



Application of Computational Proteomics and Lipidomics in Drug Discovery

Nitish Kumar Mishra^{1,2*} and Mamta Shukla³

¹Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE, USA

²University of Maryland Institute for Advanced Computer Studies (UMIACS), Center for Bioinformatics & Computational Biology (CBCB), University of Maryland, College Park, MD, USA

³Immunobiology Division, CSIR-Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India

Abstract

The process of drug discovery requires integration of biochemical and genetic tests to analyze the effects of drug molecules on biological systems. Comparative proteomic/lipidomic methods have identified a large number of differentially expressed novel proteins and lipids that can be used as prominent biomarkers for disease classification and drug resistance. Lipidomics or proteomics are not only used for target identification and deconvolution but also for analysis of off-targets and for studying the mode of action of drug molecules. In addition, they play significant roles in toxicity and preclinical trials at very early stages of drug development as well as in analysis of adverse effects of existing drug molecules. Since large-scale 'omics' data are now available in the public domain, bioinformatics and statistical analysis tools are needed to decipher knowledge from this vast amount of data. This review gives a brief overview of advancements in technological and computational methods in the area of lipidomics and proteomics based drug design.

Keywords: Proteomics; Lipidomics; Mass-spectrometry; Target deconvolution; Off-target; Toxicity

Introduction

Drug discovery is a time consuming and cost intensive process. It takes around 12-15 years and costs up to €800 million in order to bring a new drug in market [1,2]. Most drugs exert their therapeutic effect by binding to and regulating the activity of a particular therapeutic target. Identification and validation of such targets is a first and important step in drug discovery. Currently therapeutic targets can be identified using both structural and sequential approaches. Selectivity and specificity are not only major challenges in drug design but also important factors for withdrawing drug molecules from the market [2,3]. There are numerous proteins in the human body and it is practically impossible to ascertain whether these drug molecules bind with high affinity only to intended target proteins or also interact with other off-targets.

The drug discovery process requires several biochemical and genetic assays in order to delineate the effects of drug candidates on cellular systems and model organisms [4]. State-of-the-art proteomics/lipidomics techniques measure the changes in proteins/lipids and their isoforms quantitatively upon drug exposure, and are important tools at various stages in small drug molecule discovery. Advancement in high-throughput technology and better understanding of biology is helpful in this aspect. The 'omics' technology utilizes high-throughput techniques for generating vast amount of data allowing new directions in drug discovery [5]. In general the '-omics' suffix has been used to denote the study of the entire set of entities in a class. 'Omics' data provide comprehensive descriptions of nearly all components and interactions within a system that are required to enable a system level understanding [6]. Genomics, proteomics, toxicogenomics, lipidomics, pharmacogenomics, metabolomics and other areas of 'omics' have become handy tools in modern drug discovery. These 'omics' technologies are very popular in disease biomarker identification [4,7], drug target identification [8-11], and profiling of drug molecules [12-14].

Proteomics based methods are now popular in drug target identification as well as in off-target analysis. Recent technological

advancement in mass spectrometry (MS) and rapid improvements in chromatographic techniques have led to the rapid expansion of the proteomics and lipidomics. Recent development in computational database search algorithm [15], pathway mapping [16] give new dimension in area of biomarker and target identification. Different software for MS data processing and analysis are already available but they give a lot of false positive and false negative hit, so integration of new component to overcome this problem is still needed [6-18]. This review encompasses an overview of applications of lipidomics and proteomics in drug designing.

Proteomics in Drug Designing

Proteins are the principal targets of small chemical drug molecules. Common applications of proteomics in the drug discovery include target identification and validation, identification of toxicity biomarkers, efficacy estimation and understanding the mode of action of the drug molecules and their toxicity. MS based proteomics technologies are ideally suited for the discovery of biomarkers in the absence of prior knowledge of quantitative and qualitative changes in proteins. Following are the major areas in drug discovery where proteomics have become popular.

Deconvolution of Drug Targets

Drug target deconvolution is a process involving identification of complete spectrums of proteins that are associated with the bioactive chemical drug molecules [4,19]. Information about spectrum of target

***Corresponding author:** Nitish Kumar Mishra, Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE, USA, E-mail: nitishimtech@gmail.com

Received August 30, 2013; **Accepted** January 06, 2014; **Published** January 13, 2014

Citation: Mishra NK, Shukla M (2014) Application of Computational Proteomics and Lipidomics in Drug Discovery. J Theor Comput Sci 1: 105. doi: 10.4172/2376-130X.1000105

Copyright: © 2014 Mishra NK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

proteins against bioactive drug molecule helps in drug toxicity research by identification of off-targets, leading to drug prioritization. It also aids in identification of additional unexplored targets of existing drug molecules. It is worthwhile to add detailed target deconvolution in each drug discovery process. The deconvolution of therapeutic drug targets should be done in consecutive steps whereby experiments are repeated at least twice and only resultant consensus proteins are considered to be valid. Proteins that observed repeatedly in every independent experiment using unrelated drugs and/or with matrices without immobilized drugs are removed from list. Finally, the most frequent proteins those are present in several cell lines, also known as 'core-proteome' proteins are also removed from the final list [20].

The direct way to identify the molecular target of a drug candidate involves immobilizing the drug molecules on solid matrix, e.g. agarose, sepharose or streptavidin magnetic beads [21-24]. Use of complex protein mixtures such as cell or tissue or organ lysates, with matrix-bound drug molecule, captures the target proteins. Matrix and linker molecule are selected on the basis of little or no unspecific binding of proteins. This also includes a control that involves beads with linker without drug molecule. These controls are included for every experiment to identify unspecific binders. In past chemoproteomic approach based target deconvolution based on classical drug affinity chromatography has been successfully used in identification of molecular target of immunosuppressants [25,26] and inhibitors of histone deacetylation [27]. Protein kinases are major therapeutic targets and their involvement in cancer and inflammation has been well explored. Although several successful cancer drugs are associated with the well-defined protein kinase target profiles, such as Imanitib or Dasatinib, several off-targets have been identified for these drugs [28,29]. The use of single immobilized kinase inhibitors allows the capturing of specific target proteins [30,31] as well as off-targets.

