JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
PUBLISHED BY ELSEVIER ON BEHALF OF THE AMERICAN
COLLEGE OF CARDIOLOGY FOUNDATION

VOL. 68, NO. 25, 2016
ISSN 0735-1097/$36.00
http://dx.doi.org/10.1016/j.jacc.2016.10.031

FOCUS SEMINAR: GENETICS

STATE-OF-THE-ART REVIEW

Proteomics Research in Cardiovascular Medicine and Biomarker Discovery

Maggie P.Y. Lam, PhD,[a] Peipei Ping, PhD,[a] Elizabeth Murphy, PhD[b]

ABSTRACT

Proteomics is a systems physiology discipline to address the large-scale characterization of protein species within a biological system, be it a cell, a tissue, a body biofluid, an organism, or a cohort population. Building on advances from chemical analytical platforms (e.g., mass spectrometry and other technologies), proteomics approaches have contributed powerful applications in cardiovascular biomedicine, most notably in: 1) the discovery of circulating protein biomarkers of heart diseases from plasma samples; and 2) the identification of disease mechanisms and potential therapeutic targets in cardiovascular tissues, in both preclinical models and translational studies. Contemporary proteomics investigations offer powerful means to simultaneously examine tens of thousands of proteins in various samples, and understand their molecular phenotypes in health and disease. This concise review introduces study design considerations, example applications and use cases, as well as interpretation and analysis of proteomics data in cardiovascular biomedicine. (J Am Coll Cardiol 2016;68:2819–30) Published by Elsevier on behalf of the American College of Cardiology Foundation.

Over the past 20 years, there has been accelerating growth in the application of proteomics in cardiovascular biomedicine, building on pioneering studies with translational significance (1,2) and mechanistic insights (3–8). The majority of human genes functions to create proteins, which are the molecular workhorses that carry out virtually all metabolic, signaling, and physiological functions in life. Many human diseases may be characterized by the elicited changes in proteome configurations. In complex late onset diseases in particular, considerable environmental components exist, such that individual genetic variants are often poorly predictive of disease states. Telltale molecular signatures of acquired cardiovascular diseases may instead manifest through intermediate phenotypes, including transcript and protein abundance. Beyond what may be learned from gene and transcript information alone, molecular changes at the protein layer, in particular, can provide independent insights into disease mechanisms in the following 3 areas.

First, it has gradually emerged from large-scale studies that transcript and protein changes in a system may, at times, correspond poorly—with transcript levels explaining as little as 10% to 30% of variations in protein abundance. Although the exact contribution of transcripts to protein-level abundance is debated, it is evident that in the heart and other organs, a large number of post-transcriptional modulators can alter disease-driver protein expression and function during pathogenesis, without changes in transcript abundance (9,10). These modulators include components of the nonsense-mediated decay pathway, long noncoding RNAs, and microRNAs (11–14).

Second, plasma proteins provide an accessible readout of the status of potentially all tissues, and

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

From the [a]NIH BD2K Center of Excellence and Department of Physiology, Medicine and Bioinformatics, University of California, Los Angeles, California; and the [b]Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland. This work was supported by NIH K99-HL127302 to Dr. Lam; NIH R37-HL063901 and NIH U54-GM114833 to Dr. Ping; and NIH ZIA-HL006059 and NIH ZIA-HL002066 to Dr. Murphy.

Manuscript received August 3, 2016; revised manuscript received October 20, 2016, accepted October 21, 2016.

ABBREVIATIONS AND ACRONYMS

AP-MS = affinity purification-mass spectrometry

FDR = false discovery rate

MS = mass spectrometry

PDH = pyruvate dehydrogenase

PTM = post-translational modification

RNA-seq = ribonucleic acid-sequencing

TMT = tandem mass tag

are the sources of many current biomarkers in use (Table 1). Clinically useful biomarkers can include circulating proteins that have been secreted or leaked directly into the plasma from resident cells (e.g., myocytes) following diseases or injuries (15), and are thus spatially uncoupled from the transcript change at the tissue of origin. The tissues of origin of circulating proteins may be unknown, inaccessible, or invasive to procure, making it impractical to hunt for these potential biomarkers via transcript measurements, instead of protein measurements.

