

For reprint orders, please contact: [reprints@future-science.com](mailto:reprints@future-science.com)

## Bioanalysis for precision medicine

Shuang Yang<sup>1</sup>, Zeren Yang<sup>2</sup> & Perry G Wang<sup>\*,3</sup>

<sup>1</sup>Laboratory of Bacterial Polysaccharides, Division of Bacterial, Parasitic and Allergenic Products, Office of Vaccines Research and Review, FDA, Silver Spring, MD 20901, USA

<sup>2</sup>Department of Biological Engineering, University of Maryland, Baltimore County, MD 21236, USA

<sup>3</sup>Division of Food Ingredients, Office of Food Additive Safety, Center for Food Safety and Applied Nutrition, FDA, College Park, MD 20074, USA

\*Author for correspondence: [Perry.wang@fda.hhs.gov](mailto:Perry.wang@fda.hhs.gov)

**“In comparison to the ‘one-size-fits all’ treatment (or same treatment for all patients who have similar diseases), precision medicine (PM) has become an important strategy to identify improved diagnostic and targeted therapeutic treatment”**

First draft submitted: 24 April 2019; Accepted for publication: 9 May 2019; Published online: 28 June 2019

**Keywords:** biomarkers • personalized genomics • precision medicine

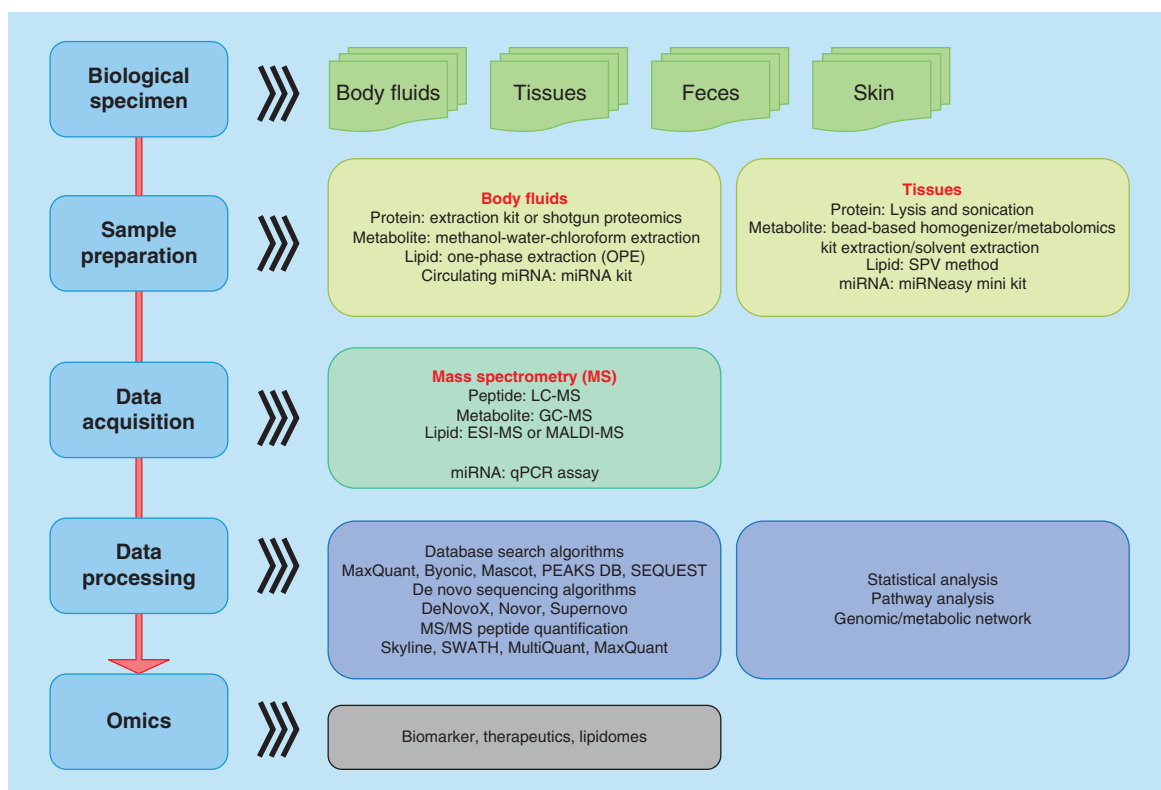
Human genetic variation may lead to different responses from patients who are subjected to the same treatments. The efficacy and potency of a drug may vary because of genetic differences among individuals. In comparison to the ‘one-size-fits all’ treatment (or same treatment for all patients who have similar diseases), precision medicine (PM) has become an important strategy to identify improved diagnostic and targeted therapeutic treatment.