Flesischer et al. [32] used affinity-based proteomics to identify nicotinamide phosphoribosyltransferase as a target of the potent and selective cytotoxic agent CB30865. Huang et al. [33] used chemoproteomic approach to identify tankyrases as the target of the small molecule XAV939. Filippakopoulos et al. [34] demonstrated that the small molecule, JQ1, displaces BET proteins from the chromatin; hence this compound is efficient in patient-derived xenograft having squamous carcinoma. Dawson et al. [35] used a multitier proteomic strategy to characterize BET-dependent histone binding of various protein complexes including the super elongation complex (SEC) and polymerase-associated factor complex. These success stories suggest that chemoproteomic approach enables the identification of direct target of drug molecules and provides insights into regulatory mechanism depending on protein-protein interactions.

Selectivity Profiling

Binding mode centric profiling based on the binding/activity of a small drug molecule against proteins of particular protein target class may help in selectivity and specificity analysis of drug molecules. The affinity of a given compound to all members of a target class is determined by quantifying the amount of proteins captured by the affinity matrix. Precisely, inhibition of binding curves is obtained and used for the calculation of apparent K_d value [36-38]. This is a robust and reliable approach as proteins are assayed under physiological conditions. In addition, the multiplexing capability of MS for protein identification can provide ranked affinities of a compound against all members of the target class in a single experiment.

In case of protein kinases, the conserved ATP-binding site has been used to generate nonselective ATP-competitive affinity matrices (e.g. 'kinobeads') that allows the determination of IC₅₀ values and the selectivity and specificity of drug molecules for up to 150 kinase target proteins in a single experimental run [28,29,39]. Such 'kinobeads' have been successfully applied in selectivity profiling of clinical BCR-ABL inhibitors in the chronic myeloid leukemia cell line K562 [28], EGFR inhibitors in HeLa cells [37], and 13 other multi-kinase inhibitors in chronic lymphoid leukemia cells under clinical investigation [29]. Wu et al. [40] used immobilized kinase inhibitors to identify targets in head and neck cancer by analyzing the kinase complement across 34 squamous cell carcinoma cell lines established from patients.

Mode of Action of Drug and Target Validation

Chemoproteomic based target deconvolution of lead molecules does not necessarily identify well annotated and characterized proteins. Hence an initial challenge is to link these proteins to disease biology and to elucidate the mode of action of drug molecules for generating the observed response phenotype.

Building of protein-protein interaction networks by affinity proteomic approaches can help in characterization of functional roles of proteins under experimental conditions. In an ideal condition, placing protein into an interaction network identifies a protein directly as a player in the disease process under investigation. In addition, protein-protein interaction studies can be used to shed light on mechanism other than direct inhibition or activation by which a drug can modulate target activity. Differential protein complex formation with and without compound treatment, either in cell lysate or during the purification procedure, allows the identification of compound sensitive protein-protein interaction [41]. The generation of large-scale protein interaction maps also enables the identification of more favorable drug target candidates.

Global Proteomic Profiling of Post-translational Modifications in Drug Resistance

Proteomic approaches are becoming an important tool to characterize the mode of action of enzymes that modulate drug compounds. It sheds light on post-translational modification of substrate proteins such as phosphorylation, acetylation and ubiquitination. Differential phosphorylation proteomic analysis, using selective small molecule inhibitors of particular kinases, has been used to identify substrates in human cells and characterize the effects of kinase inhibition on signaling. Chemogenetic kinase trapping approach allows for direct and unequivocal identification of kinase substrates. We use genetically engineered ATP-binding pocket that can bind an unnatural bulky ATP analog. This analog cannot bind with wild type kinase and hence cannot transfer its phosphate group to substrate proteins. The use of thio-ATP followed by a covalent capture step and identification of modified peptides by MS has been successfully applied for the characterization of human CDK1 and CDK proteins [42].

Another current focus of drug discovery effort is identification of epigenetic targets that modulate the posttranslational modification state of histones. Quantitative proteomics have been successfully used to study the effect of small drug molecules by monitoring protein acetylation and methylation. Application of proteomic approaches are not restricted to the identification of the mode of action but could also be applied to the identification of cellular mechanism of drug resistance [43].

Proteomics for Biomarker Discovery

Proteomic based biomarker discovery has gained substantial attention in recent years. The identification of prominent biomarkers of disease, drug efficacy and drug toxicity is important in drug discovery and disease diagnosis. The overall goal of biomarker profiling is to identify the list of proteins, which differentially expressed in disease as compared to normal cells. For example, Korolainen et al. [44] identified 26 proteins which show statistically significant changes in Alzheimer's disease.

The identification of a mechanistic biomarker of drug efficacy can be achieved by monitoring the levels of PTMs as phosphorylation of kinase proteins, protein acetylation and deacetylation or protein fragment by protease activity, quantitative and qualitative proteomic analysis by using global proteome profiling. The power of MS-based proteomics defines its ability to discover these modifications at a large scale and monitors their responses to drug treatment. It also estimates quantitative change in the level of proteins by other system perturbations [45]. For example, the output of an enzymatic activity as pharmacodynamics biomarker is used for monitoring global protein levels as a parameter for the effect of applied treatment.

Proteomic studies on drug selectivity and mode-of-action could provide appropriate molecular toxicity biomarkers. Liver toxicity is particularly one of most common problems. Global proteomic profiling of human hepatocytes or rodent livers treated with a drug could be used to identify proteins which undergo abundant changes in response to drugs that may be useful as surrogate pharmacodynamic biomarkers [46]. It is important to translate such findings from cell line to relevant animal models of disease and, eventually, in human context.

Proteomics are being applied to identify the biomarkers in cancer and drug resistance, thereby leading to personalized therapeutic strategies of cancer patients. Besada et al. [47] used comparative proteomic analysis of the breast tumor xenografts, which are sensitive and resistant to tamoxifen. It was observed that twelve proteins are up regulated and nine were down regulated. Umar et al. [48] performed comparative proteomic analyses on LCM-purified human breast tumor cells, which are both sensitive and resistant to tamoxifen. They found a set of biomarkers such as extracellular matrix metalloproteinase inducer; ENPP1, EIF3E and GNB4 are associated with tamoxifen resistance. Recent research suggests that several proteins such as annexin IV and claudin-4 are involved in modulating response of cisplatin in ovarian cancer are potential biomarker of treatment response.

