11-27-2018

IMass time: The future, in future!

Qingwei Ma

Eric Adua
Edith Cowan University

Mary C. Boyce
Edith Cowan University

Xingang Li
Edith Cowan University

Guang Ji

See next page for additional authors

10.1089/omi.2018.0162
This is an Author's Accepted Manuscript of: Ma, Q., Adua, E., Boyce, M. C., Li, X., Ji, G., & Wang, W. (2018). IIMass time: The future, in future!. *OMICS: a Journal of Integrative Biology*, *22*(11), 679-695. Final publication is Available from Mary Ann Liebert, Inc., publishers at: https://doi.org/10.1089/omi.2018.0162

IMass Time: the future, in future!

Qingwei Ma[2#], Eric Adua[1 #], Mary C. Boyce[3], Xingang Li[1], Guang Ji [4,5] and Wei Wang[1, 6]

[1]School of Medical and Health Sciences, Edith Cowan University, Perth, Australia

[2]Bioyong (Beijing) Technology Co., Ltd., Beijing, China

[3]School of Science, Edith Cowan University, Perth, Australia

[4]China-Canada Centre of Research for Digestive Diseases, University of Ottawa, Ontario, Canada

[5]Institute of Digestive Diseases, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[6]Taishan Medical University, Tai'an, China

Running title: *IMass Time: the future, in future!*

#Qingwei Ma, Eric Adua contributed equally to this work

Correspondence:

Professor Wei Wang, MD, PhD, FFPH, FRSB, FRSM
School of Medical and Health Sciences
Edith Cowan University
270 Joondalup Drive, Joondalup, WA 6027, Perth Australia
Tel: +61 (0) 418469913. Email: wei.wang@ecu.edu.au

Professor Guang Ji, MD, PhD, Chief Doctor
Institute of Digestive Diseases, Longhua Hospital, Shanghai University of Traditional Chinese Medicine
725 South Wanping Road, 200032, Shanghai, Shanghai, China
Tel: +86 (21) 64385700. Email: jiliver@vip.sina.com

Keywords: Mass spectrometry, chronic diseases, artificial intelligence, proteomics, glycomics, lipidomics, metabolomics, innovation

Abstract

Joseph John Thomson discovered and proved through a series of experiments the existence of electrons. His work earned him a Nobel Prize in 1906 and initiated the era of mass spectrometry (MS). In the intervening time, other researchers have also been awarded the Nobel Prize for significant advances in MS technology. The development of soft ionization techniques was central to the application of MS to large biological molecules and led to an unprecedented interest in the study of biomolecules such as proteins (proteomics), metabolites (metabolomics), carbohydrates (glycomics) and lipids (lipidomics), allowing a better understanding of the molecular underpinnings of health and disease. The interest in large molecules drove improvements in MS resolution and now the challenge is in data deconvolution, intelligent exploitation of heterogeneous data and interpretation, all of which can be ameliorated with a proposed IMass technology. We define IMass as a combination of mass spectrometry (MS) and artificial intelligence (AI) with each performing a specific role. IMass will offer advantages such as improving speed, sensitivity and analyses of large data that are presently not possible with MS alone. Here, we present an overview of the MS considering historical perspectives and applications, challenges as well as insightful highlights of IMass.

HISTORICAL PERSPECTIVES OF MS

Mass spectrometry (MS) is an analytical technique that ionizes chemical species and sorts the ions based on their mass-to-charge ratio (De Hoffmann & Stroobant, 2007; Sheynkman et al., 2016). By analyzing the ions, information including molecular mass, chemical structure and fragmentation pattern of a molecule is obtained (De Hoffmann & Stroobant, 2007; Lehmann, 2016). Although the pioneering work of MS has been attributed to Joseph John Thompson, history has shown that other scientists such as Eugen Goldstein (1886) and Johann Wilhelm Hittorf (1869) began such experiments long before him. Since then, several important events initiated by different scientists have taken place (Cooks et al., 2006; Danikiewicz, 2013;

Fridriksson et al., 1999; Hurst et al., 1998; Karas & Hillenkamp, 1988; Kauppila et al., 2006; Salih et al., 1998; Veenstra, 1999; Wilkins et al., 1999), many of which are shown in **Table 1.** Most notable in the context of this review is the 2002 Nobel Prize to Fenn and Tanuka for their work on soft ionization methods, electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI), which expanded the capability of MS to include large biological molecules. In ESI, a buffer of an analyte is dispersed by an electrospray into an aerosol. The aerosol undergoes repeated solvent evaporation, and as the electric field increases, offspring droplets are formed which split into multiply charged ions (Hoffmann & Stroobant, 2001). MALDI on the other hand, involves the application of laser pulse to analyte-matrix crystals causing them to sublime into gaseous ions. The gaseous ions then migrate to the analyser under the influence of an electric field (Hoffmann & Stroobant, 2001). Surface enhanced laser desorption ionisation (SELDI) is a variant form of MALDI where proteins of interest bind to a surface before subsequent laser ionisation and MS analyses. The interaction with the surface can be specific to a protein/peptide and hence effective for pre-fractionation of protein mixtures (Hutchens & Yip, 1993; Poon, 2007).

Mass analyzers

As MS technology continued to evolve, there was a pressing need for instruments that improved the analyses of organic molecules in a more accurate and precise manner. Consequently, magnetic sector double focusing, time of flight (TOF), quadrupole and the *Fourier* transform ion cyclotron resonance (FT-ICR) mass analyzers were developed. Ernest Lawrence introduced the cyclotron in 1932 where charged particles were accelerated under magnetic and radiofrequency fields. After two decades, Hipple and his colleagues applied the cyclotron's principle to design a mass analyzer with better trapping and detection called the Omegatron. In this technique, only ions with specific m/z were accelerated and was later called the ion cyclotron resonance (ICR) (Russell & Siuzdak, 2003). William E. Stephens developed

the TOF mass analyzer in 1946 where ions moved at different velocities towards a collector (Russell & Siuzdak, 2003). In the 1950s, Paul Wolfgang developed the quadrupole mass analyzer for which he received the Nobel Prize in 1989 (Paul, 1990). Here, an electric field triggered the formation of ions in a quadrupole. As the polarity was altered, the ions oscillated and those with specific m/z passed to the detector and were analyzed (Paul, 1990).

Melvin Comisarow and Alan Marshall in 1974 further explained that accelerated ions generate currents as they hit the detector and that this current is *Fourier* transformed into a frequency spectrum, hence the name *Fourier* transform ion cyclotron resonance (FT-ICR) (Comisarow & Marshall, 1974). The FT-ICR is undoubtedly powerful and provides good resolving power and mass accuracy, but while it can operate in MS/MS mode it is not very sensitive due to lower pressure and hence poor collision efficiency of the collision cell (Makarov et al., 2006). These MS/MS experiments are completed very effectively using triple quadrupole MS technology. In it most classical mode, selected reaction monitoring (SRM), selected precursor ions pass through the first quadrupole and into the second quadrupole or collision cell where they are fragmented. The fragmented ions are resolved and directed to the detector for analysis. In some cases, the third quadruple is replaced with a linear ion trap (LIT), which allows further MS/MS experiments on the fragments (Makarov, 2000).

The most recent innovation in MS is arguable the orbitrap, invented by Makarov. The orbitrap utilizes static electrostatic fields to cause the back and forth movement of ions around a spindle and the m/z ratios are obtained from harmonic axial oscillations. The axial motion of the ions generates currents which are detected while the accompanying signals undergoes *Fourier* transformation into a mass spectrum (Hardman & Makarov, 2003; Makarov, 2000; Makarov et al., 2006; Scigelova & Makarov, 2006). The Orbitrap has several advantages including: mass resolution of 240,000 FWHM at *m/z* 200; mass range of 50-6000 amu, large dynamic range, greater than 5000:1; and high mass accuracy of 1-3 ppm (Hu et al., 2005).

A plethora of hybrid mass spectrometers have been developed to extend the capability and functionality of instruments and include: quadrupoles in combination with TOF and orbitraps (Downard, 2007; Griffiths, 1997; Hardman & Makarov, 2003; Hevesy, 1948; Hu et al., 2005; Scigelova & Makarov, 2006; Thomson, 1906).

INSERT TABLE 2 HERE

The timeline of Nobel Prize awarded in MS technology field (Downard, 2007; Gault & McClenaghan, 2013; Griffiths, 1997; Hargittai, 2007; Hevesy, 1948; Konijnenberg, Butterer, & Sobott, 2013; Paul, 1990; Stahl, Steup, Karas, & Hillenkamp, 1991; Tanaka et al., 1988; Thomson, 1906) is shown in **Table 3.**

INSERT TABLE 3 HERE

Applications of MS

MS technology was first used in gas analysis, while measuring stable isotopes of chemical elements and then applied to the analysis of complex hydrocarbon mixtures in petroleum fractions in the early 20th century. Early use of MS verified that positive and repeatable mass spectra can be obtained in MS analysis of organic molecules, enlightening researchers to elucidate the structure of organic compounds (Hsieh et al., 2015; Lehmann, 2016). From the mid-20th century, MS technology has been used in the life sciences and clinical medicine fields, for profiling and detecting complex macromolecules such as proteins, carbohydrates, lipids and metabolites (Aretz & Meierhofer, 2016; Blanksby & Mitchell, 2010). In fact, the MS technology has become the tool of choice in providing detailed information on how these biomolecules interact with each other during normal and diseased states (Hoffmann & Stroobant, 2001). Therefore, a larger part of this review will address the application of MS in three major fields of systems biology: proteomics, glycomics and lipidomics.

1) Mass spectrometry in proteomics

The proteome, introduced in 1994 (Wilkins et al., 1996), is defined as a collection of proteins expressed by the genome, cells and tissues at a given period or under a specified condition, and proteomics is the study of the large scale of proteins including their structure, functions and their role in health and diseases (Patterson & Aebersold, 2003). Proteins are abundant in mammalian cells (~2-4 million), occupying nearly half of the total cell mass and forming complex networks that control cell signalling, define cell function and interact with other molecules to manifest cell phenotype (Aebersold & Mann, 2016; Clancy & Hovig, 2014). Protein complexity is increased by a myriad of post-translational modification(s) (PTM) including: phosphorylation, ubiquitylation, glycosylation, sialylation, nitrosylation, lipidation, acetylation and methylation (Adua et al., 2017; Wang, 2016). For example, there are nearly 19,000 sites of ubiquitylation on almost 5,000 proteins (Kim et al., 2011; Larance & Lamond, 2015). Additionally, splice variants, protein stability/instability and dynamism, and transient protein interactions collectively make the protein more complex (Larance & Lamond, 2015; Mallick & Kuster, 2010).

