8 Proteomics and Metabolomics in Cardiovascular Medicine

ROBERT GERSZTEN

NOVEL TECHNOLOGIES IN THE IDENTIFICATION OF BIOMARKERS

The limitations of currently available biomarkers for screening or prognostic use underscore the importance of identifying "uncorrelated" or "orthogonal" biomarkers associated with novel disease pathways. Most current cardiovascular biomarkers have derived from extensions of targeted physiologic studies investigating known pathways such as tissue injury, inflammation, or hemostasis. By contrast, emerging technologies now enable the systematic, unbiased characterization of variation in proteins and metabolites associated with disease conditions.

INTRODUCTION TO PROTEOMICS AND METABOLOMICS

Of the emerging platforms for biomarker discovery, perhaps none have garnered more recent attention than proteomics and metabolomics. Proteomics aims to catalogue the entire protein products of the human genome. By contrast, metabolomics attempts to systemically capture smaller biochemical compounds, including simple amino acids and related amines, as well as lipids, sugars, nucleotides, and other intermediary metabolites. Although still in their infancy with respect to other approaches, proteomics and metabolomics offer insight into the full complexity of a given disease (Fig. 8.1). Because proteins and metabolites are downstream of genetic variation and transcriptional changes, they provide instantaneous "snapshots" of the state of a cell or organism. They can change rapidly in response to environmental stressors such as exercise or directly by the ingestion of foods or other compounds. Although the effects of catecholamines and natriuretic peptides on cardiovascular homeostasis are well-established, a growing body of literature suggests unanticipated roles of small proteins and metabolites in the control of biologic functions such as blood pressure and energy homeostasis.[1] Thus metabolomics and proteomics may not only identify novel biomarkers but also provide information on biology and highlight potential therapeutic targets.

The term *proteome* was coined in the 1990s with the increasing realization that, although all cells of a given organism contain an equivalent genomic content, their protein content does not represent all possible proteins that the genome can express. Selective gene expression during development and differentiation and in response to external stimuli results in each cell producing only a subset of the encoded proteins at any given time. One can speak not only of the general human proteome but also more specifically about the proteome of tissues such as the heart, of specific cells such as cardiac myocytes, and even of subproteomes that correspond to particular organelles or biologic compartments, such as mitochondria.

The proteome thus provides information beyond the messenger RNA (mRNA) expression profile of a particular genome. Studies suggest that gene expression often correlates poorly with protein levels. Protein expression depends not only on transcription but also on mRNA stability and rates of protein synthesis and degradation, so the presence or absence of mRNA may not accurately reflect levels of the corresponding protein. Following transcription and translation, proteins may undergo one or more of dozens of potential post-translational modifications (such as phosphorylation, glycosylation, acetylation, or sulfation) at multiple sites which modulate protein function. Subsequent enzymatic and nonenzymatic alterations greatly expand the number of simultaneously existing protein species.

When compared with proteomics techniques, metabolomics technologies focus on smaller compounds, generally less than 2 kDa in size. Metabolites are usually easily separated from protein constituents by simple extraction techniques and precipitation and removal of the proteins. As early as the 1970s, Arthur Robinson and Linus Pauling postulated that the quantitative and qualitative pattern of metabolites in biologic fluids reflected the functional status of the complex biologic system from which they were derived. The term "metabolic profiling" was introduced to describe data obtained from gas chromatographic analysis of a patient sample. This emerging approach to quantitative metabolic profiling of large numbers of small molecules in biofluids was ultimately termed "metabonomics" by Nicholson and colleagues and "metabolomics" by others. Recently, more focused analyses of specific metabolite families or subsets have given rise to new terms such as "lipidomics."

In terms of applications to human diagnostics, seminal studies of inborn errors of metabolism in infants have served as a key springboard. Millington and colleagues pioneered the use of mass spectrometry (MS)-based methods for monitoring fatty acid oxidation, as well as organic and selected amino acids.[2] Their work culminated in neonatal screening for metabolic disorders, thereby enabling the identification of infants with fatty acid oxidation disorders, organic acidemias, and aminoacidopathies. In certain situations, rapid identification of these disorders triggers intervention in the form of dietary modulation, conferring therapeutic benefits. A global metabolomic or proteomic analysis of more indolent, complex diseases such as atherosclerosis might similarly spotlight pathways for dietary or drug modulation.

