifikanter positiver Prognosemarker war ABAT im ER-positiven Mammakarzinoms sowie auch im ER-negativen Mammakarzinoms.

Zusammengefasst zeigte diese Studie zum ersten Mal, dass das (ER⁺)-Mammakarzinom und das (ER⁻)-Mammakarzinom zusätzlich zu dem bekannten starken Genexpressionsunterschieden, starke Unterschiede im Stoffwechsel besitzen, die sich in voneinander verschiedenen Metabolitenmustern manifestieren. Mittels einer Multiomik-Strategie konnten wir eine differentielle Regulation des Beta-Alanin-Abbaus detektieren, die Unterschiede in der Expression des ABAT-Gens und des ABAT-Proteins umfasst. Kaplan-Meier-Analyse eines großen unabhängigen Datensatzes zeigte, dass die ABAT-RNA ein prognostischer Marker für das Mammakarzinom insgesamt, sowie das (ER⁺)-Mammakarzinom und das (ER⁻)-Mammakarzinom ist.

3.6 Metabolomanalyse des Mammakarzinoms: Glutamat-Anreicherung als neuer diagnostischer Parameter

Literatur: Jan Budczies, Berit M. Pfitzner, Balazs Györfy, Klaus-Jürgen Winzer, Cornelia Radke, Manfred Dietel, Oliver Fiehn, Carsten Denkert. *Glutamate enrichment as new diagnostic opportunity in breast cancer. International Journal of Cancer, 2015; 136(7):1619-28, <http://dx.doi.org/10.1002/ijc.29152>.*

Die dritte Studie zur Metabolomanalyse des Mammakarzinom ist einer genauen Analyse des Glutamine-Stoffwechsels gewidmet. Motivation für diese fokussierte Analyse ist, dass für viele Tumorzellen - zusätzlich zu Glukose - Glutamin eine wichtige Energiequelle und einen wichtigen molekularen Baustein darstellt. Der erste Schritt der zellulären Verwendung von Glutamin ist seine Umwandlung in Glutamat, zum Beispiel durch eine Glutaminase. Hier setzen drei in den letzten Jahren vorgestellte neue spezifische Inhibitoren der Glutaminase 1 (GLS1) an: BPTES, 968 und CB-839 [20, 21, 22].

Im Vergleich Mammakarzinom zu Mammanormalgewebe fanden wir eine starke Erhöhung von Glutamat (fold change = 6,9), während Glutamin unverändert war. Ei-

ne positive Korrelation von Glutamat und Glutamin in den Normalgeweben ($R=0,56$) änderte sich zu einer negativen Korrelation ($R=-0,46$) in den Krebsgeweben. Zwischen (ER-)-Mammakarzinom und (ER+)-Mammakarzinom detektierten wir eine Erhöhung von Glutamat (fold change = 1,5) und eine Erniedrigung von Glutamin (fold change = -2,1). Alle dieser Änderungen waren hochsignifikant im Trainingsdatensatz und konnten im Validierungsdatensatz validiert werden.

Im Hauptteil der Studie wurde der Quotient aus Glutamat- und Glutaminkonzentration, das *glutamate/glutamine ratio* (GGR), untersucht. Das GGR war signifikant höher in (ER+)-Tumoren im Vergleich zu Normalgeweben (fold change = 5,6) und signifikant höher in (ER-)-Tumoren im Vergleich zu (ER+)-Tumoren (fold change = 3,2). Es wurde ein Schwellenwert von $GGR = 2$ festgelegt und alle Gewebe mit Metabolitenkonzentrationen unter diesem Schwellenwert als Glutamat-normal und alle Gewebe über diesem Schwellenwert als Glutamat-angereichert klassifiziert. Mit dieser Festlegung waren 2% der Normalgewebe, 56% der (ER+)-Tumore und 88% der (ER-)-Tumore Glutamat-angereichert. Weiter zeigt sich eine signifikante Korrelation des GRR mit dem Tumorgrading, aber keine signifikante Korrelation mit anderen klinisch-pathologischen Parametern wie Tumorstadium, Nodalstatus und HER2-Status. Sowohl in einer univariaten als auch einer multivariaten Analyse war das GRR ein positiver prognostischer Marker für das Mammakarzinom.

Ferner untersuchen wir die Korrelation des GGR mit dem Metabolomik-Gesamtdatensatz und den genomweiten Genexpressiondaten. In einem unüberwachten hierarchischen Clustering der Metabolomikdaten der Tumore kristallisierten sich zwei Hauptcluster heraus: Das eine der beiden Cluster bestand überwiegend aus Glutamat-angereicherten Tumoren, während das andere Cluster mit Glutamat-normalen Tumoren angereichert war. In einer überwachten Analyse konnte ein 25-Metaboliten-Klassifikator (ohne Glutamin und Glutamat) für die Trennung der Glutamat-angereicherten von den Glutamat-normalen Tumoren validiert werden mit einer Fläche unter der ROC-Kurve von 0.92. Hingegen konnte mit einem im Genexpressiondatensatz optimierten 100-Gene-Klassifikator nur eine Fläche unter der ROC-Kurve von 0.84 erreicht werden. Zusammengefasst zeigen diese Daten, dass sich die Höhe des GGR sehr stark mit dem globalen Metabolitenmuster im GC-TOFMS-Datensatz korreliert. Hingegen zeigt der Ansatz, einen Surrogat-Klassifikator aus Genexpressiondaten zu erstellen, nur begrenzte Genauigkeit. Insbesondere zeigte das GGR keine signifikante Korrelation mit den beiden Glutaminasen GLS und GLS2 und eine zwar signifikante, aber nur mäßige negative Korrelation ($R = -0,28$) mit der Glutamat-Ammonium-Ligase GLUL.

In dieser Studie wurde eine neue Klassifikation des Mammakarzinoms in Glutamat-angereicherte und Glutamat-normale Tumore mittels des GRR vorgestellt. Die Hochregulation des GRR in Mammakarzinomen im Vergleich zu Normalgeweben ist im Einklang mit einer speziellen Verwendung von Glutamin in Krebszellen, die für viele Tumorarten beschrieben ist. An dieser Stelle setzen die in den letzten Jahren entdeckten neuen Glutaminase-Inhibitoren an, die darauf zielen, den ersten Schritt der Glutaminverwendung durch die Krebszelle zu blockieren. Unsere Studie zeigte eine Erhöhung des GGR in 88% der (ER-)-Tumore und in 56% der (ER+)-Tumore. Sie unterstützt damit Ergebnisse aus Zellkultur und Tiermodellen, die eine Auswertung der Glutaminase-Inhibitoren im triple-negativen Mammakarziom priorisieren [22]. Allgemeiner zeigten unsere Ergebnisse keine Abhängigkeit der Metabolitenprofile vom HER2-Status und lassen eine Auswertung der Glutaminaseinhibitoren in (ER-)-Tumoren und ausgewählten (ER+)-Tumoren interessant erscheinen.

4 Diskussion

Im Jahr 2006 haben wir mit unserer Studie zu Ovarialtumoren eine der ersten Metabolomikstudien an einer großen Kohorte von Tumorgeweben vorgestellt [45]. In dieser Studie zeigten wir, dass eine reproduzierbare Analyse von Metaboliten in Gefriergegeweben mittels GC-TOFMS möglich ist. Die resultierenden Metabolitenmuster trennten invasive Ovarialkarzinome von Borderlinetumoren des Ovars mit hoher Sensitivität und Spezifität. Die Metabolomforschung setzten wir in einer Studie zum Kolonkarzinom und vier Studien zum Mammakarzinom fort. Letztere wurden in den Verbundprojekten METAcancer und ARROWS von der EU-Kommission gefördert. Etwa seit der Jahrtausendwende ist die Metabolomik als jüngstes Fach zu den anderen -omik-Wissenschaften hinzugetreten mit einer bis heute anhaltenden starken jährlichen Zunahme von Studien insgesamt und insbesondere von Krebsstudien (Abbildung 1, S. 12). Die Metabolanalyse von Krebsgeweben ist aus mehreren Gründen interessant:

Erstens können Metabolomuntersuchungen bei der Identifikation von therapeutischen Zielstrukturen helfen. Auf Grund des markant veränderten Stoffwechsels von Krebszellen im Vergleich zu Normalzellen existiert ein therapeutisches Fenster für in den Zellstoffwechsel eingreifende Substanzen. Vergleichende Metabolomanalysen an Krebs- und Normalgeweben können helfen die Veränderungen des Zellstoffwechsels zu detektieren und zu charakterisieren. Hierzu haben wir eigene Studien am Kolonkarzinom und am Mammakarzinom durchgeführt [72, 73]. Eine seit langem bekannte metabolische Veränderung in Krebszellen ist der Warbureffekt, der zu einer verschiedenen Metabolisierung von Glukose in Krebszellen im Vergleich zu Normalzellen führt. Daher sind verschiedene Substanzen entwickelt worden, die die an der Glykolyse beteiligte Enzyme wie z.B. Hexokinasen und Pyruvatkinasen hemmen. Einige dieser Substanzen werden derzeit in klinischen Studien getestet [74]. Neben Glukose wird auch Glutamat von vielen Krebszellen anders metabolisiert als von Normalzellen. Unsere Metabolomik-Analyse zeigte ein erhöhtes Glutamat-Glutamine-Verhältnis im Mammakarzinom im Vergleich zum Normalgewebe [75].

Das zweite Ziel der Metabolomuntersuchungen an Tumorgeweben ist die Identifikation und Validierung neuer diagnostischer, prognostischer und prädiktiver Biomarker. Da eine immer größere Zahl von zielgerichteten systemischen Therapien zu Verfügung steht, werden molekulare Marker in der heutigen Krebsmedizin dringender denn je benötigt. Die Entwicklung einer therapiebegleitenden Diagnostik (companion diagnostics) wird von den Pharmafirmen in der Regel parallel zur Entwicklung des Wirkstoffes vorangetrieben, um eine sichere und effektive Nutzung des Medikaments zu ermöglichen. Ein Beispiel für die Erfolgsgeschichte von Biomarkern ist die Klassifikation des Mammakarzinoms in molekulare Subtypen. Die Bestimmung von Hormonrezeptor und HER2-Status beim Mammakarzinom ist heute weltweiter Standard und mit den Behandlungsmöglichkeiten einer endokrinen Therapie (z.B. mit Tamoxifen) bzw. einer anti-HER2-Therapie (z.B. mit Trastuzumab) verknüpft.

4.1 Diagnostische, prognostische und prädiktive Marker

In den vorgelegten Originalarbeiten haben wir diagnostische Marker vorgestellt, die Ovarialkarzinome von Borderlinetumoren, Kolonkarzinome von Kolonnorgewebe und Mammakarzinome von Normalgewebe der Mamma trennen [45, 72, 73]. Trotz guter Er-

gebnisse werden Metabolomanalysen den Goldstandard einer histopathologischen Gewebeanalyse sicher nicht ersetzen, denn letztere trägt mit der optische-mikroskopische Begutachtung der räumlichen Inhomogenität und Komplexität einer Gewebeprobe in idealer Weise Rechnung. Denkbar ist eine unterstützende Funktion der Metabolomik, wenn eine histopathologische Klassifikation nicht einfach ist (Trennung Ovarialkarzinom von Borderlinetumoren) oder eine schnelle Diagnose nötig ist (Bewertung der Qualität von Biopsien, Bewertung der Malignität von Gewebeproben im Operationssaal).

Bei der vergleichenden Analyse des hormonrezeptor-negativen Mammakarzinoms mit dem hormonrezeptor-positiven Mammakarzinom identifizierten wir die Regulation des Beta-Alanin-Abbaus als prädiktiven Marker [76]. Die Regulation zeigte sich in einer Herabregulation des ABAT-Transkripts, die mit einer Akkumulation von Beta-Alanin korrelierte. Bei ABAT handelt es sich um eine Aminotransferase, die den ersten Schritt des Abbaus von Beta-Alanin katalysiert. In einer Analyse eines großen externen Genexpressionsdatensatzes war ABAT ein positiver prognostischer Marker für das Mammakarzinom insgesamt, wie auch für die beiden molekularen Subtypen des Hormonrezeptor-positiven und des Hormonrezeptor-negativen Mammakarzinoms.