Third, disease processes may be mediated not by the altered abundance of gene products, but rather by other functional parameters of the proteome, including protein post-translational modifications (PTMs), protein-protein interactions, and protein degradation, which take place after the protein molecules are synthesized and cannot be predicted from genetic information a priori.

Proteomics technologies allow researchers to measure protein function on a large scale, and hence interrogate the molecular layer that closely abuts physiological phenotypes (16). However, the complexity of the human proteome also presents a daunting analytical challenge for large-scale characterization. For example, circulating proteins in human plasma, found at a concentration of ~70 mg/ml, are estimated to comprise at least 10,000 distinct protein species (17) over a concentration range of >10 orders of magnitude (10-billion-fold differences) (18). Approximately 3,000 proteins in plasma may be viewed as classical resident plasma proteins, which tend to occupy the higher end of the abundance spectrum. These resident proteins include extremely abundant species, such as albumin, immunoglobulins, transferrins, and fibrinogens, which are found in the milligram/milliliter concentration range and account for >90% of plasma protein content. Other plasma proteins come from secretion or leakage from various organs before being diluted into the 5 liters of peripheral blood, and are found in the microgram/milliliter to nanogram/milliliter range. Very low-abundance proteins, including interleukin-6 and -12, tumor necrosis factor-alpha, and other cytokines, circulate in the picogram/milliliter range, and are typically masked by higher-abundance proteins in untargeted proteomics analyses (Figure 1). Adding to this complexity, plasma proteins may be proteolytically processed, creating numerous degradation product species, and repetitive sequences in proteins can give a spurious appearance of higher abundance. These challenges are compounded by the absence of a protein equivalent of the polymerase chain reaction, such that amplification of minute samples is not possible.

In the following 3 segments of this review, we will discuss 3 aspects of proteomics applications in cardiovascular biomedical research: 1) study design for biomarker research and discovery, with discussions on approaches to circumvent analytical challenges; 2) characterization of multidimensional protein parameters in disease mechanism research; and 3) interpretation and validation of proteomics data or methodologies.

TABLE 1 Selected Protein Biomarkers for Cardiovascular Diseases

Protein Biomarker	Abbreviation	Disease Relevance	Required Assay Sensitivity (Estimate)	Discovery Period*	Ref. #
Apolipoprotein A-I	APOA	Cardiovascular event risk	~1 mg/ml	Mid-1980s	(96)
Apolipoprotein B	APOB	Cardiovascular event risk	~1 mg/ml	Mid-1980s	(97)
B-type natriuretic peptide	BNP	Heart failure, acute coronary syndrome	~100 pg/ml	Early 2000s	(98)
C-reactive protein	CRP	Cardiovascular event risk	~10 μg/ml	Late 1990s	(99)
Creatine kinase-myocardial band	CK-MB	Acute myocardial infarct, myocardial necrosis	~1 ng/ml	1960s to 1970s	(100)
Cystatin-C	CST3	Cardiovascular event risk	~1 μg/ml	2000s	(101)
Fibrinogen	FBN	Cardiovascular event risk	~1 mg/ml	1980s	(99)
Lipoprotein-associated phospholipase A2	Lp-PLA2	Coronary heart disease risk	~100 ng/ml	2000s	(102)
Myeloperoxidase	MPO	Ischemic heart disease; acute coronary syndrome	~10 ng/ml	1980s	(103)
Myoglobin	MYO	Myocardial infarction, necrosis	~10 ng/ml	Late 1970s	(104)
Serum amyloid A	SAA	Coronary artery disease	~10 μg/ml	1990s	(105)
Troponin I	cTnI	Myocardial injury, myocardial infarction	~10 pg/ml	1970s to 1990s	(106)
Troponin T	cTnT	Myocardial injury, myocardial infarction	~10 pg/ml	1970s to 1990s	(107)

*Periods of initial discovery or assay development (108,109).