In the future, PM may also inform personalized gene therapy treatments. As there are approximately 19,000–20,000 human protein-coding genes [1], and possibly hundreds of those genes may harbor variations contributing to human illness; hence, it is essential to develop gene-based PM. Gene therapies have already found clinical successes in the treatment of several noncancer diseases, such as severe combined immunodeficiency-X1, hemophilia B and blindness caused by retinitis pigmentosa. Many high mortality diseases including cancers and heart failure have not yet been successfully treated by gene therapies alone. Thus, gene therapy holds promise as an ideal and potentially effective strategy for cancer treatment by incorporating proteomic, lipidomic and metabolomic methodologies.

### Gene expression profiling for personalized genomics

Gene expression can be measured at protein level by western blot, at DNA level by microarray analysis, or at mRNA level by reverse-transcription PCR (RT-PCR). Western blot is a widely-used technique that estimates changes in protein level. DNA microarray analysis, often conducted using a DNA chip or biochip, is a collection of microscopic DNA spots attached to a solid surface. It is used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. A microarray usually contains >1000 nucleic acids to quantify the relative abundance of nucleic acid sequences in a mixture via hybridization and subsequent detection of the hybridization events [2]. The quantitation of gene expression by next-generation DNA sequencing (NGS) and proteomics is accomplished via mass spectrometry (MS). Efforts have been devoted to developing a robust and efficient method for achieving high-throughput and reproducible measurements of global gene expression patterns, such as using p-aminophenyl trimethoxysilane, self-contained and fully integrated microarray, and enhanced specificity of nucleic acid hybridization. RT-PCR and quantitative PCR (qPCR) are techniques applied together to amplify and quantify the production of targeted DNA molecules. RT-PCR/qPCR instrumentation detects signals using nonspecific fluorescent dyes or sequence-specific DNA probes. The recently developed droplet digital PCR system enables the processing of 2 million PCR reactions using conventional TaqMan assays with a 96-well plate high-throughput workflow. This system can accurately measure copy number variation, detect mutant DNA in a 100,000-fold excess of wild-type background and quantify the absolute quantity of circulating DNA from cell-free plasma [3]. The accuracy and flexibility of this platform highlights its usefulness in detecting differences in gene expression to inform PM treatment options.

newlands  
press



**Figure 1.** An overview of 'omics'.

ESI-MS: Electrospray ionization mass spectrometry; MALDI-MS: Matrix-assisted laser desorption/ionization mass spectrometry; SPV: Sulfo-phospho-vanillin.

### Bioanalysis techniques for personalized 'omics'

PM using an individual genomics has been applied to medical diagnostics and the treatment of individual patients. However, other 'omics' methods, such as proteomics, lipidomics and metabolomics can also be used to determine a patient's response to disease and treatment [4]. Advances in analytical and diagnostic techniques have enabled the development of portable medical devices. Such advancements make PM accessible to smaller medical facilities, and thus, expand access to more patients. Bioanalysis techniques are ubiquitous in the development of tools for personalized 'omics' analysis, from sample preparation to data acquisition. Different specimens are also required for the use of various bioanalytical platforms.

An overview of bioanalysis techniques through different stages of sample analysis is summarized in Figure 1. The stages include the collection of biological specimens, preparation of samples, data acquisition, data processing and data output for decision-making. Many types of samples may be analyzed, including body fluids, tissues, feces and skin. For diagnostics and prognostics, the preferred medical practice is to collect samples noninvasively. Therefore, body fluids are more likely to be sampled for disease diagnosis and early biomarker discovery compared with tissues or other specimens. Body fluids (including blood, serum, urine, saliva and lung bronchoalveolar lavage) can be collected and analyzed for their constituent genes, proteins, lipids and metabolites. In general, we can extract proteins, metabolites, lipids/glycolipids and circulating miRNAs using methods specific to a given fluid sample. For example, serum proteins can be extracted using lysis buffer or 8 M urea, followed by tryptic digestion; urine proteins are isolated using desalting kits and saliva proteins can be extracted using trichloroacetic acid (TCA)/acetone. It is noteworthy that there are many innovative methods for analyzing specific subsets of proteins, such as the solid-phase chemoenzymatic method for protein *N*- and *O*-glycosites and occupancy [5,6], *N*-glycans, acetylation, phosphorylation and ubiquitylation. Metabolites from body fluids are often isolated using methanol-water-chloroform phase extraction, in which the metabolites dissolve in the polar phase, proteins in the interphase and other molecules such as lipids in the nonpolar phase [7]; global lipids can be extracted using

one-phase extraction (OPE) and circulating miRNAs are isolated using miRNA kits, such as Applied BioSystem's TaqMan miRNA Reverse Transcription Kit or Qiagen's miRNeasy Mini Kit.