Bei der Analyse des Glutaminmetabolismus im Mammakarzinom identifizierten wir das Glutamate/Glutamine-Verhältnis (glutamate/glutamine ratio, GGR) als möglichen prädiktiven Marker für eine Therapie mit Glutaminase-Inhibitoren [75]. Das GGR ist im Mammakarzinom im Vergleich zu Normalgeweben der Mamma erhöht. Ein Cutoff-Punkt für das GGR wurde in der Weise festgelegt, dass alle Normalgewebe im Trainingsdatensatz unter diesem Cutoff-Punkt lagen. Mit dieser Definition waren 88% der Hormonrezeptor-negativen und 56% der Hormonrezeptor-positiven Mammakarzinome Glutamate-angereichert. Dieses Ergebnis generiert die Hypothese, dass vor allem Hormonrezeptor-negative und ausgewählte Hormonrezeptor-positive Mammakarzinompatientinnen von einer Therapie mit Glutaminase-Inhibitoren profitieren könnten. Es ist im Einklang mit Ergebnissen aus Zellkulturmodellen und einem Mausmodell, in denen Modelle des tripel-negativen Mammakarzinoms am besten auf die Therapie mit CB-839 ansprachen [22].

Methodisch handelt es sich bei den hier vorgestellten Metabolomikanalysen um retrospektive Studien an Gefriergewebekohorten. Die Analysen beruhen auf der Korrelation der Metabolomikdaten mit klinisch-pathologischen Parametern, der Korrelationsanalyse innerhalb der Metabolomikdaten und der Korrelation der Metabolomikdaten mit anderen -omik-Daten oder immunhistologischen Daten. Somit untersuchen die Studien Korrelationen, aber keine Kausalitäten und sind hypothesengenerierend in Bezug auf die zugrunde liegenden molekularen Mechanismen oder die Wirkung von Therapien. Beispielsweise lassen unsere Daten es wahrscheinlich erscheinen, dass es im Hormonrezeptor-negativen Mammakarzinom durch eine Herabregulierung der Expression von ABAT zu einer Akkumulation von Beta-Alanin kommt. Als weiteres Beispiel lässt die Erhöhung des Glutamat/Glutamin-Verhältnisses in fast allen Hormonrezeptor-negativen Tumoren und einigen Hormonrezeptor-positiven Tumoren eine Evaluation der neuen Glutaminaseinhibitoren in Patienten mit solchen Tumoren besonders erfolgversprechend erscheinen. Die Validierung dieser Hypothesen kann jedoch grundsätzlich nur in Systemen durchgeführt werden, die ein experimentelles Eingreifen zulassen, also entweder in Modellsystemen (Zellkultur oder Tiermodell) oder in randomisierten klinischen Studien.

4.2 Bioinformatische Methoden

Für die Auswertung und Interpretation von Metabolomikdaten können einerseits bioinformatische Methoden verwendet werden, die zur Analyse anderer -omik-Datenarten entwickelt wurden, aber von universeller Anwendbarkeit sind und solche, die spezifisch auf die Analyse von Metabolomikdaten zugeschnitten sind. Nicht weiter eingegangen werden kann an dieser Stelle auf die Verarbeitung der Rohdaten, bei GC-TOFMS-Analysen aus den Ionenflugzeiten und den Massenspektren bestehen. Die Zusammenfassung der Rohdaten zu Intensitäten definierter Metaboliten wurde für die hier vorgestellten GC-TOFMS-Studien am Fiehn Lab mit Hilfe der spektralen Datenbank Binbase durchgeführt [77].

Durch den Einsatz der gleichen Gewebemengen (z.B. von Proben gleichen Gewichts) kann versucht werden, aus der GC-TOFMS-Messung zwischen den Proben vergleichbare Intensitäten zu erhalten. Da jedoch bei der Wägung der Anteil von Wasser in den Proben schwer kontrolliert werden kann und wegen der Inhomogenität der Gewebeproben im Hinblick auf verschiedene Arten von malignen und nicht-malignen Zellen, kann jedoch trotzdem nicht sichergestellt werden, dass für jede Probe die gleiche Anzahl an Tumorzellen zur Analyse kommt. In allen unseren Studien haben wir daher eine Skalierungsnormalisierung durchgeführt, bei der die Intensität jedes Metaboliten durch die Summe der Intensitäten aller gemessenen und auf eine wohldefinierte chemische Struktur abgebildeten Metaboliten dividiert wird. In der Tat haben wir in der Studie zum Kolonkarzinom gezeigt, dass die Muster von differentiellen Metaboliten zwar konsistent zwischen normalisierten und nicht-normalisierten Daten sind, eine Datennormalisierung jedoch zu einer höheren Sensitivität bei der Detektion differentiellen Metabolitenkonzentrationen führt [72].

4.2.1 Differentiell veränderte Metaboliten

Zur Detektion differentiell veränderter Metaboliten zwischen zwei klinisch-pathologisch unterschiedlichen Gruppen von Geweben verwendeten wir in allen drei Studien Welchs t-Test. Es hat sich gezeigt, dass die Metabolitenwerte der GC-TOF-Daten auf logarithmischer Skala in guter Näherung normalverteilt sind. Zum Beispiel wiesen in den Tumoren des Mammakarzinomdatensatzes von 162 Metaboliten nur 21 (13%) eine Shapiro-Wilk-Statistik $W < 0.9$ und nur 3 (2%) eine Shapiro-Wilk-Statistik $W < 0.8$ auf. Eine Verwendung der p-Werte aus der Shapiro-Wilk-Statistik oder ähnlichen Statistiken zur Entscheidung, ob ein parametrischer oder nicht-parametrischer Test verwendet werden soll, halten wir für wenig hilfreich. Die Signifikanz dieser p-Werte hängt von der Fallzahl ab und ein solches Vorgehen würde daher zu einer unsinnigen Abhängigkeit der Entscheidung für einen nicht-parametrischen Test von der Fallzahl führen. Die Verwendung von Welchs t-Tests in den hier vorgestellten Studien wird durch die Beobachtung unterstützt, dass der t-Test sehr robust auch gegen starke Abweichungen von der Normalverteilung ist [78]. Kritisch ist hingegen eine Anwendung von Students t-Test bei gleichzeitig verschiedener Varianz und verschiedener Fallzahl in den zu vergleichenden Gruppen [78]. Klinisch-pathologisch relevante Tumorsubtypen z.B. des Mammakarzinoms besitzen unterschiedliche Inzidenz, so dass es auch in unseren Kohorten zu Gruppenvergleichen mit stark unterschiedlichen Fallzahlen kam. Daher haben wir für unsere Analysen durchgehend Welchs t-test und nicht Students t-test verwendet. Der Welch-Test verwendet im Gegensatz zum Student-Test keine gepoolte Schätzung der Varianz und ist mit un-

terschiedlichen Varianzen und Fallzahlen in den zu vergleichenden Gruppen verträglich [79].

Praktisch validieren konnten wir den von uns verwendeten Test in der Studie zum Mammakarzinom, in der prospektiv eine Aufspaltung in einen Trainingsdatensatz (2/3) und einen Validierungsdatensatz (1/3) vorgenommen wurde. Beim Vergleich Tumor gegen Normalgewebe wurden 427 signifikant ($p < 0.05$, ohne Korrektur für multiples Testen) veränderte Metabolitenpeaks detektiert, von denen 368 (86%) validiert werden konnten [73]. Beim Vergleich von Hormonrezeptor-positivem Brustkrebs mit Hormonrezeptor-negativem Brustkrebs wurden 65 signifikant veränderte Metabolitenpeaks detektiert, von denen 40 (62%) validiert werden konnten [76]. Die Tatsache, dass in diesen beiden Analysen nicht alle Metaboliten validiert werden konnten, ist wahrscheinlich zum größten Teil eine Folge von falsch-positiven Detektionen im Rahmen des multiplen Testens. Die Verwendung einer p -Wertschwelle von 0.05 bei einer Gesamtzahl von 468 Metabolitenpeaks lässt eine Gesamtzahl von 23 falsch positiven Detektionen erwarten. Dies entspricht ziemlich genau der Zahl von nicht-validierbaren Detektionen bei der Analyse des Hormonrezeptorstatus. Bei der vergleichenden Analyse zwischen Tumor- und Normalgewebe war die Zahl der nicht-validierbaren Detektionen größer. Hier könnten Gewebehomogenitäten und insbesondere eine ungleiche Zusammensetzung der Normalgewebe in Trainings- und Validierungsdatensatz einen Teil der nicht-validierbaren Detektionen erklären. Bei den Krebsgeweben war ein Mindesttumorzellgehalt von 40% eine Voraussetzung für die Integration in die Studie. Sehr inhomogen war die Zusammensetzung der Normalgewebe, die z.B. einen Anteil von 40% oder mehr an normalem Brustgewebe aufweisen konnten oder aber komplett oder zu über 60% aus Fettgewebe bestehen konnten.

Eine Korrektur für multiples Testen der in der Biomedizin üblichen Signifikanzschwelle $p < 0.05$ wurde in der Studie zum Ovarialkarzinom, in der Studie zum Kolonkarzinom und in den Studien zum Mammakarzinom in unterschiedlicher Weise durchgeführt: In der Ovarstudie legen wir eine Signifikanzschwelle von $p < 0.01$ fest und detektierten damit 51 signifikante Änderungen in einem Datensatz von 291 Metabolitenpeaks. In einer Monte-Carlo-Analyse generierten wir 1000 permutierte Datensätze, bei denen die Proben zufällig in Ovarialkarzinome und Borderlinetumore eingeteilt wurden. In dieser Weise modellierten wir die Nullverteilung und schätzten aus der Zahl der signifikanten Findungen in den permutierten Datensätzen die false discovery rate (FDR) ab. Für die Signifikanzschwelle $p \leq 0.01$ ergab sich eine moderate FDR von 7.8%. In der Studie zum Kolonkarzinom untersuchten wir in einem Datensatz von 206 Metabolitenpeaks die drei verschiedenen Schwellenwerte $p = 0.05$, 0.01 und 0.0024, von denen der letztgenannte Schwellenwert einer Bonferronikorrektur entspricht. Für jede dieser Schwellen führten wir eine Permutationsanalyse durch und erhielten die Abschätzungen FDR=8.2%, 2.3% und 0.2%. In der Studie zum Mammakarzinom forderten wir im Trainingsdatensatz, wie auch im unabhängigen, prospektiv definierten Validierungsdatensatz eine Signifikanz von $p < 0.05$. (Genauer gesagt forderten wir für die Validierung eine Regulation in dieselbe Richtung und eine Signifikanz $p < 0.05$ beim einseitigen Test.) Bei 468 gemessenen Metabolitenpeaks führt dieses Vorgehen dazu, dass (wegen $468 \times 0.05 \times 0.05 = 1.4$) unter der Nullhypothese maximal mit 1-2 falsch positiven Findungen zu rechnen ist. Zusammengefasst liegt die Zahl der untersuchten Moleküle in einer typischen Metabolomikstudie um zwei Größenordnungen unter der Zahl der in einer genomweiten Expressionsstudie. Dies führt dazu, dass die p -Wert-Korrektur in Metabolomikstudien weniger streng gewählt

werden kann als in Microarraystudien. Daher ist in Metabolomikstudien der Verlust an Detektionssensitivität durch Korrektur für multiples Testen in der Regel moderat.

4.2.2 Unüberwachte Mustererkennung

Zusätzlich zur sukzessiven Analyse der einzelnen Moleküle erlauben -omik-Daten das Studium der Korrelationsstruktur einerseits des gesamten Metaboloms und andererseits der gesamten Patientenkohorte sowie die Verwendung von Mustern aus mehreren oder vielen Metaboliten als Biomarker. Einerseits können Methoden des unüberwachten Lernens verwendet werden, um klinisch relevante Gruppierungen von Patienten oder Metaboliten neu zu entdecken. Andererseits können Methoden des überwachten Lernens dazu verwendet werden, Metabolitenmuster zu identifizieren, die mit einem Krankheitsparameter oder dem klinischen Verlauf der Krankheit korrelieren, also zum Beispiel prognostisch für das klinische Ergebnis oder prädiktiv für das Ansprechen auf eine Therapie sind. Unüberwachte Methoden unterscheiden sich von überwachten Methoden grundsätzlich dadurch, dass erstere auf die Detektion von Strukturen ohne Hinzunahme von externen Parametern zielen, während zweitere die Prädiktion eines klinischen Parameters aus den -omik-Daten zum Ziel haben, der mit in die Analyse einbezogen wird.