Many proteomics approaches, particularly those concerned with cell signalling and biomarker studies, are possible either via bottom up or top down MS analyses. In the bottom up approach, proteins are first digested with enzymes (e.g. trypsin) into peptides following which peptides are subjected to chromatographic separations. After ionisation and fragmentation, spectral analysis and structural assignments are performed by appropriate database matching (Hutchens & Yip, 1993; Karas et al., 1991;

Smith et al., 1991) **(Figure 1).** The top down approach aims to measure intact proteins (Aebersold & Mann, 2016; Liu et al., 2010; Patterson & Aebersold, 2003).

INSERT FIGURE 1 HERE

Both approaches have allowed studies on how proteins change in response to environmental influences or pathophysiological conditions (Shi et al., 2016). A typical illustration is the examination of peptides in type II diabetes mellitus (T2DM) patients and healthy controls (Meng et al., 2016). Briefly, serum samples were prepared from whole blood following which proteins were isolated and purified. After repeated washing of the supernatant in the presence of magnetic beads, the final samples were analysed on a MALDI-TOF-MS to generate MS peptide profiles. After data interpretation, amino acids of the candidate peptides were identified as being significant and potential biomarkers for T2DM (Wang et al., 2017).

Similarly, MALDI-TOF-MS was used to identify serum peptides as potential biomarkers for colorectal cancer (Wang et al., 2017-In press). MALDI-TOF-MS was also used to profile ageing related proteins in the plasma (Lu et al., 2012). The study identified 44 peptides that were differentially expressed among the age groups and observed significant associations between age and three proteins: fibrinogen alpha (FGA), albumin (ALB) and apolipoprotein A-I (ApoA1). These proteins could be important biomarkers for ageing (Lu et al., 2012).

In another study, SELDI-MS was used to profile peptides in the plasma of patients with ovarian cancer and age and gender-matched controls. With a sensitivity and specificity of 84% and 89% respectively, the authors detected protein peaks that were only expressed in the patients but not controls (Wu et al., 2006).

Of late, it is becoming increasingly apparent that a single MS technology is not sufficient to comprehensively perform all proteomic measurements, especially for proteomics studies involving large populations (Gika et al., 2014; Patterson &

Aebersold, 2003). Coupling MS with separation techniques such as liquid chromatography (LC) and gas chromatography (GC) reduces the complexity of the sample by separating the proteins/peptides by time. For example, in a 2-year observational study, LC in tandem with a high resolution Orbitrap MS was used to profile peptides produced in acellular bronchoalveolar lavage (BAL) fluids of patients suspected to be suffering from lung cancer. After extensive interrogation of the LC-MS data and the literature, the researchers identified 133 differentially expressed proteins that could be potential biomarkers for lung cancer (Carvalho et al., 2017).

For complex samples, the fragmentation pattern of the analytes is necessary to aid identification. Several MS/MS acquisition modes have been employed to capture and elucidate the complexity of the sample. One approach, data-dependant acquisition (DDA) involves preselecting the most intense ions and sending them sequentially to the collision cell, the resulting fragment ions are then analysed by the detector (Bauer et al., 2014; Porter & Bereman, 2015). A limitation of this approach is that only the more abundant peptides are selected for MS/MS fragmentation, while an advantage is that the MS/MS pattern, or fragment ions, can be directly linked to the precursor ion. Another approach is data independent acquisition (DIA) where all the precursor ions at a given time are sent to the collision cell, or where the precursor ions within selected mass ranges are sequentially sent to the collision cell (Porter & Bereman, 2015). The advantage of this approach is more complete coverage of the peptide profile as all precursor ions are fragmented (Porter & Bereman, 2015). However, the data processing, or deconvolution of the data, is more complex as the relation between parent ion and MS/MS is lost.

The DIA approach was employed to profile proteins in hepatitis B virus-associated hepatocellular carcinoma (HCC) and non-tumour cells (Gao et al., 2017). They identified and quantified a total of 4,216 proteins of which 338 were differentially expressed between the groups. In addition, 191 and 147 proteins were up-regulated and down-regulated respectively, in tumour cells. Further, the study identified important

metabolic pathways that were altered in HCC including the pentose phosphate pathway (PPP), glycolysis, gluconeogenesis, fatty acid synthesis and β-oxidation as well as other metabolic enzymes including glucose-6 phosphate dehydrogenase (G6PD) and phosphoenolpyruvate carboxykinase (PCK) (Gao et al., 2017). Similarly, the DIA approach was applied to profile proteins in the saliva of nasopharyngeal carcinoma (NPC) and healthy controls. Among the 1,414 proteins identified, 29 were differential expressed (Luo et al., 2017).

MS technology has improved biomarker discovery; however, proteomics analyses and interpretation of the raw MS data would not have been possible without dedicated processing software and online resources. Examples of such are Census, BioworksBrowser (Meng et al., 2016), Mascot, SEQUEST (Elias, Haas, Faherty, & Gygi, 2005), COMPASS, MaxQuant and Skyline (Cifani et al., 2017) amongst others, all of which have enabled peptide identification and quantification. These web resources are supported by statistical packages such as Statistical Package for Social Sciences (SPSS) and the R Software for analysis while visualisation and annotation are also possible with the Database for Annotation, Visualisation and Integrated Discovery (DAVID) (Huang et al., 2009).

2) Mass spectrometry in metabolomics

The metabolome refers to the collection of metabolites in cells, tissues and organs of an organism and a metabolite is defined as a substrate, an intermediate or a product of an enzyme catalysed biochemical reaction (Dunn et al., 2011; Mathew & Padmanaban, 2013; Tebani et al., 2016).

Metabolites are abundant with 2,000 plus existing in mammals that range from 50 to 1500 Da (Aretz & Meierhofer, 2016). Metabolites are important for the synthesis of adenosine triphosphate (ATP) and are key intermediates in cell signalling and regulation (Dunn et al., 2011). They exhibit different physicochemical properties such as solubility, half-life, molecular weight, acidity, basicity, hydrophobicity and hydrophilicity (Dunn et al., 2011). Moreover, unlike the transcriptome and the proteome

whose turn over occurs in minutes to hours, the synthesis and degradation of metabolites occur in seconds (Doerr, 2017).

Compared to proteomics, metabolomics' biological significance cannot be overemphasized. For example, unlike genes and proteins that can be influenced epigenetically and post-translationally respectively, metabolites are direct measures (substrates/products) of a biological activity and better linked with an individual's phenotype (Patti et al., 2012). Additionally, they act as indicators of genetic and environmental change. As a result, information derived from metabolome is of several orders of magnitude greater than the genome, transcriptome and the proteome (Patti et al., 2012).

Metabolomics refers to the system-wide study and analysis of the structure and function of a collection of metabolites and this can be achieved in two strategies (Li et al., 2016): targeted and non-targeted. The targeted approach involves measuring specific metabolites of interest and this is normally applied in pharmacologic research where the interest lies on therapeutic compounds. Untargeted metabolomics however, involves the measurement of the entire metabolome in a biological system (Patti et al., 2012; Savolainen et al., 2015).

Like proteomics, metabolomics analysis requires sophisticated analytical techniques that are accompanied by dedicated software for data handling and interpretation. MS has been the driver of modern metabolomics because of its ability to analyse complex metabolites with high performance (Ghaste et al., 2016). Usually, metabolomics profiling is performed using MS interfaced with a separation mode (GC, LC or capillary electrophoresis, CE) rather than MS alone (Savolainen et al., 2015; Zhang et al., 2016). Furthermore, for LC separations, samples are routinely separated using two columns, a reversed phase column to profile the non-polar metabolites and a hydrophilic interaction liquid chromatography (HILIC) column to capture the polar metabolites (Manaf et al., 2018-In press). Another approach used to optimally capture the metabolome, and which is also used in proteomics, is to use more than one MS

platform. For example, using both GC-MS and LC-MS or CE-MS to analyse the samples provides a more complete profile of the metabolome (González-Peña et al., 2016; Psychogios et al., 2011).

Recently, two platforms, GC-MS and nuclear magnetic resonance (NMR) were used to profile metabolites from the serum of patients with unstable angina (UA), ST-elevation myocardial infarction (STEMI) and healthy controls. The study found 19 unique metabolites that could be potential biomarkers for acute coronary syndrome (Ali et al., 2016). In another example, two complementary techniques, LC-QTOF-MS and GC-TOF-MS were used to examine urinary and blood metabolites in patients with obstructive sleep apnea, simple snorers and healthy controls. Here, 56 different metabolites including 4-, glycochenodeoxycholate-3-sulfate, arabinose, hydroxypentenoic acid, xanthine, isoleucine, serine, and xanthine, amongst others, were identified. Of the 56 metabolites, 21 were expressed in simple snorers and 31 in obstructive sleep apnea individuals. Interestingly, 24 of the detected metabolites were always higher or lower among the two groups when compared with controls, harnessing these metabolites can promote the diagnosis of polysomnography (PSG)-associated obstructive sleep apnea (Xu et al., 2016). Further, GC-MS was used to profile metabolites and examine abnormalities in the sera of traumatic brain injury patients with cognitive defects and without cognitive defects and healthy controls. They found nine metabolites including galactose, phenylalanine, linoleic acid, pyroglutamic acid, citric acid, palmitic acid, and 2, 3, 4-trihydroxybutyrate and arachidonic acids that were distinguishable among the three groups and this could advance the understanding on the mechanisms that underpin cognitive impairment (Yi et al., 2016). In another study, LC-MS was used to profile metabolites in the urine of patients with esophageal squamous cell carcinoma and healthy controls. Here, 19 of the 83 identified metabolites were potential biomarkers for esophageal squamous cell carcinoma diagnosis. Further, LC-MS helped to reveal purine and pyrimidine alterations in esophageal squamous cell carcinoma (Xu et al., 2016).