Any biomarker discovery can produce lengthy list of candidate proteins that are detected differentially in case vs controls which requires further verification in large number of samples. Verification of these candidate proteins requires targeted, multiplexed assays to screen and quantify proteins in patient plasma sample with high sensitivity and specificity. Because there is no any quantitative assay for the majority of human proteins, assays (like enzyme-linked immunosorbent assays (ELISA) must be developed for de novo for clinical testing of candidate protein biomarkers, and de novo assay development is very expensive for testing large number of candidate biomarker. Recent advances in proteomics have become an integral part of biomarker discovery, quantification and validation of candidates in bodily fluids [49,50]. Selected reaction/multiple reaction monitoring (SRM/MRM) mass spectrometry holds the promise to overcome this bottleneck. SRM/MRM MS technology has high reproducibility across complex samples. Keshishian et al. [51] quantify six biomarkers in serum which was previously reported by ELISA. Later on Whiteaker et al. [52]

reported fabulin-2 as a marker for breast cancer later Nicol et al. [53] reported carcinoembryonic antigen as a marker of lung cancer by using MRS-MS. Recently Muraoka et al. [54] identified and quantified 5122 proteins with high confidence in 18 breast cancer patient tissue sample by using shotgun proteomics coupled with the isobaric tag for relative quantification (iTRAQ) and SRM/MRM. A total of 61 proteins were found to be altered by 2-fold or more between high and low-risk breast cancer tissues and 49 of these proteins were subsequently verified with targeted proteomics using SRM/MRM. Twenty-three proteins were shown to be differentially expressed between high and low-risk group. Narumi et al. [55] performed large-scale differential phosphoproteome analysis coupled iTRAQ technique and subsequent validation by SRM/MRM of human breast cancer tissues in high and low-risk recurrence groups. They successfully quantified 15 probable cancer biomarker phosphopeptides by SRM using stable isotope peptides.

Lipidomics in Drug Designing

Lipid molecules within human body are enormously complex and they are the fundamental component of biological membranes. They also play multiple important roles in biological systems such as, formation of cellular membranes, storage of energy and cell signaling, these could be expected to reflect much in health and disease. Lipidomics is a metabolomics approach targeted on lipids that aims for comprehensive analysis of lipids in biological systems. Lipidomics research involves the identification and quantification of the thousands of cellular lipid molecular species and their interactions with other lipids, proteins, sugars and other metabolites. Recently, lipidomics caught attention due to the well-recognized roles of lipids in numerous human diseases, such as diabetes, obesity, atherosclerosis, Alzheimer's disease etc. Application of lipidomics would not only provide insights into the specific roles of lipid molecular species in health and disease, but would also assist in identifying the potential biomarkers for establishing preventive or therapeutic approaches for human health. The major objective of lipidomics is to link the lipid metabolites and/or lipid metabolic pathways in complex biological systems and to interpret the changes in the lipid metabolism or in the regulation of these pathways in metabolic and inflammatory diseases from a physiological and/or pathological perspective. Lipidomics is usually focused on the measurement of alterations of lipids at system-level indicative of disease or due to environmental perturbations or in response to diet, drugs and toxins as well as genetics [56].

Recent advancements in MS and innovations in chromatographic technologies have largely driven the advancement in lipidomics. The major biological significance of lipidomics is the achievement of the traditional lipid research in two major point: (i) how to link metabolites and/or lipid metabolic pathways in complex biological systems to individuals metabolic health; (ii) how to interpret the changes in the lipid metabolism or in the regulation of these pathways linked to metabolic and inflammatory diseases from the pathophysiological perspectives. For this reason, lipidomic investigation usually focus on the measurement of alterations of lipid at systems level indicative disease, environmental perturbations or response to diet, drug and toxins as well as genetics [56]. Often the lipid profiles in clinical investigations related to person that are in disease state or have specific genetic profiles become the basis for detection of the potential biomarkers related to disease or specific gene expression compared to control [57,58].

Usually, lipidomic analyses of given sample are performed by shotgun and/or targeted approaches depending on the question

raised by researcher. Shotgun technology is an analysis of multiple lipid classes in one run where lipid extracts are infused directly into a mass spectrometer. The advantage of shotgun approach is that it enables the identification and quantification of hundreds of lipids in less than 30 min/sample, making it suitable for initial screening. Most important the shotgun approach has been demonstrated to be highly reproducible, matching suitable for good laboratory practice (GLP) requirements [59]. In targeted lipidomics, lipid extracts are primarily separated by liquid chromatography before monitoring by online MS [60]. A lipidomic approach is applicable to all therapeutic area, including cardiovascular disease, autoimmune, diabetes, neurological disease, cancer, as well as inflammatory diseases [61]. Following are the major areas in drug discovery where proteomics have become popular.

Lipidomics for Biomarker Discovery

Lipid metabolic disorder or abnormalities is involved in several human diseases such as inborn disease/syndrome, coronary heart disease, brain injuries, cancer including all other discussed in last paragraph. For example, obesity is very common and most vital risk factors of heart diseases and diabetes. High level of low-density lipoproteins (LDL) and triacylglycerol and decreased level of high-density lipoprotein (HDL) are common indicators of abdominal obesity. Therefore, monitoring of alteration in lipid metabolites in biological samples would be helpful for the identification of lipid metabolites indicative of metabolic disorders or disease. Quehenberger et al. [62] described MS-based lipidomic tools, which were developed by the LIPID MAPS Consortium [63] and used for the systematic identification and quantification of human lipidome. They presented plasma concentration of more than 500 different lipid species from six main lipid categories [64,65]. Jung et al. [66] developed high-throughput, anticipate that this toolkit will contribute to basic research, nutritional research and promote the discovery of new disease biomarkers, disease related mechanisms of actions and drug targets. Min et al. [67] used qualitative and quantitative profiling of six different categories of urinary phospholipids from patients with prostate cancer to develop an analytical method for discovery of candidate biomarker by using shotgun lipidomics. They used nanoflow chromatography-electrospray ionization-tandem mass spectrometry and identified that one phosphatidylcholine, one phosphatidylethanolamines, six phosphatidylserines and one phosphatidylinositol show significant differences between control and cancer patients.