Heatmaps mit hierarchischen Clustering von Genen und Patienten sind die verbreitetste Methode zur unüberwachten Analyse von Genexpressionsdaten und zur Visualisierung von Genexpressionmustern. Auch für Metabolomikdaten eignen sich Heatmaps einerseits zur Visualisierung von differenziell regulierten Metaboliten, andererseits aber auch als unüberwachte Methode zur Entdeckung neuer Krankheitssubtypen. In unserer GC-TOFMS-Studie führte ein hierarchisches Clustering über alle Metaboliten zur Aufspaltung der Mammakarzinome in zwei deutlich getrennte Patientencluster [75]. Ein molekularer Unterschied zwischen den Geweben in den beiden Clustern bestand in ihrem unterschiedlichen Aminosäuregehalt: Die Gewebe des ersten Clusters wiesen ein viel höheres Glutamate-Glutamine-Verhältnis (GGR) als die Gewebe des zweiten Clusters auf. Die hier beschriebene Clusteringstruktur fanden wir im Trainingsdatensatz des METAcancer-Projekts und konnten sie im Validierungsdatensatz validieren.

4.2.3 Überwachte Mustererkennung

Methoden des überwachten Lernens (im Folgenden synonym auch als „Klassifikation“ bezeichnet) können dazu verwendet werden mittels -omik-Daten einen molekularen Test zu konstruieren. Ziel eines solchen Tests ist es, basierend auf einem molekularen Muster eine Diagnose zu stellen, eine Aussage über die Prognose zu gewinnen oder das Ansprechen auf eine Therapie vorherzusagen. Wie in der Einleitung diskutiert, ist es bei der Entwicklung solcher Tests von großer Bedeutung durch die Anwendung angemessener bioinformatischer Methoden einerseits *overfitting* und andererseits eine Überinterpretation der Klassifikationsergebnisse streng zu vermeiden. Zu diesem Zweck lehnten wir die Validierung molekularer Tests in unseren Metabolomikstudien an das *multiple random cross-validation protocol* an, das im Jahr 2005 in einer vielbeachteten Studie von Michiels et al. vorgestellt wurde [69]. In diesem Protokoll wird eine große Zahl von Trainings- und Testdatensätzen zufällig gezogen und analysiert. Es handelt sich damit um eine vollständigere Validierung als die klassische Kreuzvalidierung und um die gründlichste Art der Validierung überhaupt. Mit dieser Methode validierten wir die Trennbarkeit invasiver Ovarialkarzinome von Borderlinetumoren [45] und die Trennbarkeit von Tumor- und Nor-

malgewebe des Kolonkarzinoms [72]. Basierend auf dem multiple random validation protocol entwickelten wir das R-Paket cancerclass zur Entwicklung und Validierung molekularer Tests [61]. In den Mammakarzinomstudien verwendeten wir cancerclass, um nach Surrogat-Klassifikatoren aus Metaboliten oder Genen für das Glutamat/Glutamin-Verhältnis (GGR) zu suchen [75]. Es stellte sich heraus, dass eine Klassifikation in Glutamat-angereicherte und Glutamat-normale Mammakarzinome statt mit dem GGR sehr gut mit einem 25-Metaboliten-Klassifikator, jedoch weniger gut mit einem 100-Gen-Klassifikator durchgeführt werden kann. Insgesamt hat sich das multiple random validation protocol und seine Implementation im R-Paket cancerclass inklusive der Visualisierungsmöglichkeiten als sehr geeignet und hilfreich für die überwachte Analyse von Metabolomikdaten herausgestellt.

Es gibt eine große Zahl von Algorithmen für Probleme des überwachten Lernens, von denen viele bereits auf molekulare Hochdurchsatzdaten angewendet wurden. In der Ovarialstudie und in der Kolonstudie verglichen wir jeweils die folgenden vier Methoden miteinander: lineare Diskriminanzanalyse (LDA), *linear support vector machines* (SVM), *nearest mean classification* (NMC) und *nearest centroid classification* (NCC). Für LDA und SVM verwendeten wir dabei die Implementierungen in den R Paketen MASS und e1071. Für NMC und NCC verwendeten wir eigene Implementierungen wie in den Referenzen [57] und [69] beschrieben. Alle Methoden erlauben eine gute Trennung von Ovarialkarzinomen und Borderlinetumoren und von Kolonkarzinomen und Kolonnormalgewebe. Das beste Ergebnis wurde in beiden Datensätzen mit NCC erzielt, das zweitbeste mit NMC. Die hervorragenden Ergebnisse der Algorithmen NCC und NMC sind im Einklang mit den Erfahrungen anderer Autoren, dass einfache Klassifikationsverfahren für molekulare Hochdurchsatzdaten oft ebenso gut oder sogar besser abschneiden als komplexere Verfahren [57, 58]. Zusätzlich untersuchten wir den Einfluß einer *feature selection* auf die Klassifikationsergebnisse. Einerseits kam es im Ovar- und im Kolondatensatz durch *feature selection* nicht zu einer Verbesserung der Klassifikationsergebnisse im Vergleich zur Klassifikation mit allen 291 bzw. 206 Metaboliten. Andererseits reichten in beiden Datensätzen bereits 10 Metaboliten für ein optimales Ergebnis aus. Zusammengefasst war für die beiden GC-TOFMS Datensätze mit wenigen hundert Metaboliten anders als bei Genexpressiondatensätzen mit zehntausenden Genen eine feature selection nicht essentiell, aber möglich.

4.2.4 Funktionelle Analyse

PROFILE Clustering ist eine von uns neu entwickelte Methode zur Visualisierung und Interpretation von Metabolitenänderungen in Kontext der biochemischen Stoffwechselwege [72]. Das Akronym PROFILE leitet sich von *projection from interaction lattice* ab und bezeichnet die Projektion des komplexen Netzwerks der Stoffwechselreaktionen auf ein hierarchisches Clustering einer eindimensionalen Anordnung von Metaboliten. Die Projektion entsteht dadurch, dass die Metaboliten bezüglich ihrer Distanz im Netzwerk der biochemischen Reaktionen geclustert werden. Wichtig ist dabei, dass für jede der biochemischen Reaktionen die Hauptreaktionspartner bekannt sind, da eine Distanz von eins zwischen zwei Metaboliten nur dann vorliegen soll, wenn eine enzymatische Reaktion existiert, in der das eine Metabolit Substrat und das andere Metabolit Produkt ist. Unsere Implementation des PROFILE Clustering verwendet dazu die Information aus der Datenbank KEGG RPAIR. Für die funktionelle Analyse von Metaboliten-

Veränderungen in GC-TOFMS-Daten erwies sich PROFILE Clustering sehr nützlich: Fast alle identifizierten Metaboliten waren in der KEGG annotiert, so dass die gesamten Metabolitenänderungen auf einen Blick visualisiert und funktionell klassifiziert werden konnten. In den hier vorgelegten Studien haben wir PROFILE-Clusterings der Metabolitenunterschiede zwischen Tumor- und Normalgewebe des Kolons und der Mamma, sowie zwischen Glutamat-angereicherten und Glutamat-normalen Mammakarzinomen erstellt [72, 73, 75].

Mit METAtarget stellten wir eine Methode des *reverse engineering* vor, die zum Ziel hat, aus Metabolitenänderungen Hypothesen über die zugrunde liegenden Änderungen der Enzymexpression zu generieren [80]. Zu diesem Zweck werden Änderungen des Gleichgewichts zwischen Produkt und Substrat einer enzymatischen Reaktion untersucht. Für den GC-TOFMS-Datensatz zum Mammakarzinom war die Analyse von 91 Produkt-Substratpaaren möglich, von denen 13 signifikant verschieden zwischen (ER-)-Mammakarzinom und (ER+)-Mammakarzinom waren. Unter den signifikanten Reaktionen war die Umwandlung von Glukose zu Glukose-6-Phosphat, die den ersten Schritt der Glykolyse darstellt, und die Umwandlung von Glutamin zu Glutamate, die den ersten Schritt der Glutaminolyse darstellt. Beide Reaktionsgleichgewichte waren im (ER-)Mammakarzinom zum Produkt hin verschoben. Für die Umwandlung von Glukose zu Glukose-6-Phosphat wurden 4 Reaktionen und 6 katalysierende Enzyme detektiert, während für die Umwandlung von Glutamin zu Glutamat 10 Reaktionen und 13 katalysierende Enzyme detektiert wurden. In einem externen Genexpressionsdatensatz waren zwischen (ER-) und (ER+)-Mammakarzinom die Hexokinase 3 (HK3) differentiell expremiert (Umwandlung Glukose zu Glukose-6-Phosphat) sowie 11 Enzyme, die Glutamin in Glutamat umwandeln. Insbesondere war dabei die Glutaminase 1 (GLS) heraufreguliert im (ER-)-Mammakarzinom, während Glutaminase 2 (GLS2) und die Glutamin-Synthetase (GLUL) herunterreguliert waren. Zusammengefasst konnte in dem GC-TOFMS-Datensatz wegen der begrenzten Überdeckung des humanen Metaboloms eine begrenzte Anzahl von Substrat-Produktpaaren untersucht werden. Interessante Perspektiven eröffnen sich durch den zweiten Schritt der METAtarget-Analyse, bei dem für jedes der signifikanten Produkt-Substratpaare alle relevanten Reaktionen und die entsprechenden katalysierenden Enzyme aufgelistet werden. Die identifizierten Enzyme können in Genexpression- oder Proteomikstudien weiter untersucht werden und stellen potentielle Targets für eine therapeutische Intervention dar.

4.3 GC-TOFMS-Analyse verschiedener Karzinome

Die vorgestellten Metabolomikstudien umfassen den Vergleich von Ovarialkarzinomen mit Borderline-Tumoren des Ovars, von Kolonkarzinomen mit Kolonnormalgewebe, von Mammakarzinomen mit Mammanormalgewebe und von hormonrezeptor-negativen mit hormonrezeptor-positiven Mammakarzinomen. Die Anzahl der detektierten Metabolitenpeaks betrug 291 in der Ovarstudie, 206 in der Kolonstudie und 468 in der Mammastudie, von denen 116, 107 und 162 auf definierte Metaboliten abgebildet werden konnten. In der Mammastudie, die etwa 5 Jahre später als die beiden anderen Studien durchgeführt wurde, konnte die Zahl der identifizierten Metaboliten ungefähr um das Anderthalbfache gesteigert werden. Die Verbesserung der Abdeckung liegt vor allem an einer größeren Datenbank von Referenzspektren, die in dieser Zeit kontinuierlich erweitert wurde.

Metabolit	Ovarialkarzinom vs. Borderline- Tumor		Kolonkarzinom vs. Kolonnor- malgewebe		Mammakarzinom vs. Mammanor- malgewebe		ER- Mammakar- zinom vs. ER+ Mammakarzinom	
	fc	p	fc	p	fc	p	fc	p
Uracil	4,2	6,2E-06	2,0	3,1E-05	2,4	8,5E-30	1,6	2,2E-08
Beta-Alanin	1,5	0,27	4,9	5,8E-13	2,6	1,1E-34	2,4	1,5E-20
Glutamin	2,0	0,0083	-2,1	0,00089	-1,1	0,4	-2,2	2,7E-06
Glutamat	3,0	0,00018	1,6	0,0012	6,5	1,6E-40	1,5	1,2E-05
Palmitinsäure	-1,3	0,022	-1,4	0,0037	-2,7	1,7E-36	1,0	0,58
Stearinsäure	-1,4	0,0096	-1,3	0,06	-2,6	2,5E-34	1,0	0,61

Tabelle 1: Vergleich der Metabolitenänderungen in vier GC-TOFMS-Studien. Die Tabelle zeigt fold changes (fc) und Signifikanzen (p, ohne Korrektur für multiples Testen) für sechs besonders interessante Metaboliten.