While providing insights on the biomarker potential of specific metabolites, this section cannot end without recognising databases or platforms that support data handling and interpretation. Examples are MetaboAnalyst (Xia et al., 2012; Xia et al., 2015), MeltDB (Costa et al., 2015; Kessler et al., 2013) and Metabolite Set Enrichment Analysis (MSEA) (Xia & Wishart, 2010). Others include HMDB,' 'Metlin,' 'XCMS' and 'MzMine (Pluskal et al., 2010).

3) Mass spectrometry in glycomics

The glycome comprises a collection of glycan structures in the cell. The glycan structure is complex, and it is made up of least ten monosaccharides that join each other stereochemically (Igl et al., 2011). They are dynamic and post-translationally bind to proteins in glycosylation (Zoldoš et al., 2013; Sebastian et al., 2016, Liu et al., 2018; Liu et al., 2018b). Glycosylation affects nearly 8,000 proteins altering their function, secretion, folding, degradation and clearance (Adua et al., 2017; Fiedler & Simons, 1995; Helenius & Aebi, 2001, 2004; PARODI, 2000). Thus far, four main types of glycans are recognized and these are *N*-glycans that bind to asparagine (Asn) residues in Asn-X-threonine [Thr]/Serine [Ser], sequon; *O*-glycans that bind to Ser and Thr residues; glycosaminoglycans (GAGs) that attach to proteins in a sequon (Gly)-X-Gly (X≠proline) and C-glycans that bind to peptides via carbon-carbon bonds in a tryptophan (Trp)-X-X-Trp sequon (X≠proline). While recognizing all these glycan types, this review will centre on *N*-glycans (Adua et al., 2017; Wang, 2016).

N-glycosylation is an ordered event that occur via the interplay of glycosidases and glycotransferases as they transverse the secretory pathway. This process is not template driven, and not directly encoded by genes. As such, *N*-glycans are products of

an extensive protein modification whose structures are many orders of magnitude more complex than the proteins (Adua et al., 2017; Wang, 2016; Yu et al., 2016). *N*-glycans regulate and control cellular function as well as an underlying disease (Ge et al., 2018; Hebert & Molinari, 2007; Liu et al., 2018a; Russell et al., 2018). Therefore, the study of glycan structure and function (glycomics) has been the focus of a cutting-edge research in the post-genomics era. *N*-glycans are stable, up to a year and only change in response to external stimuli. It is therefore not unexpected that *N*-glycans are altered in many chronic diseases such as cancer (Bones et al., 2010; Lau & Dennis, 2008; Llop et al., 2016; Mereiter et al., 2016), diabetes (Itoh et al., 2007; Testa et al., 2015), metabolic syndrome (Lu et al., 2011; McLachlan et al., 2016), hypertension (Wang et al., 2016), Alzheimer's (Gizaw et al., 2017) and Parkinson's disease (Russell et al., 2017).

Like proteomics, the successful application of glycomics in many of these chronic diseases have been powered by MS. MS allows structural assignments of constituent monosaccharides within a complex oligosaccharide, and in conjunction with exoglycosidase digestions, the full structural and linkage analysis are made (**Figure 2**). It is however important to realize the type of *N*-glycans under investigation since acidic and neutral glycans require distinct ways of analyses. While a 2, 5-dihydroxybenzoic acid (2, 5-DHB) matrix in positive ion mode is sufficient to ionize neutral glycans and generate a good MS spectrum, acidic glycans are better ionized in a negative ion mode with either 2, 4, 6-trihydroxyacetophenone (THAP) or 6-aza-2-thiothymine matrices (Snovida & Perreault, 2007; Wada et al., 2007).

INSERT FIGURE 2 HERE

Employing the state-of-the-art MALDI-TOF-MS and hydrophilic interaction liquid chromatography (HILIC) technology, a study investigated the *N*-glycosylation and IgG profiles in the plasma of mothers and the umbilical cord of their newborns (Jansen et al., 2016). Briefly, IgG was isolated from the plasma with IgG Fc beads, after which 20µL was transferred onto a 96 well plate. Proceeding this were multiple washing steps, while the eluent was transferred onto another plate and centrifuged. IgG *N*-glycans and plasma *N*-glycans were released in a stepwise manner using peptide *N*-glycase F (PNGase F) in phosphate buffered saline (PBS). This was followed by esterification while the samples were separated and purified on HILIC. The purified and esterified samples were added to an MTP anchor Chip 800/384 MALDI plate and measured on MALDI-TOF-MS. The generated spectrum was then processed and analysed using MassyTools software. Utilising this method, the authors quantified 37 IgG *N*-glycans and 45 total plasma *N*-glycans in the plasma of mothers and their newborns. Additionally, the study observed a reduced sialylation, and galactosylation and an increased fucosylation in plasma of the umbilical cord (Jansen et al., 2016).

In another study involving the simultaneous application of four methods for analysing immunoglobulin G (IgG) *N*-glycans, it was shown that MS methods (MALDI-TOF-MS and nanoLC-ESI-MS) when compared to UPLC combined with fluorescence detection and capillary gel electrophoresis with laser induced florescence detection, had a higher throughput, was suitable for site specific glycosylation and provided detailed structural information with highly sensitivity (Huffman et al., 2014).

In glycomics, the software supports for data handling and interpretation include: GlycoBase (Hizal et al., 2014), GlycoExtractor (Artemenko et al., 2010), EUROCarbDB (von der Lieth et al., 2011), GLYCOSCIENCE.de (Lütteke et al., 2006), GlycomeDB (Ranzinger et al., 2011), MassyTools (Jansen et al., 2015) and GlycoWorkBench (Ceroni et al., 2008; Yu et al., 2013).

4) Mass spectrometry in lipidomics

The lipidome comprises a collection of lipids in the cells of an organism (Sandra & Sandra, 2013; Shevchenko & Simons, 2010). Lipids are abundant, and it is estimated that the eukaryotic cell comprises 10,000 to 100,000 lipid species from different lipid classes (Van Meer et al., 2008). These lipid classes are sphingolipids [e.g. sphingosine phosphate and ceramides (CER)] and glycerophospholipids [e.g. phosphatidylcholines (PC), lyso-phosphatidylcholines (LPC), phosphatidylserines (PSs) and phosphatidylethanolamines (PEs)], glycerolipids [(monoacylglycerols (MAGs), diacylglycerols (DAGs) and triacylglycerols (TAGs)] (Shevchenko & Simons, 2010; Wenk et al., 2015). Taken together, these lipids make up the matrix of the cell membrane and are responsible for many cellular processes including membrane trafficking, biological reproduction, cell division, cellular architecture, signalling, cell-cell interaction, efficient fuelling and energy schemes for the cell (German et al., 2007; Kontush & Chapman, 2010; Van Meer et al., 2008). Also, lipids are responsible for maintaining subcellular compartmentalization, generation of 1^{st} and 2^{nd} messengers during signal transduction and for ensuring a balance in electrochemical gradient (Hu

et al., 2009; Rai & Bhatnagar, 2017).

Lipidomics refers to the system-wide study of lipids including their structures, regulation and function within the cell (Blanksby & Mitchell, 2010; Schwudke et al., 2011). Introduced in 2003, lipidomics has been a frontier in recent research owing to their agricultural, pharmaceutical and clinical relevance (Wenk, 2005). Like proteomics, lipidomics analysis can be achieved by two main strategies: targeted and non-targeted lipidomics. While targeted lipidomics is restricted to identifying and quantifying known single lipid species using specific methods, non-targeted lipidomics on the other hand, relies on appropriate methods for the simultaneous identification and relative quantification of all lipids in a system (Sethi & Brietzke, 2016).

Lipidomics analysis is challenging and regardless of the strategy employed, a powerful analytical tool is required. GC with flame ionization detection (FID) was widely used, and in fact, is still used for separating targeted well identified lipids such as the fatty acids because it is relatively cheap and simple to operate (Wenger, 2014). NMR has emerged as a useful tool because it can complete analysis within a short analytical run and importantly, did not require extensive sample pre-separation prior to detection. However, NMR can only detect the most abundant lipid species/metabolite within the sample (Gika et al., 2014).

However, MS based techniques are superior both in coverage and identification. In addition to the already mentioned MS techniques for proteomics and glycomics, lipidomics can also be performed using MALDI-FT-ICR-MS (Stübiger et al., 2012). However, regardless of the technique employed, MS characterizes lipid species in two

ways. The first stage involves the determination of the intact mass of the molecules while the second involves collision of lipid ions with inert gases followed by dissociation into structural fragments. Comprehensive information such as classification, nomenclature and identification are obtained with the help of online resources such as the LIPID maps, Lipid Profiler (Ejsing et al., 2006), Lipid Inspector (Schwudke et al., 2006), LipidXplorer and Lipid Data Analyser (Hartler et al., 2011).

Taken together, advances in MS technology and the accompanying software have revealed the role of lipids in various chronic conditions. For example, a study profiled lipid in the plasma of early stage cancer patients and individuals with benign breast disease (Chen et al., 2016). Briefly, lipids were isolated from plasma and centrifuged. Following this was the addition of internal standards and proceeded by injection into the LC system. Lipid profiling was then accomplished by LC-ESI-MS/MS while data analysis was performed using Applied Biosystems Analyst. With this technology, the study was able to quantify 15 lipid species that could be potential biomarkers for breast cancer (Chen et al., 2016). In another study, the MALDI-TOF-MS technique was used to quantify 157 lipid species in plasma, 171 in high density lipoprotein cholesterol (HDL), 182 in low density lipoprotein (LDL) and 148 in very low density lipoproteins (VLDL)(Serna et al., 2015).

INSERT FIGURE 3 HERE

Similarly, while performing a global lipidomics profiling in prostate cancer (PCa) and benign prostate hyperplasia (BPH) patients using LC-MS, a total of 350 lipid

species comprising 6 cholesterol ester (CE), 7 DAG, 9 hexosylceramide (HexCer), 24 free fatty acids, 10 Cer, 10 LPE and 10 LPC, amongst others, were identified, all of which could be potential biomarkers for prostate cancer (Li et al., 2016). Another study combined LC-ESI-SRM and MALDI-QIT-TOF-MS/MS to conduct a targeted lipidomics analyses in 13 patients with familiar combined hyperlipidaemia (FCH) and found many significant associations between atherogenic LPC species and VLDL **(Figure 3)** (Stübiger et al., 2012). Similarly, a targeted lipidomics profiling in the sera of chronic hepatitis B (CHB) patients, hepatitis B virus associated cirrhosis (HCV) and hepatocarcinoma (HCC) using UPLC-MS and identified 140 lipid species which could be potential biomarkers (Wu et al., 2017).