Isolation of molecules from tissue specimens is more challenging. Tissues are typically dissected then processed by an ultrasonic homogenizer. One promising method uses mixed solvents to simultaneously extract proteins, metabolites and lipids [7]. After tissue homogenization, samples are dissolved in a mixture solution consisting of methyl tert-butyl ether, methanol and water. The mixture is centrifuged to form three layers: an organic phase (methyl tert-butyl ether-methanol) at the top, a polar phase in the middle (water) and a pellet. Lipids partition into the organic phase and the pellet contains proteins; whereas, the polar phase contains primary metabolites (minority) and secondary metabolites (majority). Generally, PM requires rapid and robust processing of a sample. Thus, a high-throughput platform is beneficial for sample preparation and analysis. For example, a platform that integrates a bead-based homogenizer with a solvent extraction protocol is suitable for the rapid and reproducible extraction of metabolites from multiple tissues [8]. A platform with a capacity of up to 384-AutoTips has the potential to analyze the glycoprotein and glycan components of many clinical specimens [9]. Lipids in the tissue can be extracted using a colorimetric sulfo-phospho-vanillin method performed in a 96-well microplate throughout the entire assay, providing advantages over other approaches such as automation and accurate quantification [10]. Meanwhile, miRNAs from formalin-fixed paraffin-embedded (FFPE) tissue can also be isolated by commercially available miRNA kits, including the RNeasy FFPE kit, absolutely RNA FFPE kit, high pure RNA micro kit, PureLink RNA isolation kit and RecoverAll total nucleic acid isolation kit [11].

Data acquisition is one of the critical steps for analyzing biological changes of specimens. With the exception of miRNAs, which are usually measured by RT-PCR/qPCR [12], other molecules can be analyzed by MS-based techniques. This is because of advanced progress in MS instrumentation rendering MS a powerful tool for the in-depth analysis of different types of molecules. MS can be readily interfaced with different front-end sample inlet techniques, for example, GC for metabolites, LC for peptides or different ionization methods such as ESI for lipids and peptides, or matrix-assisted laser desorption/ionization for lipids and glycans. MS shows great potential for diagnostic applications using GC-MS based metabolite profiling, offering a good combination of sensitivity and specificity over nuclear magnetic resonance (NMR) [9]. LC-MS based platforms have been widely established for the analysis of peptides, including post-translationally modified peptides, which enable the identification of >10,000 peptides with mass accuracy as low as the p.p.b. level [13]. This allows identification over 20,000 human proteins, making it possible to analyze the complete human proteome [14].

### Bioinformatics for personalized 'omics'

A variety of software programs have been developed for data processing. Genomic data can be analyzed using commercial software, such as ingenuity pathway analysis, from which the genomic network can be established. Metabolic networks can also be analyzed using ingenuity pathway analysis to explore the roles of upstream gene regulators. Discovery-phase biomarker studies typically generate a tremendous number of MS spectra that require reliable software for identification and quantification. Two major types of algorithms have been used for the analysis of MS spectrum data: database search algorithms (DSAs) [15] and *de novo* sequencing algorithms (DnSAs) [16]. DSAs generally use known databases including protein, metabolite or lipid libraries (e.g., NCBI or Uniprot). For example, if the acquired MS spectra are intact glycopeptides, protein and glycan databases can be used to match the spectra through library matching. In this workflow, intact glycopeptides are searched using Byonic to study sialic linkages in human serum. DSAs include MaxQuant, Mascot, PEAKS DB and SEQUEST; DnSAs use tandem mass spectra to determine peptide amino acid sequences. DnSAs include PEAKS, DeNovoX, Novor and Supernovo.

