Metabolites as Biomarkers

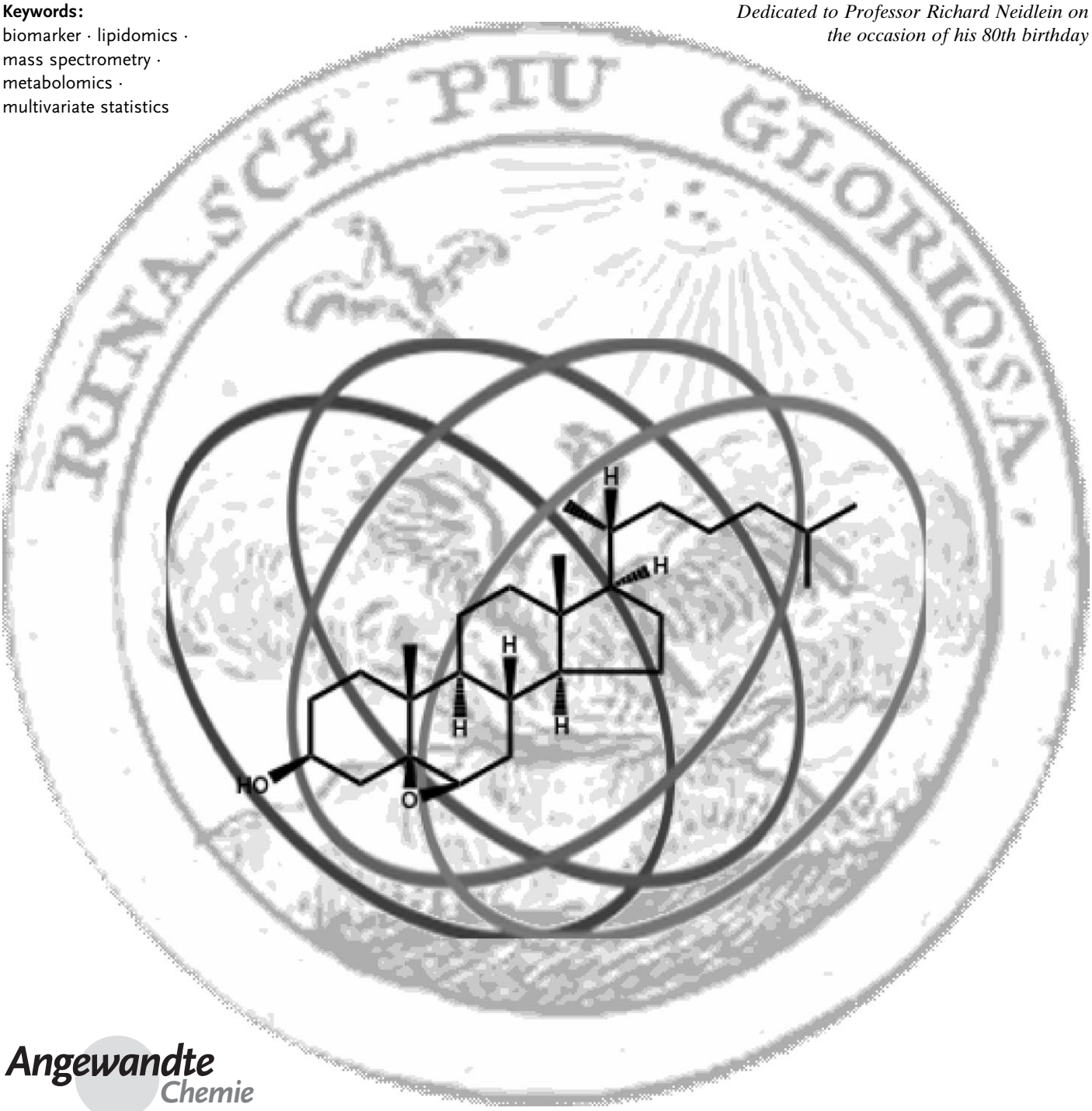
DOI: 10.1002/anie.200905579

Targeted Metabolomics for Biomarker Discovery

*William J. Griffiths, Therese Koal, Yuqin Wang, Matthias Kohl, David P. Enot, and Hans-Peter Deigner**

Keywords:
biomarker · lipidomics · mass spectrometry · metabolomics · multivariate statistics

Dedicated to Professor Richard Neidlein on the occasion of his 80th birthday



Angewandte International Edition Chemie

Metabolomics is a truly interdisciplinary field of science, which combines analytical chemistry, platform technology, mass spectrometry, and NMR spectroscopy with sophisticated data analysis. Applied to biomarker discovery, it includes aspects of pathobiochemistry, systems biology/medicine, and molecular diagnostics and requires bioinformatics and multivariate statistics. While successfully established in the screening of inborn errors in neonates, metabolomics is now widely used in the characterization and diagnostic research of an ever increasing number of diseases. In this Review we highlight important technical prerequisites as well as recent developments in metabolomics and metabolomics data analysis with special emphasis on their utility in biomarker identification and qualification, as well as targeted metabolomics by employing high-throughput mass spectrometry.

From the Contents

1. Introduction

Metabolomics is a rather recent and ambitious concept in bioanalytics that aims to quantify not less than the totality of small molecules (molecular weight less than 1500 Daltons) present in any biological system or any specific physiological state. Considering metabolomics as the rebirth of a science might be slightly pretentious, but the field has gone from using and adapting well-proven analytical platforms to eventually drive the development of new technologies in less than a decade to fulfill the increasing appetite for more compounds.[1–9] Beside the rise of "omics" and the semantic association to the proteome and genome, measuring a broad range of metabolites in biofluid or tissue samples provides a signature of the functional metabolic phenotype that contributes, completes, and also drives our understanding of diseases and toxicological effects at the end of a biological cascade.

Assessing the origin of metabolomics brings us far back into the history of bioanalytical sciences before the terms "metabolome",[10] "metabonomics",[11] and "metabolomics" were finally accepted by the scientific community. The science of investigating the metabolome, although the term "metabolomics" was not used, started in the late 1960s, but has received much attention during the past five years. The journey started in 1966 when Dalgliesh et al. combined gas chromatography (GC) with flame ionization detection (FID) for investigations on metabolism.[12] In 1971, Pauling, Robinson et al. conceived the core idea that information-rich data reflecting the functional status of a complex biological system resides in the quantitative and qualitative pattern of metabolites in body fluids.[13] At the same time, Horning and Horning described the gas chromatographic output from a patient sample as metabolic profiling,[14] and Mamer et al. investigated urinary acids by GC-MS.[15]

Within the last decade the number of scientific publications on the simultaneous measurement of endogenous compounds has increased at a truly exponential rate thanks mainly to two analytical techniques: nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).[16–26] As the number of publications (Figure 1) indexed in PubMed indicates, MS-based techniques have been largely preferred over NMR since 2005 (ratio of 2:1). Whereas machine accessibility, established data handling, and the nondestructive nature of the analysis favor NMR spectroscopic methods, the clear shift towards MS-based strategies can be explained by its higher sensitivity, improved metabolite discrimination, coverage of the metabolome space, and modularity to perform compound-class-specific analysis. Among the 2169 publications related to metabolomics that

[*] Dr. T. Koal, Dr. D. P. Enot, Priv.-Doz. Dr. H.-P. Deigner
BIOCRATES Life Sciences AG
Innrain 66A, 6020 Innsbruck (Austria)
Fax: (+43) 512-579-823-4270
E-mail: hans-peter.deigner@biocrates.com
Homepage: http://www.biocrates.com
Dr. W. J. Griffiths, Dr. Y. Wang
Swansea University, School of Medicine
Institute of Mass Spectrometry
Singleton Park, Swansea SA2 8PP (UK)
Dr. M. Kohl
University Bayreuth, Department of Mathematics
95440 Bayreuth (Germany)

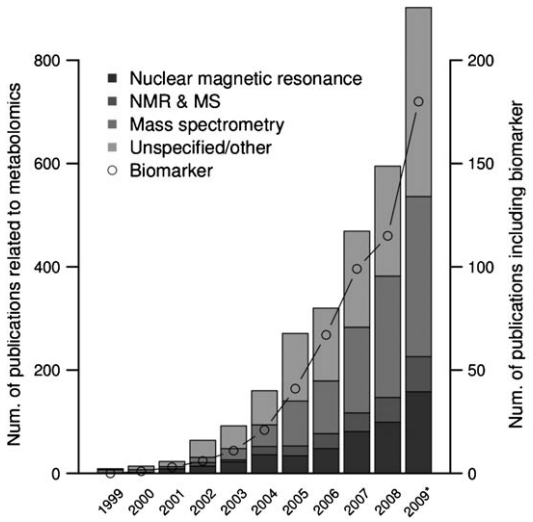


Figure 1. Number of papers in PubMed covering the following search terms: metabolomic*, metabonomic*, "metaboli* profiling", "metaboli* fingerprinting", and/or lipidomics (left vertical axis) compared to the following set of terms in the title and in the abstract, as well as the corresponding MeSH keywords: "nuclear magnetic resonance/NMR", "mass spectrometry/MS" and "biological marker/biomarker" (right vertical axis).

are indexed with Medical Subject Headings (MeSH) in MEDLINE, over 90% of the literature concentrates on living organisms, with particular attention on mammalian biofluids and plants.

The discovery of biomarkers by metabolomics, especially by targeted metabolomics (Figure 1), has a high potential for success in the detection of biomarkers as well as for the validation of biomarkers.[27–40] In this Review we attempt to cover and provide our personal views on recent issues of high-throughput quantitative mass spectrometry and the latest developments in metabolomics. We focus on aspects that have relevance to the identification of biomarkers including technological trends, analytical methods, biomedical applications, and data analysis.

Hans-Peter Deigner graduated in Pharmaceutical Chemistry at Heidelberg University, Germany, and after a position as research associate at Harvard Medical School became group leader and senior lecturer/ Associate Professor at Heidelberg. He has extensive experience in biomarker research in biotechnology companies. 2004–2006 he was Professor in Biomedicinal Chemistry at the School of Chemical Sciences and Pharmacy, University of East Anglia, Norwich. Since 2007 he has been directing biomarker research in molecular diagnostics. His research interests include bioanalytics, the integration of omics data, and the combinatorial application.

2. Why Targeted Metabolomics?

There is no single analytical technique that is suited to the precise and accurate identification and quantification of all the metabolites of interest. Regardless of the analytical set up, metabolomics studies can be divided into two different types: targeted and nontargeted approaches, depending at which stage the metabolite identification is performed during the data processing. Nontargeted metabolomics is used for global metabolome analysis, that is, comprehensive analysis of all the measurable analytes in a sample (including analyte identification of unknown signals) and must be coupled to chemometric methods to compress the data into a small set of signals, which then undergo annotation by means of in silico (library and database searches) or experimental (ionization experiments) routines. In a targeted metabolomics strategy, predefined metabolite-specific signals (by selected reaction monitoring (SRM) by tandem mass spectrometry (MS/MS), or selected ion recording (SIR) by GC-MS) are often used to determine precisely and accurately relative abundancies and concentrations of a limited number of preknown and expected endogenous metabolites.

3. Mass Spectrometric Analysis for Targeted Metabolomics

Opting for a nontargeted strategy potentially offers de novo target discovery since the exploration of the chemical space is only limited by the sample preparation and the characteristics of the analytical technique (sensitivity and coverage). The main drawbacks lie in the restricted number of established protocols for processing raw information, platform dependency, and in the challenges for unequivocally elucidating the chemical identity of the potential targets. In contrast, a targeted strategy is inevitably biased by the a priori knowledge of the metabolome and the chemical content of the sample. For the discovery of biomarkers, it pays off in regard to optimizing the sample preparation for the detection of low abundance molecules, analytical artefacts are prevented from entering subsequent analysis, and finally metabolite changes can be directly mapped onto biological knowledge or integrated with other "omics" data. In their original forms, both strategies suffer from a poor direct comparability with data measurement originating from multiple platforms and experiments, and remain a critical drawback for long-term biomarker validation studies. This can be overcome by determining absolute metabolite concentrations by means of calibration to internal (added to the sample before extraction) or external standards (added to the sample after extraction) so that individual variance between the preparation of the samples and, more importantly, matrix effects and other inferences in the sample are minimized. Although its implementation can be daunting in a high-throughput profiling context, quantification of metabolites is clearly the ultimate goal.

Mass spectrometry in targeted metabolomics is used in three main formats: gas chromatography–mass spectrometry (GC-MS), flow-injection analysis–tandem mass spectrometry

(FIA-MS/MS), and HPLC–tandem mass spectrometry (HPLC-MS/MS) where the variable is the method of sample presentation.[1–4,6,16–21,23,41–44] Each of these formats can be utilized with varying degrees of mass resolution, for example, from unit mass on quadrupole instruments to higher resolution techniques such as Fourier transform mass spectroscopy (FT-ICR-MS) and Orbitrap instruments. The mass spectrometer can be employed as a mass analyzer alone, in which case the ion abundance is plotted against the m/z value, or as a tandem mass analyzer, where, for example, in the acquisition of a product-ion spectrum the first analyzer (MS_1) operates to select ions of a given m/z value and these ions are then fragmented, usually by collision with an inert gas in a collision-induced dissociation (CID) process within a collision cell. The second analyzer (MS_2) is then used to measure the m/z value of the resultant fragment ions to give a MS/MS or MS^2 spectrum. This process can be extended further by utilizing a third analyzer (MS_3), where a fragment ion selected by MS_2 is dissociated and its MS/MS/MS or MS^3 spectrum recorded. Mass analyzers can be arranged in series in space, such as on hybrid instruments (for example, tandem quadrupole, Q-TOF instruments), or in time, such as with ion traps. Ion traps are particularly adept at recording MS^3 and further MS^n spectra. Triple quadrupole instruments offer other advantages such as SRM "scans", where MS_1 is "parked" on an m/z value of interest and MS_2 on the m/z value(s) of a known fragment ion. No scanning occurs here, and thus ion transmission to the detector is maximized. Thus, when a component with the m/z value defined by MS_1 elutes from the column and fragments to give ions defined by MS_2, a signal is recorded. Many SRMs can be programmed to occur in a given "scan", in which case the analyzers jump from one SRM to another. The result is a multiple reaction monitoring (MRM) "scan". In a chromatographic analysis, the SRMs can be time-programmed so as to coincide with the approximate time of elution of the components of interest, thereby avoiding "scanning" redundant m/z values during analyte elution and maximizing the sensitivity (scheduled MRM). Similar "scans" can be recorded on MS-only instruments, where MS_1 is "parked" on the m/z value of interest and selected ion recording or selected ion monitoring (SIM) chromatograms generated.