While MS technology has been a propeller for many "OMICS" studies, there are some prevailing challenges. Therefore, the remainder of this manuscript is dedicated to revealing these challenges and highlighting the need for innovation.

Summary of existing problems in MS

Having provided a comprehensive review of the MS technology and its applications, there is the need to recap the main limitations that characterize this technology. For example, the MALDI-TOF-MS is limited by signal suppression effects and narrow dynamic ranges of detectors (Poon, 2007). These problem seem to be ameliorated with the advent of the SELDI-TOF MS, however, SELDI-TOF has the following limitations: poor resolution for large proteins (e.g. markers with molecular weight > 20kDa); enormous amount of time required for purifying all significant SELDI peaks; lack of reproducibility; 4) serum/plasma proteins < μg/ml cannot be

detected; susceptibility to identifying false-significant biomarkers (Poon, 2007). These challenges have largely been addressed by exploring multiple hyphenated MS platforms and other techniques (e.g. NMR). Although these complementary approaches have improved sensitivity, accuracy, run times and analytical coverage, there are still problems relating to identification of low abundant analytes, sequencing speed (e.g. proteomics), deconvolution, handling and performing large data analysis, inadequate information from spectra to produce sequence identifications, data standardization and data fusion across multiple analytical platforms (Gika et al., 2014; Liu et al., 2010).

Given the prevailing challenges, it is time to shift towards extreme automation powered by information technology (IT) or computer science. IT promises to solve many complex problems using conventional algorithms and neural networks. Moreover, IT can enable intelligent exploitation of heterogeneous data, translational research and provide solution for implantation science (Combi, 2017; Özdemir & Hekim, 2018). While it is beyond the scope of this review to discuss in detail these concepts, a brief introduction and definition of terms will be necessary.

IMass

IMass is a term coined by us that incorporates traditional MS technology and Artificial intelligence (AI). Here, experimental methods are accomplished with MS while AI performs data processing and analysis. The IMass will help to answer complex scientific questions and generate ideas that were not previously possible with MS technology alone.

1) Information technology

IT is a generic term that encompasses information management, retrieval, data manipulation and processing. Information systems and application software are designed, exploited, installed and carried out by computer science and communication technology. IT includes sensor technology, computer technology, AI technology, communication technology and internet technology.

Nowadays, IT has improved tremendously. Internet of Things (IoT) and cloud computing are typical representations (Özdemir & Hekim, 2018). IoT Internet of things are wireless connected network of objects or devices that interact with each other via embedded systems and ubiquitous intelligence (Xia et al., 2012). IoT is an important part of the new generation IT, changing information and communication based on Internet and extending the user sides to stereo dimension. By using information sensing devices including radio frequency identification (RFID) infrared sensors, global positioning system (GPS), laser scanner amongst others, internet of things is able to connect relayed events according to agreed protocols that in turn facilitate the identification, location, monitoring and management of objects (Kopetz, 2011; Sun, Song et al., 2016; Weber & Weber, 2010; Xia et al., 2012).

Cloud computing is defined as an internet-based computing that allows the storage and sharing of data and resources with other computers and devices. It shares characteristics with grid computing, parallel computing and utility computing (Armbrust et al., 2010; Buyya et al., 2009; Dikaiakos et al., 2009). Multiple PCs with relatively low cost are integrated to a system with powerful computing capacity. Results

of the system are then fedback to end users through advanced business model such as SaaS, PaaS, IaaS and MSP (Lenk et al., 2009).

2) Artificial intelligence

AI is an important branch of computer science. It was officially put forward by McCarthy in Dartmouth Society in 1956 as "artificial system with certain degree of intelligent behavior that applies studies in human intelligence activity (Hamet & Tremblay, 2017; Nilsson, 2014; Patel et al., 2009; Russell & Norvig, 1995)". AI has undergone many developmental stages since the 1950s. Alan Turing proposed the theoretical model which established the theoretical foundation of modern computers, and the famous Turing guidelines, which has been the most important standard of intelligent machine (Castelfranchi, 2013). The era of AI gave rise to the idea that instead of programming computers to perform tasks, they could teach themselves or learn to perform tasks without being explicitly programmed to do so. This concept was later referred to as machine learning (ML). ML applies set of rules called algorithms to solve a problem and it easily adapts to changes in data, scalable and efficient compared to those programmed by humans. An extension of ML is deep learning (DL) which is based on data representations and designed to learn from an input data and apply to other data (Cohen & Feigenbaum, 2014; Nilsson, 2014).

Knowledge-based expert systems have been developing rapidly and its application range has provided enormous benefits across all human fields. Nowadays, AI has advanced into large-scale distributed expert cooperative systems, parallel deduction and multiple expert system development tools (Wenger, 2014). The year 2016 was known

as “The primary era of AI” because of massive developments of AI which was also characterized by several landmark events. For instance, AI program AlphaGo designed by Google DeepMind defeated human Go champion Lee Chang-ho in March 2016. IBM WATSON HealthCare passed the United States Medical Licensing Examination (USMLE) and got medical license in the same year. Moreover, the first self-driving car designed by EasyMile (France) and Citymobil2 (European Union) was tested in 2016 **(Figure 4).** With the rapid advances in software programming, electronic speed and capacity, it is obvious that IT will not only impact society and daily life but also, the intelligence of computers will someday surpass that of humans (Hamet & Tremblay, 2017). However, the compelling question is how the concepts and applications of AI or IoT can be impactful in MS and OMICS analysis?

3) Application of IMass to MS

The MS systems resulted in an increased operational complexity and overwhelming experimental data were obtained from a single analysis. This increasing and enormous amount of information demanded an optimization of the instrument’s operational conditions to acquire the most significant data (Place, 1995; Wong, 1984). This triggered the need to incorporate AI to MS instrumentation during the early 1970s and several developments have followed over the years (**Figure 4**). The first-generation intelligent software, alongside AI algorithms including Bayesian algorithm, Vector Support Machine, Decision Tree, Random Forest and Artificial Neural Network were all developed to support MS (Kondrat et al., 1978; RA et al., 1979). Expect system, described as computer program that houses a high quality and specialized knowledge

to solve complex problems was developed. It was designed to build upon the problem-solving abilities of human experts and in some cases, apply complex reasoning to solve problems that is beyond the capabilities of human experts. Neural networks that mimic the pattern-recognition and parallel processing of human experts were also developed (Place, 1995; Wong, 1984).

Over the years, it has been shown that AI has enormous potential to transform the efficiency of MS. The incorporation of MS to AI and expert system, not only will it adapt and respond to situations quickly, but also it will utilize the power of reasoning and inference to effectively perform tasks and interpret correct the data to meet the researcher's experimental targets. AI was applied for tuning the triple quadrupole mass spectrometer (TQMS). This was to allow analytical chemists to tune the instrument over small mass ranges to increase the sensitivity for each ion. Here, two approaches were employed; an expert system and algorithmic approach. It was shown that AI powered tuning increased the sensitivity, was faster compared to manual tuning methods which is time consuming. Further, while analysing 12 sulphur compounds using TQMS, it was shown that more time was required by a human expert to manually tune for each parent/daughter ion compared to an AI tuning system (Brand and Wong, 1986).

For example, in 1990, expert system alongside Bayesian algorithm and AI software were employed to detect low resolution mass spectra (Scott, 1991). The expert system obtained a canonical representation of structures and a new heuristic for incorporating constraints of the mass spectrum (Sridhar et al., 1991). The MS combined

with AI and expert system were utilized for optimizing and controlling MS performance, detector signal collection as the function of m/z and improved overall data analysis.

With different combination metrics, AI can transform the information extracted from the MS, interpret data from spectra and present such data in a visual and symbolic form. It can be useful for identifying pre-analytical and analytical errors; AI can be applied to assess the performance of an MS instrument and can reveal whether the instrument's performance is within algorithmically pre-determined specification; AI can enable the prediction of protein complexes and can provide machine learning algorithms such as in protein-protein interactions to reveal therapeutic targets (Hamet & Tremblay, 2017); Stabilization of spectrogram is the consistency of spectrograms of repeated experiments from the same sample. For MS, spectrograms of repeated experiments are usually unidentical, with relatively large coefficient of variation (French et al., 2014; Muddiman & Oberg, 2005; Ramakrishnan, Nair, & Rangiah, 2016). Data analysis of MS belong to multi-index evaluation system, characteristics of each index are different, and the dimensions differ by several orders of magnitude (Du et al., 2016; Edmands et al., 2014; Hamm et al., 2012). Because of variation in the indices, there is a possibility of over-highlighting some while weakening others, especially those with relatively smaller numbers. However, with the higher computational power and the higher statistical strengths by AI, these problems can be solved (Du et al., 2016; Edmands et al., 2014; Hamm et al., 2012; Liao et al., 2014); Resolution of spectrum library is the ability to differentiate one spectral peak from another. Resolution of spectrum library of traditional MS is low, not only because of the limited computing

power to support high resolution and high accuracy, but also because only few spectrograms are generated in a single-instrument model. For example, traditional MS technology has been applied on microorganism identification studies (Carbonnelle et al., 2011; Seng et al., 2010). However, the results are not always consistent or reproducible and often unstable, as well as variations that cannot be accepted. By incorporating AI technology, IMass will be able to overcome these shortcomings and produce more stable results with less variation. For example, enabled by artificial neural network analysis of the mass spectra, streptomycetes was identified it was shown to be rapid, reliable and cost-effective (Howells et al., 1992; Chun, 1993).

Over the years, other significant improvements have been made. Bayesian algorithm was introduced to the protein search engine to identify protein from protein databases by MS data mapping (Zhang et al., 2000). The vector machine was applied to classify peptide MS/MS spectra and SEQUEST score (Anderson, 2003). In 2004, Random Forest, also called decision forest, developed to standardize the mass spectra and for sample classification (Satten et al., 2004; Tong et al., 2004).