ANALYTIC CHALLENGES FOR PROTEOMICS AND METABOLOMICS

The many classes of proteins and chemicals present analytic challenges, particularly as applied to searching for biomarkers in blood. Many different types of cells contribute to the plasma proteome and metabolome, thus increasing their complexities and presenting
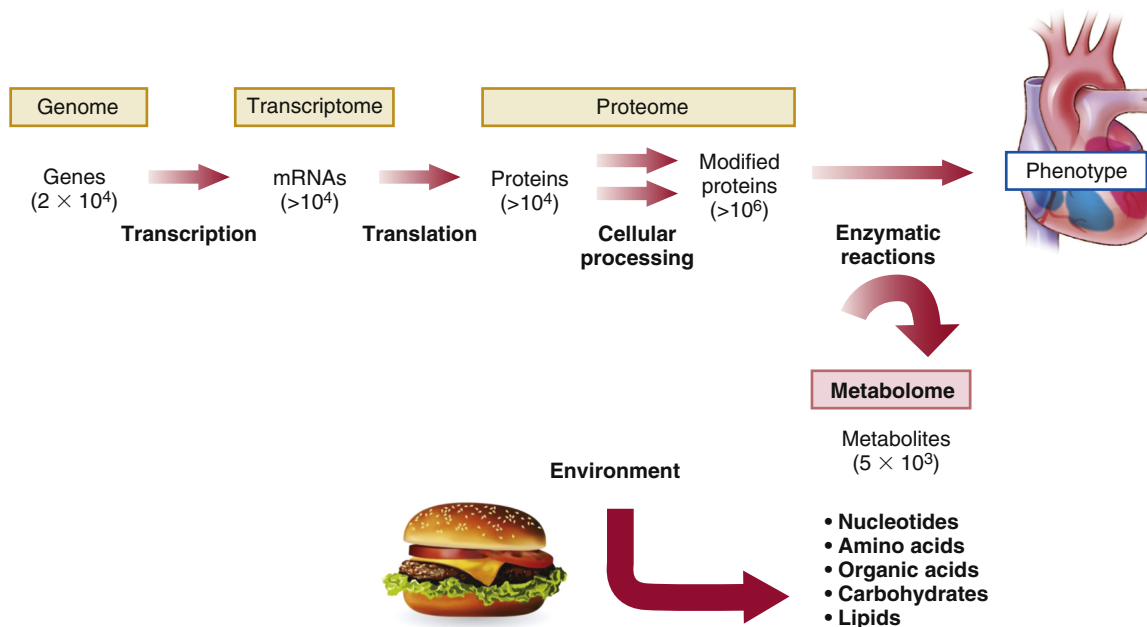
FIGURE 8.1 The conceptual relationship of the genome, transcriptome, proteome, and metabolome. Informational complexity increases from genome to transcriptome to proteome. The estimated number of entities of each type of molecule in humans is indicated in parentheses.

challenges to interpretation of the data that emerge. In the case of the blood proteome, the 22 most abundant proteins, including albumin and the immunoglobulins, account for 99% of the total proteome mass (Fig. 8.2). Many of the biologically interesting molecules relevant to human disease occur in low abundance. Cardiac markers such as troponin circulate in the nanomolar range, insulin in the picomolar range, and tumor necrosis factor in the femtomolar range. Plasma contains tens of thousands of unique protein species in concentrations spanning a range of more than 10 orders of magnitude. Indeed, some suggest that the plasma proteome might encompass the entire set of human polypeptide species resulting from splice variants and post-translational modifications because the protein content of plasma unexpectedly includes proteins of all functional classes and from apparently all cellular localizations. Many low-abundance proteins in plasma are intracellular or membrane proteins that are present in plasma as a result of cellular turnover. There is also an increasing appreciation of gut-derived small molecules and peptides (i.e., from the "microbiome") present in blood that appear to activate signaling pathways in leukocytes or metabolically active tissues.[3] Recent estimates suggest that the human metabolome consists of fewer molecular entities than the human proteome[4] and thus may be somewhat more tractable to analyze and systematize than the human proteome.