Recently Zhou et al. [68] identified plasma lipid biomarkers for prostate cancer by using lipidomics and bioinformatics. They used identified 15 lipid candidate marker which can classify disease and normal sample with accuracy 97.3%, which demonstrate the power of lipidomics in disease biomarker field. Drug toxicity marker analysis is another high potential area in high-throughput lipidomics. Ximelagatran, an oral thrombin inhibitor was withdrawn from market owing to increased risk of severe liver damage with an unknown cause after Sergent et al. [69] lipidomic analysis. Based on their results, the investigators concluded that the lipid changes led to the loss of membrane integrity and leakage of cellular proteins. Their research identified distinct molar phospholipid ratios as novel biomarkers for hepatotoxicity of ximelagatran drug. Recently Jänis et al. [70] reported lipid biomarkers of drug efficacy. Several proprotein convertase subtilisin/kexin Type 9 (PCSK9) inhibitors are currently being developed by pharmaceutical companies because these compound have been identified to be a potent lowering drug. Lipidomic analysis human carrying a well-characterized PCSK9 loss-of-function mutation observed that PCSK9 inhibition lowered plasma concentration of

certain cholesteryl esters and short chain sphingolipid species much more efficiently than did LDL cholesterol. The authors suggested that these specific lipid species could be utilized for the characterization of novel PCSK9 inhibitors and as sensitive efficacy markers of PCSK9 inhibition.

Lipidomics in Target Discovery

Lipid play vital role in several biological function, differential change in concentration of different lipids can be used as probes of functionality of various metabolic pathways in disease. This area is still unexplored but we believe that integration of gene expression, flux lipidomics and other omics data can play vital role target identification in future.

Bioinformatics in Proteomics and Lipidomics

Large amount of proteomics and lipidomics data are now available in public domain. Omics bioinformatics is, thus, emerging as well as challenging for proteomics and lipidomics. Protein/lipid concentration changes in living biological systems reflects regulation at multiple spatial and dynamic scales, e.g., cellular biochemical reactions, intracellular trafficking of proteins and lipids, cell membrane composition change, protein biosynthesis and degradation and lipid metabolism and lipid oxidation. In order to address protein/lipid regulation, following are the steps required in bioinformatics: (a) data processing and identification, (b) statistical analysis of the data, and (c) pathway analysis.

Preprocessing of data

Specific workflow of proteomics data processing depends on the specific biological problems. Data pre-processing and identification are methods that ameliorate turning raw omics data from experiments into a final proteomics/lipidomics dataset that can be interpreted and analyzed. This may include tools for automatic data processing, identification and mining. Current proteomics dominated by MS based approaches, use direct infusion techniques of liquid chromatography coupled by MS (LC/MS). For background correction and data processing, several free and commercial softwares are available but R [71] based e.g. msProcess, PROcess and MATLAB (<http://www.mathworks.com/>) based e.g. Backcor are most popular. There are several freely available software for processing of mass spectrometry data.

OpenMS

OpenMS is an open source framework for LC-MS based proteomics [72]. OpenMS offers data structures and algorithms for the processing of mass spectrometry data. The library is written in C++ and it will work on all major platforms as Windows XP/7/8, Linux, MacOS. OpenMS is freely downloadable at <http://open-ms.sourceforge.net/>.

MZmine

MZmine 2 is improved version of popular MZmine [73] is framework for differential analysis of mass spectrometry data, is an open-source software for mass-spectrometry data processing, with the main focus on LC-MS data [74]. MZmine 2 is freely available at <http://mzmine.sourceforge.net/>. MZmine 2 can read and process both unit mass resolution and exact mass resolution data in both continuous and centroided modes, including fragmentation scans. Web can visualize raw data together with peak picking and identification results, which is very useful for evaluating different peak detection methods.

Peak detection in MZmine 2 is performed in a three-step manner;

first mass values are detected within each spectrum. In the second step, a chromatogram is constructed for each of the mass values which span over certain time range. Finally, deconvolution algorithms are applied to each chromatogram to recognize the actual chromatographic peaks. MZmine 2 can report the quantification results in table form in comma separated value (CSV) or using charts, we can download CVS result file. There are several modules for further processing of peak detection results, including deisotoping, filtering and alignment. Peak identification can be performed by searching a custom database or by connecting to PubChem Compound database [75]. MZmine 2 also contains basic methods for statistical analysis of processed data.

OpenChrom

OpenChrom is an open source software for chromatography and mass spectrometry based on the Eclipse Rich Client Platform (RCP). Mass spectrometry data generated, for example, by GC/MS, LC/MS, HPLC-MS, ICP-MS or MALDI-MS may be imported directly, without prior conversion, for subsequent visualization and evaluation. The focus is to handle data files from different GC/MS systems and vendors. OpenChrom support of various vendor data formats, data may also be imported in common formats such as NetCDF, csv or mzXML. All data format converters are provided as separate plug-ins. OpenChrom have adaptable graphical user interface and is available for various operating systems, e.g. Windows, Linux, Solaris and Mac OS X which is freely available at <https://www.openchrom.net/>

ProteoWizard

The ProteoWizard Library and Tools are a set of modular and extensible open-source, cross-platform tools and software libraries that facilitate proteomics data analysis. The libraries enable rapid tool creation by providing a robust, pluggable development framework that simplifies and unifies data file access, and performs standard chemistry and LCMS dataset computations. It can read major vendor raw data format and other as mzML, mzXML, MGF etc. and convert in different file format. ProteoWizard is freely available at <http://proteowizard.sourceforge.net/>

XCMS

XCMS is a popular R [71] based Bioconductor [76] package developed for processing and visualization of LC-MS and GC-MS data [77]. The Xcms package reads full-scan LC/MS data from AIA/ANDI format NetCDF, mzXML, and mzData. All data to be analyzed by must be converted to one of those file formats. All NetCDF/mzXML/mzData format exported file put in same place throughout the analysis. During peak identification, Xcms uses a separate line for each sample to report the status of processing. It outputs have two numbers separated by a colon. The first number is the m/z it is currently processing, and second number is the number of peaks that have been identified so far. XCMS have several advanced tools for processing, peak detection, filling the missing data, retention time correction, analysis and visualization of results, selecting and visualizing peaks.