INSERT FIGURE 4 HERE

4) IMASS and Big Data analysis

Thus far, research has advanced and multi-OMICS studies involving large population are increasing. Consequently, there is the influx of Big Data or large "OMICS-based" datasets that are often characterised by large volumes, variety, veracity, valorisation and velocity (Özdemir & Hekim, 2018). Moreover, Big Data is becoming complex comprising demographic, biomedical signals, genetic data, and clinical

pathways (Combi, 2017). Additional sources of challenge for OMICS analysis are consequence of poor experimental design, low signal-to noise ratio and high analytical variance. Other challenges also relate to real-time data gathering, mining storage, predictive data analytics, and visualization. The full potential of such Big Data can only be harnessed after real-time data analysis and cannot be possible with the old scientific practices. AI and IoT can be employed to translate Big Data into knowledge-based innovations (Özdemir & Hekim, 2018). For example, it can support the translation of science into a more decision-based tasks such as patient stratification, disease prediction, diagnosis, and therapy (Combi, 2017; McCudden, 2017). Also, ML methods can draw meaningful conclusions from relatively small data; transforming data into visualizable images (Grapov et al., 2018). ML methods may provide efficient integration of omics data integration, thereby amplifying significant biological variants and enabling full interrogation of biological systems.

The size and complexity of data put a significant pressure on computers for analysis. IMass supported by cloud computing (Hoopmann & Moritz, 2013; Mohammed et al., 2012; Schadt et al., 2010), makes it more suitable to analyze complex and large sample data. Cloud computing resources can be expanded and can provide additional power for data analysis, facilitate sharing of large data (Hoopmann & Moritz, 2013; Mohammed et al., 2012; Schadt et al., 2010). This will be an important step towards MS innovation; presently, comparing MS data to medical reference libraries will require data to be inputted manually (Cho et al., 2015; Schilling et al., 2011), which is often time consuming, labor intensive and less accurate. IMass can be connected to

medical and biological components information libraries directly, data can be combined, shared and stored for high throughput analysis. This innovation can significantly save time, energy and at the same time, increase accuracy.

However, the main challenge for IMASS is that OMICS data are not amenable to certain assumptions of DL. As a result, domain specific approaches are usually required to deal with unrelated biological variance. In the context of protein-protein interaction, it will be challenging to represent peptide and protein in a meaningful way. Advanced studies will be required to ascertain novel ways to represent protein sequence information. Moreover, high-quality data, large amount of correctly labelled training data and high parameter neural networks will be required for IMASS and achieving this is expensive and time consuming. Further, IMASS is complex and understanding specific neural networks and interpreting models will be challenging for amateur researchers. In addition, how to apply or use deep learning methods to yield a more realistic model and how to employ the estimation method to develop the prior knowledge of the deep neural network is can be a challenging task. Nonetheless, IMASS will be highly relevant in the pursuit of translating MS based research into clinical diagnosis and prognosis of diseases.

Conclusion and future directions

It should be clear to readers by now that the advent of the MS technology has revolutionized medical research to a larger degree. It has provided an insight on the molecular intricacies that underlie many chronic diseases and has made it possible to perform large scale “OMICS” analyses. All the application of MS technology for

population-based studies cited in this review have mainly focused on single OMICS data. However, it is about time that scientists shift focus from single OMICS to integrated or multiple OMICS data. This has become necessary because the molecular complexities associated with many diseases cannot be unraveled with just a single "OMIC" data (Hasin et al., 2017). Additionally, single OMICS analysis only provides correlations between traits and diseases with limited information on causative changes. In theory, beside the central dogma for DNA, RNA and proteins, the compelling question that is worth asking is what happens to the regulation of lipids and sugars and their glycosylation and how they are involved in health and diseases? Is there a para-central dogma for encoding these functions? For practice, integration of different OMICS data will promote a comprehensive understanding on the flow of information that underpins chronic diseases (Hasin et al., 2017). For example, one study used genomics and glycomics data to show that hepatocyte nuclear factor alpha (HNF1α) is the master regular of fucosylation (Lauc et al., 2010). Similarly, another study combined glycomics and lipidomics data to shed light on the association between glycans and lipids in four European populations. Further, this study provided information on the interactive metabolic pathways that exist between glycans and lipids (Igl et al., 2011).

While the MS technology has been pivotal in medical research, it has been limited in certain aspects. In theory, beside the central dogma for DNA, RNA and Proteins, what happens to the regulation of lipids and sugars and their glycosylation? Is there a para-central dogma for encoding these functions? In practice, from the past years to

date, then towards the future, the goals for MS technology must change from high resolution and high sensitivity to high stability and high repeatability, from focusing on single peaks to multiple peaks within a complex spectrogram, and from solving simple questions to solving more complex ones. Actualizing these demands require some innovations to the present MS technology. The incorporation of MS with AI is highly feasible and have broad prospects. For instance, incorporating MS technology with AI system platforms like AliCloud and IBM WATSON is such a great opportunity for researchers using MS to work out experimental data and use related algorithms exploited by AI technology to store and perform complex data analysis and processing. This is what researchers are expecting, and exactly what we are continuously working hard on and hoping it will come into fruition rather sooner than envisaged.

Acknowledgement

The study was supported by the grants from National Natural Science Foundation of China (NSFC 81771207, 81673247, 81573215, 81370083,) and the Joint Project of the Australian National Health and Medical Research Council and the National Natural Science Foundation of China (NHMRC APP1112767-NSFC 81561128020)

Conflict of interest

None declared

Abbreviations

A-I	Apolipoprotein
AI	Artificial intelligence
ALB	Albumin
ATP	Adenosine triphosphate
CE	Capillary electrophoresis
DAG	Diacylglycerols
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DDA	Data-dependant acquisition
DHB	Dihydroxybenzoic acid
DIA	Data independent acquisition
ESI	Electrospray ionization
FGA	Fibrinogen alpha
FID	Flame ionization detection
FT-ICR	Fourier transform ion cyclotron resonance
G6PD	Glucose-6 phosphate dehydrogenase
GC	Gas chromatography
GPS	Global positioning system
HDL	High density lipoprotein cholesterol
HILIC	Hydrophilic interaction liquid chromatography
ICR	Ion cyclotron resonance
IoT	Internet of things
IT	Information technology
LC	Liquid chromatography
LIT	Linear ion trap
LPC	Lyso-phosphatidylcholines
MAG	Monoacylglycerols
MALDI	Matrix assisted laser desorption ionization
ML	Machine learning
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NPC	Nasopharyngeal carcinoma

PC	Phosphatidylcholines
PCK	Phosphoenolpyruvate carboxykinase
PE	Phosphatidylethanolamine
PPP	Pentose phosphate pathway
PS	Phosphatidylserine
PTM	Post-translational modification
RFID	Radio frequency identification
SELDI	Surface enhanced laser desorption ionisation
SPSS	Statistical Package for Social Sciences
SRM	Selected reaction monitoring
T2DM	Type II diabetes mellitus
TAG	Triacylglycerol
TOF	Time of flight
UPLC	Ultra-performance liquid chromatography

References

Adua E, Russell A, Roberts P, et al. (2017). Innovation Analysis on Postgenomic Biomarkers: Glycomics for Chronic Diseases. OMICS 21(4), 183-196.

Aebersold R and Mann M (2016). Mass-spectrometric exploration of proteome structure and function. Nature 537(7620), 347-355.

Ali SE, Farag MA, Holvoet P, et al. (2016). A Comparative Metabolomics Approach Reveals Early Biomarkers for Metabolic Response to Acute Myocardial Infarction. Sci Rep 6, 1-14.

Anderson DC, Li W, Payan DG and Noble WS. (2003). A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. J Proteome Res 2(2), 137-146.

Aretz I, and Meierhofer D (2016). Advantages and pitfalls of mass spectrometry based metabolome profiling in systems biology. Int J Mol Sci 17(5), 1-14.

Armbrust M, Fox A, Griffith R et al. (2010). A view of cloud computing. Commun. ACM 53(4), 50-58.

Artemenko NV, Campbell MP and Rudd PM. (2010). GlycoExtractor: a web-based

interface for high throughput processing of HPLC-glycan data. J Proteome Res, 9(4), 2037-2041.

Bauer M, Ahrné E, Baron AP, et al. (2014). Evaluation of data-dependent and-independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. J Proteome Res 13(12), 5973-5988.

Blanksby SJ and Mitchell TW (2010). Advances in mass spectrometry for lipidomics. Annu Rev Anal Chem 3, 433-465.

Bones J, Mittermayr S, O'Donoghue N, Guttman AS and Rudd PM (2010). Ultra performance liquid chromatographic profiling of serum N-glycans for fast and efficient identification of cancer associated alterations in glycosylation. Anal Chem 82(24), 10208-10215.

Brand, H. R., & Wong, C. M. (1986, April). Application of knowledge based systems technology to triple quadrupole mass spectrometry (TQMS). In Proceedings of the Fifth National Conference on Artificial Intelligence (pp. 812-819).

Buyya R, Yeo, CS, Venugopal S., Broberg J and Brandic I (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Gener Comp Syst 25(6), 599-616.

Carbonnelle E, Mesquita C, Bille E et al. (2011). MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. Clin Biochem 44(1), 104-109.

Carvalho AS, Cuco CM, Lavareda C et al. (2017). Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer. Sci Rep, 1-13.

Castelfranchi C (2013). Alan Turing's "Computing Machinery and Intelligence". Topoi, 32(2), 293-299.

Ceroni A, Maass K, Geyer H Geyer, et al. (2008). GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J Proteome Res 7(4), 1650-1659.

Chen X, Chen H, Dai M et al. (2016). Plasma lipidomics profiling identified lipid biomarkers in distinguishing early-stage breast cancer from benign lesions. Oncotarget, 7(24), 36622-36631.

Cho JY, Lee HJ, Jeong SK et al. (2015). Combination of multiple spectral libraries improves the current search methods used to identify missing proteins in the chromosome-centric human proteome project. J Proteome Res 14(12), 4959-

4966.

Chun J, Atalan E, Ward AC and Goodfellow M (1993). Artificial neural network analysis of pyrolysis mass spectrometric data in the identification of Streptomyces strains. FEMS Microbiol Lett. 107(2-3), 321-326.