Several features contribute critically to the success of proteomic or metabolomic technologies. First, the technique must have the capability of identifying a wide breadth of proteins or metabolite analytes within complex biologic samples across a broad range of physical characteristics, including size and charge. Second, the technologies must be sensitive enough to probe the proteome or metabolome to adequate "depths"—that is, to provide resolution of biologically active compounds of the lowest abundance. Frequently, the least abundant entities play critical regulatory roles in the response to physiologic stressors. Third, tools must also work across a broad dynamic range, a notion underscored in Figure 8.2—they must be able to simultaneously identify both more abundant and less abundant proteins in the same complex mixture. Unfortunately, many analytic techniques apply well only across concentrations of several orders of magnitude. Finally, the ideal technology should be stable and reproducible, an attribute necessary for minimizing artifacts during initial discovery, validation, and testing for clinical applications.

Robust, searchable databases for validation of identified proteins or metabolites represent an increasingly crucial support for biomarker discovery. The scope of investigation addressable by these techniques has widened immeasurably since completion of the Human Genome Project. At present, the human databases are the largest and easiest to use, which will help accelerate translational investigation. Genomic databases collectively provide a catalog of all known or theoretical proteins expressed in organisms for which databases exist. Software that can search through databases for identification of candidates has proven essential to interpretation of the data; much of this software is available on the Internet. Collaborative efforts have recently begun to catalog both the human proteome and the plasma metabolome.

OVERVIEW OF THE DISCOVERY PROCESS

Figure 8.3 summarizes the essential elements of the discovery approach by using a proteomics experiment as an example. Biologic samples consist of a complex mixture containing intact and partially degraded proteins and metabolites of various molecular weights, modifications, and solubility. The chance of identifying proteins or metabolites in a mixture increases as the complexity of the mixture decreases. As suggested by Liebler, the problem of complexity and how to deal with it resembles the process of printing a book. Printing all the words on a single page could be accomplished quickly, but the resulting page would be illegibly black with ink; dividing the text into multiple pages reduces the complexity to reveal organized text. Samples can be analogously enriched for certain components through fractionation or affinity depletion columns, but all preparative procedures—including solubilization, denaturation, and reduction processes—should be compatible with the constraints of subsequent analysis steps. The quest to reduce complexity requires careful balance against the possibility that each additional step might also introduce undesired protein or metabolite modifications or loss.

Several analytic techniques can serve to identify metabolites or proteins, although MS instrumentation offers an unrivaled ability to provide several layers of complementary information, which has benefited tremendously from whole-genome analysis and the genomics revolution. MS provides accurate mass detection of peptides from proteolytic digests of complex protein mixtures or small metabolites derived from tissues or blood. The set of peptide or metabolite mass measurements can be searched in databases to obtain definitive identification of the parent proteins or metabolites of interest. Favorably compared against other proteomics and metabolomics technologies,