Two other important scan modes utilized on triple quadrupole instruments are precursor ion scans and neutral loss scans. In precursor ion scans, MS_2 is set to transmit the m/z value of a known fragment, and MS_1 is scanned to identify ions dissociating to give this fragment. In the neutral loss scan, MS_1 and MS_2 are scanned in parallel and offset by a set m/z value which corresponds to the loss of a defined neutral fragment.

4. Sample Preparation for FIA-MS/MS and HPLC-MS/MS

Sample preparation prior to mass spectrometry analysis is a key step in all metabolomic studies whether the goal is one of global (nontargeted) or targeted analysis. There is a school of thought amongst the metabolomics community that sample preparation should be kept to a minimum. This is derived from the belief that sample-handling steps will introduce uncontrolled analyte loss, and by their very nature are selective and thus discriminating in a global analysis. These suppositions are both true, but in the first case can be corrected for (by the use of appropriate internal standards), and in the second can be used to the analysts advantage in a targeted approach. In fact, there are applications in metabolomics where a minimum sample preparation is all that is required, for example, for analysis of human bile, where simple sample dilution is all that is required prior to FIA-MS/MS analysis,[45] or in the analysis of urine from patients with cholestatic liver disorders,[46] where FIA-MS/MS analysis can be performed after a simple C_{18} solid-phase extraction (SPE). However, the complexity of the global metabolome and its extreme dynamic range (for example, $mg\,mL^{-1}$ to $pg\,mL^{-1}$ in plasma) dictates that sample preparation steps must be performed prior to analysis in most cases. In the case of the analysis of tissue metabolomes, extraction is a necessary prerequisite, except in cases where analysis is to be performed directly on the tissue, as is the case with matrix-assisted laser desorption/ionization (MALDI) imaging[47–49] or desorption electrospray ionization (DESI).[50]

The enormous heterogeneity in the physical and chemical properties combined with wide concentration ranges of endogenous metabolites present in a biological system has so far precluded attempts of whole metabolome analysis with just one or a few analytical methods.[1,2] Additional sample preparation steps, for example, derivatization or SPE, are necessary for the analysis of lower concentrations of endogenous metabolites to improve the detection limit. The method of sample preparation is dependent on the starting matrix, the analytical method, and whether the analytical goal is one of a targeted analysis on a specific class of compounds or on the global metabolome. In this Review we focus on targeted approaches and describe some examples of sample preparation methods towards this end.

Most HPLC-MS metabolomic studies are conducted using electrospray ionization (ESI).[51,52] In regard to metabolomics, the two key properties of this ionization method are dynamic: the range of the ESI process itself and the differing ionization cross-section of different classes of analytes. These features dictate that the most abundant and readily ionized components will be discriminated for during analysis of a crude extract. So, for example, phospholipids which are abundant in blood can be analyzed by FIA-ESI-MS/MS following a simple protein precipitation with methanol,[53] or Bligh/Dyer-like extraction (chloroform/methanol)[54] without the need for prior HPLC separation. On the other hand, metabolites that are less abundant or ionize less well in the ESI source require more detailed separation steps so as to allow their presentation to the ESI source at a concentration that is sufficiently high for ionization.

Before we discuss sample preparation for targeted metabolite analysis, it is of value to consider briefly approaches used for global metabolomics. Siuzdak and coworkers have, for example, optimized an extraction method for serum and CSF based on the extraction of 50 μL samples into ice-cold methanol, centrifugation, followed by vacuum

drying of the supernatant and reconstitution of the extract in an LC solvent.[53,55,56] By using this method (ESI positive-mode; duration 60 min) Crews et al. were able to identify more than 4000 signals reproducibly in both plasma and CSF. A signal was regarded as a mass spectrometric signal when it had a signal/noise (S/N) ratio of greater than 10:1.[53] By using a similar extraction protocol for plasma and UPLC-ESI-MS/MS, Zelena et al. were able to identify 1600 in the positive mode and 1100 signals in the negative mode with ESI detection.[57] In contrast, Wishart et al. were able to identify only about 200 unique signals when they directly injected CSF into a HPLC-ESI-MS/MS system (combined positive and negative detection with ESI).[42] Surprisingly few of the CSF signals could be assigned to actual compounds by MS/MS: Crews et al. only identified 12 compounds and Wishart et al. only 17 compounds.

5. Targeted HPLC-MS/MS Analysis of Lipids in Biofluids

Lipids play an important role in the metabolome and thus their analysis is of particular interest. However, analytical methods that involve reconstitution of a metabolite extract into an aqueous HPLC solvent (for example, 5% acetonitrile in water,[53,56] or 100% water[57]) are likely to be strongly discriminatory against its lipid content, for the simple reason that many lipids are not soluble in these solvents. In fact, for lipids, a better policy may be to employ a dedicated or targeted approach to both sample preparation and analysis.[58]

Lipid-modifying enzymes such as phospholipases, 5-lipoxygenase, phosphoinositide 3-kinase, and sphingosine kinase control signaling lipids and form a complex network that is relevant to essential processes such as apoptosis and proliferation. They, thus contribute to the pathomechanisms of numerous serious states of disease such as cancer, diabetes, inflammation, and metabolic diseases.[59] An increasing number of key players and their interactions have been identified, and more insight into their role in pathobiochemistry has been obtained. However, additional quantitative high-throughput lipidomics methods are essential prerequisites for assessing their potential as biomarkers.

Taking the lipidomic analysis of plasma and CSF as an example, one favored strategy starts by the extraction of lipids into 99.9% ethanol, preferably in a volume ratio of at least 1:10 for plasma and 1:4 for CSF, by dropwise addition of the biofluid into ethanol in an ultrasonic bath.[60] In this way, lipid–protein associations are disrupted and free lipids are dissolved in solution. The lipids are then fractionated according to their lipophilicity on a C_{18} SPE cartridge; this is achieved by diluting the lipid-containing ethanol solution to 70% ethanol and passage through a C_{18} cartridge. The more hydrophobic lipids such as triglycerides, cholesterol, and its esters are retained on the column, while less hydrophobic lipids such as oxysterols, bile acids, steroids, and phospholipids are eluted after further addition of 70% ethanol to the SPR cartridge. If required, the more-hydrophobic lipid fraction can be eluted with 99.9% ethanol or chloroform/methanol (1:1, v/v). This separation step can be particularly important for the subsequent analysis of oxysterols and steroids, which can be formed endogenously and also as an artefact of cholesterol autoxidation.[61,62] If phospholipids are the target of analysis, the 70% eluent is now suitable for analysis by HPLC-ESI-MS/MS or FIA-ESI-MS/MS. If, on the other hand, the focus of the analysis is, for example, oxysterols, steroids, or bile acids, alternative protocols may be preferred. One strategy is to employ an anion-exchange column (for example, lipidex-DEAP, which will fractionate compounds according to acid strength) and then perform HPLC-ESI-MS/MS with negative-mode ESI.[63] ESI in the positive ion mode is usually adopted for the analysis of neutral molecules, although the ion current may be low in the absence of any basic functional groups on the target molecules. This is the situation for oxysterols and free steroids.

One approach to improve the ionization cross-section for neutral molecules when employing ESI is to derivatize the molecule so that the derivative contains an acidic or a basic group. This tactic can also be used to enhance the ionization cross-section of existing but weakly acidic or basic groups.[55,64–67] Here, two derivatization methods are discussed, although many others exist in the literature. To enhance the ionization of weak acids, for example, carboxylic acids ($pK_a \approx 6$), derivatization can be performed with aminosulfonic acids (for example, 2-aminoethanesulfonic acid) or 4-aminobenzoic acid, in the presence of a carbodiimide, for example, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). The incorporation of the strongly acidic sulfonate group ($pK_a \approx 1.5$) in the target analyte improves the ionization in the negative-ion ESI mode.[64] 2-Aminoethanesulfonic acid is available in a deuterated form ($NH_2CD_2CD_2SO_3H$), and can thus be used in quantitative studies based on stable isotope labeling (similar to the use of ICAT in proteomics).[55]

Not only does derivatization with an aminosulfonic acid improve ionization, it also directs fragmentation when MS/MS is performed. This enhances the array of structurally informative fragment ions, and also gives characteristic fragment ions of low mass (Figure 2). Charge-tagging can also be used in a chemoselective manner for other functional groups. For example, oxo groups (in aldehydes and ketones) can be derivatized with a hydrazine (for example, 2,4-dinitrophenylhydrazine). The result is improved ionization of the target analyte, MS/MS fragmentation properties, and solubility in the mobile phase for reversed-phase chromatography.[60,67–69] The disadvantage is the extra sample handling steps. Oxo groups are present in steroids, sterols, and bile acids prevalent in plasma.[45,70,71] Thus, following extraction and isolation, lipids containing an oxo group can be derivatized under mildly acidic conditions with hydrazine reagents and analyzed by HPLC-MS/MS (or MS^n) to reveal lipids containing oxo groups in, for example, plasma. We have found that the Girard P hydrazine (GP) reagent is particularly useful for enhancing the ionization, solubility, and fragmentation of oxo-containg steroids and bile acids. The following steroids and bile acids have been identified in adult and/or infant plasma by using this derivatization method: DHEA 3-sulfate, epiandrosterone 3-sulfate, androsterone 3-sulfate, etiocholanolone 3-glucuronide, 3β,5β-dihydroxy-B-norcho-

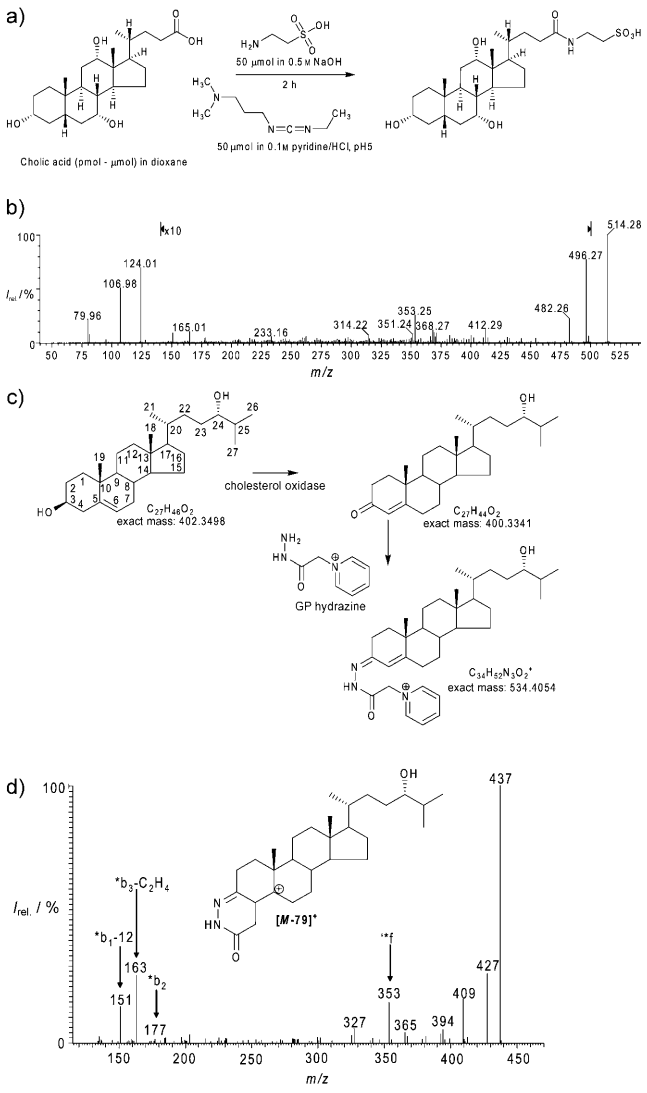


Figure 2. a) Derivatization of cholic acid with 2-aminoethanesulfonic acid. b) Negative-ion ESI-MS/MS spectrum of the resulting derivatized compound ($[M-H]^-$ *m/z* 514). c) Oxidation and derivatization of 24*S*-hydroxycholesterol. d) ESI-MS^3 of the resulting derivative in infant plasma ($[M]^+$ *m/z* 534).[194]

lestane-6β-carboxyaldehyde, 7α-hydroxy-3-oxocholest-4-en-27-oic acid, 3-oxocholest-4-en-27-oic acid, and 7α 27-dihydroxycholest-4-en-3-one.[60]

A larger part of the lipidome can be explored by incorporating an oxidation step, which converts alcohol groups to oxo groups, prior to derivatization with the Girard P hydrazine (Figure 2). This can be achieved by the use of chemical oxidation reagents or in a more specific manner by using enzymes such as cholesterol oxidase which converts 3β-hydroxy-5-ene- and 3β-hydroxy-5α-hydrogen-containing sterols and steroids into their 3-oxo-4-ene and 5-oxo-5α(H) analogues, respectively.[72] In this way, desmosterol, 7-dehydrocholesterol, cholesterol, 24*S*-, 27-, and 7α-hydroxycholesterols, 7α,27-dihydroxycholesterol, 3β-hydroxycholest-5-en-27-oic acid, and 3β,7α-dihydroxycholest-5-en-27-oic acid have also been identified in plasma.[60] In regard to the CSF and the global HPLC-MS metabolomic studies performed by Wishart et al. and Crews et al., where a total of 17 and 12 compounds were identified respectively,[42,56] Ogundare et al. have, by using the GP derivatization strategy (with and without the use of cholesterol oxidase), been able to identify nine sterols and bile acids by comparison of their HPLC-MS^3 properties with those of authentic standards. They also postulated the structure of a further eight cholesterol metabolites from their HPLC-MS^3 spectra and retention times.[73] Thus, the number of 3β-hydroxy and oxo- sterols/steroids and bile acids identified in the CSF lipidome after GP derivatization is equivalent to the total number of compounds identified by HPLC-MS (without derivatization) in the global metabolome. The advantage of a targeted approach is clear.