Tabelle 1 zeigt eine Übersicht über fünf besonders interessante Biomarker in verschiedenen Tumor- und Normalgeweben. Das Nucleotid Uracil fiel bereits in unserer ersten Metabolomikstudie als 4,2-fach und hochsignifikant erhöht in Ovarialkarzinomen gegenüber Borderlinetumoren auf und war damit das am zweistärksten regulierte Metabolit in der gesamten Studie. Im Vergleich von Kolonkarzinomen und Mammakarzinomen mit entsprechenden Normalgeweben ergab sich jeweils eine über 2-fache und hochsignifikante Hochregulation. Schließlich war Uracil im (ER-)-Mammakarzinom 1,6-fach und hochsignifikant erhöht im Vergleich zum (ER+)-Mammakarzinom. Interessanterweise stach Uracil mit diesem Konzentrationsmuster gegenüber anderen Nucleosiden und Nucleotiden heraus. In der Brustkrebsstudie konnten neben Uracil auch Uridine, Adenin, Adenosine, Guanin und Guanosine detektiert werden. Jedoch keines dieser Metaboliten reicherte sich signifikant im ER- Subtypen des Mammakarzinoms im Vergleich um ER+ Subtyp an. Unsere wie auch folgende Ergebnisse anderer Arbeitsgruppen lassen Uracil als interessanten Tumormarker erscheinen: In einer 1H -NMR-Metabolomikstudie war Uracil signifikant erhöht in den Stadien I-IV des Rektumkarzinoms im Vergleich zum entsprechendem Normalgewebe [81]. Manna et al. fanden Uracil hochreguliert im Urin eines Azoxymethan-induzierten Mausmodells des Kolonkarzinoms sowie in einer gepaarten Analyse von Tumor- und Normalgewebeprobe des humanen Kolonkarzinoms [82].

Beta-Alanin war im (ER-)-Mammakarzinom im Vergleich zum (ER+)-Mammakarzinom 2,4-fach erhöht und damit das am stärksten und am signifikantesten regulierte Metabolit in diesem Vergleich. In einer Korrelationsstudie mit Enzymexpressiondaten im Mammakarzinom detektierten wir eine starke und hochsignifikante negative Korrelation von $R = -0,62$ von Beta-Alanin mit der RNA der Aminotransferase ABAT. Da ABAT den Abbau von Beta-Alanin zu Malonate-Semialdehyde katalysiert, legen die detektierten Korrelationen nahe, dass die höhere Expression von ABAT im (ER-)-Mammakarzinom dort zu einer Akkumulation von Beta-Alanin führen. Diese Hypothese müsste in mechanistischen Studien, zum Beispiel an Zellkulturen, validiert werden. Unsere Studie zum Mammakarzinom ist die erste zu einer möglichen Rolle von Beta-Alanin und ABAT in malignen Tumoren. Ebenfalls signifikant erhöht war Beta-Alanin im Mammakarzinom (2,6-fach) und im Kolonkarzinom (4,9-fach) im Vergleich zu den entsprechenden Nor-

malgeweben.

Therapeutische Strategien, die auf den veränderten Glutaminmetabolismus in Krebszellen zielen sind zahlreich [75]. Eine neue Entwicklung in diesem Feld setzte mit der Entdeckung der Glutaminase-Inhibitoren BPTES [20] und 968 [21] ein, die beide spezifisch das Enzym Glutaminase 1 (GLS) hemmen. CB-839 [22], ein strukturelles Analog von BPTES, wird seit Anfang 2014 in klinischen Studien (derzeit Phase I) in soliden und hämatologischen Tumoren getestet. Diese pharmakologische Entwicklung hat uns motiviert Glutamin und Glutamat im Mammakarzinom näher zu untersuchen [75]: Im Vergleich von Mammakarzinom mit Normalgeweben zeigte sich eine signifikante Erhöhung von Glutamat, während Glutamin unverändert war. Der Vergleich des (ER-)-Mammakarzinoms mit dem (ER+)-Mammakarzinom zeigt eine signifikante Erhöhung von Glutamat und eine signifikante Verminderung von Glutamin. In diese Studie führten wir die Glutamat/Glutamine-Ratio (GGR) als neuen Biomarker ein und definierten Gewebe als Glutamat-angereicht, bei dem GGR über dem Level aller Normalgewebe im Trainingsdatensatz lag, während Gewebe, bei denen das nicht der Fall war, als Glutamat-normal definiert wurden. Mit dieser Definition waren 88% der (ER-)-Tumore und 56% (ER+)-Tumore und nur 2% der Normalgewebe Glutamat-angereicht. Auf dieser Studie an klinischen Mammakarzinomgeweben erscheint eine Testung der neuen Glutamat-Inhibitoren besonders interessant in (ER-)Tumoren und in ausgewählten (ER+)-Tumoren. Kompatibel zu diesem Ergebnis sind Resultate aus Zellkulturen und Mausmodellen, die auf das triple-negative Mammakarzinom als besonders geeignetes Target für CB-839 weisen [22].

Vergleichend haben wir in Tabelle 1 die Regulation von Glutamin und Glutamat auch in den anderen von uns untersuchten Tumorentitäten zusammengestellt: Im Kolonkarzinom war Glutamat im Vergleich zum Normalgewebe signifikant erhöht, während Glutamin signifikant vermindert war. Im Ovarialkarzinom waren sowohl Glutamat als auch Glutamin im Vergleich zu Borderlinetumoren signifikant erhöht. Zusammengefasst weisen alle drei untersuchten Karzinomentitäten Veränderungen im Glutaminmetabolismus auf, die eine Testung der neuen Glutaminase-Inhibitoren interessant erscheinen lassen. Ob es im Ovarial- und Kolonkarzinom ausgezeichnete Subgruppen von Tumoren gibt, bei denen eine Anwendung besonders erfolgversprechend ist, müsste in detaillierten Analysen untersucht werden.

Konsistent über die drei Tumorentitäten beobachteten wir eine Verminderung vieler freier Fettsäuren in den Karzinomgeweben im Vergleich zu Normalgeweben oder Borderlinetumoren. Beispielhaft für diese Beobachtung haben wir in Tabelle 1 Palmitinsäure (C16:0) und Stearinsäure (C18:0) aufgeführt, die in allen drei Tumorentitäten signifikant vermindert waren. Nicht verändert waren die freien Fettsäuren zwischen (ER-)-Mammakarzinomen und (ER+)-Mammakarzinomen. Ferner zeigten wir in der Studie über 271 Mammakarzinome und 98 Mammanormalgewebe, dass die freien Fettsäuren C14:0, C15:0, C16:0, C17:0, C18:0 und C20:0 stark miteinander korrelieren (alle paarweisen Pearson-Korrelationen $> 0,82$) [73]. Die Herunterregulation der freien Fettsäuren in Karzinomen scheint im Widerspruch der Hochregulation der Fettsäure-Synthase (*fatty acid synthase*, FASN) in vielen Krebsgeweben zu stehen. In der Tat zeigte eine Lipidomikstudie der METAcancer-Kohorte eine Erhöhung von vielen Membranlipiden wie Phosphatidylethanolaminen und Phosphatidylcholinen [70]. Die Daten der Metabolomikstudie zusammen mit denen aus der Lipidomikstudie zum Mammakarzinom unterstützen die Hypothese, dass die Synthese freier Fettsäuren erhöht ist, diese aber schnell weiter zu

Membran-Phospholipiden metabolisiert werden. Dazu passt unsere Beobachtung einer starken Erhöhung von Phospho-Ethanolamin und einer Verminderung von Ethanolamin im Mammakarzinom im Vergleich zu Mammanormalgewebe, d.h. einer Verschiebung der Regulation des Kennedy-Stoffwechselwegs in Richtung einer stärkeren Synthese von Membranlipiden [73].

4.4 2-Hydroxyglutarat (2-HG) als Onkometabolit

Während Warbureffekt und andere Stoffwechselveränderungen in Krebszellen schon seit über 50 Jahren bekannt sind, verstärkt sich seit einigen Jahren die Evidenz für die Beteiligung von Onkometaboliten an der Entstehung und Progression von Tumoren [83]. Diese Entwicklung nahm ihren Anfang mit der erstmaligen Identifikation des Onkometabolits 2-Hydroxyglutarat (2-HG) in einer vergleichenden Metabolomikanalyse von Isozitatdehydrogenase 1 (IDH1) mutierten Zellen mit IDH1 wildtypischen Zellen [84]. In dieser Arbeit wurde gezeigt, dass eine Mutation in der IDH1 von Arginine 132 zu Histidin zu einem gain of function führt, der die Reduktion von Alpha-Ketoglutarat zu 2-HG erlaubt. Die Motivation zum Studium von IDH1-Mutationen kam aus der Beobachtung, dass IDH1 oder IDH2 in ungefähr 10% der Glioblastome und mehr als 70% der Grad II-III Gliome mutiert sind [85, 86]. Zudem sind mittlerweile weitere Tumorentitäten bekannt, in denen IDH1 oder IDH2 Mutationen vorkommen, so z.B. in etwa 20% aller AML-Patienten [83]. Im Mammakarzinom wurden hingegen überhaupt keine Mutationen in IDH1 oder IDH2 gefunden [86, 87].

Systematisch gehört 2-HG zum Glutamin-Metabolismus und ordnet sich im PROFILE-Clustering in das entsprechende Cluster ein [73]: Der Vergleich von Mammakarzinomen zu Mammanormalgeweben neben der Erhöhung von Glutamate und Oxoproline und einer Verminderung von Alpha-Ketoglutarat eine Erhöhung von 2-HG. Im Vergleich von (ER-)-Tumoren mit (ER+)-Tumoren zeigte sich neben einer Erhöhung von Glutamat und einer Erniedrigung von Glutamin eine Erhöhung von 2-HG [76]. In einer großen amerikanischen Brustkrebsstudie wurden ebenfalls stark erhöhte 2-HG Konzentrationen in den Gewebeproben eines Teils der Patienten gefunden [49]. In der amerikanischen Arbeit wurden sehr umfangreiche Untersuchungen sowohl zur Analyse der Ursachen der 2-HG-Anreicherung als auch zur Analyse ihrer Auswirkungen durchgeführt: Einerseits waren Tumore mit 2-HG-Anreicherung stärker methyliert und wiesen ein mit schlechter Prognose assoziiertes Methylierungsmuster auf. Andererseits wurden verschiedene molekulare Veränderungen gefunden, die zu einer Akkumulation von 2-HG beitragen könnten: Zunächst korrelierten 2-HG-Akkumulation und die durch eine Genexpressionsignatur gemessene MYC-Aktivierung mit einer hohen odds ratio $OR=11,3$. Ferner korrelierte die 2-HG-Anreicherung mit einer höheren Expression des Proteins der Glutaminase 1 (GLS1). Schließlich ergaben sich aus Zellkulturexperimenten Hinweise, dass die Hochregulation jedes der Enzyme IDH2, D2HGDG und AHFE1 zur Produktion von 2-HG beitragen kann. Weitere Studien werden nötig sein, um diese molekularen Zusammenhänge genauer zu erforschen und um den Marker 2-HG ggf. in die Diagnostik des Mammakarzinoms zu integrieren und therapeutisch auszunutzen. Neue therapeutische Ansätze könnten sich z.B. aus der Anwendung der neuen Glutaminase-Inhibitoren oder von DNA-demethylierenden Substanzen entwickeln.

4.5 Ausblick

Metabolitenänderungen stellen die Endantwort der Zelle auf ihre genetische Programmierung oder Veränderungen der Zellumgebung dar. Innerhalb der Informationsverarbeitung in der Zelle, die von DNA über Epigenetik und RNA zu den Proteinen führt, befinden sich Metabolitenflüsse ganz am Ende der Kausalkette. Andererseits stehen Metabolitenprofile im Vergleich zu anderen Molekülprofilen Phänotyp und Physiologie der Zelle am nächsten. Metabolomprofile können daher als Surrogat für funktionelle Veränderungen der Zelle verwendet werden, wo diese funktionellen Assays nicht zugänglich sind. Ein faszinierender Aspekt an Metaboliten ist, dass sie nicht wie andere Biomoleküle als eindimensionale Nukleotid- oder Aminosäuresequenz vorliegen, sondern eine riesige Vielfalt an dreidimensionalen räumlichen Strukturen bilden. Diese Komplexität hat zur Folge, dass eine befriedigende Abdeckung des humanen Metaboloms experimentell-analytisch sehr schwierig ist.