Chun J, Atalan E, Kim SB, et al. (1993). Rapid identification of streptomycetes by artificial neural network analysis of pyrolysis mass spectra. FEMS Microbiol Lett. 114(1), 115-119.

Cifani P, Shakiba M, Chhangawala S and Kentsis A (2017). ProteoModlR for functional proteomic analysis. BMC Bioinformatics, 18(1), 1-7.

Clancy T and Hovig E (2014). From proteomes to complexomes in the era of systems biology. Proteomics 14(1), 24-41.

Cohen PR and Feigenbaum EA (2014). *The handbook of artificial intelligence* (Vol. 3): Butterworth-Heinemann.

Combi C (2017). Editorial from the new Editor-in-Chief: Artificial Intelligence in Medicine and the forthcoming challenges. Artif Intell Med, *76*, 37.

Comisarow, MB and Marshall AG (1974). Fourier transform ion cyclotron resonance spectroscopy. Chemical Physics Letters 25(2), 282-283.

Cooks RG, Ouyang Z Takats Z and Wiseman JM (2006). Ambient mass spectrometry. Science 311(5767), 1566-1570.

Costa C, Maraschin M and Rocha M (2015). An Integrated Computational Platform for Metabolomics Data Analysis. Paper presented at the 9th International Conference on Practical Applications of Computational Biology and Bioinformatics. 2015 (pp. 37-47). Springer, Cham.

Danikiewicz W (2013). Keith R. Jennings (Ed.): A history of European mass spectrometry. Anal Bioanal Chem 405(10), 3011-3012.

De Hoffmann E and Stroobant V (2007). *Mass spectrometry: principles and applications*: John Wiley & Sons.

Dikaiakos MD, Katsaros D, Mehra P, et al. (2009). Cloud computing: Distributed internet computing for IT and scientific research. *IEEE Internet computing, 13*(5).

Doerr A (2017). Global metabolomics. Nature Methods 14(1), 32-32.

Downard KM (2007). Francis William Aston: the man behind the mass spectrograph. Eur J Mass Spectrom 13(3), 177-190.

Du Y, M Hu Y, Xia Y and Ouyang Z (2016). Power Normalization for Mass Spectrometry

Data Analysis and Analytical Method Assessment. Anal Chem 88(6), 3156-3163.

Dunn WB, Broadhurst D, Begley P et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature Protocols *6*(7), 1060-1083.

Dunn WB, Broadhurst DI, Atherton HJ, et al. (2011). Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev 40(1), 387-426.

Edmands WM, Ferrari P and Scalbert A (2014). Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. Anal Chem 86(21), 10925-10931.

Ejsing CS, Duchoslav E, Sampaio J, et al. (2006). Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. Anal Chem 78(17), 6202-6214.

Elias JE, Haas W, Faherty BK, et al. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nature Methods 2(9), 667-675.

Fiedler K and Simons K (1995). The role of N-glycans in the secretory pathway. Cell, 81(3), 309-312.

French D, Sujishi KK, Long-Boyle JR, et al. (2014). Development and Validation of a Liquid Chromatography–Tandem Mass Spectrometry Assay to Quantify Plasma Busulfan. Ther Drug Monit 36(2), 169-174.

Fridriksson EK, Baird B and McLafferty FW (1999). Electrospray mass spectra from protein electroeluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. J Am Soc Mass Spectrom 10(5), 453-455.

Gao Y, Wang X, Sang Z et al. (2017). Quantitative proteomics by SWATH-MS reveals sophisticated metabolic reprogramming in hepatocellular carcinoma tissues. Sci Rep 7, 1-12.

Gault, VA, and McClenaghan NH (2013). *Understanding bioanalytical chemistry: principles and applications*: John Wiley & Sons.

Ge S, Wang Y, Song M, et al. (2018). Type 2 Diabetes Mellitus: Integrative Analysis of Multiomics Data for Biomarker Discovery. OMICS 22(7), 514-523.

German JB, Gillies LA, Smilowitz JT, et al. (2007). Lipidomics and lipid profiling in metabolomics. Curr Opin Lipidol 18(1), 66-71.

Ghaste M, Mistrik R and Shulaev V (2016). Applications of Fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics. Int J Mol Sci 17(6), 1-22.

Gika HG, Wilson, ID and Theodoridis GA (2014). LC–MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives. J Chromatogr B 966, 1-6.

Gizaw ST, Ohashi T, Tanaka M, et al. (2016). Glycoblotting method allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum and cerebrospinal fluid towards potential biomarker discovery. Biochim Biophys Acta Gen Subj 1860(8), 1716-1727.

Grapov D, Fahrmann J, Wanichthanarak K and Khoomrung S (2018). Rise of Deep Learning for Genomic, Proteomic, and Metabolomic Data Integration in Precision Medicine. OMICS 1, 1-7.

Griffiths IW (1997). JJ. Thomson— the Centenary of His Discovery of the Electron and of His Invention of Mass Spectrometry. Rapid Commun Mass Spectrom 11(1), 2-16.

Hamet P and Tremblay J (2017). Artificial intelligence in medicine. Metabolism *69*, S36-S40.

González-Peña D, Dudzik D, Colina-Coca C, et al. (2016). Multiplatform metabolomic fingerprinting as a tool for understanding hypercholesterolemia in Wistar rats. Eur J Nutr 55(3), 997-1010.

Hamm G, Bonnel D, Legouffe R, et al. (2012). Quantitative mass spectrometry imaging of propranolol and olanzapine using tissue extinction calculation as normalization factor. J Proteomics *75*(16), 4952-4961.

Hardman M and Makarov AA. (2003). Interfacing the orbitrap mass analyzer to an electrospray ion source. Anal Chem 75(7), 1699-1705.

Hargittai I. (2007). Gerald E. Brown and Chang-Hwan Lee, Hans Bethe and His Physics. Struct Chem 18(5), 723-724.

Hartler J, Trötzmüller M, Chitraju C, et al. (2011). Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. Bioinformatics, 27(4), 572-577.

Hasin Y, Seldin M and Lusis A (2017). Multi-omics approaches to disease. Genome Biology, 18(1), 83.

Hebert DN and Molinari M (2007). In and out of the ER: protein folding, quality control, degradation, and related human diseases. Physiol Rev 87(4), 1377-1408.

Helenius A and Aebi M (2001). Intracellular functions of N-linked glycans. Science 291(5512), 2364-2369.

Helenius A and Aebi M (2004). Roles of N-linked glycans in the endoplasmic reticulum. Ann Rev Biochem 73(1), 1019-1049.

Hevesy, G (1948). Francis William Aston. 1877-1945. *Obituary Notices of Fellows of the Royal Society* 5(16), 635-650.

Hizal DB, Wolozny D Colao J, et al. (2014). Glycoproteomic and glycomic databases. Clin Proteomics 11(1), 15.

Hoffmann ED and Stroobant V (2001). Mass spectrometry: principles and applications: England: Wiley.

Hoopmann MR and Moritz RL (2013). Current algorithmic solutions for peptide-based proteomics data generation and identification. Curr Opinion Biotechnol 24(1), 31-38.

Howells SL, Maxwell RJ, Peet AC, et al. (1992). An Investigation of Tumor ^{1}H Nuclear Magnetic Resonance Spectra by the Application of Chemometric Techniques. Magnetic Resonance in Medicine. 28(2), 214-236.

Hsieh KT, Liu PH and Urban PL (2015). Automated on-line liquid–liquid extraction system for temporal mass spectrometric analysis of dynamic samples. Analytica Chimica Acta 894, 35-43.

Hu C, van der Heijden R, Wang M, et al. (2009). Analytical strategies in lipidomics and applications in disease biomarker discovery. J Chromatogr B 877(26), 2836-2846.

Hu Q, Noll RJ, Li H, et al. (2005). The Orbitrap: a new mass spectrometer. J Mass Spectrom 40(4), 430-443.

Huang DW, Sherman BT, and Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37(1), 1-13.

Huffman JE, Pučić-Baković M, Klarić L, et al., (2014). Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in

genetic and epidemiological research. Mol Cell Proteomics 13(6), 1598-1610.

Hurst GB, Weaver K, Doktycz MJ, et al. (1998). MALDI-TOF analysis of polymerase chain reaction products from methanotrophic bacteria. Anal Chem 70(13), 2693-2698.

Hutchens TW and Yip TT. (1993). New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Commun Mass Spectrom 7(7), 576-580.

Igl W, Polašek O, Gornik O et al. (2011). Glycomics meets lipidomics—associations of N-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations. Mol BioSyst 7(6), 1852-1862.

Itoh N, Sakaue S, Nakagawa H et al. (2007). Analysis of N-glycan in serum glycoproteins from db/db mice and humans with type 2 diabetes. Am J Physiol Endocrinol Metab 293(4), E1069-E1077.

Jansen BC, Bondt A, Reiding KR, et al. (2016). MALDI-TOF-MS reveals differential N-linked plasma-and IgG-glycosylation profiles between mothers and their newborns. Sci Rep 6, 1-11.

Jansen BC, Reiding KR, Bondt A, et al. (2015). MassyTools: a high-throughput targeted data processing tool for relative quantitation and quality control developed for glycomic and glycoproteomic MALDI-MS. J Proteome Res 14(12), 5088-5098.

Jin X, Su T, Kong J, et al. (2018.) State-of-the-Art mobile intelligence: enabling robots to move like humans by Estimating Mobility with Artificial Intelligence. Applied Sci 8(3), 379.

Karas M, Bahr U and Gießmann U (1991). Matrix‐assisted laser desorption ionization mass spectrometry. Mass Spectrom Rev 10(5), 335-357.

Karas M and Hillenkamp F (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60(20), 2299-2301.

Kauppila TJ, Talaty N, Salo PK, et al. (2006). New surfaces for desorption electrospray ionization mass spectrometry: porous silicon and ultra‐thin layer chromatography plates. Rapid Commun Mass Spectrom 20(14), 2143-2150.

Kessler N, Neuweger H, Bonte A, et al. (2013). MeltDB 2.0–advances of the metabolomics software system. Bioinformatics 29(19), 2452-2459.

Kim W, Bennett EJ, Huttlin et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell 44(2), 325-340.

Kizuka Y, Kitazume S and Taniguchi N (2017). N-glycan and Alzheimer's disease.