6. Sample Preparation for GC-MS Analysis

GC and GC-MS have been used extensively for metabolome analysis since the mid-1960s.[13,74,75] However, the very nature of GC dictates that it is only suitable for volatile compounds, or those derivatized to be volatile. Thus, involatile and thermally labile molecules are analyzed better by HPLC-ESI-MS (or by one of the other atmospheric pressure ionization (API) methods) or by MALDI-MS. Nevertheless, the application of suitable derivatization methods allows a significant proportion of the metabolome to be observed by GC-MS. Despite the need for derivatization, GC-MS offers a number of advantages over HPLC-MS. GC columns still offer higher resolution than HPLC columns, and spectra obtained by using an electron ionization (EI) detector contain a wealth of structural information. The existence of extensive libraries of EI spectra (for example, http://www.nist.gov/srd/nist1a.htm) and reproducible retention indices often makes compound identification easier by GC-MS than by HPLC-ESI-MS. A disadvantage of GC-MS when using EI is that the molecular ions may be minor or absent from the spectra, thus information on the molecular weight of the starting molecule is absent. This can be overcome by recording chemical ionization (CI) spectra, where $[M+H]^+$ or $[M+NH_4]^+$ ions, for example, tend to dominate. In fact, exquisite sensitivity can be achieved, for example, 50 pg mL^{-1} for some steroids in CSF, by utilizing perfluorinated derivatives and negative-ion CI.[76] Although GC-MS can not be used alone for a truly global analysis of the metabolome, the application of targeted derivatization of multiple functional groups coupled with adequate sample extraction can maximixe the available portion of the metabolome. In approaches designed for the global analysis of, for example, biofluids, extraction is often achieved by dissolution in methanol and removal of the precipitated protein by centrifugation.[42,77]

The next question is which functional groups to derivatize and with which reagent. Oxo groups are common functional groups and can easily be derivatized with methoxyamine to methyloximes, while alcohol groups can be converted into TMS ethers with a variety of reagents.[78,79] By using this extraction and derivatization strategy for human plasma Jiye et al. were able to resolve 500 signals by GC-MS and quantify

32 endogenous compounds at levels varying from 2 mM (800 $\mu g\,mL^{-1}$) for non-esterified cholesterol to 24 μM (4.32 $\mu g\,mL^{-1}$) for inositol.[77] Wishart et al. also found by using a similar procedure that CSF was much less complex than plasma. Of the approximately 50 detected GC signals they were able to identify 41 compounds in the concentration range 3.8 mM (230 $\mu g\,mL^{-1}$) for urea to <1 μM (75 $ng\,mL^{-1}$) for glycine.[42]

7. GC-MS Analysis of Cholesterol Metabolites in Biofluids

The first step in cholesterol metabolism is oxidation to an oxysterol. Oxysterols may be metabolized further via pregnenolone to steroid hormones, or after 7α- or 27-hydroxylation to bile acids. Oxysterols are bioactive molecules, and their activity depends on their exact structure.[80,81] Similarly, bile acids and other intermediates in their biosynthetic pathway(s) are also biologically active, with specific compounds acting as ligands to nuclear receptors.[82,83] Furthermore, disruption of the steroid and bile acid biosynthetic pathways leads to disease, and plasma levels of intermediates offer biomarkers to pathology.[84]

Björkhem and co-workers have performed extensive studies on the significance of blood-borne cholesterol metabolites to atherosclerosis and neurodegenerative diseases,[85,86] with particular attention paid to CSF. They used GC-MS in the SIR detection mode as their final detector to achieve maximum sensitivity. Sample preparation involves alkaline hydrolysis of (oxy)sterol esters, followed by Folch-like extraction (KOH in ethanol, chloroform, NaCl solution) and drying of the resultant organic phase. The residue is then dissolved in toluene and the oxysterols separated from excess cholesterol by normal-phase SPE. Oxysterols elute after cholesterol and are derivatized to trimethylsilyl ethers for GC-MS analysis.[87] The following oxysterols have been identified in plasma by this method and quantified with the aid of stable isotope labeled standards: 7α-hydroxycholesterol ($43 \pm 48\,ng\,mL^{-1}$), 7β-hydroxycholesterol ($3 \pm 5\,ng\,mL^{-1}$), 7-oxocholesterol ($22 \pm 14\,ng\,mL^{-1}$), 5α,6α-epoxycholesterol ($6 \pm 8\,ng\,mL^{-1}$), 5β,6β-epoxycholesterol ($26 \pm 18\,ng\,mL^{-1}$), cholestane-3β,5α,6β-triol ($27 \pm 39\,ng\,mL^{-1}$), 24-hydroxycholesterol ($64 \pm 24\,ng\,mL^{-1}$), 25-hydroxycholesterol ($2 \pm 3\,ng\,mL^{-1}$), and 27-hydroxycholesterol ($154 \pm 43\,ng\,mL^{-1}$). It should be noted that the sum of these values correspond to the sum of the free oxysterol and its fatty acid esters.[87]

The research groups of Björkhem and Lütjohann have also applied this method to the analysis of CSF, where they measured oxysterol levels in control groups and patients with neurodegenerative disorders.[88–91] The oxysterol levels determined in control CSF were as follows: 7α-hydroxycholesterol ($1.50\,ng\,mL^{-1}$), 7β-hydroxycholesterol ($1.11\,ng\,mL^{-1}$), 7-oxocholesterol ($0.86\,ng\,mL^{-1}$), 24-hydroxycholesterol ($1.54\,ng\,mL^{-1}$), and 27-hydroxycholesterol ($0.8\,ng\,mL^{-1}$). Again it should be noted that these values correspond to the sum of free oxysterols and their fatty acid esters.[88] Unsurprisingly, none of these compounds were detected in global GC-MS metabolomic studies,[42,77] as alkaline hydrolysis is not usually carried out in global studies and the levels of free oxysterol in CSF are on the order of $20\,pg\,mL^{-1}$ or 0.05 nM.[73]

8. Data Analysis

Data analysis plays a central role in a biomarker project. The ultimate goal is to derive the most salient set of metabolites that could undergo further validation studies or clinical applications. Despite the variety of experimental design, the differing MS data characteristics, and the range of statistical tools available, the data analysis workflow follows a fixed set of steps, many of which are optional. These include signal processing, data normalization, transformation, and assessment as well as the application of statistical methods for comparison of groups and the construction of predictive models.[92,93] In this Review emphasis is placed on aspects we feel are crucial to the discovery of biomarkers and validation with metabolomics data rather than detailing all the possible methods of data analysis.

The general purpose of data preprocessing, also called low-level analysis, is to translate crude signals—the detector response—into a discrete and more useable quantity that characterizes the metabolite.[94,95]

Despite the distinctive characterization of the spectral and chromatographic features of an analyte in a targeted metabolomics setting, optimal signal processing should not be underestimated for obtaining biologically meaningful and reproducible results. This becomes more evident when data extraction is performed automatically and measurements are spanned over long periods of time. Unwanted interfering signals are caused by various sources intrinsic to the mass spectrometry analysis (loss of sensitivity, drift of the analytical system) or to factors related to the sample heterogeneity, reagent batch effects, or laboratory conditions. The preprocessing of mass spectrometry data typically consists of baseline correction, filtering or decomposition, signal extraction (that is, peak detection and quantification), and integration.[94,95] The baseline, or background, spectrum is the mass spectrum observed when no sample is intentionally introduced into the mass spectrometer.[96] The baseline can be flat or curved with a positive or negative (for example, linear) trend.[97] There are quite a large number of approaches that consist of either nonparametric methods or parametric estimates which arise out of models based on the physics of the mass spectrometer. An overview is given by Shin and Markey.[98] Besides the background noise there is also a high-frequency noise component in mass spectra. The procedures used to reduce this high-frequency noise are smoothing filters, such as the Gaussian filter or moving average filter, deconvolution filters, or decomposition methods such as the wavelet transform.[97,98] After reduction of the background and the high-frequency noise, methods are then applied to automatically detect and quantify the signals in the noise-reduced mass spectrum. Again, there exists quite a large number of different algorithms which, for example, are based on the signal-to-noise ratio.[94–98]

After extraction of the analyte measurement, a data-adjustment step, known as normalization, is usually undertaken to make metabolite profiles generated on different days, on different machines, at various dilutions etc. comparable. In practice, normalization procedures are based on some underlying assumptions regarding the data and the experimental design, and thus have to be adjusted accordingly. The most common methods aim at determining a samplewise bias factor (also known as the dilution or scaling factor) from the data themselves.[99] This scaling factor can be derived from the sum (or some related quantity) of all the measurements across the sample or can be computed by more sophisticated strategies.[92,100,101] Alternatively, sample concentrations are commonly adjusted by using information from the biological context, such as creatinine concentration, urine volume, or osmolality, at the time of sampling.[102] Despite its biological foundations, this kind of normalization is not appropriate if the study addresses biochemical processes that may have a dramatic impact on the scaling parameter values (such as in the case of kidney impairment).

The most sensible approach to surmount undesirable discrepancies between sample measurements is to determine the absolute metabolite concentrations by means of calibration with one or more internal standards. As a result, inter-individual variance is minimized and data sets originating from multiple sites and experiments are directly comparable.[103] However, the implementation of this approach can be quite difficult in a high-throughput metabolite-profiling context. The chemical diversity can make it unlikely that a single internal standard is sufficient, and instead a set of internal standards is needed.

We recently introduced a new concept that relies on endogenous housekeeping/reference metabolites to normalize metabolomics data. This approach was adapted from real-time quantitative PCR, where the use of housekeeping/reference genes is the gold standard.[104] Our data show that the use of endogenous housekeeping/reference metabolites can increase the power of statistical analyses, that is, reduces technical errors.[105] This can be illustrated by a Venn diagram that compares the adjusted p values and the area under the curve (AUC) with and without using housekeeper (HK) normalization (Figure 3).

The data set consists of 66 samples (21 controls, 45 pneumonia) and 197 metabolites, with two housekeeper metabolites selected for normalization. The features identified with housekeeper-normalized data, by using an adjusted p value of <0.01 and an AUC of >0.8 as thresholds, are in good agreement with those found with the original data. In addition, we observe an increase in the number of significant differences and higher AUC values. Besides the improved statistical power, several metabolites have been identified as candidates for reference metabolites under several experimental conditions.[105]

Metabolites are characterized by the huge dynamic range of their concentration, and typically the associated variance is larger at higher concentrations. For the purpose of statistical analysis, a variance-stabilizing transformation or transformation to normality, as for example, by taking the logarithm or generalized logarithm, or using power transformations, must

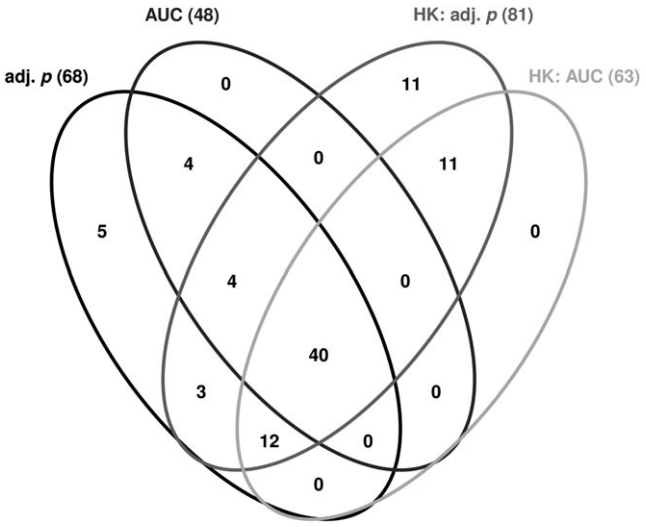


Figure 3. Venn diagram showing the number of metabolites with adjusted p values (adj. p) <0.01 and AUC >0.8, respectively, to differentiate samples (plasma) taken from control patients from those of patients with pneumonia. The results were obtained with and without using housekeeper (HK) normalized data. The sum of the numbers in each ellipse represents the number of metabolites chosen by the corresponding criterion, while the sum of the numbers at the exterior represents the number of metabolites not selected by this criterion. Numbers in overlapping areas represent metabolites having more the one criterion in common. For further details see the text.

be applied to alleviate the dependency of the variance on the metabolite concentration.[106,107] The motivation behind these transformations is that many statistical procedures rely on homoscedasticity or distributional assumptions, which as a rule are not fulfilled for untransformed mass spectrometry data. Alternative methods such as centering (for example, at the mean or median) and/or scaling (for example, by standard deviation or median absolute deviation) tend either to be inefficient or to overweight low-intensity measurements.[92,108]

Since the generation of high-quality data depends on a number of factors, the steps described above must be accompanied by rigorous assessment of the data quality. As such, the implementation of independent samples for quality control, rather than the data themselves, is necessary to identify potential flaws during the pre-processing workflow. Typical tools include a combination of diagnostic statistics and representations such as histograms, similarity plots in the form of heat maps or dendrograms, or multivariate analysis such as principal component analysis.

For most metabolomics biomarker studies, it is rarely the case that data mining methods can uncover an underlying biological mechanism without introducing a priori knowledge of the pathophysiological state of the patient. In this context, statistical analysis essentially focuses on univariate tests to compare groups of samples (class comparison) and advanced algorithms for classifying subjects by their clinical outcome (class prediction). Whereas clinical studies are traditionally restricted to a set of a few parameters, data mining strategies for metabolomics data must be adapted to cope with a larger number of measurements (up to a thousand).

In class comparison, it is common to treat each metabolite separately, with the corresponding univariate statistical analysis consisting of a statistical test such as the t test or Mann–Whitney test, or some sort of regression analysis such as linear or generalized linear models. In practice, one ends up with at least one (unadjusted) p value for each metabolite and one has performed several hundred tests (at least as many as there are metabolites) in parallel. In classical statistical hypothesis testing, one typically restricts the type I error α of a single test to 1% or 5%. However, as one now considers several hundred tests simultaneously (that is, multiple testing), the chance of getting a false-positive test result can be expected to be much larger than α. If there are N independent statistical tests and a proportion p of true null hypotheses, the chance of getting at least one false-positive test is $1-(1-\alpha)^{pN}$, which is larger than α if $pN>1$; for example, $\alpha=5\%$, $N=100$, $p=20\%$ leads to a 64.1% chance (called familywise error) of getting at least one false-positive test result. Hence, it is highly recommended to adjust the original p values to keep control over the type I error for the whole family of tests and to obtain reliable and reproducible results. Two main approaches exist for this purpose, which are called familywise error rate (FWER) and false discovery rate (FDR). The FWER is the probability that there is at least one false-positive test result, while the FDR is the expected proportion of false-positive test results. In general, the FWER is more conservative than the FDR, and should be chosen if high confidence in all the positive test results is desired. In contrast, the FDR is preferable if a certain proportion of false-positive tests is acceptable. A detailed overview of the available approaches is given in the review by Dudoit and van der Laan.[109] As highlighted by the MAQC (microarray quality control) study,[110] it is best to combine the adjusted p values with fold changes to further increase the reproducibility of the results.