Die hier vorgestellten Studien und zahlreiche Publikationen anderer Autoren, z.B. [48, 88, 49] zeigen, dass eine Analyse des Metaboliteninhalts in Gefriergeweben reproduzierbar möglich ist und zur Entwicklung von Biomarkern verwendet werden kann. Bei der von uns benutzten GC-TOFMS handelt es sich um eine Hochdurchsatzmethode, mit der hunderte von Geweben zu (im Vergleich zu anderen -omik-Plattformen) niedrigen Kosten von deutlich unter 100€ pro Probe analysiert werden können. Die von uns verwendete Plattform erlaubte die Detektion von etwa 500 Metabolitenpeaks, von denen etwa 150 auf bekannte Metaboliten abgebildet werden konnten. Die detektierten Metaboliten umfassten u.a. fast alle Aminosäuren, einige Zwischenprodukte der Glykolyse und des TCA-Zyklus, einige Nukleotide und -side, sowie die wichtigsten Zucker und freien Fettsäuren. Folgende Methoden können dabei helfen, die Zahl der detektierten Metaboliten in zukünftigen Studien zu erhöhen: 1. Erweiterung der spektralen Bibliotheken durch Messung bisher nicht enthaltener Standards 2. Strukturprädiktion aus Spektren mit chemoinformatischen Methoden 3. Gleichzeitige Verwendung verschiedener Metabolomik- bzw. Lipidomik-Plattformen. 4. Entwicklung spezialisierter Plattformen zur Analyse bisher nicht erfasster Molekülklassen.

Systembiologie bzw. Systempathologie haben zum Ziel ein integriertes Bild eines biologischen Systems bzw. einer Krankheit auf allen molekularen Ebenen zu entwickeln. Zu betonen ist, dass der Informationsgewinn aus einer integrierten Analyse von zwei oder mehr -omik-Datensätzen viel höher ist als der Informationsgewinn aus dementsprechenden Einzelanalysen. Das liegt daran, dass in der integrierten Analyse die Kreuzkorrelationen zwischen den verschiedenen Datensätzen mit einbezogen werden. Ein Beispiel für die erfolgreiche Anwendung dieser Strategie ist die integrierte Analyse der humanen Blutmetabolom und des Genotyps von über 2.000 Individuen in einer genomweiten Assoziationsstudie [89]. Seit der Entwicklung der Biochemie in der Mitte des 19. Jahrhunderts kam es zu einer sich bis heute fortsetzenden Wissenszunahme über die Funktion von Enzymen und Metaboliten und ihre Anordnung in biochemischen Stoffwechselwegen. Geordnet und gesammelt werden die Informationen über biochemische Umwandlungen und Pathways in Datenbanken, wie zum Beispiel KEGG, HumanCyc und Reactome [65, 66, 67]. Es ist sehr wünschenswert, dieses enorme Wissen für die Krebsforschung auszunutzen. Mit dem PROFILE-Clustering haben wir eine Methode entwickelt, die die Anordnung der Metaboliten in den Pathways auswertet und es erlaubt differentielle Metabolitenprofile übersichtlich zu visualisieren und funktionell zu interpretieren [72].

Die Stärke einer integrierten Analyse von Metabolomikdaten, Genexpressionsdaten und Wissen über metabolische Reaktionen haben wir in einer der hier vorgelegten Mammakarzinomstudien demonstriert [76].

Die Metabolitenflüsse in der Krebszelle werden unmittelbar durch die Konzentrationen vorhandener Stoffwechsellzyme determiniert. Besonders interessant für eine systembiologische Analyse ist daher eine integrierte Analyse von Metabolomik- und Proteomikdaten. Die Proteomanalytik hat sich in den letzten Jahren stark weiterentwickelt und wurde bereits vielfach zur Analyse von Krebsgeweben und Körperflüssigkeiten eingesetzt [90, 88]. Mittels globaler, nicht-zielgerichteter Proteomik können derzeit typischerweise etwa 1000 Proteine detektiert werden und nur bei einem Bruchteil dieser handelt es sich um Stoffwechselproteine. Hingegen werden in solchen globalen Analysen oft für die Krebsforschung uninteressante Strukturproteine mit hoher Intensität und Genauigkeit identifiziert [91]. In Zukunft wäre es daher interessant, zielgerichtete und gleichzeitig sensitivere Proteomikmethoden mit Metabolomik zur Profilierung ausgewählter metabolischer Pathways zu kombinieren. Derzeit gibt es Fortschritte in der zielgerichteten quantitativen Proteomik klinischer Proben bei denen sich ein Paradigmenwechsel vom selected reaction monitoring (SRM) zum parallel reaction monitoring (PRM) abzeichnet [92].

In den nächsten Jahren ist zu erwarten, dass sich der Trend einer steigenden Verwendung von Metabolomik in der Medizin weiter fortsetzt. Die Bedeutung für die Krebsforschung liegt dabei in der Entwicklung von diagnostischen, prognostischen und prädiktiven Biomarkern, von Biomarkern zur Kontrolle der Wirksamkeit einer Therapie sowie in der Identifikation von Zielstrukturen für therapeutische Interventionen. In Zukunft wird durch eine Erweiterung und Zusammenführung von spektralen Datenbanken eine dichtere und vollständigere Abdeckung des Zellmetabolismus durch Metabolomik möglich sein. Außerdem werden durch die Weiter- und Neuentwicklung spezialisierter Analyseplattformen bisher nicht berücksichtigte Metabolitearten und Stoffwechselwege auf ihre Rolle in der Onkologie untersucht werden können. Eine interessante Perspektive ist die Verwendung der Metabolomik in der Systembiologie bzw. Systempathologie, d.h. in integrativen Analysen zusammen mit Mutations-, Genexpressions- und Proteomikdaten.

5 Zusammenfassung

Wegen der Vielfalt an räumlichen Strukturen und biologischen Funktionen bilden Metaboliten eine faszinierende Klasse von Biomolekülen. Metabolitenänderungen stellen die Endantwort der Zelle auf ihre genetische Programmierung oder Veränderungen der Zellumgebung dar. Innerhalb der Informationsverarbeitung in der Zelle, die von DNA über mRNA zu den Proteinen führt, befinden sich Metabolitenflüsse ganz am Ende der Kausalkette, aber nah an Physiologie und Phänotyp der Zelle. Die Metabolomik, definiert als systematische Analyse des Metaboliteninhalts einer Zelle, eines Gewebes oder eines Organismus, hat sich in den letzten 15 Jahren rapide weiterentwickelt. Die hier vorgelegten Studien befassen sich mit der Entwicklung von bioinformatischen Methoden zur funktionellen Analyse von Metabolomikdaten und mit der GC-TOFMS-Analyse von Ovarialtumor-, Kolonkarzinom- und Mammakarzinomgeweben.

Die von uns entwickelten bioinformatischen Methoden umfassen das R-Paket `cancerclass` und die beiden Methoden `PROFILE-Clustering` und `METAtarget` zur funktionellen Analyse von Metabolomikdaten. Bei `cancerclass` handelt es sich um ein R-Paket zur Entwicklung und Validierung von diagnostischen Tests aus molekularen Hochdurchsatzdaten, das u.a. das von Michiels et al. vorgestellte *multiple random cross-validation protocol* [69] implementiert. `PROFILE-Clustering` ermöglicht einen Metabolomikdatensatz im Kontext der biochemischen Stoffwechselwege zu analysieren und zu interpretieren. Dazu werden die gemessenen Metaboliten entsprechend ihrer Distanz im Netzwerk der enzymatischen Reaktionen angeordnet. Untersucht man nun Metabolitenveränderungen zwischen zwei Krankheitszuständen so werden diese durch `PROFILE-Clustering` übersichtlich nach den verschiedenen Stoffwechselwegen gruppiert. `METAtarget` ist ein Algorithmus des *reverse engineering*, der zum Ziel hat, aus Metabolitenveränderungen Hypothesen über die zugrunde liegenden Änderungen der Enzymexpression zu generieren. Bei einer Anwendung auf das Mammakarzinom identifizierte `METAtarget` - mit der Umwandlung von Glukose in Glukose-6-Phosphat und mit der Umwandlung von Glutamin in Glutamat - Regulationen an Schlüsselstellen von Glykolyse und Glutaminolyse. Die Software erlaubt es im zweiten Schritt, alle für diese Umwandlungen potentiell verantwortlichen Enzyme zu identifizieren, die mittels Genexpressionsanalysen oder Proteomik weiter untersucht werden können.

Die vergleichende Analyse von Kolonkarzinom und Mammakarzinom mit entsprechenden Normalgeweben, sowie von Ovarialkarzinomen mit Borderlinetumoren des Ovars zeigte eine sehr gute Trennbarkeit der malignen von den benignen Geweben mittels Metabolitenmustern. Der Goldstandard zur Malignitätsbestimmung ist mikroskopische Histopathologie, die der räumlichen Heterogenität von humanen Geweben in idealer Weise Rechnung trägt. Die hauptsächliche Bedeutung der Metabolomikanalysen liegt darin, systematisch die Änderungen vieler Metaboliten zu erfassen und dadurch zum Verständnis der Tumorbiologie und zur Detektion therapeutischer Zielstrukturen beizutragen.

Die Metabolomanalyse der molekularen Subtypen des Mammakarzinoms ergab eine starke Abhängigkeit des Metabolitenprofils vom Hormonrezeptorstatus, jedoch keine Abhängigkeit vom HER2-Status. Insgesamt wurden 19 zwischen (ER⁻)- und (ER⁺)-Mammakarzinom signifikant verschiedene Metaboliten gefunden, von denen Beta-Alanin mit einer 2,4-fachen Hochregulation in den (ER⁻)-Tumoren die stärkste Veränderung aufwies. Auffällig waren darüber hinaus Unterschiede im Glutamin-Metabolismus mit einer

Hochregulation von Glutamat und 2-HG und einer Herunterregulation von Glutamin. Die Stärke einer Multi-Omik-Analyse zeigte die Korrelationsanalyse Metabolomikdaten mit genomweiten Expressiondaten: Im (ER-)-Mammakarzinom kam es zu einer Verminderung der Genexpression und der Proteinexpression der ABAT-Transaminase, die mit einer Blockade des Beta-Alanin-Abbaus korrelierte.

Die aktuelle Entwicklung neuer spezifischer Glutaminaseinhibitoren, von denen CB-839 in ersten klinischen Studien getestet wird, motivierte uns, den Glutaminmetabolismus im Mammakarzinom genauer zu untersuchen. Wir stellten das Glutamat-Glutamin-Verhältnis (GGR) als einen neuen Biomarker vor, der im Hormonrezeptor-negativen Mammakarzinom signifikant erhöht im Vergleich zum Hormonrezeptor-negativen Mammakarzinom war. Verwendete man das Niveau in Normalgeweben als Bezugsgröße, so wiesen 88% der (ER-)-Tumore und 56% der (ER+)-Tumore ein erhöhtes GGR auf. Gross et al. kamen in Zellkultur- und Tiermodellstudien zu dem Ergebnis, dass insbesondere das tripel-negative Mammakarzinom eine interessante Zielentität für Glutaminaseinhibition darstellt. Unsere Studien unterstützen dieses Ergebnis, fügen ihm aber hinzu, dass auch HER2-positive (ER-)-Mammakarzinome und ausgewählte (ER+)-Mammakarzinome für eine Therapie mit Glutaminaseinhibitoren in Frage kommen könnten.

Insgesamt unterstreichen die hier vorgestellten Studien die Perspektive einer wichtigen Rolle der Metabolomik in der translationalen Tumorforschung. Die Krebsmetabolomik wird in den nächsten Jahren durch weiteren Fortschritt in der Analytik mit besserer Überdeckung des humanen Metaboloms und durch eine zunehmende Bedeutung von systembiologischen Studien mit der Integration mehrerer -omik-Datenarten geprägt sein.