Disease-related changes in protein abundance can be determined using stable isotope labeling and label-free methods. To compare the relative abundance of proteins in biological samples, it has become a routine practice to label tryptic peptides using stable isotope tags such as isobaric tag for relative and absolute quantitation, isotope-coded affinity tag or tandem mass tag. The isotope tag allows for the simultaneous injection of multiple samples for LC-MS analysis, maintaining the same conditions for LC separation, ionization and detection. Similarly, *N*-glycans can be quantified using recently developed isobaric tags, enabling the quantitative analysis of up to four biological samples in an analytical set [17,18]. Protein quantification can be also achieved through peptide tandem mass spectra generated by LC-MS/MS. This strategy has no limitation on how many biologicals samples can be analyzed. The MS spectra can be analyzed using several software programs including Skyline, MultiQuant and MaxQuant.

### 'Omics': biomarker, therapeutic & lipidome

The analysis of individual 'omics' can be used to create profiles of genes, proteins, metabolites and lipids, which in turn, may be used as part of a PM approach to the diagnosis, treatment and clinical management of diseases. Since early detection of many diseases can substantially reduce mortality, the identification of biomarkers from patients can be implemented through 'omics' analysis. For example, biomarkers for personalized oncology have provided more accurate and specific prediction of disease by the use of molecularly targeted therapeutics that benefit the individual patient [19]. Meanwhile, the human plasma lipidome, consisting of over 600 distinct molecular species, is correlated with certain clinical outcomes. For example, the lipid profile of an individual could indicate increased breast cancer risks that in turn may be associated with dietary fat consumption [20].

### Conclusion

Personalized 'omics' profiling reveals dynamic and medical phenotypes. Therefore, the one-size-fits-all therapeutic strategy has limitations with respect to efficacy and benefit to individual patients. PM could offer accurate and specific treatment through 'multi-omics' analysis of specimens derived from individual patients. Various bioanalytical techniques have been developed across 'omics' fields, serving as the keys to uncover the onset and occurrence of diseases. The continued development of bioanalytical techniques and PM devices have the potential to significantly improve human health and quality of life.

### Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

### References

1. Ezkurdia I, Juan D, Rodriguez JM *et al.* Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Gen.* 23(22), 5866–5878 (2014).
2. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. *Curr. Protoc. Mol. Biol.* 101(1), 22.1.1–22.1.11 (2013).
3. Hindson BJ, Ness KD, Masquelier DA *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83(22), 8604–8610 (2011).
4. Duarte TT, Spencer CT. Personalized proteomics: the future of precision medicine. *Proteomes* 4(4), 29 (2016).
5. Zhang H, Li X-J, Martin DB, Aebersold R. Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21(6), 660 (2003).
6. Yang S, Onigman P, Wu WW *et al.* Deciphering protein *O*-glycosylation: solid-phase chemoenzymatic cleavage and enrichment. *Anal. Chem.* 90(13), 8261–8269 (2018).
7. Sapcariu SC, Kanashova T, Weindl D, Ghelfi J, Dittmar G, Hiller K. Simultaneous extraction of proteins and metabolites from cells in culture. *MethodsX* 1, 74–80 (2014).
8. Römisch-Margl W, Prehn C, Bogumil R, Röhrling C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8(1), 133–142 (2012).
9. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* 1(1), 387 (2006).
10. Cheng YS, Zheng Y, Vanderghelyst JS. Rapid quantitative analysis of lipids using a colorimetric method in a microplate format. *Lipids* 46(1), 95–103 (2011).
11. Doleshal M, Magotra AA, Choudhury B, Cannon BD, Labourier E, Szafranska AE. Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* 10(3), 203–211 (2008).
12. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3(6), 1101 (2008).
13. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26(12), 1367 (2008).
14. Kim M-S, Pinto SM, Getnet D *et al.* A draft map of the human proteome. *Nature* 509(7502), 575 (2014).
15. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74(20), 5383–5392 (2002).

16. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18), 3551–3567 (1999).
17. Yang S, Yuan W, Yang W *et al.* Glycan analysis by isobaric aldehyde reactive tags and mass spectrometry. *Anal. Chem.* 85(17), 8188–8195 (2013).
18. Yang S, Wang M, Chen L *et al.* QUANTITY: an isobaric tag for quantitative glycomics. *Sci. Rep.* 5, 17585 (2015).
19. Kalia M. Biomarkers for personalized oncology: recent advances and future challenges. *Metab. Clin. Exp.* 64(3), S16–S21 (2015).
20. Bougnoux P, Giraudeau B, Couet C. Diet, cancer, and the lipidome. *Cancer Epidemiol. Biomarkers Prev.* 15(3), 416–421 (2006).