Biochim Biophys Acta Gen Subj 1861(10),2447-54.

Kondrat RW and Cooks RG (1978). Direct analysis of mixtures by mass spectrometry. Anal. Chem. 50, 81–92A.

Konijnenberg A, Butterer A and Sobott F (2013). Native ion mobility-mass spectrometry and related methods in structural biology. Biochim Biophys Proteins Proteomics 1834(6), 1239-1256.

Kontush A and Chapman MJ (2010). Lipidomics as a tool for the study of lipoprotein metabolism. Curr Atheroscler Rep 12(3), 194-201.

Kopetz H (2011). Real-time systems: design principles for distributed embedded applications: Springer Science & Business Media.

Larance M and Lamond AI (2015). Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16(5), 269-280.

Lau KS and Dennis JW (2008). N-Glycans in cancer progression. Glycobiology 18(10), 750-760.

Lauc G, Essafi A, Huffman JE, Hayward C et al. (2010). Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1-α as a master regulator of plasma protein fucosylation. PLoS Genet 6(12), e1001256.

Lehmann WD (2016). A timeline of stable isotopes and mass spectrometry in the life sciences. Mass Spectrom. Rev. 36(1), 58-85.

Lenk A, Klems M, Nimis J, et al. (2009). *What's inside the Cloud? An architectural map of the Cloud landscape.* Paper presented at the Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing.

Li B, Tang J, Yang Q et al. (2016). Performance Evaluation and Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis. Sci Rep 6, 1-13.

Li J, Ren S, Piao Hl, et al. (2016). Integration of lipidomics and transcriptomics unravels aberrant lipid metabolism and defines cholesteryl oleate as potential biomarker of prostate cancer. Sci Rep 6, 1-11.

Liao HW, Tsai IL, Chen GY, et al. (2014). Quantification of target analytes in various biofluids using a postcolumn infused-internal standard method combined with matrix normalization factors in liquid chromatography–electrospray ionization mass spectrometry. J Chromatogr A, 1358, 85-92.

Liu D, Chu X, Wang H et al. (2018). The changes of immunoglobulin G N-glycosylation

in blood lipids and dyslipidaemia. J Transl Med 16(1), 1-10.

Liu D, Zhao Z, Wang A, et al. (2018). Ischemic stroke is associated with the pro-inflammatory potential of N-glycosylated immunoglobulin G Neuroflammation, 15(1), 123.

Liu X, Inbar Y, Dorrestein PC et al. (2010). Deconvolution and database search of complex tandem mass spectra of intact proteins a combinatorial approach. Mol Cell Proteomics *9*(12), 2772-2782.

Llop E, Ferrer-Batallé M, Barrabés S, et al. (2016). Improvement of Prostate Cancer Diagnosis by Detecting PSA Glycosylation-Specific Changes. Theranostics 6(8):1190-204.

Lu JP, Knezevic A, Wang YX, et al. (2011). Screening novel biomarkers for metabolic syndrome by profiling human plasma N-glycans in Chinese Han and Croatian populations. J Proteome Research 10(11), 4959-4969.

Lu J, Huang Y, Wang Y, et al. (2012). Profiling plasma peptides for the identification of potential ageing biomarkers in Chinese Han adults. PloS One, 7(7), e39726.

Luo Y, Mok TS, Lin X, et al. (2017). SWATH-based proteomics identified carbonic anhydrase 2 as a potential diagnosis biomarker for nasopharyngeal carcinoma. Sci Rep 7. 1-11.

Lütteke T, Bohne-Lang A, Loss A, et al. (2006). GLYCOSCIENCES. de: an Internet portal to support glycomics and glycobiology research. Glycobiology, 16(5), 71R-81R.

Makarov A (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem 72(6), 1156-1162.

Makarov A, Denisov E, Kholomeev A, et al. (2006). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem 78(7), 2113-2120.

Mallick P and Kuster B (2010). Proteomics: a pragmatic perspective. Nature Biotechnol 28(7), 695-709.

Mathew AK and Padmanaban V (2013). Metabolomics: the apogee of the omics trilogy. Int J Pharm Pharm Sci 5(2), 45-48.

McCudden C (2017). The future of artificial intelligence and interpretative specialization in clinical biochemistry. Clin Biochem 50(6), 253-254.

McLachlan F, Timofeeva M, Bermingham M, et al. (2016). A Case-control Study in an Orcadian Population Investigating the Relationship between Human Plasma N-glycans and Metabolic Syndrome. J Glycomics Lipidomics 6(139), 2153-0637.

Meng Q, Ge S, Yan W, et al. (2016). Screening for potential serum‐based proteomic biomarkers for human type 2 diabetes mellitus using MALDI‐TOF MS. Proteomics Clin Appl 11(3-4), 1600079.

Mereiter S, Magalhães A, Adamczyk B et al. (2016). Glycomic analysis of gastric carcinoma cells discloses glycans as modulators of RON receptor tyrosine kinase activation in cancer. Biochim et Biophys Acta Gen Subj 1860(8), 1795-1808.

Mohammed Y, Mostovenko E, Henneman AA, et al. (2012). Cloud parallel processing of tandem mass spectrometry based proteomics data. J Proteomics Res 11(10), 5101-5108.

Muddiman DC and Oberg AL (2005). Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem 77(8), 2406-2414.

Nilsson NJ (2014). *Principles of artificial intelligence*: Morgan Kaufmann.

Özdemir V and Hekim N (2018). Birth of industry 5.0: Making sense of big data with artificial intelligence,“the internet of things” and next-generation technology policy. OMICS 22(1), 65-76.

PARODI AJ (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. Biochem J 348(1), 1-13.

Patel VL, Shortliffe EH, Stefanelli M, et al. (2009). The coming of age of artificial intelligence in medicine. Artif Intell Med 46(1), 5-17.

Patterson SD and Aebersold RH (2003). Proteomics: the first decade and beyond. Nature Genet 33, 311-323.

Patti GJ, Yanes O, and Siuzdak G (2012). Innovation: Metabolomics: the apogee of the omics trilogy. Nature Rev Mol Cell Biol 13(4), 263-269.

Paul W (1990). Electromagnetic traps for charged and neutral particles. Rev Mod Phys 62(3), 531.

Place J F, Truchaud A, Ozawa K, et al. (1994). Use of artificial intelligence in analytical systems for the clinical laboratory. Clinica chimica acta, 231(2), S1-S34.

Pluskal T, Castillo S, Villar-Briones A, et al. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics*, *11*(1), 395.

Poon TC (2007). Opportunities and limitations of SELDI-TOF-MS in biomedical research:

practical advices. Expert Rev Proteomics 4(1), 51-65.

Porter CJ and Bereman MS (2015). Data-independent-acquisition mass spectrometry for identification of targeted-peptide site-specific modifications. Anal Bioanal Chem 407(22), 6627-6635.

Psychogios N, Hau DD, Peng J et al. (2011). The human serum metabolome. *6*(2), e16957.

Rai S, and Bhatnagar S (2017). Novel Lipidomic Biomarkers in Hyperlipidemia and Cardiovascular Diseases: An Integrative Biology Analysis. OMICS 21(3), 132-142.

Ramakrishnan P, Nair S, and Rangiah K (2016). A method for comparative metabolomics in urine using high resolution mass spectrometry. J Chromatogr A, 1443, 83-92.

Ranzinger R, Herget S, von der Lieth CW, et al. (2011). GlycomeDB—a unified database for carbohydrate structures. Nucleic Acids Res 39(suppl 1), D373-D376.

Russell A, Adua E, Ugrina I, et al. (2018). Unravelling Immunoglobulin G Fc N-Glycosylation: A Dynamic Marker Potentiating Predictive, Preventive and Personalised Medicine. Int J Mol Sci 19(2) 1-18.

Russell AC, ŠM, Garcia MT, Novokmet M, et al. (2017). The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. Glycobiology, 27(5), 501-510.

Russell H and Siuzdak G (2003). A Mass Spec Timeline. *Today'S Chemist AT Work*.

Russell S and Norvig P (2010). *Artificial intelligence: a modern approach.* 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Salih B, Masselon C and Zenobi R (1998). Matrix‐assisted laser desorption/ionization mass spectrometry of noncovalent protein–transition metal ion complexes. J Mass Spectrom 33(10), 994-1002.

Sandra K and Sandra P (2013). Lipidomics from an analytical perspective. Curr Opin Chem Biol 17(5), 847-853.

Satten GA, Datta S, Moura H, et al. (2004). Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. 20(17), 3128-3136.

Savolainen OI, Sandberg AS, and Ross AB (2015). A simultaneous metabolic profiling and quantitative multimetabolite metabolomic method for human plasma using gas-chromatography tandem mass spectrometry. J Proteome Res 15(1),

259-265.

Schadt EE, Linderman MD, Sorenson J, et al. (2010). Computational solutions to large-scale data management and analysis. Nature Rev Genet 11(9), 647-657.

Schilling O, auf dem Keller U and Overall CM (2011). Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. Gel-Free Proteomics: Methods and Protocols 257-272.

Schwudke D, Oegema J, Burton L et al. (2006). Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. Anal Chem 78(2), 585-595.

Schwudke D, Schuhmann K, Herzog R, et al. (2011). Shotgun lipidomics on high resolution mass spectrometers. Cold Spring Harb Perspect Biol 3(9), a004614.

Scigelova M and Makarov, A (2006). Orbitrap mass analyzer–overview and applications in proteomics. Proteomics 6(S2), 16-21.

Scott D R (1991). Expert system for estimates of molecular weights of volatile organic compounds from low-resolution mass spectra. Analytica Chimica Acta 246(2), 391-403.

Sebastian A, Alzain MA, Asweto CO, et al. (2016). Glycan biomarkers for rheumatoid arthritis and its remission status in Han Chinese patients. Omics: a journal of integrative biology, 20(6), 343-351.

Seng P, Rolain JM, Fournier PE, et al. (2010). MALDI-TOF-mass spectrometry applications in clinical microbiology. Future Microbiol 5(11), 1733-1754.

Serna J, García-Seisdedos D, Alcázar A, et al. (2015). Quantitative lipidomic analysis of plasma and plasma lipoproteins using MALDI-TOF mass spectrometry. Chem Phys Lipids 189, 7-18.