PrepMS

PrepMS is a simple-to-use graphical application for MS data preprocessing, peak detection, and visual data quality assessment. PrepMS is a compiled stand-alone application, which are written in MATLAB. PrepMS is freely available at <http://sourceforge.net/projects/prepms/>

Trans-Proteomic Pipeline (TPP)

TPP is a mature suite of tools for mass-spec (MS, MS/MS) based proteomics: statistical validation, quantitation, visualization, and converters from raw MS data to our open mzXML format. <http://sourceforge.net/projects/sashimi/>

Isobar

Isobar is a tool for analysis and quantitation of isobarically tagged MS/MS proteomics data. Isobar provides methods for preprocessing, normalization, and report generation for the analysis of quantitative mass spectrometry proteomics data labeled with isobaric tags, such as iTRAQ and TMT. Isobar is Bioconductor [76] package freely available at <http://bioconductor.org/packages/release/bioc/html/isobar.html>

Target search

This package provides a targeted pre-processing method for GC-MS. TargetSearch can currently read only NetCDF files. Target scan have some advanced features as baseline correction, peak identification, retention index correction, normalization, library search, metabolite profiling, peak and spectra visualization. TargetSearch software is freely available at <http://www.bioconductor.org/packages/release/bioc/html/TargetSearch.html>

MassSpecWavelet

MassSpecWavelet is R package aimed to process MS data mainly based on Wavelet Transforms [78]. The current version only supports the peak detection based on Continuous Wavelet Transform (CWT). More functions covering baseline removal, smoothing, alignment will be added in the future versions. The algorithms have been evaluated with low resolution mass spectra (SELDI and MALDI data), we believe some of the algorithms can also be applied to other kind of spectra. MassSpecWavelet is freely available at <http://bioconductor.org/packages/release/bioc/html/MassSpecWavelet.html>

MetAlign

MetAlign is tool for preprocessing of LCS-MS and GC-MS data [79]. It is capable of automatic format conversions, accurate mass calculations, baseline corrections, peak-picking, saturation and mass-peak artifact filtering, as well as alignment of up to 1000 data sets. MetAlign software output is compatible with most multivariate statistics programs.

MSPtool

Mass Spectra Preprocessing tool (MSPtool), a user-friendly versatile tool for preprocessing MS data [80]. MSPtool provides the user with a wide set of MS preprocessing steps by means of an easy-to-use graphical interface. Also, this tool has been embedded in a time-series-based framework for MS data clustering.

Other packages

There are several other R packages for mass spectrometry as msProcess, PROcess, caMassClass, FTICRMS, RProteomics, caBIG etc.

Software for Identification

Several specialized database and software are available for lipid, peptide identification.

SEQUEST

SEQUEST is a database searching algorithm match experimental spectra with theoretical spectra which are generated from peptide

sequences in silico, and then calculate scores to evaluate how well they match [81]. Then it selects a proportion of top candidate peptides based on the rank of preliminary score for cross-correlation analysis. So, for each candidate peptide identification, several scores and rankings are determined. To distinguish correct identifications from incorrect identifications, filters using a set of database searching scores are applied

Mascot

Mascot is a probability based scoring method for MS data searching, which has a number of advantages; (i) a simple rule can be used to judge whether a result is significant or not (ii) scores can be compared with those from other types of search, such as sequence homology (iii) search parameters can be readily optimized by iteration [82].

PeptideProphet

PeptideProphet is another database search tool, made on establishing statistical analysis methods to determine the possibility of positive identifications [83]. Employing the expectation maximization algorithm, the analysis learns to distinguish correct from incorrect database search results, computing probabilities that peptide assignments to spectra are correct based upon database search scores and the number of tryptic termini of peptides

ProFound, PepFrag

ProFound (<http://prowl.rockefeller.edu/prowl-cgi/profound.exe>) is a tool for searching a protein sequence collections with peptide mass maps. A Bayesian algorithm is used to rank the protein sequences in the database according to their probability of producing the peptide map. PepFrag (<http://prowl.rockefeller.edu/prowl/pepfrag.html>) is a tool for identifying proteins from a collection of sequences that matches a single tandem mass spectrum.

InsPecT

InsPecT is a tool to identify posttranslationally modified peptides from tandem mass spectra [84]. InsPecT constructs database filters that proved to be very successful in genomics searches. InsPecT uses peptide sequence tags as efficient filters that reduce the size of the database by a few orders of magnitude while retaining the correct peptide with very high probability. In addition to filtering, InsPecT also uses novel algorithms for scoring and validating in the presence of modifications, without explicit enumeration of all variants.

LIMSA

LIMSA (Lipid Mass Spectrum Analysis) is a program for quantitative analysis of mass spectra of complex lipid samples. LIMSA can do peak finding, integration, assigning, isotope correction and quantitation with internal standards. In LIMSA we can search lipids by single search or by batch analyze and summarize results. Source code of LIMSA is freely available at <http://www.helsinki.fi/science/lipids/software.html>.

Fatty acid analysis tool (FAAT)

FAAT is an algorithm based on Fourier transform mass spectral data analysis of from lipid extracts has been developed [85]. FAAT is Microsoft Visual Basic based rapid tool it generally takes tens second to interpret multiple. FAAT can reduce data by scaling, identifying monoisotopic ions, and assigning isotope packets. Unique features of FAAT is : (1) it can distinguished overlapping saturated and unsaturated lipid species, (2) known ions are assigned from a user-

defined library including species that possess methylene heterogeneity, (3) and isotopic shifts from stable isotope labeling experiments are identified and assigned. FAAT can determine abundance differences between samples grown under normal and stressed conditions.

Pathways Analysis

Similar to other omics data, high dimensional proteomics and lipidomics/proteomics data needs accurate statistical analysis. Several statistical methods such as, principle components, correlation, and multivariate analysis are used commonly for getting co-regulated lipid and proteins. Various R based free packages are available for these statistical analyses. Cluster analysis provides a statistical framework to get proteins/lipids that separate different sample groups from each other, and/or co-vary in a specific study. The major goal of clustering method is to group sample, variables, or both into a homogenous group. Several freely available R based software are available for both supervised and unsupervised clustering such as MASS, PLAS-DA, AMORE, hclust, PLS, PLSR etc.

Statistical methods alone just provide information about key metabolites affected within a specific group of samples. Pathway analysis takes this information further to identify affected metabolic pathway. Such analysis proceeds by combining different omics data, as proteomics, lipidomics, transcriptomics etc. KEGG [86], LIPID-MAPS [87], human metabolite database (HMDB) [88], human proteome research database (HPRDB) [89], plasma proteome database (PPD) [90], PubChem [75], and DrugBank [91,92], ChemProt [93] provides information of global metabolic schemes, metabolites, enzymes and their respective links to drug. Time is ripe to integrate these individual components in biological network for advanced drug designing. These days, several visualization tools and plugins are available for Cytoscape, which can be used for biological network construction. Knowledge based and genome-scale pathway reconstruction methods are thus needed, which can deal with large-scale metabolites data and biochemical reactions.