The main focus in the development of biomarkers is to generate a predictive model that is capable of assigning persons to their correct pathological group with the highest possible accuracy. The high dimensionality of metabolomic information means that most classification methods cannot be applied directly, or at least only in a non-efficient manner: increasing the number of measurements results in distances between the subjects becoming similar, and consequently the classification algorithm is unable to establish decision boundaries. This phenomenon, known as the curse of dimensionality, is even magnified when noisy and irrelevant features are incorporated. Hence, one major aspect in the discovery of biomarkers is to reduce the initial data set so that computations are tractable, model predictive power is improved, and the biochemical interpretation can focus on a small set of relevant metabolites. Finding a model that combines a reasonable subset of salient features and an optimum accuracy can be approached from different angles: by filtering features according to some sort of relevance measure before applying a classification algorithm or "wrapping" the subset search with the classification algorithm. Metabolites with very low variability are often considered as un-informative, and thus may be removed in a first step by a so-called unspecific filtering.[111,112] Specific filtering approaches use certain criteria, usually in the form of an output from class comparison or correlation-based metrics, to assess the discriminative power of individual features.[113] These techniques are very efficient, but neglect to integrate potential interactions and correlations between features explicitly, so that they are very likely to fail in situations where individual features are only informative in combination. Wrappers rely on an iterative search of candidate subsets in combination with a scoring metric (usually given in terms of classification accuracy) that grades subsets of features. Since exhaustive exploration of the feature space is generally impractical (the number of possible subsets is on the order of 2^d, where d is the number of measurements), mathematical optimization strategies are implemented to locate a reasonable set of biomarkers. The application of a learning algorithm for each subset evaluation, the efficiency of the search strategy, and the additional steps required for estimating the scoring metric results in the wrappers being the most time-consuming of all the feature-selection strategies. The computational costs in discovering a biomarker often pays off, since interactions between the features are in principle preserved and the final predictive signatures often show higher accuracies while being of shorter length.[114]

After feature selection, the classifier has to be computed, typically by a (deterministic) function that maps a multi-dimensional vector to a binary (or n-ary) outcome variable. For this purpose, a large number of supervised learning algorithms are available, for example, logistic regression, (diagonal) linear or quadratic discriminant analysis, shrunken centroids regularized discriminant analysis, random forests, neural networks, support vector machines, generalized partial least squares, partitioning around medoids, recursive partitioning and regression trees, K-nearest neighbor classifiers, bagging, boosting, naive Bayes, and many more.[115–117] Class discovery methods (that is, unsupervised learning methods) such as clustering are not appropriate for classifying new samples.[118]

In a next step, the computed predictive signature has to be validated by using an appropriate performance measure. As there are n correct decisions in n-ary classification, one typically considers a set of metrics; for example, a pair of metrics such as sensitivity and specificity, or positive and negative predictive value, is frequently employed in the case of binary classification. In general, the use of a single, that is, one-dimensional, performance metric, for example, odds ratio, to measure the predictive performance can be quite misleading as it represents only a one-dimensional aspect of an n-dimensional problem (n correct decisions). If a classifier, for example, predicts 50 of 100 (50%) samples of a first group and 99 of 100 (99%) samples of a second group correctly the odds ratio is $0.5/0.5\times0.99/0.01=99$. However, the classifier is no better than chance in predicting the membership of the first group. A comprehensive overview of the possibilities for binary classification challenges (for example predictive values) is given in reference [119]. Typically, statistical measures such as variance or confidence intervals should be given for the performance measures to assess their statistical significance and to get a more realistic impression of the performance of the classifier in practice. Such statistical

measures can be obtained by using one or more rounds of cross-validation, bootstrap, or some split-sample approach in the training-test step.[116] This part of the classifier development is also called internal validation. To avoid any bias in the estimation of the classification error test, data are formed from observations that are not involved in any decision during all the modeling phases, including data pre-processing and feature selection.[120,121] Preliminary results from microarray studies have so far indicated that, since biological effects are of significant magnitude and classifiers are of adequate accuracies, the choice of the pre-processing algorithms has only a minor influence.[122,123]

In addition to accuracy, a classifier must fulfil several criteria: prevalence has to be taken into account, the predictive signature should be backed up with biochemical evidence, and the clinical or therapeutic relevance has to be demonstrated. The clinical utility of a classifier means that it should provide added value to clinical predictors already in use.[124] This is the case if either it is superior to existing clinical predictors of the phenotype or its combination with existing clinical predictor(s) is superior to the clinical predictor(s) alone. For this purpose, likelihood ratio tests should be applied.[125] Moreover, for every day use, the predictive molecular signature must be highly reproducible between laboratories, which can be confirmed by processing the samples in other laboratories. Clinical utility, therapeutic relevance, and inter-laboratory reproducibility should be addressed, in particular, by so-called external validations. This means that the predictive power of the trained classifier is assessed by predicting the class label of novel unlabeled cases from independent datasets. These independent datasets can be from the same clinical center, but should ideally also come from another clinical center. The external validation may also include the transfer of the classifier to the platform that would be used for clinical application.

A summary of recommendations for the best approach to the analysis and reporting of biomarker discovery and validation based on metabolomics is given in Table 1.[126,127]

Table 1: Recommendations for metabolomics data analysis/statistics.

No.	Recommendations
1	domain of validity of the biomarker: provides a detailed description of the cohort
2	adequate hypothesis testing: take into account potential bias such as population stratification, technical batches, and confounding factors
3	use an appropriate indicator for uncertainty and measurement error (interval of confidence) and for calculation of sample-size effects
4	correct for multiplicity of statistical tests
5	do not rely solely on (adjusted) *p* values: combine with fold changes or other criteria
6	validate predictive signatures internally and externally
7	estimate predictive abilities of multivariate models on test set(s) independent to the one used during the training/modeling phase
8	judge quality of predictive models on the basis of several measures and compare it to the gold standard routinely employed in practice
9	place the predictive set of markers in both pathophysiological and clinical contexts

9. Application in Disease Areas: Biomarker and Metabolomics

Since the metabolome is sensitive to age, gut microbial composition, and lifestyle, metabolomics is ideal for the characterization of dietary and therapeutic interventions, metabolism, and metabolism-related disorders (see also a recent review by Oresic).[128] In fact, the application of metabolomics now covers a wide area including nutrition and environmental research as well as an increasing number of biomarker studies in diagnostics and drug research (Figure 1). We focus here on new developments in early diagnostics and its use in established or emerging therapeutic interventions.

Although much progress has been made in environmental metabolomics, there are numerous additional potential applications. Some interesting recent ones include the use of so-called "sentinel species" of vertebrate and invertebrate animals which are indicative of the status of the environment (for example, fish species have been used to assess the affects of pollutants, climate change, or biodiversity and also their altered biomarker pattern for risk assessment of pharmaceuticals and industrial chemicals). Earthworms (*Eisenia veneta*) have been exposed to pollutants such as nitrophenol[129] and fluorinated anilines,[130] and fish and fish embryos (*Oryzias latipes*) have been used to investigate the effects of trichloroethylene and the pesticide dinoseb on their development.[131,132]

A total number of approximately 1 133 000 biomarkers of interest has been estimated, of which the genome accounts for about 25 000 to 30 000,[133] the transcriptome 100 000,[134] the proteome 1 000 000,[135] and the metabolome about 2500 to 3000.[134,136] Although, exploitation of metabolites in this context started later, because of technological developments, an ever increasing number of studies are now dealing with metabolites as a way to identify pathomechanisms of complex diseases, to characterize phenotypes, and to develop diagnostic biomarkers as markers for patient stratification and companion diagnostics.[137] Moreover, promising applications in toxicology and the profiling of drug candidates has led to the expectation that metabolomics will become established as indispensable for drug identification and development.[138]

For many years specific metabolites have been measured in body fluids to diagnose particular diseases such as diabetes, by measurement of glucose, and vascular diseases by determination of cholesterol. Metabolomics, with its impressive and ever increasing coverage of endogenous compounds and its intrinsic high-throughput capacity, now provides a much more comprehensive assessment of a patient's health status

and can be utilized in the identification, qualification, and development of biomarkers.

Historically, metabolomics was early on applied successfully to screen newborn babies with inborn errors of metabolism, such as for 30 defects in enzymes that catalyze the metabolism of carbohydrates, amino acids,[139] fatty acids, and nucleic acids, as well as the urea cycle.[140] The screening of newborn babies by mass spectrometry is now widely established.[141–146] One reason for this success is that the relationship between the (predominantly monogenetic) disease state, metabolic biomarkers, and genetics is easily understood in many inborn errors,[139,147] such as defects in cytochrome P450 enzymes that catalyze the synthesis of cholesterol,[148] bile acids,[149] steroids,[150–153] and vitamin D3.[154] Further indispensable uses include the analysis of metabolic aberrations in lysosomal storage disorders that cause lysosomal accumulation of the enzyme-specific sphingolipid substrate[139,147] in GM_1 and GM_2 gangliosidosis, Gaucher's disease, Niemann–Pick disease, Fabry's disease, and many others. Accordingly, some of the first attempts to determine disease biomarkers by metabolic profiling have been applied in pediatrics research and diagnosis.[155] Other promising areas naturally comprise disorders associated with concentration changes of endogenous low-molecular-weight agents such as lipids, sugars, nucleotides, organic acids, and amino acids, for example, in vascular and cardiovascular diseases.

Therefore, metabolic-, age-, and obesity-related disorders such as diabetes mellitus, metabolic syndrome, atherosclerosis are natural fields for metabolomics applications. Four areas were selected as having the highest priority by the Metabolic Disorders Steering Committee of The Biomarkers Consortium (http://www.biomarkersconsortium.org/), a major public-private biomedical research partnership. These four areas were atherosclerosis, beta cell function, diabetes microvascular complications, and functional changes in aging.[156]

However, metabolomics is not confined to diseased states that are known or expected to involve metabolic disorders, but could be tested successfully in many other areas. An analysis of the current distribution of metabolomics-related publications across main disease areas (performed by us in October 2009, data not shown) indicates the main activities concern neoplastic diseases, metabolic and cardiovascular diseases, as well as disorders of the nervous, endocrine, and digestive systems. The distribution is likely to be influenced by the number of incidences as well as the medical relevance and need, but it also reflects the expected (or actual) likelihood of identifying relevant metabolites involved in the etiology and pathophysiology of target diseases. An ever increasing number of disease areas that already cover a wide spectrum are now being addressed by metabolomics techniques to improve diagnosis, measure disease progression, guide treatment, accelerate drug development, and target therapies. In the following we highlight some recent representative examples to demonstrate the potential of metabolites as biomarkers in several areas.

Metabolomics was used to monitor 18 patients without evidence of ischemia (control group) and 18 patients with inducible ischemia (case group) undergoing exercise stress testing. Plasma samples were collected before and after testing and analyzed with a high-sensitivity ESI triple-quadrupole mass spectrometer under SRM conditions. Whereas, as expected, the level of lactic acid increased after exercise in both cases, there was a significantly different regulation of 6 other metabolites, which could be used to distinguish ischemic patients from control subjects.[157,158] There is a need for improvement in the diagnostics of Crohn's disease, an inflammatory bowel disease with unclear etiology that is characterized by chronic inflammation of the gastrointestinal tract. Ion-cyclotron resonance-Fourier transform mass spectrometry (FT-ICR-MS), which enables ultra-high mass resolution, was used to discern the masses of thousands of metabolites in faeces collected from identical pairs of twins.[159] Markers differentially regulated in diseased and control groups are involved in the metabolism of amino acids, fatty acids, bile acids, and arachidonic acid, with significant differences found in the types and number of metabolites of tyrosine and phenylalanine metabolism as well as bile acid and fatty acid biosynthesis. The pathways identified may also help in understanding the pathobiochemistry underlying Crohn's disease, and could lead to future drug targets.

Human hepatopathies are challenging for diagnostics, since many distinct diseases present similar clinical signs and laboratory findings. Whitfield et al. used a canine hepatic disease model and metabolomics to successfully probe the diagnostic utility of metabolomics.[160] Metabolomics has also found application in the study of brain diseases and neurodegenerative diseases. Given the error rate of clinical parameters in Parkinson's disease, biomarkers for the diagnosis and monitoring of disease (progression) is of great clinical relevance. Bogdanov et al.[161] identified 8-hydroxy-2-deoxyguanosine, uric acid, and glutathione concentrations as a way to discriminate between patients and controls. They also highlighted rate metabolomics as a promising method for the identification of biomarkers for both the diagnosis and monitoring of disease progression.

Pre-eclampsia is a multisystem disorder of pregnancy that has major maternal and perinatal implications. Genetic programming was used to generate a model from GC-TOF-MS data of metabolites to discriminate pre-eclampsia samples from normal pregnant controls.[162] Metabolomics profiling also proved superiority to standard clinical chemistry parameters in the diagnosis of renal, liver, and heart transplant rejections[163,164] and kidney toxicity.[165]

It is known that tumor tissue, compared to healthy tissue, varies in terms of the concentrations of distinct metabolites (for example, in energy metabolism) and possess a distinct metabolic phenotype. Acordingly, metabolomics has been applied in oncology.[7,166] Most of these studies are, however, focused on ^{1}H NMR or PET-based investigations and the analysis of tissue and biopsies. So far, there has been no convincing report of quantitative differences in the periphery, that is, in plasma or serum and no convincing demonstration of an association with tumor incidence or size. We are of the opinion that differences in the concentration of some peripheral metabolites in plasma or serum (likely being quantitated by sensitive mass spectrometry) may also be used

in some cancer types in the near future to differentiate subjects with neoplastic diseases from healthy controls. Thus, this will have the promise for the early diagnose of such individuals, at least for some types of cancer. In addition, markers for monitoring treatment efficacy are urgently needed, and may be identified by quantitative metabolomics.

Longitudinal studies (with intrinsic normalization) on patients provide very valuable data in the discovery of diagnostic biomarkers. However, bias introduced, for example, by variation in food intake along with changing physiological conditions and other variables is still possible. Physiological stress also clearly affects the metabolome, which makes it very advisable to include hospitalized subjects not suffering from the target disease selected for marker identification. In this context, we have examined the effect of bypass surgery on metabolites of a fraction of the plasma metabolome. Quantitative analysis of over 200 metabolites was applied to 17 patients undergoing bypass surgery and monitored for a short time before and after.[193] The data clearly indicate that numerous significant changes in metabolite concentrations can be detected and that the time-points can be distinguished on the bases of quantitated metabolites (Figure 4).