NORMAL RANGE ABUNDANCIES
LOG$_{10}$ CONCENTRATION IN pg/mL
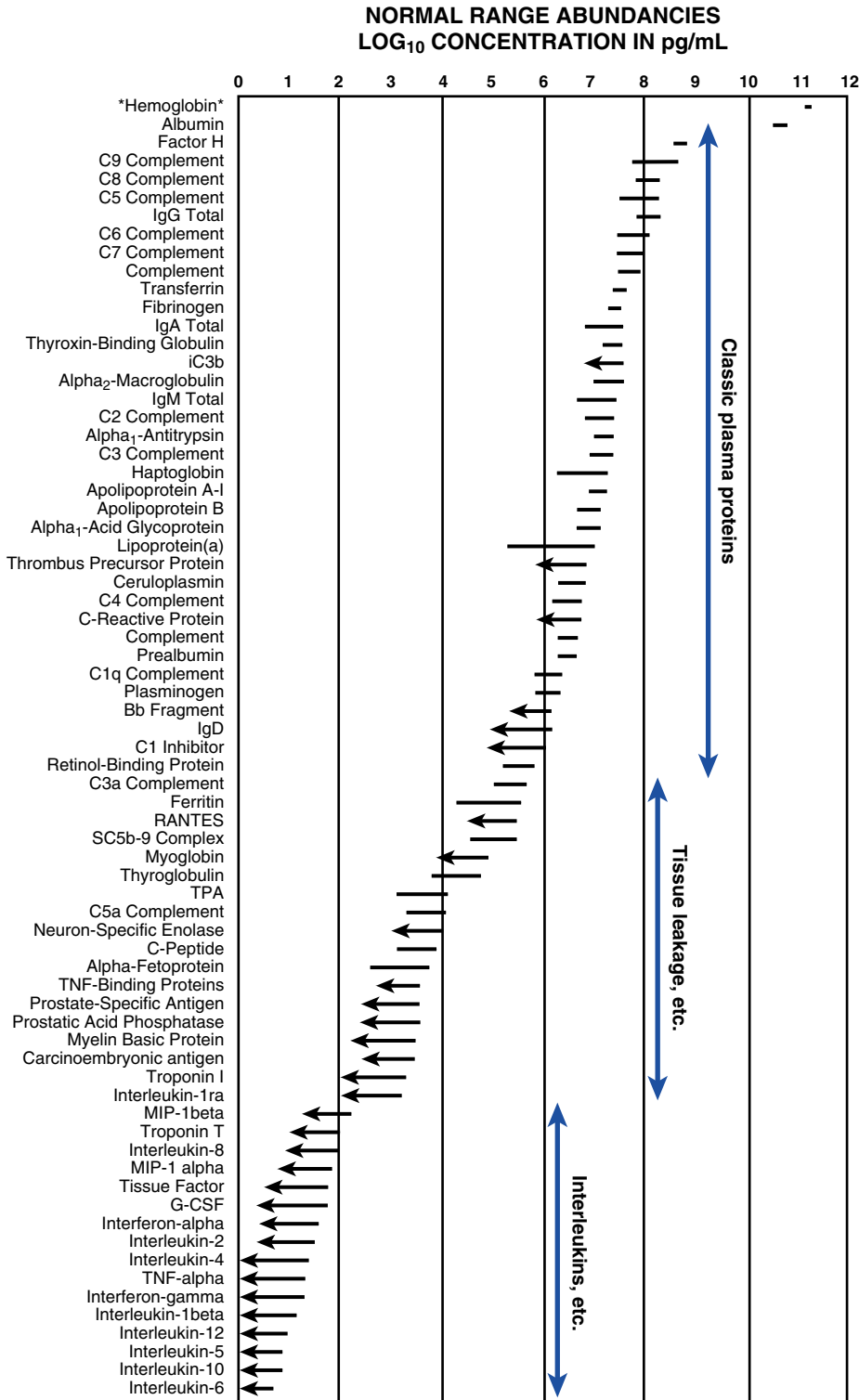
FIGURE 8.2 Reference concentration for representative protein analytes in plasma. Protein abundance is plotted on a log scale spanning 12 orders of magnitude. When only an upper limit is quoted, the lower end of the interval line shows an *arrowhead*. The classic plasma proteins are clustered to the left (high abundance), the tissue leakage markers (e.g., enzymes, troponins) are clustered in the center, and the cytokines are clustered to the right (low abundance). *G-CSF*, Granulocyte colony-stimulating factor; *MIP*, macrophage inflammatory protein; *RANTES*, regulated on activation, T cell expressed and secreted; *TNF*, tumor necrosis factor; *TPA*, tissue plasminogen activator. (From Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 2003;2:50.)

MS offers high sensitivity and amenability to automation, thus promoting high-throughput processing. MS has a wide range of applicability and not only detects metabolites and proteins but also characterizes any post-translational modifications.