There are so many tools for MS data processing and analysis, so it's very difficult to conclude which one is best. Each software have some advantage and disadvantage, it would be better to use different software for different step of analysis rather than single one. In general it observed that MZmine 2, XCMS perform better than any other for data processing and Wavelet transform based MassSpecWavelet for peak identification. For identification, LIMSA for lipid identification and SEQUEST and PeptideProphet for peptide identification. In case of downstream analysis R based free software such as Limma, hclust, PLS, PLSR are best.

Conclusions

Target identification and validation involves identifying proteins, whose expression levels or activities change in disease states. These proteins may serve as potential therapeutic targets or may be used to classify patients for clinical trials. Proteomics technologies may also help in identifying protein-protein interactions that influence either the disease state or the proposed therapy. Efficient biomarkers are used to assess whether target modulation has occurred or not. They are used for the characterization of disease models and to assess the effects and mechanism of action of lead candidates in animal models. Toxicity (safety) biomarkers are used to screen compounds in pre-clinical studies for target organ toxicities and followed by their employment during clinical trials.

The use of proteomic approaches contributes significantly to our

understanding of the potential biomarker, drug target identification and deconvolution, mode of action of drug molecules and mechanism of drug resistance. Chemotherapeutic drug resistance in one of major problems and advancement in proteomic approaches can play major role in cancer drug resistance in near future. With the use of sensitivity of analytical method, future research needs to focus on the use of these qualitative and quantitative proteomic/lipidomics data of cell lines on animal models as well as on humans. Similarly, now genotypic and phenotypic data of different human ethnic populations are publically available in HapMap database [94,95]. Using the information in the HapMap and other genotype and phenotype data, researchers will be able to find genes that affect health, disease [96], and individual responses to medications [97,98] and environmental factors. It is high time for the integration of these genotypic and phenotypic data with global proteomics & lipidomics for the development of better understanding of disease cause, mode of action of drug molecules, adverse and toxicity effect of drugs in the area of advanced drug designing. Fortunately we have free computational tools which can help in integration of such data e.g. MixOmics, canonical correlation analysis (CCA).

Diseases often occur in only few cells. Therefore, direct whole proteomic analysis by MS can be difficult because the biomarker signal is diluted by the presence of other components of cell. There is an urgent need for development and implication of existing statistical methods for background noise correction to extract maximum information. At present, the existing quantitative proteomics and lipidomics methods are not up to mark. An improvement in the existing methods and development of new robust methods is the need of time. Recently, single cell proteomics gives a new insight about various differentially dynamic proteins in individual cells. Cellular response to drugs is a highly dynamic process and the overall effect of drug molecules is an ensemble of proteome dynamics in individual cells, both spatially and temporally. Single cell proteomics provides a way for understanding of how seemingly identical cells show different responses to signals and drugs. It can be an immense aid in designing better and improved drug molecules.

There are so many available computational tools for lipidomics/proteomics data analysis, improvement in these software is still needed in order to reduce the number of false positive and false negative. In recent past people try to solve these problems, but still lot more things have to do for sensitivity, specificity improvement [15]. One major problem in lipidomics/proteomics area is that each machine will provide different mass spectra for same sample, developing new robust computational algorithm which can overcome this problem and make these data comparable is still needed. Machine learning techniques have great potential to recognize pattern in complex dataset, it's high time to utilize these techniques in lipidomics/proteomics based drug designing. Change in expression of different metabolites of various metabolic pathways in disease can be used to identify druggable target enzymes to control the pathway of interest [61]. It is often useful to integrate lipidomics and gene expressions might be useful for better understanding of multiple changes in complex pathways [16,99]. Metabolic tracing experiments (FLUX lipidomics) enables the quantitative measurement of molecular metabolism, including synthesis and degradation in real time can reveal the kinetics of individual molecules. In future we need advanced bioinformatics tools for comparative metabolomics, lipidomics [100] and pathway analysis [101]. Pathway mapping combined with gene expression analysis and flux experiments will help to reveal insights into metabolism that might be future of target discovery.

References

1. Gasteiger J, Engel T (2003) *Chemoinformatics: A Textbook*. (1stedn.) Wiley-VHC, Weinheim, Germany.
2. Mishra NK (2011) Computational modeling of P450s for toxicity prediction. *Expert Opin Drug Metab Toxicol* 7: 1211-1231.
3. Mishra NK, Raghava GP (2011) Prediction of specificity and cross-reactivity of kinase inhibitors. *Lett Drug Des Discov* 8: 223-228.
4. Schirle M, Bantscheff M, Kuster B (2012) Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* 19: 72-84.
5. Altman RB, Rubin DL, Klein TE (2004) An "omics" view of drug development. *Drug Development Research* 62: 81-85.
6. Joyce AR, Palsson BØ (2006) The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7: 198-210.
7. Li XH, Li C, Xiao ZQ (2011) Proteomics for identifying mechanisms and biomarkers of drug resistance in cancer. *J Proteomics* 74: 2642-2649.
8. Roti G, Stegmaier K (2012) Genetic and proteomic approaches to identify cancer drug targets. *Br J Cancer* 106: 254-261.
9. Wierzbka K, Muroi M, Osada H (2011) Proteomics accelerating the identification of the target molecule of bioactive small molecules. *Curr Opin Chem Biol* 15: 57-65.
10. Chan JN, Vuckovic D, Sleno L, Olsen JB, Pogoutse O, et al. (2012) Target identification by chromatographic co-elution: monitoring of drug-protein interactions without immobilization or chemical derivatization. *Mol Cell Proteomics* 11: M111.
11. Li Z, Wang RS, Zhang XS (2011) Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC Syst Biol* 5 Suppl 1: S11.
12. Bantscheff M, Hopf C, Kruse U, Drewes G (2007) Proteomics-based strategies in kinase drug discovery. *Ernst Schering Found Symp Proc* : 1-28.
13. Bantscheff M, Scholten A, Heck AJ (2009) Revealing promiscuous drug-target interactions by chemical proteomics. *Drug Discov Today* 14: 1021-1029.
14. Bantscheff M, Drewes G (2012) Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg Med Chem* 20: 1973-1978.
15. Jiang X, Jiang X, Han G, Ye M, Zou H (2007) Optimization of filtering criterion for SEQUEST database searching to improve proteome coverage in shotgun proteomics. *BMC Bioinformatics* 8: 323.
16. Gupta S, Maurya MR, Merrill AH Jr, Glass CK, Subramaniam S (2011) Integration of lipidomics and transcriptomics data towards a systems biology model of sphingolipid metabolism. *BMC Syst Biol* 5: 26.
17. Kuerschner L, Ejsing CS, Ekroos K, Shevchenko A, Anderson KI, et al. (2005) Polyene-lipids: a new tool to image lipids. *Nat Methods* 2: 39-45.
18. Palsson S, Hickling TP, Bradshaw-Pierce EL, Zager M, Jooss K, et al. (2013) The development of a fully-integrated immune response model (FIRM) simulator of the immune response through integration of multiple subset models. *BMC Syst Biol* 7: 95.
19. Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2010) Discovery of new cholesteryl ester transfer protein inhibitors via ligand-based pharmacophore modeling and QSAR analysis followed by synthetic exploration. *Eur J Med Chem* 45: 1598-1617.
20. Raida M (2011) Drug target deconvolution by chemical proteomics. *Curr Opin Chem Biol* 15: 570-575.
21. Rix U, Superti-Furga G (2009) Target profiling of small molecules by chemical proteomics. *Nat Chem Biol* 5: 616-624.
22. Knockaert M, Gray N, Damiens E, Chang YT, Grellier P, et al. (2000) Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors. *Chem Biol* 7: 411-422.
23. Hong J, Lee J, Min KH, Walker JR, Peters EC, et al. (2007) Identification and characterization of small-molecule inducers of epidermal keratinocyte differentiation. *ACS Chem Biol* 2: 171-175.