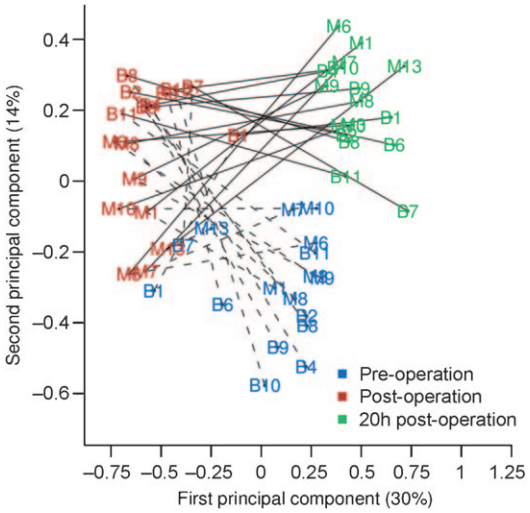


Figure 4. Mapping of targeted metabolomics profiles of 17 patients on the first two principal components (PC). For each individual, plasma samples were collected at the time of surgery as well at the end and 20 h following operation. By transforming the high-dimensional data to the coordinate system generated by the first two PCs, in sum representing 44% of the variability present in the data, one can follow the time-course of each patient, where each time point forms its own cluster.

The heatmap in Figure 5 shows that metabolite concentrations cluster at distinct times of sampling (before and after treatment) and numerous metabolite concentrations are differentially regulated by surgical trauma. Principal component analysis shows that most of the metabolite variability arises from the sampling time and physiological stress arising

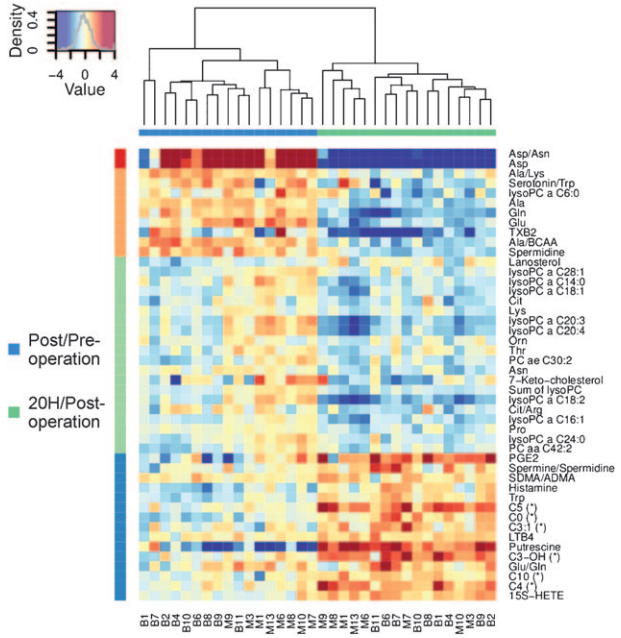


Figure 5. Data mining revealed that 46 features were altered significantly (at adjusted p value <0.01) around and following the operation. Visualization of the most relevant metabolite changes (each row represents a metabolite) by displaying actual concentration ratios after and before the operation (Pre Op.) and between 20 h after the operation and immediately after operation (Post Op.) for each patient (each column represents a patient). The dendrogram above the image shows the result of a hierarchical cluster analysis (based on the selected features and used for sorting the patients) which leads to a perfect separation of the two groups. Red and blue cells indicate high and low concentrations, respectively, in the sample from the second time point (see diagram in the top left corner).

from surgery, for example, medication (narcosis). Thus, sensitive responses of metabolite concentrations determined in body fluids in response to altered conditions and various treatments of a (diseased) subject have to be addressed through the use of appropriate controls to avoid selection of (not directly disease-related) biomarker candidates.

We are convinced that the value of targeted metabolomics will be greatest in the area of disease diagnosis. This is already evident with the use of mass spectrometry in clinical screening of neonates for metabolic disorders.[167,168] At the other end of the age spectrum, targeted lipidomics is potentially of major importance in the early diagnosis of neurodegenerative disease.

Almost one quarter of the body's cholesterol is found in the brain.[169] The major route for cholesterol export from the human brain is by metabolism to 24*S*-hydroxycholesterol, which unlike cholesterol can traverse the blood–brain barrier.[170,171] 24*S*-Hydroxycholesterol is formed from cholesterol in a reaction catalyzed by neuron-specific CYP46A1,[172] and exported from human brain at a rate of 6–7 mg day^{-1}.[170,171] In contrast to 24*S*-hydroxycholesterol, 27-hydroxycholesterol is imported into the brain at a rate of 5 mg day^{-1},[173] where it is metabolized further to 7α-hydroxy-3-oxocholest-4-en-27-oic acid and exported back across the blood–brain barrier at a rate of 2 mg day^{-1}.[174]

As CSF surrounds the brain and acts as a fluid to receive the metabolic products of brain cells, the oxysterol content of CSF has been suggested to represent a marker of cholesterol homeostasis within the brain.[175] Similarly, CSF levels of 24*S*-hydroxycholesterol should reflect the number of metabolically active neurons and also the availability of cholesterol for metabolism (by CYP46A1). During neurodegeneration, the number of metabolically active neurons must decrease; however, the concentration of cholesterol available for 24*S*-hydroxylation is likely to increase as it becomes free from membranes during neurodegeneration. The measurement of oxysterol levels (free oxysterol plus fatty acid esters) in CSF by GC-MS appears to reflect the rate of neuronal degeneration rather than the number of metabolically active neurons,[86] since levels of 24*S*-hydroxycholesterol are found to be significantly higher (ca. 2 ng mL^{-1}) in patients with Alzheimer's disease (AD), vascular dementia (VD), and mild cognitive impairment (MCI) than controls (< 1.5 ng mL^{-1}).[89,90,176] Levels of 27-hydroxycholesterol in CSF measured by GC-MS are also elevated in AD and MCI patients (1.8 ng mL^{-1}) over controls (1 ng mL^{-1}),[90] but this is suggested to be a consequence of a malfunctioning blood–brain barrier.[86] The above discussion demonstrates the potential value of targeted metabolomics. However, the target in the GC-MS studies was rather narrow. One way to broaden the target is to employ a less-selective analytical method.

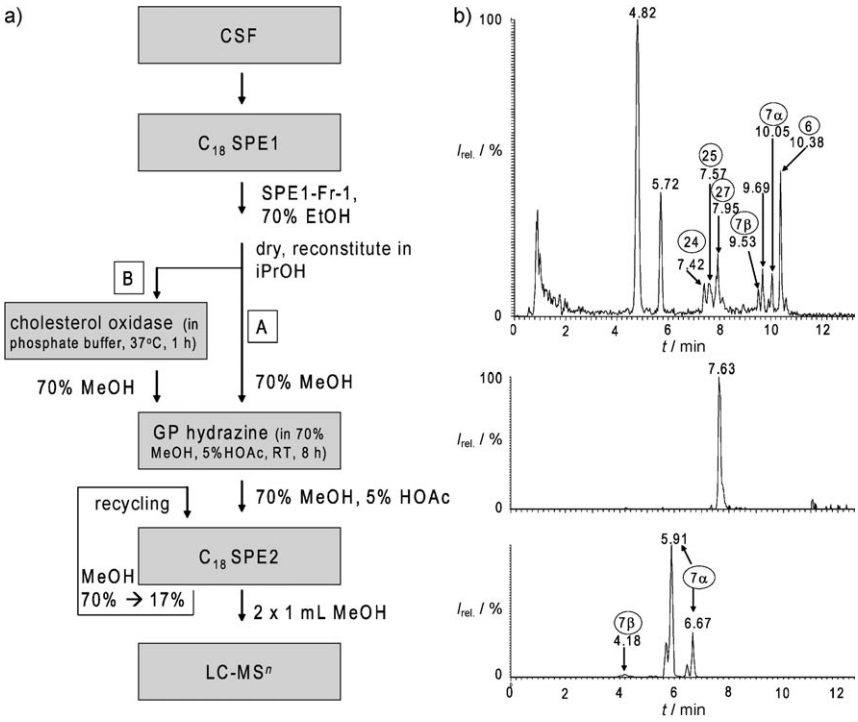


Figure 6. a) Sample preparation of bile acids and oxysterols by GP derivatization for LC-ESI-MS analysis. b) RICs for the $[M]^+$ ions of GP-derivatized monohydroxycholesterols (*m/z* 534.4; upper panel), hydroxycholestenoic acids (*m/z* 548.3847; central panel), and/or hydroxy-3-oxocholest-4-enoic acids (*m/z* 564.3796; lower panel).

The analytical window can be expanded to include acidic as well as neutral lipids by using a reversed-phase SPE cartridge rather than a normal-phase SPE one during the preparation of the CSF sample,. Furthermore, the use of GP derivatization with and with out incorporation of a cholesterol oxidase step (Figure 2) results in steroids, sterols, and bile acids containing both oxo and 3β-hydroxy-5-ene and 3β-hydroxy-5α-hydrogen moieties becoming visible by LC-ESI-MS. The following oxysterols and bile acids have been identified in human CSF by using this method: 24*S*-hydroxycholesterol (0.018 ± 0.009 ng mL^{-1}), 25-hydroxycholesterol (0.030 ± 0.010 ng mL^{-1}), 27-hydroxycholesterol (0.029 ± 0.012 ng mL^{-1}), 7β-hydroxycholesterol (0.009 ± 0.005 ng mL^{-1}), 7-oxocholesterol (0.034 ± 0.017 ng mL^{-1}), 3β-hydroxycholest-5-en-27-oic acid (0.416 ± 0.193 ng mL^{-1}), 3β,7β-dihydroxycholest-5-en-27-oic acid (0.165 ± 0.155 ng mL^{-1}) and 7α-hydroxy-3-oxocholes-4-en-27-oic acid (7.17 ± 2.826 ng mL^{-1}; Figure 6).[73] These values are for free concentrations, as no hydrolysis step was employed.

Metabolomics can be applied in drug research and development for the identification of new drug targets, prioritizing lead compounds, and assessing toxicity, thus contributing to the development of novel and safer drugs. With a need to demonstrate progress in terms of efficacy and tolerance, drug development faces a demanding challenge to minimize toxicity and predict toxicity as early as possible in the drug-development process. The application of metabolomics to the molecular toxicology of a drug candidate can contribute to the understanding of the underlying mechanisms and reduce failure, both in early and in late-stage drug development. Needless to say, the identification of biomarkers of drug toxicity for use in preclinical and clinical studies[177–181] is, therefore, one of the major areas of pharmaceutical research.

Metabolomics thus has the potential to reduce both the cost and time associated with the development of a new chemical compound, particularly since half of them currently fail in phase III clinical testing. Toxicity or adverse reactions (ADR) are, however, also issues for registered drugs. For example, ADR occurs in approximately 5% of all treated patients; it has been estimated that 2–20% of all hospital admissions are due to ADR, and about 10% of all hospitalized patients experience ADR during their hospital stay.[182] Among these, metabolic disorders represent 10.2% of ADR related to cardiovascular drug therapy in Germany[183] and about 12% in cardiology hospitals in France.[184] Interestingly, human embryonic stem cells capable of differentiating into a variety of cell types, may, as claimed by Stemina (http://www.stemina.com/), offer a promising alternative for toxicity screening, for example, cardiomyocyte progenitors could be used in predictive toxicology.

It certainly is too early to provide a thorough assessment of the role of metabolic biomarkers among other types of

biomolecules used in diagnostics. However, we have attempted to summarize some relevant features, advantages, and disadvantages of (the most commonly used) omics data and markers in regard to biomarker discovery, development, and diagnostic application (Table 2). It must be pointed out that this also only relates to applications in the mentioned contexts and does not intend any rating of the scientific value of the respective omics technology. Although an evaluation always remains incomplete, reflects the notions of the authors, and thus may be biased by personal views, we believe that some potential advantages of metabolites as markers are without question.

Metabolites (and their differential concentrations) denote functional endpoints and are thus likely to indicate the actual

Table 2: Summary of features of the discovery and application of RNA, protein, and metabolic biomarkers and comparative assessment.

Feature	Funct. genomics, transcriptomics		proteomics		Metabolomics	
species independency related to models and human subjects	species dependent	−	species dependent	−	cellular metabolic pathways are highly conserved among species; advantage in screening in disease models	++
coverage of whole -ome in screening	pan-genomic, $>10^6$ targets, splice variants; upcoming high-throughput sequencing;	+++	limited with 2D electrophoresis, $>10^6$ including posttranslat. modifications but poor accessibility; however, recent advances in MS technology	+	some way to a more complete quantitative coverage ($\approx 10^3$) requires application of multiple analytical platforms/methods	−−
time to response/ turnaround time/ sample processing	>2.5 h, PCR or microarrays; several steps of sample processing required	−	for example, antibody-based test, ELISA, array, >3 h; multistep sample prep. for MS	+/−	<2 h possible for smaller panel; direct analysis of plasma, serum, blood-spots	++
sample size/sensitivity	>2 mL blood without amplification, 100 ng total RNA with amplification	−	small (pg) for antibody tests (ELISA), arrays ng	+	for example, 20 μL plasma/ serum or 50 mg tissue for several methods long-term stability is an open question	++
reflecting actual state (of organism, cell etc.)	more a preview; majority of transcriptional changes require at least 1 h; utility depends on speed of disease progress	+/−	majority of changes in protein concentrations (w/o sec. modifications) requires at least 1 h	−	metabolites change rapidly in response to physiologic perturbations: minutes or less	+++
response to stimulus—dynamic range	transcriptional changes up to 100-fold or more	++	concentration changes about 10-fold; much information hidden in secondary modifications	+/−	metabolite concentrations affected by (enzymatic) amplification; affected by environment, nutrition etc.	++ evaluation depends on study objective
non-invasive option for biomarker assay	requires cell extraction	−−	usually requires cell/body liquid extraction, direct determination possible for serum/plasma using antibodies	−	possible for urine, saliva, faeces	++
detection of inborn errors	multigenetic alterations more likely to be reflected by transcriptomics	++	functionality required or sequence details, not well described by protein concentrations	−	common in pediatrics, less likely for multigenetic causes	++
use in toxicology	useful; however, functional consequences hard to assess	+	functional consequences hard to assess	+/−	provides functional readout; relevance has to be confirmed	+
data interpretation in context of systems biology, models pathways, etc.	numerous but frequently indirect interactions; also indirect evidence of interactions	++	limited by coverage of proteome	+	limited by coverage and database knowledge; disease dependent; good knowledge of pathways	+/−

state of an organism without much delay in signaling. This does not necessarily mean that metabolic marker candidates are, for example, more valuable as prognostic biomarkers; however, they could be the biomarkers of choice especially for very fast progressing diseases. Other factors here are the relatively small amount of sample—a highly relevant consideration in pediatrics—and the short time to response. On the other hand, the coverage of the metabolome with targeted methods to afford quantitative data is still relatively poor and thus somewhat limits the chances for discoveries in the early random screening of markers. However, this situation is improving rapidly.