Sethi S and Brietzke E (2016). Recent Advances in Lipidomics: Analytical and Clinical Perspectives. Prostaglandins Other Lipid Mediat. 128, 8-16.

Shevchenko A and Simons K (2010). Lipidomics: coming to grips with lipid diversity. Nature Rev Mol Cell biol 11(8), 593-598.

Sheynkman GM, Shortreed MR, Cesnik AJ, et al. (2016). Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. Annu Rev Anal Chem 9, 521-545.

Shi T, Song E, Nie S, et al. (2016). Advances in targeted proteomics and applications to biomedical research. Proteomics 16(15-16), 2160-2182.

Smith RD, Loo JA, Loo RRO, et al. (1991). Principles and practice of electrospray ionization? mass spectrometry for large polypeptides and proteins. Mass Spectrom Rev 10(5), 359-452.

Snovida SI and Perreault H (2007). A 2, 5‐dihydroxybenzoic acid/N, N‐dimethylaniline matrix for the analysis of oligosaccharides by matrix‐assisted laser desorption/ionization mass spectrometry. Rapid Comm Mass Spectrom 21(22), 3711-3715.

Stahl B, Steup M Karas M and Hillenkamp F (1991). Analysis of neutral oligosaccharides by matrix-assisted laser desorption ionization mass spectrometry. Anal Chem 63(14), 1463-1466.

Stübiger G, Aldover-Macasaet E, Bicker W, et al. (2012). Targeted profiling of atherogenic phospholipids in human plasma and lipoproteins of hyperlipidemic patients using MALDI-QIT-TOF-MS/MS. Atherosclerosis 224(1), 177-186.

Sun Y, Song H, Jara AJ, et al. (2016). Internet of things and big data analytics for smart and connected communities. IEEE Access*, 4*, 766-773.

Tanaka K, Waki H, Ido Y, et al. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time‐of‐flight mass spectrometry. Rapid Comm Mass Spectrom 2(8), 151-153.

Tebani A, Schmitz-Afonso I, Rutledge DN, et al. (2016). Optimization of a liquid chromatography ion mobility-mass spectrometry method for untargeted metabolomics using experimental design and multivariate data analysis. Analytica Chimica Acta 913, 55-62.

Testa R, Vanhooren V, Bonfigli AR, et al. (2015). N-Glycomic Changes in Serum Proteins in Type 2 Diabetes Mellitus Correlate with Complications and with Metabolic Syndrome Parameters. PloS One 10(3), e0119983.

Thomson JJ (1906). LXX. On the number of corpuscles in an atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 11*(66), 769-781.

Tong W, Xie Q, Hong H, et al. (2004). Using decision forest to classify prostate cancer samples on the basis of SELDI-TOF MS data: assessing chance correlation and prediction confidence. 112(16), 1622-1627.

Van Meer G, Voelker DR and Feigenson GW (2008). Membrane lipids: where they are

and how they behave. Nature Rev Mol Cell Biol 9(2), 112-124.

Veenstra TD (1999). Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. Biochem Biophys Res Commun 257(1), 1-5.

von der Lieth CW, Freire AA, Blank D et al. (2011). EUROCarbDB: an open-access platform for glycoinformatics. Glycobiology*, 21*(4), 493-502.

Wada Y, Azadi P, Costello et al. (2007). Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology*,* 17(4), 411-422.

Wang H, Luo C, Zhu S, et al. (2017). Serum peptidome profiling for the diagnosis of colorectal cancer: discovery and validation in two independent cohorts. Oncotarget, 8(35), 59376-59386.

Wang Y, Klarić L, Yu X, et al. (2016). The association between glycosylation of immunoglobulin G and hypertension: a multiple ethnic cross-sectional study. Medicine*, 95*(17), 1-11.

Wang Y, Adua E, Russell AC, et al. (2016). Glycomics and its application potential in precision medicine in *Precision medicine in China.* Science/AAAS, Washington, DC, 36 - 39. doi.org/10.1126/science.354.6319.1601-bdoi.org/10.1126/science.354.6319.1601-b

Weber RH and Weber R (2010). *Internet of things* (Vol. 12): Springer.

Wenger E (2014). *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*: Morgan Kaufmann.

Wenk MR (2005). The emerging field of lipidomics. Nat Rev Drug Discov 4(7), 594-610.

Wilkins MR, Gasteiger E, Gooley AA, et al., (1999). High-throughput mass spectrometric discovery of protein post-translational modifications11Edited by R. Huber. J Molecular Biol 289(3), 645-657.

Wilkins MR, Sanchez JC, Gooley AA, et al. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13(1), 19-50.

Wong CM., Crawford RW, Kehler TP., et al. (1984). Application of artificial intelligence to triple quadrupole mass spectrometry (TQMS). IEEE Transactions on Nuclear Science, 31(1), 804-810.

Wu SP, Lin YW, Lai HC, et al. (2006). SELDI-TOF MS profiling of plasma proteins in

ovarian cancer. Taiwan J Obstet Gynecol, 45(1), 26-32.

Wu T, Zheng X, Yang M, et al. (2017). Serum lipid alterations identified in chronic hepatitis B, hepatitis B virus-associated cirrhosis and carcinoma patients. Sci Rep *7,* 1-12.

Xia F, Yang LT, Wang L, et al. (2012). Internet of things. Int J Commun Syst 25(9), 1101-1102.

Xia J, Mandal R, Sinelnikov IV, et al. (2012). MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. Nucleic Acids Res 40(W1), W127-W133.

Xia J, Sinelnikov IV, Han B, et al. (2015). MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res 43(W1), W251-W257.

Xia J and Wishart DS (2010). MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res *38*(suppl 2), W71-W77.

Xu H, Zheng X, Qian Y, et al. (2016). Metabolomics Profiling for Obstructive Sleep Apnea and Simple Snorers. Sci Rep 6.1-9.

Xu J, Chen Y, Zhang R, et al. (2016). Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging. Sci Rep 6, 35010.

Yi L, Shi S, Wang Y, et al., (2016). Serum metabolic profiling reveals altered metabolic pathways in patients with post-traumatic cognitive impairments. Sci Rep *6*. 1-11.

Yost RA, Enke CG, McGilvery DC, et al. (1979). High efficiency collision-induced dissociation in an RF-only quadrupole. Int. J. Mass Spectrom. Ion Phys. 30. 127–36.

Yu CY, Mayampurath A and Tang H (2013). Software tools for glycan profiling. Mass Spectrometry of Glycoproteins: Methods and Protocols, 269-276.

Yu X, Wang Y, Kristic J, et al., Profiling IgG N-glycans as potential biomarker of chronological and biological ages. Medicine, 95(28), e4112.

Zhang ZH, Vaziri ND, Wei F, et al. (2016). An integrated lipidomics and metabolomics reveal nephroprotective effect and biochemical mechanism of Rheum officinale in chronic renal failure. *6*, 22151, 1-18.

Zhang W and Chait BT (2000). ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. 72(11). 2482-2489.

Zhao YY, Miao H, Cheng XL, et al. (2015). Lipidomics: Novel insight into the biochemical mechanism of lipid metabolism and dysregulation-associated disease. Chem Biol Interact 240 220-238.

Zoldoš V, Horvat T and Lauc G (2013). Glycomics meets genomics, epigenomics and other high throughput omics for system biology studies. Curr Opin Chem Biol 17(1), 34-40.

Table 1. Timeline of important events in development of MS technology

Time	Event
1886	E. Goldstein observes canal rays
1897	W. Wien demonstrates that canal rays can be deflected using strong electric and magnetic fields
	J.J Thompson measures the mass-to-charge ratio of electrons
1901	W. Kaufmann uses a mass spectrometer to measure the relativistic mass increase of electrons
1905	J.J. Thompson begins the study of positive rays
1913	J.J. Thompson is able to separate particles of different mass-to-charge ratios
1919	F. Aston constructs the first velocity focussing MS with mass resolution of 130
1931	E.O. Lawrence invents the cyclotron
1934	J. Mattauch and R. Herzog develop the double-focussing MS
1936	A.J Dempster develops the spark ionisation source
1937	F. Aston constructs a mass spectrograph with resolution of 2000
1942	E.O. Lawrence develops the cyclotron for uranium isotope separation
1943	Westinghouse markets its MS and proclaims it to be " A new electronic method for fast, accurate gas analysis"
1946	William Stephens ? presents the concept of a time-of-flight MS
1954	A.J.C. Nicholson proposes a hydrogen transfer reaction that will come to be known as
	the McLafferty rearrangement
1959	Researchers at Dow Chemical interface a gas chromatograph to a MS
1964	British MS society establishes as first dedicated MS society. It holds its first meeting in 1965 in London
1966	F.H. Field and M.S.B. Munson develop chemical ionisation
1968	M. Dole develops electrospray ionisation
1969	H.D. Beckey develops field desorption
1974	M.B. Comisarow and A.G. Marshall develop Fourier Transform Ion Cyclotron Resonance MS
1976	R. MacFarlane et al., develop plasma desorption MS
1984	J.B. Fenn et al., use electrospray to ionize biomolecules
1985	F. Hillenkamp et al., describe and coin the term matrix-assisted laser desorption ionization
1987	K. Tanaka uses the "ultra-fine metal plus liquid matrix method" to ionize intact proteins
1999	A. Makarov presents the Orbitrap MS
2004	Z. Takats et al., develop the Desorption Electrospray Ionisation (DESI) method
2004	D.F Hunt et al., develop Electron Transfer Dissociation (ETD) method
2005	R.B. Cody and J.A Laramee develop the Direct Analysis in Real Time (DART) ion source

Table 2: Timeline of Nobel Prize awarded in MS technology field

Time	Event
1906	J.J. Thomson is awarded the Nobel Prize in Physics “in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases”
1922	F. Aston is awarded the Nobel Prize in Chemistry “for his discovery, by means of his MS, of isotopes, in a large number of non-radioactive elements, and for his enunciation of the whole-number rule”
1939	E.O. Lawrence is awarded the Nobel Prize in Physics for the cyclotron
1989	W. Paul is awarded the Nobel Prize in Physics “for the development of the ion trap technique”
2002	J.B. Fenn and K. Tanaka are awarded the Nobel Prize in Chemistry “for the development of soft desorption ionization methods for MS analyses of biological macro-molecules”