24. Brehmer D, Greff Z, Godl K, Blencke S, Kurtenbach A, et al. (2005) Cellular targets of gefitinib. *Cancer Res* 65: 379-382.
25. Brown EJ, Albers MW, Shin TB, Ichikawa K, Keith CT, et al. (1994) A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* 369: 756-758.
26. Harding MW, Galat A, Uehling DE, Schreiber SL (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* 341: 758-760.
27. Taunton J, Hassig CA, Schreiber SL (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272: 408-411.
28. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, et al. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 25: 1035-1044.
29. Kruse U, Pallasch CP, Bantscheff M, Eberhard D, Frenzel L, et al. (2011) Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells. *Leukemia* 25: 89-100.
30. Rix U, RemsingRix LL, Terker AS, Fernbach NV, Hantschel O, et al. (2010) A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells. *Leukemia* 24: 44-50.
31. Hantschel O, Rix U, Schmidt U, Bürckstümmer T, Kneidinger M, et al. (2007) The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib. *Proc Natl Acad Sci U S A* 104: 13283-13288.
32. Fleischer TC, Murphy BR, Flick JS, Terry-Lorenzo RT, Gao ZH, et al. (2010) Chemical proteomics identifies Namp1 as the target of CB30865, an orphan cytotoxic compound. *Chem Biol* 17: 659-664.
33. Huang SM, Mishina YM, Liu S, Cheung A, Stegmeier F, et al. (2009) Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* 461: 614-620.
34. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468: 1067-1073.
35. Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, et al. (2011) Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* 478: 529-533.
36. Bantscheff M, Hopf C, Savitski MM, Dittmann A, Grandi P, et al. (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol* 29: 255-265.
37. Sharma K, Weber C, Bairlein M, Greff Z, Kéri G, et al. (2009) Proteomics strategy for quantitative protein interaction profiling in cell extracts. *Nat Methods* 6: 741-744.
38. Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, et al. (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* 46: 350-358.
39. Schirle M, Petrella EC, Brittain SM, Schwalb D, Harrington E, et al. (2012) Kinase inhibitor profiling using chemoproteomics. *Methods Mol Biol* 795: 161-177.
40. Wu J, Katrekar A, Honigberg LA, Smith AM, Conn MT, et al. (2006) Identification of substrates of human protein-tyrosine phosphatase PTPN22. *J Biol Chem* 281: 11002-11010.
41. Smith KT, Martin-Brown SA, Florens L, Washburn MP, Workman JL (2010) Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. *Chem Biol* 17: 65-74.
42. Blethrow JD, Glavy JS, Morgan DO, Shokat KM (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A* 105: 1442-1447.
43. Gioia R, Leroy C, Drullion C, Lagarde V, Etienne G, et al. (2011) Quantitative phosphoproteomics revealed interplay between Syk and Lyn in the resistance to nilotinib in chronic myeloid leukemia cells. *Blood* 118: 2211-2221.
44. Korolainen MA, Nyman TA, Aittokallio T, Pirttilä T (2010) An update on clinical proteomics in Alzheimer's research. *J Neurochem* 112: 1386-1414.
45. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325: 834-840.
46. Ortiz PA, Bruno ME, Moore T, Nesnow S, Winnik W, et al. (2010) Proteomic analysis of propiconazole responses in mouse liver: comparison of genomic and proteomic profiles. *J Proteome Res* 9: 1268-1278.
47. Besada V, Diaz M, Becker M, Ramos Y, Castellanos-Serra L, et al. (2006) Proteomics of xenografted human breast cancer indicates novel targets related to tamoxifen resistance. *Proteomics* 6: 1038-1048.
48. Umar A, Kang H, Timmermans AM, Look MP, Meijer-van Gelder ME, et al. (2009) Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. *Mol Cell Proteomics* 8: 1278-1294.
49. Whiteaker JR, Lin C, Kennedy J, Hou L, Trute M, et al. (2011) A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* 29: 625-634.
50. Thingholm TE, Bak S, Beck-Nielsen H, Jensen ON, Gaster M (2011) Characterization of human myotubes from type 2 diabetic and nondiabetic subjects using complementary quantitative mass spectrometric methods. *Mol Cell Proteomics* 10: M110.
51. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* 6: 2212-2229.
52. Whiteaker JR, Zhang H, Zhao L, Wang P, Kelly-Spratt KS, et al. (2007) Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J Proteome Res* 6: 3962-3975.
53. Nicol GR, Han M, Kim J, Birse CE, Brand E, et al. (2008) Use of an immunoaffinity-mass spectrometry-based approach for the quantification of protein biomarkers from serum samples of lung cancer patients. *Mol Cell Proteomics* 7: 1974-1982.
54. Muraoka S, Kume H, Watanabe S, Adachi J, Kuwano M, et al. (2012) Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples. *J Proteome Res* 11: 4201-4210.
55. Narumi R, Murakami T, Kuga T, Adachi J, Shiromizu T, et al. (2012) A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *J Proteome Res* 11: 5311-5322.
56. Han X, Gross RW (2005) Shotgun lipidomics: multidimensional MS analysis of cellular lipidomes. *Expert Rev Proteomics* 2: 253-264.
57. Yetukuri L, Katajamaa M, Medina-Gomez G, Seppänen-Laakso T, Vidal-Puig A, et al. (2007) Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis. *BMC Syst Biol* 1: 12.
58. Hu C, van Dommelen J, van der Heijden R, Spijksma G, Reijmers TH, et al. (2008) RPLC-ion-trap-FTMS method for lipid profiling of plasma: method validation and application to p53 mutant mouse model. *J Proteome Res* 7: 4982-4991.
59. Heiskanen LA, Suoniemi M, Ta HX, Tarasov K, Ekroos K (2013) Long-term performance and stability of molecular shotgun lipidomic analysis of human plasma samples. *Anal Chem* 85: 8757-8763.
60. Ogiso H, Suzuki T, Taguchi R (2008) Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Anal Biochem* 375: 124-131.
61. Vihervaara T, Suoniemi M, Laaksonen R (2013) Lipidomics in drug discovery. *Drug Discov Today* .
62. Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, et al. (2010) Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res* 51: 3299-3305.
63. Schmelzer K, Fahy E, Subramaniam S, Dennis EA (2007) The lipid maps initiative in lipidomics. *Methods Enzymol* 432: 171-183.
64. Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CR, et al. (2009) Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 50 Suppl: S9-14.
65. Fahy E, Subramaniam S, Brown HA, Glass CK, Merrill AH Jr, et al. (2005) A comprehensive classification system for lipids. *J Lipid Res* 46: 839-861.
66. Jung HR, Sylvänne T, Koistinen KM, Tarasov K, Kauhanen D, et al. (2011) High throughput quantitative molecular lipidomics. *Biochim Biophys Acta* 1811: 925-934.