Literatur

- [1] Guido Kroemer and Jacques Pouyssegur. Tumor cell metabolism: cancer's achilles' heel. *Cancer Cell*, 13(6):472–82, Jun 2008.
- [2] Daniel A Tennant, Raúl V Durán, Houda Boulahbel, and Eyal Gottlieb. Metabolic transformation in cancer. *Carcinogenesis*, 30(8):1269–80, Aug 2009.
- [3] Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, 324(5930):1029–33, May 2009.
- [4] Almut Schulze and Adrian L Harris. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature*, 491(7424):364–73, Nov 2012.
- [5] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–74, Mar 2011.
- [6] Georg Löffler, Petro E Petrides, and Peter C Heinrich. *Biochemie & Pathobiochemie*, 8. Auflage. Springer Medizin Verlag Heidelberg, 2007.
- [7] O Warburg, K Posener, and E Negelein. Über den Stoffwechsel der Tumoren. *Biochemische Zeitschrift*, 152:319–344, 1924.
- [8] Thi Bui and Craig B Thompson. Cancer's sweet tooth. *Cancer Cell*, 9(6):419–20, Jun 2006.
- [9] Ralph J DeBerardinis, Julian J Lum, Georgia Hatzivassiliou, and Craig B Thompson. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab*, 7(1):11–20, Jan 2008.
- [10] Kehinde Adekola, Steven T Rosen, and Mala Shanmugam. Glucose transporters in cancer metabolism. *Curr Opin Oncol*, 24(6):650–4, Nov 2012.
- [11] Gary J Kelloff, John M Hoffman, Bruce Johnson, Howard I Scher, Barry A Siegel, Edward Y Cheng, Bruce D Cheson, Joyce O'shaughnessy, Kathryn Z Guyton, David A Mankoff, Lalitha Shankar, Steven M Larson, Caroline C Sigman, Richard L Schilsky, and Daniel C Sullivan. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res*, 11(8):2785–808, Apr 2005.
- [12] S S Gambhir, J Czernin, J Schwimmer, D H Silverman, R E Coleman, and M E Phelps. A tabulated summary of the FDG PET literature. *J Nucl Med*, 42(5 Suppl):1S–93S, May 2001.
- [13] Ralph J DeBerardinis, Anthony Mancuso, Evgueni Daikhin, Ilana Nissim, Marc Yudkoff, Suzanne Wehrli, and Craig B Thompson. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci U S A*, 104(49):19345–50, Dec 2007.

- [14] Weiqin Lu, Helene Pelicano, and Peng Huang. Cancer metabolism: is glutamine sweeter than glucose? *Cancer Cell*, 18(3):199–200, Sep 2010.
- [15] Chi V Dang. Rethinking the warburg effect with myc micromanaging glutamine metabolism. *Cancer Res*, 70(3):859–62, Feb 2010.
- [16] Jing Fan, Jurre J Kamphorst, Robin Mathew, Michelle K Chung, Eileen White, Tomer Shlomi, and Joshua D Rabinowitz. Glutamine-driven oxidative phosphorylation is a major atp source in transformed mammalian cells in both normoxia and hypoxia. *Mol Syst Biol*, 9:712, 2013.
- [17] Harry eagle. Nutrition needs of mammalian cells in tissue culture. *Science*, 122(3168):501–14, Sep 1955.
- [18] Lifeng Yang, Tyler Moss, Lingegowda S Mangala, Juan Marini, Hongyun Zhao, Stephen Wahlig, Guillermo Armaiz-Pena, Dahai Jiang, Abhinav Achreja, Julia Win, Rajesha Roopaimoole, Cristian Rodriguez-Aguayo, Imelda Mercado-Uribe, Gabriel Lopez-Berestein, Jinsong Liu, Takashi Tsukamoto, Anil K Sood, Prahlad T Ram, and Deepak Nagrath. Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol Syst Biol*, 10:728, 2014.
- [19] Jaekyoung Son, Costas A Lyssiotis, Haoqiang Ying, Xiaoxu Wang, Sujun Hua, Matteo Ligorio, Rushika M Perera, Cristina R Ferrone, Edouard Mullarky, Ng Shyh-Chang, Ya’an Kang, Jason B Fleming, Nabeel Bardeesy, John M Asara, Marcia C Haigis, Ronald A DePinho, Lewis C Cantley, and Alec C Kimmelman. Glutamine supports pancreatic cancer growth through a kras-regulated metabolic pathway. *Nature*, 496(7443):101–5, Apr 2013.
- [20] Mary M Robinson, Steven J McBryant, Takashi Tsukamoto, Camilo Rojas, Dana V Ferraris, Sean K Hamilton, Jeffrey C Hansen, and Norman P Curthoys. Novel mechanism of inhibition of rat kidney-type glutaminase by bis-2-(5-phenylacetamido-1,2,4-thiadiazol-2-yl)ethyl sulfide (BPTES). *Biochem J*, 406(3):407–14, Sep 2007.
- [21] Jian-Bin Wang, Jon W Erickson, Reina Fuji, Sekar Ramachandran, Ping Gao, Ramani Dinavahi, Kristin F Wilson, Andre L B Ambrosio, Sandra M G Dias, Chi V Dang, and Richard A Cerione. Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell*, 18(3):207–19, Sep 2010.
- [22] Matt I Gross, Susan D Demo, Jennifer B Dennison, Lijing Chen, Tania Chernov-Rogan, Bindu Goyal, Julie R Janes, Guy J Laidig, Evan R Lewis, Jim Li, Andrew L Mackinnon, Francesco Parlati, Mirna L M Rodriguez, Peter J Shwonek, Eric B Sjogren, Timothy F Stanton, Taotao Wang, Jinfu Yang, Frances Zhao, and Mark K Bennett. Antitumor activity of the glutaminase inhibitor CB-839 in triple-negative breast cancer. *Mol Cancer Ther*, Mar 2014.
- [23] Claudio R Santos and Almut Schulze. Lipid metabolism in cancer. *FEBS J*, 279(15):2610–23, Aug 2012.
- [24] Grace Medes, Alice Thomas, and Sidney Weinhouse. Metabolism of neoplastic tissue. iv. a study of lipid synthesis in neoplastic tissue slices in vitro. *Cancer Res*, 13(1):27–9, Jan 1953.

- [25] J I Epstein, M Carmichael, and A W Partin. OA-519 (fatty acid synthase) as an independent predictor of pathologic state in adenocarcinoma of the prostate. *Urology*, 45(1):81–6, Jan 1995.
- [26] A Rashid, E S Pizer, M Moga, L Z Milgraum, M Zahurak, G R Pasternack, F P Kuhajda, and S R Hamilton. Elevated expression of fatty acid synthase and fatty acid synthetic activity in colorectal neoplasia. *Am J Pathol*, 150(1):201–8, Jan 1997.
- [27] L Z Milgraum, L A Witters, G R Pasternack, and F P Kuhajda. Enzymes of the fatty acid synthesis pathway are highly expressed in in situ breast carcinoma. *Clin Cancer Res*, 3(11):2115–20, Nov 1997.
- [28] Georgia Hatzivassiliou, Fangping Zhao, Daniel E Bauer, Charalambos Andreadis, Anthony N Shaw, Dashyant Dhanak, Sunil R Hingorani, David A Tuveson, and Craig B Thompson. ATP citrate lyase inhibition can suppress tumor cell growth. *Cancer Cell*, 8(4):311–21, Oct 2005.
- [29] Kristine Glunde, Zaver M Bhujwala, and Sabrina M Ronen. Choline metabolism in malignant transformation. *Nat Rev Cancer*, 11(12):835–48, Dec 2011.
- [30] J R Griffiths, E Cady, R H Edwards, V R McCready, D R Wilkie, and E Wiltshaw. ³¹P-NMR studies of a human tumour in situ. *Lancet*, 1(8339):1435–6, Jun 1983.
- [31] W T Evanochko, T C Ng, M B Lilly, A J Lawson, T H Corbett, J R Durant, and J D Glickson. In vivo ³¹P NMR study of the metabolism of murine mammary 16/c adenocarcinoma and its response to chemotherapy, x-radiation, and hyperthermia. *Proc Natl Acad Sci U S A*, 80(2):334–8, Jan 1983.
- [32] P F Daly, R C Lyon, P J Faustino, and J S Cohen. Phospholipid metabolism in cancer cells monitored by ³¹P NMR spectroscopy. *J Biol Chem*, 262(31):14875–8, Nov 1987.
- [33] E O Aboagye and Z M Bhujwala. Malignant transformation alters membrane choline phospholipid metabolism of human mammary epithelial cells. *Cancer Res*, 59(1):80–4, Jan 1999.
- [34] Eugenia E Calle and Rudolf Kaaks. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*, 4(8):579–91, Aug 2004.
- [35] Aya Kuchiba, Teppei Morikawa, Mai Yamauchi, Yu Imamura, Xiaoyun Liao, Andrew T Chan, Jeffrey A Meyerhardt, Edward Giovannucci, Charles S Fuchs, and Shuji Ogino. Body mass index and risk of colorectal cancer according to fatty acid synthase expression in the nurses’ health study. *J Natl Cancer Inst*, 104(5):415–20, Mar 2012.
- [36] David S Wishart, Timothy Jewison, An Chi Guo, Michael Wilson, Craig Knox, Yifeng Liu, Yannick Djoumbou, Rupasri Mandal, Farid Aziat, Edison Dong, Souhaila Bouatra, Igor Sinelnikov, David Arndt, Jianguo Xia, Philip Liu, Faizath Yalou, Trent Bjorndahl, Rolando Perez-Pineiro, Roman Eisner, Felicity Allen, Vanessa Neveu, Russ Greiner, and Augustin Scalbert. Hmdb 3.0—the human metabolome database in 2013. *Nucleic Acids Res*, 41(Database issue):D801–7, Jan 2013.

- [37] J K Nicholson, J C Lindon, and E Holmes. 'metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11):1181–9, Nov 1999.
- [38] O Fiehn, J Kopka, P Dörmann, T Altmann, R N Trethewey, and L Willmitzer. Metabolite profiling for plant functional genomics. *Nat Biotechnol*, 18(11):1157–61, Nov 2000.
- [39] Lloyd W Sumner, Pedro Mendes, and Richard A Dixon. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, 62(6):817–36, Mar 2003.
- [40] L M Raamsdonk, B Teusink, D Broadhurst, N Zhang, A Hayes, M C Walsh, J A Berden, K M Brindle, D B Kell, J J Rowland, H V Westerhoff, K van Dam, and S G Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol*, 19(1):45–50, Jan 2001.
- [41] S Mazurek, W Zwerschke, P Jansen-Dürr, and E Eigenbrodt. Metabolic cooperation between different oncogenes during cell transformation: interaction between activated RAS and HPV-16 E7. *Oncogene*, 20(47):6891–8, Oct 2001.
- [42] John R Griffiths and Marion Stubbs. Opportunities for studying cancer by metabolomics: preliminary observations on tumors deficient in hypoxia-inducible factor 1. *Adv Enzyme Regul*, 43:67–76, 2003.
- [43] Kunle Odunsi, Robert M Wollman, Christine B Ambrosone, Alan Hutson, Susan E McCann, Jonathan Tammela, John P Geisler, Gregory Miller, Thomas Sellers, William Cliby, Feng Qian, Bernadette Keitz, Marilyn Intengan, Shashikant Lele, and James L Alderfer. Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *Int J Cancer*, 113(5):782–8, Feb 2005.
- [44] Huanwen Chen, Zhengzheng Pan, Nari Talaty, Daniel Raftery, and R Graham Cooks. Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. *Rapid Commun Mass Spectrom*, 20(10):1577–84, 2006.
- [45] Carsten Denkert, Jan Budczies, Tobias Kind, Wilko Weichert, Peter Tablack, Jalid Sehouli, Silvia Niesporek, Dominique Könsgen, Manfred Dietel, and Oliver Fiehn. Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res*, 66(22):10795–804, Nov 2006.
- [46] Vladimir Shulaev. Metabolomics technology and bioinformatics. *Brief Bioinform*, 7(2):128–39, Jun 2006.
- [47] 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. *Anal Chem*, 81(16):6656–67, Aug 2009.