Although biochemical interpretation of tissue metabolite data is relatively straightforward, plasma and serum concentrations are affected by multiple parameters from many compartments. However, although scientifically interesting and of potential value for further discoveries, we are convinced that biological plausibility is not an absolute prerequisite for any biomarker, such as a current lack of biological interpretation, and not an obstacle for further development. Conversely, this acceptance criterion would imply that we fully understand the biology as well as the kinetics and interactions of a system (and could model them)—a situation from which we are still very far removed. Cumulative biological evidence for the involvement of distinct pathways may, however, provide additional evidence and encouragement for further investigation in cases of insufficient statistical significance.

As a part of the FDA's Critical Path Initiative, intended to improve the predictability and efficiency of drug development, the FDA is preparing guidelines for the qualification and use of biomarkers. Metabolomics is an FDA-identified Critical Path Opportunity and has the potential to play a vital role in many areas: the development of better biomarkers, safety biomarkers related to kidney, liver, heart, and vascular damage, disease models for lead optimization and toxicity assessment, the identification of safety biomarkers and biomarkers of efficacy and toxicity to facilitate clinical trial designs, patient stratification, diagnostic monitoring of patient response to drug treatment, and many more.[185–187]

However, although platform compatibilities and processing of data as well as issues of (clinical) studies, for example, in microarray technology, have been addressed in several projects and many articles to provide a means for quality control, to develop guidelines (for microarray) data analysis, to evaluate the advantages and disadvantages of various data analysis methods, to reach consensus on the "best practices", and to establish quality control metrics and thresholds for assessing the performance of various (microarray) platforms, there is still a long way to go in metabolomics. In parallel with the more advanced state of functional genomics and the microarray technology, quality control methods and metrics for genomic data submissions as well as guidance are currently more developed. For example, definitions of genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories have been established.[188,189]

10. Beyond Data Integration

Global metabolic profiling has long been used to determine biomarkers to aid in the assessment of the pathophysiological health status of patients. The emergence of genomics and proteomics technologies has generated plausible mechanisms that correlate with the diagnostic and prognostic metabolic biomarkers in the health to disease continuum.[190,191] Global metabolic profiling has been referred to as either metabolomics[9] or metabonomics,[11] although both identify metabolic alterations under very different conditions. We have probed the statistical combination of biomarker sets from different types of biomolecules, independent of data integration and biochemical interpretation, to combined diagnostic signatures (combining several types of biomolecules). This has been performed on a purely statistical basis by applying various classification methods, and we have achieved encouraging results. It is important to note that this is distinct from approaches in which an integrative multidimensional analysis is utilized and combined, for example, genomes, epigenomes, and transcriptomes.[192] Such analyses attempt to analyze biological relationships between different omics data by various means.

We have employed support vector machines with a linear core to predict ischemia/hypoxia in rats (7 animals with ischemia versus 9 controls). By combining the two top-ranked (based on a t test and fold changes) metabolites (out of 275) with two mRNAs (SDF1 and VEGF), the estimated accuracy of the classifiers could be increased from 50% and 68.75%, respectively, to 75%. The corresponding ROC (receiver operating characteristic) curves are shown in Figure 7.

Our method thus combines measurements of at least two different types of biomolecules (RNA, noncoding RNA/microRNA, proteins, metabolites) on a mere statistical basis, irrespective of whether there is a known or unknown biological relationship of any kind, links, or apparent

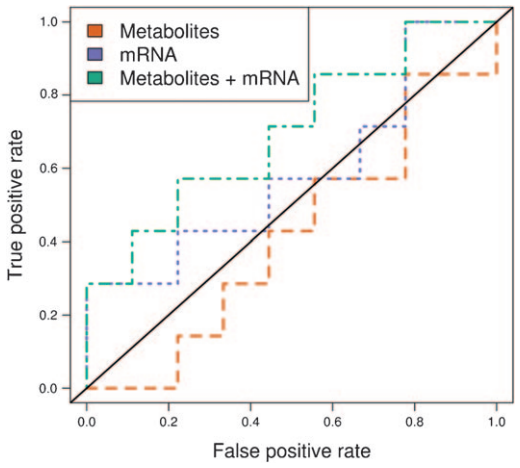


Figure 7. ROC (receiver-operating characteristic) curves for LOO (leave-one-out) cross-validated classifiers showing the superior accuracy of predictive signatures by combining distinct types of biomolecules.

biological plausibility, to afford a combined biomarker composed of several types of biomolecules.

The example presented here demonstrates that a diagnostic method and disease-state-specific classifier composed of at least two types of biomolecules with the combined biomolecules belonging to at least two distinct types may best describe the respective state of cells, a tissue, an organ, or an organism. The example delineated that molecular signatures and classifiers can be superior to a composition of molecules of the same type and afford a biomarker with high sensitivities and specificities, particularly in a complex disease such as cancer. A general statement, however, is subject to further confirmation, as is the biological rationale.

11. Analytical Challenges for Biomarkers

As long as reasonable experimental assumptions are respected, most of the metabolomics technologies can now routinely predict potential biomarkers. Alleviation of the bottleneck of taking a large number of candidates or the wrong set of candidates to validation begins at the discovery step. Several aspects must be taken into account. Data generation require optimized sample acquisition and handling (metabolite stability), optimized sample preparation (extraction efficiency, sample clean-up), and preservation to avoid sample aging to obtain validated analysis methods (accuracy, precision) with sufficient sensitivity, selectivity, and minimized matrix and carryover effects. Coupled with an adequate signal processing, efforts in validating chemical assays are paid off with respect to reducing technical variance and improving the number of true analytes that enter downstream analyses.

If careful design, extended randomization, proper selection of the subject, and adequate statistical treatment are absolute requirements to address potential biases and limit erroneous conclusions, assessment of the sample quality during collection and handling is often overlooked in a metabolomics experiment. Several studies report the effect of preservatives, shipping conditions, repeated freeze/thaw cycles, and glassware contamination as well as storage temperature and duration (up to 6 months) on metabolite levels through changes in the NMR and MS spectra. From the limited information available (platform sensitivities, no reproducibility of these experiments), physicochemical properties of metabolomics profile can be dramatically affected to such a point that handling-associated variance becomes a non-negligible masking factor. It is recommended that experimental protocols should be strictly controlled, be consistent during pretreatment prior to analysis, prevent enzymatic and chemical reaction by deep freezing, and restrict sample manipulation to a bare minimum. To date, little is known with regard to the effect of long-term storage on the (bio)chemical transformation of small molecules in complex matrices, thus raising questions about reproducible measurements from human samples stored in biobanks and prospective epidemiologic studies.

12. Conclusion/Trends in Metabolomics

To date, metabolomics could be regarded as still in its infancy since several issues have not been fully addressed for metabolite profiling to keep pace with genomics technologies. If different needs, experimental design, and machine availability will keep forms of fingerprinting metabolomics based on NMR spectroscopy or mass spectrometry in use for a long time, there will be a large shift toward more comprehensive, targeted, and quantitative analysis of the metabolome. Mass spectrometry is the most likely tool, with the potential to untangle the complex mammalian metabolome, which is estimated to be more than 3000 compounds.

Coupled with efficient chromatography and ad hoc data analysis, technologies that rely on FT-ICR-MS, Orbitrap, and asymmetric waveform ion mobility analyzers are emerging as dominant analytical methods for metabolomic studies because of the accuracy, high-throughput, and coverage (>1000 unique metabolites) that can be achieved. Despite providing accurate mass that may reduce the number of potential molecular formulas down to a few candidates, these strategies cannot alone provide structural information. Recently, the combination of metabolite-specific databases in relation to species and biological context (restricting potential candidates) with spectral libraries (behavior of chemical standards) has allowed definitive identification or at least putative identification. On-going exchange of chemical and analytical information must be encouraged for metabolomics to expand.

Despite clearly differentiated chromatographic and spectral features, the chemical origins of many analytes from large-scale metabolome analyses remain uncertain or unknown. The characterization of novel metabolites, whether endogenous or exogenous, is routinely ignored and there is little doubt that elucidation is the next obstacle to overcome to expand our collection of quantifiable targets. Quantitative metabolomics is expected to increasingly gain favor over the currently dominant chemometric-based methods that rely on raw intensities and peak areas, since it provides the most sensible approach to circumvent artefacts arising from matrix effects and to allow meaningful comparable measurements across datasets. The benefits are, however, limited by the cost and availability of labeled compounds as well as the time-consuming procedures for validating quantitation methods. In addition to these, restriction of sample volumes (with small animal models or neonates) may not permit extensive panels of chemical assays to be run to cover a large part of the metabolome.

Attention is now turning to the use of surrogate matrixes or indirect quantification by using representative labeled compounds to expand the chemical coverage. We believe that these practices should remain at the discovery stage and common analytical procedures (biomarker verification) must be enforced before claiming a MS signal to be a potential biomarker and likely to meet clinical demands. There is, however, convincing evidence that the use of metabolomics as a screening or a diagnostics method, or metabolites as a part of a diagnostic pattern, will become an established part of diagnostics in the near future. The proof of principle of the

suitability of metabolites as disease markers was supplied decades ago with the successful use of single markers of metabolic disorders. It can, however, be expected that future applications will include many diseases where links to alterations in metabolite concentrations are not that evident. Apart from metabolomics, the future focus here is clearly on the combination of several markers (including also different types of biomolecules) as it can not be expected that useful information on complex diseases may be accumulated with single molecules. We are convinced that the performance of future diagnostics in terms of sensitivities and specificities can thus be further improved significantly.

Abbreviations

AD	Alzheimer's disease
ADR	Adverse drug reaction
API	Atmospheric pressure ionization
AUC	Area under the ROC curve
CSF	Cerebrospinal fluid
CI	Chemical ionization
CID	Collision-induced dissociation
CYP46A1	Cytochrome P450, family 46, subfamily A, polypeptide 1
DESI	Desoption electrospray ionization
DHEA	Dehydroepiandrosterone
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EI	Electron impact
ESI-MS	Electrospray ionization mass spectrometry
FIA-MS	Flow injection analysis mass spectrometry
FID	Flame ionization detection
FT-MS	Fourier-transform mass spectrometry
GC	Gas chromatography
GC-MS	Gas chromatography–mass spectrometry
GP	Girard P hydrazine
HPLC	High-performance liquid chromatography
ICAT	Isotope-coded affinity tag
MALDI	Matrix-assisted laser desorption/ionization
MCI	Mild cognitive impairment
MS	Mass spectrometry
MS/MS or MS^2	Tandem mass spectrometry
NMR	Nuclear resonance spectroscopy
PET	Positron emission tomography
SPE	Solid-phase extraction
SRM	Selected reaction monitoring
TMS	Trimethylsilyl
TOF-MS	Time of flight mass spectrometry
UPLC	Ultraperformance liquid chromatography
VD	Vascular disease

Received: October 6, 2009
Published online: July 13, 2010

[1] N. Blow, *Nature* **2008**, *455*, 697 – 700.
[2] C. Curtin, *Genome Technol.* **2008**, 31 – 35.
[3] R. P. Horgan, O. H. Clancy, J. E. Myers, P. N. Baker, *BJOG* **2009**, *116*, 173 – 181.
[4] D. Ryan, K. Robards, *Anal. Chem.* **2006**, *78*, 7954 – 7958.
[5] J. K. Nicholson, J. C. Lindon, *Nature* **2008**, *455*, 1054 – 1056.
[6] J. C. Lindon, E. Holmes, J. K. Nicholson, *FEBS J.* **2007**, *274*, 1140 – 1151.
[7] J. L. Spratlin, N. J. Serkova, S. G. Eckhardt, *Clin. Cancer Res.* **2009**, *15*, 431 – 440.
[8] D. B. Kell, *Curr. Opin. Microbiol.* **2004**, *7*, 296 – 307.
[9] O. Fiehn, *Plant Mol. Biol.* **2002**, *48*, 155 – 171.
[10] H. Tweeddale, L. Notley-McRobb, T. Ferenci, *J. Bacteriol.* **1998**, *180*, 5109 – 5116.
[11] J. K. Nicholson, J. C. Lindon, E. Holmes, *Xenobiotica* **1999**, *29*, 1181 – 1189.
[12] C. E. Dalgliesh, E. C. Horning, M. G. Horning, K. L. Knox, K. Yarger, *Biochem. J.* **1966**, *101*, 792 – 810.
[13] L. Pauling, A. B. Robinson, R. Teranishi, P. Cary, *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 2374 – 2376.
[14] E. C. Horning, M. G. Horning, *Clin. Chem.* **1971**, *17*, 802 – 809.
[15] O. A. Mamer, J. C. Crawhall, S. S. Tjoa, *Clin. Chim. Acta* **1971**, *32*, 171 – 184.
[16] K. Hollywood, D. R. Brison, R. Goodacre, *Proteomics* **2006**, *6*, 4716 – 4723.
[17] E. J. Want, A. Nordström, H. Morita, G. Siuzdak, *J. Proteome Res.* **2007**, *6*, 459 – 468.
[18] W. Lu, B. D. Bennett, J. D. Rabinowitz, *J. Chromatogr. B* **2008**, *871*, 236 – 242.
[19] W. B. Dunn, N. J. C. Bailey, H. E. Johnson, *Analyst* **2005**, *130*, 606 – 625.
[20] X. Feng, X. Liu, Q. Luo, B. F. Liu, *Mass Spectrom. Rev.* **2008**, *27*, 635 – 660.
[21] K. Dettmer, P. A. Aronov, B. D. Hammock, *Mass Spectrom. Rev.* **2007**, *26*, 51 – 78.
[22] J. Nielsen, S. Oliver, *Trends Biotechnol.* **2005**, *23*, 544 – 546.
[23] E. Werner, J. F. Heilier, C. Ducruix, E. Ezan, C. Junot, *J. Chromatogr. B* **2008**, *871*, 143 – 163.
[24] Q. N. Van, T. D. Veenstra, *Genome Med.* **2009**, *1*, 5.
[25] H. J. Issaq, Q. N. Van, T. J. Waybright, G. M. Muschik, T. D. Veenstra, *J. Sep. Sci.* **2009**, *32*, 2183 – 2199.
[26] W. J. Griffiths, Y. Wang, *Chem. Soc. Rev.* **2009**, *38*, 1882 – 1896.
[27] C. Hu, R. van der Heijden, M. Wang, J. van der Greef, T. Hankemeier, G. Xu, *J. Chromatogr. B* **2009**, *877*, 2836 – 2846.
[28] R. G. Lowe, M. Lord, K. Rybak, R. D. Trengove, R. P. Oliver, P. S. Solomon, *Fungal Genet. Biol.* **2009**, *46*, 381 – 389.
[29] Y. Sawada, K. Akiyama, A. Sakata, A. Kuwahara, H. Otsuki, T. Sakurai, K. Saito, M. Y. Hirai, *Plant Cell Physiol.* **2009**, *50*, 37 – 47.
[30] R. G. Lowe, M. Lord, K. Rybak, R. D. Trengove, R. P. Oliver, P. S. Solomon, *Fungal Genet. Biol.* **2008**, *45*, 1479 – 1486.
[31] K. Hanhineva, I. Rogachev, H. Kokko, S. Mintz-Oron, I. Venger, S. Karenlampi, A. Aharoni, *Phytochemistry* **2008**, *69*, 2463 – 2481.
[32] S. J. Rochfort, V. C. Trenerry, M. Imsic, J. Panozzo, R. Jones, *Phytochemistry* **2008**, *69*, 1671 – 1679.
[33] E. Altmaier, S. L. Ramsay, A. Graber, H. W. Mewes, K. M. Weinberger, K. Suhre, *Endocrinology* **2008**, *149*, 3478 – 3489.
[34] A. U. Jackson, S. R. Werner, N. Talaty, Y. Song, K. Campbell, R. G. Cooks, J. A. Morgan, *Anal. Biochem.* **2008**, *375*, 272 – 281.
[35] T. R. Koves, J. R. Ussher, R. C. Noland, D. Slentz, M. Mosedale, O. Ilkayeva, J. Bain, R. Stevens, J. R. Dyck, C. B. Newgard, G. D. Lopaschuk, D. M. Muoio, *Cell Metab.* **2008**, *7*, 45 – 56.
[36] W. J. Griffiths, K. Karu, M. Hornshaw, G. Woffendin, Y. Wang, *Eur. J. Mass Spectrom.* **2007**, *13*, 45 – 50.
[37] C. E. Wheelock, S. Goto, B. D. Hammock, J. W. Newman, *Metabolomics* **2007**, *3*, 137 – 145.