67. Min HK, Lim S, Chung BC, Moon MH (2011) Shotgun lipidomics for candidate biomarkers of urinary phospholipids in prostate cancer. *Anal Bioanal Chem* 399: 823-830.
68. Zhou X, Mao J, Ai J, Deng Y, Roth MR, et al. (2012) Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS One* 7: e48889.
69. Sergent O, Ekroos K, Lefeuvre-Orfila L, Rissel M, Forsberg GB, et al. (2009) Ximelagatran increases membrane fluidity and changes membrane lipid composition in primary human hepatocytes. *Toxicol In Vitro* 23: 1305-1310.
70. Jänis MT, Tarasov K, Ta HX, Suoniemi M, Ekroos K, et al. (2013) Beyond LDL-C lowering: distinct molecular sphingolipids are good indicators of proprotein convertase subtilisin/kexin type 9 (PCSK9) deficiency. *Atherosclerosis* 228: 380-385.
71. Dean CB, Nielsen JD (2007) Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal* 13: 497-512.
72. Sturm M, Bertsch A, Gröpl C, Hildebrandt A, Hussong R, et al. (2008) OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics* 9: 163.
73. Katajamaa M, Miettinen J, Oresic M (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22: 634-636.
74. Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11: 395.
75. Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: Integrated platform of small molecules and biological activities. *Annual Reports in Computational Chemistry* 4: 214-241.
76. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
77. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78: 779-787.
78. Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 22: 2059-2065.
79. Lommen A (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal Chem* 81: 3079-3086.
80. Gullo F, Ponti G, Tagarelli A, Tradigo G, Veltri P (2008) MSPtool: A Versatile Tool for Mass Spectrometry Data Preprocessing. In: 21st IEEE International Symposium on Computer-Based Medical Systems, Finland.
81. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5: 976-989.
82. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
83. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 74: 5383-5392.
84. Tanner S, Shu H, Frank A, Wang LC, Zandi E, et al. (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* 77: 4626-4639.
85. Leavell MD, Leary JA (2006) Fatty acid analysis tool (FAAT): An FT-ICR MS lipid analysis algorithm. *Anal Chem* 78: 5497-5503.
86. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, et al. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42: D199-205.
87. Fahy E, Sud M, Cotter D, Subramaniam S (2007) LIPID MAPS online tools for lipid research. *Nucleic Acids Res* 35: W606-612.
88. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al. (2013) HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* 41: D801-807.
89. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, et al. (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res* 37: D767-772.
90. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, et al. (2014) Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res* 42: D959-965.
91. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36: D901-906.
92. Knox C, Law V, Jewison T, Liu P, Ly S, et al. (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 39: D1035-1041.
93. Kim Kjærulff S, Wich L, Kringelum J, Jacobsen UP, Kouskoumvekaki I, et al. (2013) ChemProt-2.0: visual navigation in a disease chemical biology database. *Nucleic Acids Res* 41: D464-469.
94. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437: 1299-1320.
95. Montpetit A, Chagnon F (2006) [The Haplotype Map of the human genome: a revolution in the genetics of complex diseases]. *Med Sci (Paris)* 22: 1061-1067.
96. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, et al. (2013) GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet* 45: 362-370, 370e1-2.
97. Chung S, Low SK, Zembutsu H, Takahashi A, Kubo M, et al. (2013) A genome-wide association study of chemotherapy-induced alopecia in breast cancer patients. *Breast Cancer Res* 15: R81.
98. Landmark-Høyvik H, Dumeaux V, Nebdal D, Lund E, Tost J, et al. (2013) Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. *Genomics* 102: 278-287.
99. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9: 693.
100. Nguyen DD, Wu CH, Moree WJ, Lamsa A, Medema MH, et al. (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc Natl Acad Sci U S A* 110: E2611-2620.
101. Fehring G, Liu G, Briollais L, Brennan P, Amos CI, et al. (2012) Comparison of pathway analysis approaches using lung cancer GWAS data sets. *PLoS One* 7: e31816.