- [48] Arun Sreekumar, Laila M Poisson, Thekkelnaycke M Rajendiran, Amjad P Khan, Qi Cao, Jindan Yu, Bharathi Laxman, Rohit Mehra, Robert J Lonigro, Yong Li, Mukesh K Nyati, Aarif Ahsan, Shanker Kalyana-Sundaram, Bo Han, Xuhong Cao, Jaeman Byun, Gilbert S Omenn, Debashis Ghosh, Subramaniam Pennathur, Danny C Alexander, Alvin Berger, Jeffrey R Shuster, John T Wei, Sooryanarayana Varambally, Christopher Beecher, and Arul M Chinnaiyan. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, 457(7231):910–4, Feb 2009.
- [49] Atsushi Terunuma, Nagireddy Putluri, Prachi Mishra, Ewy A Mathé, Tiffany H Dorsey, Ming Yi, Tiffany A Wallace, Haleem J Issaq, Ming Zhou, J Keith Killian, Holly S Stevenson, Edward D Karoly, King Chan, Susmita Samanta, Darue Prieto, Tiffany Y T Hsu, Sarah J Kurley, Vasanta Putluri, Rajni Sonavane, Daniel C Edelman, Jacob Wulff, Adrienne M Starks, Yimmeng Yang, Rick A Kittles, Harry G Yfantis, Dong H Lee, Olga B Ioffe, Rachel Schiff, Robert M Stephens, Paul S Meltzer, Timothy D Veenstra, Thomas F Westbrook, Arun Sreekumar, and Stefan Ambbs. Myc-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J Clin Invest*, 124(1):398–412, Jan 2014.
- [50] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300, 1995.
- [51] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.
- [52] Falk Hlubek, Thomas Brabletz, Jan Budczies, Sabine Pfeiffer, Andreas Jung, and Thomas Kirchner. Heterogeneous expression of wnt/beta-catenin target genes within colorectal cancer. *Int J Cancer*, 121(9):1941–8, Nov 2007.
- [53] Jan Budczies, Wilko Weichert, Aurelia Noske, Berit Maria Müller, Claudia Weller, Timo Wittenberger, Hans-Peter Hofmann, Manfred Dietel, Carsten Denkert, and Volker Gekeler. Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *J Histochem Cytochem*, 59(2):146–57, Feb 2011.
- [54] C M Perou, T Sørlie, M B Eisen, M van de Rijn, S S Jeffrey, C A Rees, J R Pollack, D T Ross, H Johnsen, L A Akslen, O Fluge, A Pergamenschikov, C Williams, S X Zhu, P E Lønning, A L Børresen-Dale, P O Brown, and D Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–52, Aug 2000.
- [55] T Sørlie, C M Perou, R Tibshirani, T Aas, S Geisler, H Johnsen, T Hastie, M B Eisen, M van de Rijn, S S Jeffrey, T Thorsen, H Quist, J C Matese, P O Brown, D Botstein, P E Lønning, and A L Børresen-Dale. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98(19):10869–74, Sep 2001.
- [56] Therese Sorlie, Robert Tibshirani, Joel Parker, Trevor Hastie, J S Marron, Andrew Nobel, Shihong Deng, Hilde Johnsen, Robert Pesich, Stephanie Geisler, Janos Demeeter, Charles M Perou, Per E Lønning, Patrick O Brown, Anne-Lise Børresen-Dale,

and David Botstein. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*, 100(14):8418–23, Jul 2003.

- [57] Lodewyk F A Wessels, Marcel J T Reinders, Augustinus A M Hart, Cor J Veenman, Hongyue Dai, Yudong D He, and Laura J van't Veer. A protocol for building and evaluating predictors of disease state based on microarray data. *Bioinformatics*, 21(19):3755–62, Oct 2005.
- [58] Sandrine Dudoit, Jane Fridlyand, and Terence P Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.
- [59] Leming Shi, Gregory Campbell, Wendell D Jones, Fabien Campagne, Zhining Wen, Stephen J Walker, Zhenqiang Su, Tzu-Ming Chu, Federico M Goodsaid, Lajos Pusztai, John D Jr Shaughnessy, André Oberthuer, Russell S Thomas, Richard S Paules, Mark Fielden, Bart Barlogie, Weijie Chen, Pan Du, Matthias Fischer, Cesare Furlanello, Brandon D Gallas, Xijin Ge, Dalila B Megherbi, W Fraser Symmans, May D Wang, John Zhang, Hans Bitter, Benedikt Brors, Pierre R Bushel, Max Bylesjo, Minjun Chen, Jie Cheng, Jing Cheng, Jeff Chou, Timothy S Davison, Mauro Delorenzi, Youping Deng, Viswanath Devanarayan, David J Dix, Joaquin Dopazo, Kevin C Dorff, Fathi Elloumi, Jianqing Fan, Shicai Fan, Xiaohui Fan, Hong Fang, Nina Gonzaludo, Kenneth R Hess, Huixiao Hong, Jun Huan, Rafael A Iriazary, Richard Judson, Dilafruz Juraeva, Samir Lababidi, Christophe G Lambert, Li Li, Yanen Li, Zhen Li, Simon M Lin, Guozhen Liu, Edward K Lobenhofer, Jun Luo, Wen Luo, Matthew N McCall, Yuri Nikolsky, Gene A Pennello, Roger G Perkins, Reena Philip, Vlad Popovici, Nathan D Price, Feng Qian, Andreas Scherer, Tieliu Shi, Weiwei Shi, Jaeyun Sung, Danielle Thierry-Mieg, Jean Thierry-Mieg, Venkata Thodima, Johan Trygg, Lakshmi Vishnuvajjala, Sue Jane Wang, Jianping Wu, Yichao Wu, Qian Xie, Waleed A Yousef, Liang Zhang, Xuegong Zhang, Sheng Zhong, Yiming Zhou, Sheng Zhu, Dhivya Arasappan, Wenjun Bao, Anne Bergstrom Lucas, Frank Berthold, Richard J Brennan, Andreas Buness, Jennifer G Catalano, Chang Chang, Rong Chen, Yiyu Cheng, Jian Cui, Wendy Czika, Francesca Demichelis, Xutao Deng, Damir Dosymbekov, Roland Eils, Yang Feng, Jennifer Fostel, Stephanie Fulmer-Smentek, James C Fuscoe, Laurent Gatto, Weigong Ge, Darlene R Goldstein, Li Guo, Donald N Halbert, Jing Han, Stephen C Harris, Christos Hatzis, Damir Herman, Jianping Huang, Roderick V Jensen, Rui Jiang, Charles D Johnson, Giuseppe Jurman, Yvonne Kahlert, Sadik A Khuder, Matthias Kohl, Jianying Li, Li Li, Menglong Li, Quan-Zhen Li, Shao Li, Zhiguang Li, Jie Liu, Ying Liu, Zhichao Liu, Lu Meng, Manuel Madera, Francisco Martinez-Murillo, Ignacio Medina, Joseph Meehan, Kelci Miclaus, Richard A Moffitt, David Montaner, Piali Mukherjee, George J Mulligan, Padraic Neville, Tatiana Nikolskaya, Baitang Ning, Grier P Page, Joel Parker, R Mitchell Parry, Xuejun Peng, Ron L Peterson, John H Phan, Brian Quanz, Yi Ren, Samantha Riccadonna, Alan H Roter, Frank W Samuelson, Martin M Schumacher, Joseph D Shambaugh, Qiang Shi, Richard Shippy, Shengzhu Si, Aaron Smalter, Christos Sotiriou, Mat Soukup, Frank Staedtler, Guido Steiner, Todd H Stokes, Qinglan Sun, Pei-Yi Tan, Rong Tang, Zivana Tezak, Brett Thorn, Marina Tsyganova, Yaron Turpaz, Silvia C Vega, Roberto Visintainer, Juergen von Frese, Charles Wang, Eric Wang, Junwei Wang, Wei Wang, Frank

- Westermann, James C Willey, Matthew Woods, Shujian Wu, Nianqing Xiao, Joshua Xu, Lei Xu, Lun Yang, Xiao Zeng, Jialu Zhang, Li Zhang, Min Zhang, Chen Zhao, Raj K Puri, Uwe Scherf, Weida Tong, Russell D Wolfinger, and . The microarray quality control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol*, 28(8):827–38, Aug 2010.
- [60] Claudio Lottaz, Dennis Kostka, Florian Markowetz, and Rainer Spang. *Computational Diagnostics with Gene Expression Profiles*, pages 281–296. Humana Press, 2008.
- [61] Jan Budczies, Kosztyla Daniel, Christian von Törne, Albrecht Stenzinger, Silvia Darb-Esfahani, Manfred Dietel, and Carsten Denkert. Cancerclass: An r package for development and validation of diagnostic tests from high-dimensional molecular data. *Journal of Statistical Software*, 59(1):1–19, 2014.
- [62] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning, Second Edition*. Springer, New York, 2009.
- [63] Christophe Ambroise and Geoffrey J McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A*, 99(10):6562–6, May 2002.
- [64] J Bruce German, Bruce D Hammock, and Steven M Watkins. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics*, 1(1):3–9, Mar 2005.
- [65] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*, 42(Database issue):D199–205, Jan 2014.
- [66] Ron Caspi, Tomer Altman, Richard Billington, Kate Dreher, Hartmut Foerster, Carol A Fulcher, Timothy A Holland, Ingrid M Keseler, Anamika Kothari, Aya Kubo, Markus Krummenacker, Mario Latendresse, Lukas A Mueller, Quang Ong, Suzanne Paley, Pallavi Subhraveti, Daniel S Weaver, Deepika Weerasinghe, Peifen Zhang, and Peter D Karp. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, 42(Database issue):D459–71, Jan 2014.
- [67] David Croft, Antonio Fabregat Mundo, Robin Haw, Marija Milacic, Joel Weiser, Guanming Wu, Michael Caudy, Phani Garapati, Marc Gillespie, Maulik R Kamdar, Bijay Jassal, Steven Jupe, Lisa Matthews, Bruce May, Stanislav Palatnik, Karen Rothfels, Veronica Shamovsky, Heeyeon Song, Mark Williams, Ewan Birney, Henning Hermjakob, Lincoln Stein, and Peter D’Eustachio. The Reactome pathway knowledgebase. *Nucleic Acids Res*, 42(Database issue):D472–7, Jan 2014.
- [68] Dinesh K Barupal, Pradeep K Haldiya, Gert Wohlgemuth, Tobias Kind, Shanker L Kothari, Kent E Pinkerton, and Oliver Fiehn. MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics*, 13:99, 2012.