[38] D. J. McNally, A. J. Aubry, J. P. Hui, N. H. Khieu, D. Whitfield, C. P. Ewing, P. Guerry, J. R. Brisson, S. M. Logan. E. Soo, *J. Biol. Chem.* **2007**, *282*, 14463 – 14475.
[39] S. P. Devaiah, M. R. Roth, E. Baughman, M. Li, P. Tamura, R. Jeannotte, R. Welti, X. Wang, *Phytochemistry* **2006**, *67*, 1907 – 1924.
[40] J. C. Verdonk, V. de Ric, H. A. Verhoeven, M. A. Haring, A. J. van Tunen, R. C. Schuurink, *Phytochemistry* **2003**, *62*, 997 – 1008.
[41] U. Ceglarek, A. Leichtler, M. Brügel, L. Kortz, R. Brauer, K. Bresler, J. Thiery, G. M. Fiedler, *Mol. Cell. Endocrinol.* **2009**, *301*, 266 – 271.
[42] D. S. Wishart, M. J. Lewis, J. A. Morrissey, M. D. Flegel, K. Jeronic, Y. Xiong, D. Cheng, R. Eisner, B. Gautam, D. Tzur, S. Sawhney, F. Bamforth, R. Greiner, L. Li, *J. Chromatogr. B* **2008**, *871*, 164 – 173.
[43] J. L. Griffin, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 147 – 161.
[44] C. Gieger, L. Geistlinger, E. Altmaier, M. Hrabé de Angelis, F. Kronenberg, T. Meitinger, H.-W. Mewes, H.-E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig, K. Suhre, *PLoS Genet.* **2008**, *4*, e1000282; de Angelis, F. Kronenberg, T. Meitinger, H.-W. Mewes, H.-E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig, K. Suhre, *PLoS Genet.* **2008**, *4*, e1000282; de Angelis, F. Kronenberg, T. Meitinger, H.-W. Mewes, H.-E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig, K. Suhre, *PLoS Genet.* **2008**, *4*, e1000282.
[45] W. J. Griffiths, J. Sjövall, *J. Lipid Res.* **2009**, in press.
[46] K. D. Setchell, J. E. Heubi, *J. Pediatr. Gastroenterol. Nutr.* **2006**, *43* Suppl 1, 17 – 22.
[47] P. J. Trim, S. J. Atkinson, A. P. Princivalle, P. S. Marshall, A. West, M. R. Clench, *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1503 – 1509.
[48] K. E. Burnum, D. S. Cornett, S. M. Puolitaival, S. B. Milne, D. S. Myers, S. Tranguch, H. A. Brown, S. K. Dey, R. M. Caprioli, *J. Lipid Res.* **2009**, in press.
[49] R. C. Murphy, J. A. Hankin, R. M. Barkley, *J. Lipid Res.* **2009**, *50*, S317 – 322.
[50] C. Wu, D. R. Ifa, N. E. Manicke, R. G. Cooks, *Anal. Chem.* **2009**, *81*, 7618 – 7624.
[51] J. B. Fenn, *Angew. Chem.* **2003**, *115*, 3999 – 4024; *Angew. Chem. Int. Ed.* **2003**, *42*, 3871 – 3894.
[52] N. B. Cech, C. G. Enke, *Mass Spectrom.* **2001**, *20*, 362 – 387.
[53] E. J. Want, G. O'Maille, C. A. Smith, T. R. Brandon, W. Uritboonthai, C. Qin, S. A. Trauger, G. Siuzdak, *Anal. Chem.* **2006**, *78*, 743 – 752.
[54] X. Jiang, H. Cheng, K. Yang, R. W. Gross, X. Han, *Anal. Biochem.* **2007**, *371*, 135 – 145.
[55] G. O'Maille, E. P. Go, L. Hoang, E. J. Want, C. Smith, P. O'Maille, A. Nordstrom, H. Morita, C. Qin, W. Uritboonthai, J. Apon, R. Moore, J. Garrett, G. Siuzdak, *Spectrosc. Int. J.* **2008**, *22*, 327 – 343.
[56] B. Crews, W. R. Wikoff, G. J. Patti, H. K. Woo, E. Kalisiak, J. Heideker, G. Siuzdak, *Anal. Chem.* **2009**, *81*, 8538 – 8544.
[57] E. Zelena, W. B. Dunn, D. Broadhurst, S. Francis-McIntyre, K. M. Carroll, P. Begley, S. O'Hagan, J. D. Knowles, A. Halsall, *Anal. Chem.* **2009**, *81*, 1357 – 1364.
[58] K. Schmelzer, E. Fahy, S. Subramaniam, E. A. Dennis, *Methods Enzymol.* **2007**, *432*, 171 – 183.
[59] M. P. Wymann, R. Schneiter, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 162 – 176.
[60] W. J. Griffiths, M. Hornshaw, G. Woffendin, S. F. Baker, A. Lockhart, S. Heidelberger, M. Gustafsson, J. Sjövall, Y. Wang, *J. Proteome Res.* **2008**, *7*, 3602 – 3612.
[61] P. Liere, A. Pianos, B. Eychenne, A. Cambourg, K. Bodin, W. Griffiths, M. Schumacher, E. E. Baulieu, J. Sjovall, *J. Lipid Res.* **2009**, in press.
[62] G. J. Schroepfer, Jr., *Physiol. Rev.* **2000**, *80*, 361 – 554.
[63] S. Liu, W. J. Griffiths, J. Sjövall, *Anal. Chem.* **2003**, *75*, 791 – 797.
[64] Y. Yang, W. J. Griffiths, H. Nazer, J. Sjövall, *Biomed. Chromatogr.* **1997**, *11*, 240 – 255.
[65] G. J. Van Berkel, J. M. Quirke, R. A. Tigani, A. S. Dilley, T. R. Covey, *Anal. Chem.* **1998**, *70*, 1544 – 1554.
[66] D. W. Johnson, H. J. ten Brink, R. C. Schuit, C. Jakobs, *J. Lipid Res.* **2001**, *42*, 9 – 16.
[67] T. Higashi, K. Shimada, *Anal. Bioanal. Chem.* **2004**, *378*, 875 – 882.
[68] T. Higashi, A. Yamauchi, K. Shimada, *J. Chromatogr. B* **2005**, *825*, 214 – 222.
[69] W. J. Griffiths, S. Liu, G. Alvelius, J. Sjövall, *Rapid Commun. Mass Spectrom.* **2003**, *17*, 924 – 935.
[70] H. L. Makin, J. W. Honour, C. H. Shackleton, W. J. Griffiths, *Chromatography and Mass Spectrometry in Steroid analysis*, 2nd ed. (Eds.: H. L. J. Makin, D. B. Gower), Springer, Berlin, **2009**, 163 – 282.
[71] M. Axelson, B. Mörk, J. Sjövall, *J. Lipid Res.* **1988**, *29*, 629 – 641.
[72] J. MacLachlan, A. T. Wotherspoon, R. O. Ansell, C. J. Brooks, *J. Steroid Biochem. Mol. Biol.* **2000**, *72*, 169 – 195.
[73] M. Ogundare, S. Theofilopoulos, A. Lockhart, L. J. Hall, E. Arenas, J. Sjövall, A. G. Brenton, Y. Wang, W. J. Griffiths, *J. Biol. Chem.* **2010**, *285*, 4666 – 4679.
[74] P. Eneroth, K. Hellstroem, R. Ryhage, *J. Lipid Res.* **1964**, *5*, 245 – 262.
[75] E. C. Horning, C. J. Brooks, W. J. Vanden Heuvel, *Adv. Lipid Res.* **1968**, *6*, 273 – 392.
[76] Y. S. Kim, H. Zhang, H. Y. Kim, *Anal. Biochem.* **2000**, *277*, 187 – 195.
[77] A. Jiye, J. Trygg, J. Gullberg, A. I. Johansson, P. Jonsson, H. Antti, S. L. Marklund, T. Moritz, *Anal. Chem.* **2005**, *77*, 8086 – 8094.
[78] R. P. Evershed in *Handbook of Derivatives for Chromatography* (Eds.: K. Blau, J. M. Halket), Wiley, Chichester, **1993**, pp. 52 – 108.
[79] "Mass spectrometry for metabolite identification": Y. Wang, W. J. Griffiths in *Metabolomics, Metabonomics and Metabolite Profiling* (Ed.: W. J. Griffiths), The Royal Society of Chemistry, Cambridge, **2008**, 1 – 43.
[80] J. M. Lehmann, S. A. Kliewer, L. B. Moore, T. A. Smith-Oliver, B. B. Oliver, J. L. Su, S. S. Sundseth, D. A. Winegar, D. E. Blanchard, T. A. Spencer, T. M. Willson, *J. Biol. Chem.* **1997**, *272*, 3137 – 3140.
[81] A. Radhakrishnan, Y. Ikeda, H. J. Kwon, M. S. Brown, J. L. Goldstein, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6511 – 6518.
[82] C. Song, S. Liao, *Endocrinology* **2000**, *141*, 4180 – 4184.
[83] P. B. Hylemon, H. Zhou, W. M. Pandak, S. Ren, G. Gil, P. Dent, *J. Lipid Res.* **2009**, *50*, 1509 – 1520.
[84] K. D. Setchell, M. Schwarz, N. C. O'Connell, E. G. Lund, D. L. Davis, R. Lathe, H. R. Thompson, R. Weslie Tyson, R. J. Sokol, D. W. Russell, *J. Clin. Invest.* **1998**, *102*, 1690 – 1703.
[85] A. Babiker, S. Dzeletovic, B. Wiklund, N. Pettersson, J. Salonen, K. Nyyssönen, M. Eriksson, U. Diczfalusy, I. Björkhem, *Scand. J. Clin. Lab. Invest.* **2005**, *65*, 365 – 375.
[86] I. Björkhem, A. Cedazo-Minguez, V. Leoni, S. Meaney, *Mol. Aspects Med.* **2009**, *30*, 171 – 179.
[87] S. Dzeletovic, O. Breuer, E. Lund, U. Diczfalusy, *Anal. Biochem.* **1995**, *225*, 73 – 80.
[88] V. Leoni, D. Lütjohann, T. Masterman, *J. Lipid Res.* **2005**, *46*, 191 – 195.
[89] A. Papassotiropoulos, D. Lütjohann, M. Bagli, S. Locatelli, F. Jessen, R. Buschfort, U. Ptok, I. Björkhem, K. von Bergmann, R. Heun, *J. Psychiatr. Res.* **2002**, *36*, 27 – 32.
[90] V. Leoni, M. Shafaati, A. Salomon, M. Kivipelto, I. Björkhem, L. O. Wahlund, *Neurosci. Lett.* **2006**, *397*, 83 – 87.