- [69] Stefan Michiels, Serge Koscielny, and Catherine Hill. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*, 365(9458):488–92, Feb 2005.
- [70] Mika Hilvo, Carsten Denkert, Laura Lehtinen, Berit Müller, Scarlet Brockmöller, Tuulikki Seppänen-Laakso, Jan Budczies, Elmar Bucher, Laxman Yetukuri, Sandra Castillo, Emilia Berg, Heli Nygren, Marko Sysi-Aho, Julian L Griffin, Oliver Fiehn, Sibylle Loibl, Christiane Richter-Ehrenstein, Cornelia Radke, Tuulia Hyötyläinen, Olli Kallioniemi, Kristiina Iljin, and Matej Oresic. Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res*, 71(9):3236–45, May 2011.
- [71] Balazs Györfy, Andras Lanczky, Aron C Eklund, Carsten Denkert, Jan Budczies, Qiyuan Li, and Zoltan Szallasi. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*, 123(3):725–31, Oct 2010.
- [72] Carsten Denkert, Jan Budczies, Wilko Weichert, Gert Wohlgemuth, Martin Scholz, Tobias Kind, Silvia Niesporek, Aurelia Noske, Anna Buckendahl, Manfred Dietel, and Oliver Fiehn. Metabolite profiling of human colon carcinoma—deregulation of tca cycle and amino acid turnover. *Mol Cancer*, 7:72, 2008.
- [73] Jan Budczies, Carsten Denkert, Berit M Müller, Scarlet F Brockmöller, Frederick Klauschen, Balazs Györfy, Manfred Dietel, Christiane Richter-Ehrenstein, Ulrike Marten, Reza M Salek, Julian L Griffin, Mika Hilvo, Matej Orešič, Gert Wohlgemuth, and Oliver Fiehn. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue - a gc-tofms based metabolomics study. *BMC Genomics*, 13:334, 2012.
- [74] Daniel A Tennant, Raúl V Durán, and Eyal Gottlieb. Targeting metabolic transformation for cancer therapy. *Nat Rev Cancer*, 10(4):267–77, Apr 2010.
- [75] Jan Budczies, Berit M Pfitzner, Balazs Györfy, Klaus-Jürgen Winzer, Cornelia Radke, Manfred Dietel, Oliver Fiehn, and Carsten Denkert. Glutamate enrichment as new diagnostic opportunity in breast cancer. *Int J Cancer*, 136(7):1619–28, Apr 2015.
- [76] Jan Budczies, Scarlet F Brockmöller, Berit M Müller, Dinesh K Barupal, Christiane Richter-Ehrenstein, Anke Kleine-Tebbe, Julian L Griffin, Matej Orešič, Manfred Dietel, Carsten Denkert, and Oliver Fiehn. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *J Proteomics*, 94:279–88, Dec 2013.
- [77] Tobias Kind, Gert Wohlgemuth, Do Yup Lee, Yun Lu, Mine Palazoglu, Sevini Shahbaz, and Oliver Fiehn. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem*, 81(24):10038–48, Dec 2009.
- [78] C Alan Boneau. The effects of violations of assumptions underlying the test. *Psychol Bull*, 57:49–64, Jan 1960.

- [79] B L Welch. The generalisation of Student's problems when several different population variances are involved. *Biometrika*, 34(1-2):28–35, 1947.
- [80] Jan Budczies, Carsten Denkert, Berit M Müller, Scarlet F Brockmöller, Manfred Dietel, Jules L Griffin, Matej Oresic, and Oliver Fiehn. Metatarget - extracting key enzymes of metabolic regulation from high-throughput metabolomics data using KEGG REACTION information. *Lecture Notes in Informatics, German conference on bioinformatics*, P-173:103–112, 2010.
- [81] Huijuan Wang, Liang Wang, Hailong Zhang, Pengchi Deng, Jie Chen, Bin Zhou, Jing Hu, Jun Zou, Wenjie Lu, Pu Xiang, Tianming Wu, Xiaoni Shao, Yuan Li, Zongguang Zhou, and Ying-Lan Zhao. 1H NMR-based metabolic profiling of human rectal cancer tissue. *Mol Cancer*, 12(1):121, 2013.
- [82] Soumen K Manna, Naoki Tanaka, Kristopher W Krausz, Majda Haznadar, Xiang Xue, Tsutomu Matsubara, Elise D Bowman, Eric R Fearon, Curtis C Harris, Yatrik M Shah, and Frank J Gonzalez. Biomarkers of coordinate metabolic reprogramming in colorectal tumors in mice and humans. *Gastroenterology*, 146(5):1313–24, May 2014.
- [83] Aurélie Morin, Eric Letouzé, Anne-Paule Gimenez-Roqueplo, and Judith Favier. Oncometabolites-driven tumorigenesis: From genetics to targeted therapy. *Int J Cancer*, 135(10):2237–48, Nov 2014.
- [84] Lenny Dang, David W White, Stefan Gross, Bryson D Bennett, Mark A Bittinger, Edward M Driggers, Valeria R Fantin, Hyun Gyung Jang, Shengfang Jin, Marie C Keenan, Kevin M Marks, Robert M Prins, Patrick S Ward, Katharine E Yen, Linda M Liau, Joshua D Rabinowitz, Lewis C Cantley, Craig B Thompson, Matthew G Vander Heiden, and Shinsan M Su. Cancer-associated *idh1* mutations produce 2-hydroxyglutarate. *Nature*, 462(7274):739–44, Dec 2009.
- [85] D Williams Parsons, Siân Jones, Xiaosong Zhang, Jimmy Cheng-Ho Lin, Rebecca J Leary, Philipp Angenendt, Parminder Mankoo, Hannah Carter, I-Mei Siu, Gary L Gallia, Alessandro Olivi, Roger McLendon, B Ahmed Rasheed, Stephen Keir, Tatiana Nikolskaya, Yuri Nikolsky, Dana A Busam, Hanna Tekleab, Luis A Jr Diaz, James Hartigan, Doug R Smith, Robert L Strausberg, Suely Kazue Nagahashi Marie, Sueli Mieko Oba Shinjo, Hai Yan, Gregory J Riggins, Darell D Bigner, Rachel Karchin, Nick Papadopoulos, Giovanni Parmigiani, Bert Vogelstein, Victor E Velculescu, and Kenneth W Kinzler. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321(5897):1807–12, Sep 2008.
- [86] Hai Yan, D Williams Parsons, Genglin Jin, Roger McLendon, B Ahmed Rasheed, Weishi Yuan, Ivan Kos, Ines Batinic-Haberle, Siân Jones, Gregory J Riggins, Henry Friedman, Allan Friedman, David Reardon, James Herndon, Kenneth W Kinzler, Victor E Velculescu, Bert Vogelstein, and Darell D Bigner. *Idh1* and *idh2* mutations in gliomas. *N Engl J Med*, 360(8):765–73, Feb 2009.
- [87] Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70, Oct 2012.

- [88] Ai-hua Zhang, Hui Sun, Guang-li Yan, Ying Han, and Xi-jun Wang. Serum proteomics in biomedical research: a systematic review. *Appl Biochem Biotechnol*, 170(4):774–86, Jun 2013.
- [89] Karsten Suhre, So-Youn Shin, Ann-Kristin Petersen, Robert P Mohney, David Meredith, Brigitte Wägele, Elisabeth Altmaier, , Panos Deloukas, Jeanette Erdmann, Elin Grundberg, Christopher J Hammond, Martin Hrabé de Angelis, Gabi Kastenmüller, Anna Köttgen, Florian Kronenberg, Massimo Mangino, Christa Meisinger, Thomas Meitinger, Hans-Werner Mewes, Michael V Milburn, Cornelia Prehn, Johannes Raffler, Janina S Ried, Werner Römisch-Margl, Nilesh J Samani, Kerrin S Small, H-Erich Wichmann, Guangju Zhai, Thomas Illig, Tim D Spector, Jerzy Adamski, Nicole Soranzo, and Christian Gieger. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477(7362):54–60, Sep 2011.
- [90] Carine Steiner, Axel Ducret, Jean-Christophe Tille, Marlene Thomas, Thomas A McKee, Laura Rubbia-Brandt, Alexander Scherl, Pierre Lescuyer, and Paul Cutler. Applications of mass spectrometry for quantitative protein analysis in formalin-fixed paraffin-embedded tissues. *Proteomics*, 14(4-5):441–51, Mar 2014.
- [91] Jan Budczies, Michael Bockmayr, Bertrad Perroud, Denise Treue, Manfred Dietel, Oliver Fiehn, and Carsten Denkert. Parallel proteomics and gene expression analysis of breast cancer. *In preparation*.
- [92] Sebastien Gallien and Bruno Domon. Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods*, 81:15–33, Apr 2015.

Danksagung

Ich möchte Herr Prof. Dr. Carsten Denkert für die hervorragende und immer angenehme wissenschaftliche Zusammenarbeit danken. Seine langjährige Betreuung hat entscheidend dazu beigetragen, dass ich mein Forschungsprofil im interdisziplinären Umfeld von Pathologie, Molekularbiologie und Bioinformatik weiterentwickeln konnte. Die von ihm zusammen mit Prof. Dr. Oliver Fiehn initiierten Metabolomikanalysen von Krebsgeweben bilden die Grundlage der hier vorgelegten Arbeiten.

Mein Dank gilt weiter Herrn Prof. Dr. Manfred Dietel für die Unterstützung meines Habilitationsvorhabens und die Schaffung einer Atmosphäre des offenen wissenschaftlichen Austausches und der kollegialen Kooperation an seinem Institut. Ich habe immer gerne in diesem Umfeld gearbeitet und ohne dieses wären viele der in den letzten Jahren durchgeführten Projekte nicht möglich gewesen.

Allen Kollegen und Kolleginnen in der Arbeitsgruppe Translationale Tumorforschung danke ich für die effektive und angenehme wissenschaftliche Zusammenarbeit, insbesondere gilt mein Dank PD Dr. Silvia Darb-Esfahani, Dr. Berit Müller, Dr. Annika Lehmann, Dr. Denise Treue, Dr. Bruno Sinn, Dr. Wolfgang Schmitt, Dr. Barbara Ingold, Judit Prinzler, Ines Koch, Britta Beyer und Petra Wachs. Meinem Doktoranden Michael Bockmayr und den von mir betreuten Masterstudenten Jana Manin und Daniel Kosztyla danke ich für die gute Zusammenarbeit.

Schließlich möchte ich allen meinen externen Kooperationspartner herzlich danken, namentlich erwähnen möchte ich Herrn Prof. Dr. Wilko Weichert (Universitätsklinikum Heidelberg) für Kooperationen zum Kolon- und Lungenkarzinom und Herrn PD Dr. Falk Hlubek (LMU München) für Kooperationen zu Genexpressionsmustern im Kolonkarzinom. Ganz besonderer Dank geht an Herrn Dr. Albrecht Stenzinger (Universitätsklinikum Heidelberg) für die intensive und effektive wissenschaftliche Zusammenarbeit in den letzten Jahren.

Herr Prof. Dr. Oliver Fiehn (UC Davis, CA, USA) danke ich für die sehr gute Zusammenarbeit bei der GC-TOFMS-Analytik und seine Gastfreundschaft.

Anhang

Die hier vorgelegte kumulative Habilitationsschrift beruht auf folgenden bereits publizierten Originalarbeiten:

1. Denkert C, Budczies J, Kind T, Weichert W, Tablack P, Sehouli J, Niesporek S, Könsgen D, Dietel M, Fiehn O. Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res* 2006;66:10795-804.
2. Denkert* C, Budczies* J, Weichert* W, Wohlgemuth G, Scholz M, Kind T, Niesporek S, Noske A, Buckendahl A, Dietel M, Fiehn O. Metabolite profiling of human colon carcinoma – deregulation of TCA cycle and amino acid turnover. *Mol Cancer* 2008;7:72. (* = gemeinsame Erstautorschaft)
3. Budczies J, Denkert C, Müller BM, Brockmöller SF, Dietel M, Griffin JL, Oresic M, Fiehn O. METAtarget – extracting key enzymes of metabolic regulation from high-throughput metabolomics data using KEGG reaction information. *Lecture Notes in Informatics, German conference on bioinformatics 2010*;P-173:103-12. (Vortrag und Beitrag zum Tagungsband, peer-reviewed)
4. Budczies J, Denkert C, Müller BM, Brockmöller SF, Klauschen F, Györfy B, Dietel M, Richter-Ehrenstein C, Marten U, Salek RM, Griffin JL, Hilvo M, Orešič M, Wohlgemuth G, Fiehn O. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue – a GC-TOFMS based metabolomics study. *BMC Genomics* 2012;13:334.
5. Budczies J, Brockmöller SF, Müller BM, Barupal DK, Richter-Ehrenstein C, Kleine-Tebbe A, Griffin JL, Orešič M, Dietel M, Denkert C, Fiehn O. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *J Proteomics* 2013;94:279-88.
6. Budczies J, Pfitzner BM, Györfy B, Winzer K, Radke C, Dietel M, Fiehn O, Denkert C. Glutamate enrichment as new diagnostic opportunity in breast cancer. *Int J Cancer* 2015;136:1619-28.

Erklärung

§4 Abs. 3 (k) der HabOMed der Charité

Hiermit erkläre ich, dass

- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde,
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlern/Wissenschaftlerinnen und mit technischen Hilfskräften sowie die verwendete Literatur vollständig in der Habilitationsschrift angegeben wurden,
- mir die geltende Habilitationsordnung bekannt ist.

Ich erkläre ferner, dass mir die Satzung der Charité Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

.....
Datum

.....
Unterschrift