[91] M. Shafaati, A. Solomon, M. Kivipelto, I. Björkhem, V. Leoni, *Neurosci. Lett.* **2007**, *425*, 78 – 82.
[92] D. P. Enot, W. Lin, M. Beckmann, D. Parker, D. P. Overy, J. Draper, *Nat. Protoc.* **2008**, *3*, 446 – 470.
[93] L. W. Sumner, E. Urbanczyk-Wochniak, C. D. Broeckling, *Methods Mol. Biol.* **2008**, *406*, 409 – 436.
[94] K. R. Coombes, K. A. Baggerly, J. S. Morris in *Fundamentals of Data Mining in Genomics and Proteomics* (Eds.: M. Dubitzky, M. Granzow, D. Berrar), Kluwer, Dordrecht, **2007**, pp. 79 – 99.
[95] M. Katajamaa, M. Orešič, *J. Chromatogr. A* **2007**, *1158*, 318 – 328.
[96] A. D. McNaught, A. Wilkinson, *IUPAC. Compendium of Chemical Terminology (the "Gold Book")*, 2nd ed., Blackwell Scientific Publications, Oxford, **1997**; XML on-line corrected version: http://goldbook.iupac.org (2006–); created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins.
[97] S. G. Villas-Boas, U. Roessner, M. A. E. Hansen, J. Smedsgaard, J. Nielsen, *Metabolome Analysis: An Introduction*, Wiley, Hoboken, **2007**, p. 157.
[98] H. Shin, M. K. Markey, *J. Biomed. Inf.* **2006**, *39*, 227 – 248.
[99] G. Vivo-Truyols, J. Torres-Lapasio, A. van Nederkassel, Y. Vander Heyden, D. L. Massart, *J. Chromatogr. A* **2005**, *1096*, 133 – 145.
[100] R. J. O. Torgrip, K. M. Åberg, E. Alm, I. Schuppe-Koistinen, J. Lindberg, *Metabolomics* **2008**, *4*, 114 – 121.
[101] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *Anal. Chem.* **2006**, *78*, 4281 – 4290.
[102] B. M. Warrack, S. Hnatyshyn, K.-H. Ott, M. D. Reily, M. Sanders, H. Zhang, D. M. Drexler, *J. Chromatogr. B* **2009**, *877*, 547 – 552.
[103] M. Sysi-Aho, M. Katajamaa, L. Yetukuri, M. Orešič, *BMC Bioinf.* **2007**, *8*, 93.
[104] "Reference Gene Validation Software for Improved Normalization": J. Vandesompele, M. Kubista, M. W. Pfaffl in *Real-Time PCR: Current Technology and Applications* (Eds.: J. Logan, K. Edwards, N. Saunders), Caister Academic Press, London, **2009**, 47 – 64.
[105] M. Kohl, D. Enot, M. Keller, T. Koal, K. Weinberger, H. P. Deigner, unpublished results.
[106] W. Huber, A. von Heydebreck, H. Sueltmann, A. Poustka, M. Vingron, *Bioinformatics* **2002**, *18*(Suppl.1), 96 – 104.
[107] G. E. P. Box, D. R. Cox, *J. R. Stat. Soc. B* **1964**, *26*, 211 – 252.
[108] R. A. van den Berg, C. J. Hoefsloot Huub, J. A. Westerhuis, A. K. Smilde, M. J. van der Werf, *BMC Genomics* **2006**, *7*, 142.
[109] S. Dudoit, M. J. van der Laan, *Multiple Testing Procedures with Applications to Genomics*, Springer, Berlin, **2008**.
[110] MAQC Consortium, *Nat. Biotechnol.* **2006**, *24*, 1151–1161.
[111] R. Simon, *J. Stat. Planning Inference* **2008**, *138*, 308 – 320.
[112] F. Hahne, W. Huber in *Bioconductor Case Studies* (Eds.: F. Hahne, W. Huber, R. Gentleman, S. Falcon), Springer, Berlin, **2008**, 83 – 88.
[113] M. A. Hall, G. Holmes, *IEEE Trans. Knowl. Data Eng.* **2003**, *15*, 1437 – 1447.
[114] N. Pochet, F. De Smet, J. A. Suykens, B. L. De Moor, *Bioinformatics* **2004**, *20*, 3185 – 3195.
[115] T. M. Mitchell, *Machine Learning*, McGraw-Hill, **1997**.
[116] R. Tibshirani, T. Hastie, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, Berlin, **2009** (Springer Series in Statistics).
[117] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, Berlin, **2008** (Springer Texts in Statistics).
[118] A. Dupuy, R. Simon, *J. Natl. Cancer Inst.* **2007**, *99*, 147 – 157.
[119] http://www.ncbi.nlm.nih.gov/pubmed/8038641.
[120] S. Varma, R. Simon, *BMC Bioinf.* **2006**, *7*, 91.
[121] http://www.ncbi.nlm.nih.gov/pubmed/17392326.
[122] R. G. W. Verhaak, F. J. T. Staal, P. J. M. Valk, B. Lowenberg, M. J. T. Reinders, D. de Ridder, *BMC Bioinf.* **2006**, *7*, 105 – 120.
[123] http://www.ncbi.nlm.nih.gov/pubmed/19941644.
[124] R. Simon, *J. Natl. Cancer Inst.* **2006**, *98*, 1169 – 1171.
[125] B. J. Biggerstaff, *Stat. Med.* **2000**, *19*, 649 – 663.
[126] D. I. Broadhurst, D. B. Kell, *Metabolomics* **2006**, *2*, 171 – 196.
[127] R. Goodacre, D. Broadhurst, A. Smilde, B. S. Kristal, J. D. , Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg, F. Wulfert, *Metabolomics* **2007**, *3*, 231 – 241.
[128] M. Oresic, *Nutr. Metab. Cardiovasc Dis.* **2009**, in press.
[129] J. G. Bundy, D. Osborn, J. M. Weeks, J. C. Lindon, J. K. Nicholson, *FEBS Lett.* **2001**, *500*, 31 – 35.
[130] J. G. Bundy et al., *Environme. Toxicol. Chem.* **2002**, *21*, 1966 – 1972.
[131] M. R. Viant,, J. G. Bundy, C. A. Pincetich, J. S. de Ropp, R. S. Tjeerdema, *Metabolomics* **2005**, *1*, 149 – 158.
[132] M. R. Viant, C. A. Pincetich, D. E. Hinton, R. S. Tjeerdema, *Aquat. Toxicol.* **2006**, *76*, 329 – 342.
[133] National Human Genome Research Institute, An overview of the Human Genome Project. http://www.genome.gov/12011238. Accessed June 1, 2010.
[134] S. Vangala, A. Tonelli, *The AAPS J.* **2007**, 9, E284.
[135] P. M. Harrison, A. Kumar, N. Lang, M. Snyder, M. Gerst, *Nucleic Acids Res.* **2002**, *30*, 1083 – 1090.
[136] K. Dettmer, B. D. Hammock, *Environ. Health Perspect.* **2004**, *112*, 396 – 397.
[137] M. Oresic, A. Vidal-Puig, V. Hänninen, *Expert Rev. Mol. Diagn.* **2006**, *6*, 575 – 585.
[138] W. H. Heijne, A. S. Kienhuis, B. van Ommen, R. H. Stierum, J. P. Groten, *Expert Rev. Proteomics* **2005**, *2*, 767 – 780.
[139] A. Velázquez, M. Vela-Amieva, I. Cicerón-Arellano, I. Ibarra-González, M. E. Pérez-Andrade, Z. Olivares-Sandoval, G. Jiménez-Sánchez, *Arch. Med. Res.* **2000**, *31*, 145 – 150.
[140] W. Lehnert, D. Hunkler, *Eur. J. Pediatr.* **1986**, *145*, 260 – 266.
[141] C. Baumgartner, D. Baumgartner, *J. Biomol. Screening* **2006**, *11*, 90 – 99.
[142] C. L. Yu, X. F. Gu, *J. Peking Univ. Health Sci.* **2006**, *38*, 103 – 106.
[143] C. D. Campbell, J. Ganesh, C. Ficicioglu, *Haematologica* **2005**, *90*, ECR45.
[144] E. J. Want, B. F. Cravatt, G. Siuzdak, *ChemBioChem* **2005**, *6*, 1941 – 1951.
[145] M. Piraud, C. Vianey-Saban, C. Bourdin et al., *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3287 – 3297.
[146] M. Dott, D. Chance, M. Fierro, et al., *Am. J. Med. Genet. Part A* **2006**, *140*, 837 – 842.
[147] B. K. Burton, *Pediatrics* **1998**, *102*, e69.
[148] F. D. Porter, *J. Clin. Invest.* **2002**, *110*, 715 – 724.
[149] P. Subramaniam, P. T. Clayton, B. C. Portmann, G. Mieli-Vergani, N. Hadzić, *J. Pediatr. Gastroenterol. Nutr.* **2010**, *50*, 61 – 66.
[150] M. I. New, R. C. Wilson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12790 – 12797.
[151] K. Fujieda, T. Tajima, *Pediatr. Res.* **2005**, *57*, 62R.
[152] N. Krone, F. G. Riepe, J. Grotzinger, et al., *J. Clin. Endocrinol. Metab.* **2005**, *90*, 445 – 454.
[153] N. Krone, F. G. Rieppe, D. Gotze, et al., *J. Clin. Endocrinol. Metab.* **2005**, *90*, 3724 – 3730.
[154] N. Sawada, T. Sakaki, S. Kato, K. Inouye, *Eur. J. Biochem.* **2001**, *268*, 6607 – 6615.
[155] M. A. Constantinou, et al., *Anal. Chim. Acta* **2005**, *542*, 169 – 177.
[156] J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, et al., *Anal. Chem.* **2008**, *80*, 1280 – 1289.

[157] G. D. Lewis, A. Asnani, R. E. Gerszten, *J. Am. Coll. Cardiol.* **2008**, *52*, 117 – 123.
[158] R. Wiegand, G. F. Berriz, F. P. Roth, R. E. Gerszten, et al., *Circulation* **2005**, *112*, 3868 – 3875.
[159] J. Jansson, B. Willing, M. Lucio, A. Fekete, J. Dicksved, et al., *PLoS One* **2009**, *4*, e6386.
[160] P. D. Whitfield, P. J. M. Noble, H. Major, R. J. Beynon, R. Burrow, A. I. Freeman, A. James, *German Metabolomics* **2005**, *1*, 215 – 225.
[161] M. Bogdanov, W. R. Matson, L. Wang, T. Matson, R. Saunders-Pullman, S. S. Bressman, M. Flint, *Brain* **2008**, *131*, 389 – 396.
[162] L. C. Kenny, W. B. Dunn, D. I. Ellis, J. Myers, P. N. Baker, the GOPEC Consortium, D. B. Kell, *German Metabolomics* **2005**, *1*, 227 – 234.
[163] M. A. Silva, et al., *Liver Transplant.* **2006**, *12*, 146 – 151.
[164] D. S. Wishart, *Am. J. Transplant.* **2005**, *5*, 2814 – 2820.
[165] J. D. Bell, J. A. Lee, H. A. Lee, P. J. Sadler, D. R. Wilkie, R. H. Woodham, *Biochim. Biophys. Acta Mol. Basis Dis.* **1991**, *1096*, 101 – 107.
[166] A. Sreekumar, L. M. Poisson, T. K. Rajendiran, A. P. Khan, Q. Cao, J. Yu, B. Laxman, R. Mehra, R. J. Lonigro, Y. Li, M. K. Nyati, A. Ahsan, S. Kalyana-Sundaram, B. Han, X. Cao, J. Byun, G. S. Omenn, D. Ghosh, S. Pennathur, D. C. Alexander, A. Berger, J. R. Shuster, J. T. Wei, S. Varambally, C. Beecher, A. M. Chinnaiyan, *Nature* **2009**, *457*, 910 – 914.
[167] D. H. Chace, *Chem. Rev.* **2001**, *101*, 445 – 477.
[168] J. W. Honour, *Ann. Clin. Biochem.* **2003**, *40*, 628 – 638.
[169] J. M. Dietschy, S. D. Turley, *J. Lipid Res.* **2004**, *45*, 1375 – 1397.
[170] D. Lütjohann, O. Breuer, G. Ahlborg, I. Nennesmo, A. Sidén, U. Diczfalusy, I. Björkhem, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9799 – 9804.
[171] I. Björkhem, D. Lütjohann, U. Diczfalusy, L. Ståhle, G. Ahlborg, J. Wahren, *J. Lipid Res.* **1998**, *39*, 1594 – 1600.
[172] E. G. Lund, J. M. Guileyardo, D. W. Russell, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7238 – 7243.
[173] M. Heverin, S. Meaney, D. Lütjohann, U. Diczfalusy, J. Wahren, I. Björkhem, *J. Lipid Res.* **2005**, *46*, 1047 – 1052.
[174] S. Meaney, M. Heverin, U. Panzenboeck, L. Ekström, M. Axelsson, U. Andersson, U. Diczfalusy, I. Pikuleva, J. Wahren, W. Sattler, I. Björkhem, *J. Lipid Res.* **2007**, *48*, 944 – 951.
[175] W. J. Griffiths, Y. Wang, *J. Chromatogr. B* **2009**, *877*, 2778 – 2805.
[176] P. Schönknecht, D. Lütjohann, J. Pantel, H. Bardenheuer, T. Hartmann, K. von Bergmann, K. Beyreuther, J. Schröder, *Neurosci. Lett.* **2002**, *324*, 83 – 85.
[177] J. K. Nicholson, J. Connelly, J. C. Lindon, E. Holmes, *Nat. Rev. Drug Discovery* **2002**, *1*, 153 – 161.
[178] D. G. Robertson, *Toxicol. Sci.* **2005**, *85*, 809 – 822.
[179] D. Portilla, S. Li, K. K. Nagothu, J. Megyesi, B. Kaissling, L. Schnackenberg, R. L. Safirstein, R. D. Beger, *Kidney Int.* **2006**, *69*, 2194 – 2204.
[180] L. K. Schnackenberg, R. C. Jones, S. Thyparambil, J. T. Taylor, T. Han, W. Tong, D. K. Hansen, J. C. Fuscoe, R. D. Edmondson, R. D. Beger, Y. P. Dragan, *OMICS* **2006**, *10*, 1 – 14.
[181] P. Espandiari, J. Zhang, B. A. Rosenzweig, V. S. Vaidya, J. Sun, L. Schnackenberg, E. H. Herman, A. Knapton, J. V. Bonventre, R. D. Beger, K. L. Thompson, J. Hanig, *Toxicol. Sci.* **2007**, *99*, 637 – 648.
[182] P. A. Thürmann, *Pathologie* **2006**, *27*, 6 – 12.
[183] *S.* Teweleit, U. Kuschel, M. Hippius, M. Goettler, M. Bornschein, *Med. Klin.* **2001**, 96, 442 – 450.
[184] R. Bordet, S. Gautier, H. Le Louet, B. Dupuis, J. Caron, *Eur. J. Clin. Pharmacol.* **2001**, *56*, 935 – 941.
[185] FDA Critical Path Opportunities Report. http://www.fda.gov/oc/initiatives/criticalpath.
[186] L. K. Schnackenberg, R. D. Beger, http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ArticlesandPresentations/ucm077544.htm.
[187] L. K. Schnackenberg, *Expert Rev. Mol. Diagn.* **2007**, *7*, 247 – 259.
[188] International Conference on Harmonization (ICH)—Guidance for Industry: E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories. http://www.fda.gov/RegulatoryInformation/Guidances/ucm129286.htm).
[189] International Conference on Harmonization (ICH)—Guidance for Industry: E16: Genomic Biomarkers Related to Drug Response: Context, Structure, and Format of Qualification Submissions. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174433.pdf.
[190] O. Fiehn, *Comp. Funct. Genomics* **2001**, *2*, 155 – 168.
[191] J. B. German, D. E. Bauman, D. G. Burrin, M. L. Failla, H. C. Freake, J, C. King, S. Klein, J. A. Milner, G. H. Pelto, K. M. Rasmussen, S. H. Zeisel, *J. Nutr.* **2004**, *134*, 2729 – 2732.
[192] R. Chari, B. P. Coe, C. Wedseltoft, M. Benetti, I. M. Wilson, E. A. Vucic, C. MacAulay, R. T. Ng, W. L. Lam, *BMC Bioinf.* **2008**, *9*, 422.
[193] H. P. Deigner, D. Enot, M. Bauer, unpublished results.
[194] Some confusion can arise concerning the nomenclature used to describe 27-hydroxycholesterol. According to the rules of priority and numbering, the correct description is (25*R*)26-hydroxycholesterol; however, as the medical community uses the name 27-hydroxycholesterol, we will use this description herein.