Mass spectrometers are composed of modular elements, including an ion source, mass analyzer, and a detector/recorder (Fig. 8.4). MS instruments are classified according to the ionization source and mass analyzer used, but all process samples as gas-phase ions, the movements of which are precisely measured within an electromagnetic field. An ion source generates these gas-phase ions from the analyte through a variety of available techniques, from either the solid state by matrix-assisted laser desorption/ionization (MALDI) or directly from the liquid phase by electrospray ionization (ESI). A coupled chromatographic separation step fractionates complex sample mixtures before ESI spectroscopic analysis. The gas-phase ions then enter the mass analyzer, which resolves the peptides based on their mass-to-charge (m/z) ratio. Examples of commonly used mass analyzers include the quadrupole mass filter, ion trap mass analyzer, and time-of-flight mass analyzer. Finally, the detector records the ions via an electronic multiplier and records ion intensity versus the m/z value to create the resulting MS spectra.

These technologies can be used to characterize biologic fluids either in a targeted manner or in a pattern discovery manner. In the

former, the investigator targets a predefined set of analytes to be quantitated. For example, libraries of metabolites can be purchased and their chromatographic and MS characteristics determined empirically by "spiking" reference standards into plasma. Endogenous metabolites can then be quantified based on the information ascertained from the known standards. The targeted approach now readily permits assay of several hundred metabolites in as little as tens of microliters of plasma. In the pattern discovery experiment, by contrast, the investigator confronts a complex pattern of peaks, many of which are anonymous—the molecular identities of the species that give rise to the peaks are not generally known. Although the targeted approach is more limiting, the analysis is more straightforward because the analytes yielding the signals are already known. The untargeted or "fingerprint" approach has less inherent bias, but unambiguous identification of the peaks can prove laborious and difficult. In clinical samples, considerable care must be taken to rule out spurious associations—for example, confounding related to drug treatment.

APPLICATIONS OF MASS SPECTROMETRY–BASED DISCOVERY TO CARDIOMETABOLIC DISEASE

In an initial proof-of-principle study using a targeted metabolite profiling approach, Newgard and colleagues profiled obese versus lean

FIG. 8.3 Overview of a proteomics experiment. *MS*, mass spectrometry.

humans to gain a broad understanding of the metabolic and physiologic differences in these two disparate groups. Their studies identified a branched-chain amino acid signature that correlated highly with the metrics of insulin resistance while functional studies in model systems have highlighted a role for this pathway in disease pathogenesis.[5] Complementary studies in two large population-based cohorts demonstrated that branched-chain and aromatic amino acid concentrations associate significantly with incident type 2 diabetes up to 12 years before the onset of overt disease. Adjustment for established clinical risk factors did not substantially attenuate the strength of these associations. Furthermore, the branched-chain amino acid signature also predicts atherosclerosis even after adjusting for the metrics of insulin resistance and diabetes. For those in the top quartile of branched-chain amino acid levels, the odds for development of cardiometabolic disease exceeded any single-nucleotide polymorphism identified to date. Taken together, these findings have disclosed dysregulation of amino acid metabolism very early in the development of cardiometabolic diseases. Ongoing studies are examining the relative genetic versus environmental contributions to these findings. A recent report suggests that genetic variation in enzymes in branched-chain amino acid metabolism associate with both circulating amino acid levels and with diabetes in multiple large human cohorts, suggesting that this class of compounds also contributes to disease pathogenesis.[6] Such mendelian randomization analyses are now being performed for thousands of circulating proteins and metabolites in humans to assess for potential causal roles in cardiometabolic disease pathogenesis (see also Chapters 7 and 10).

In a translational study using nontargeted liquid chromatography–MS–based metabolite profiling applied to cardiovascular disease, Wang and associates first profiled the plasma of 75 individuals from a hospital-based cohort who experienced a myocardial infarction, stroke, or death in the ensuing 3 years and 75 age- and sex-matched controls who did not. Of 18 analytes that differed significantly between cases and controls, 3 demonstrated significant correlations among one another, thus suggesting a potential common biochemical pathway. Using complementary analytic methods, these metabolites were identified as betaine, choline, and trimethylamine-*N*-oxide, all metabolites of dietary phosphatidylcholine. Dietary supplementation of choline was sufficient to promote atherosclerosis in mice, and suppression of the intestinal bacteria responsible for the conversion of phosphatidylcholine to choline inhibited this atherogenesis. In addition to reinforcing the interaction between diet, gut bacteria, and the metabolome, this study demonstrated how metabolomic biomarker discovery can elucidate novel pathways to disease.[7]
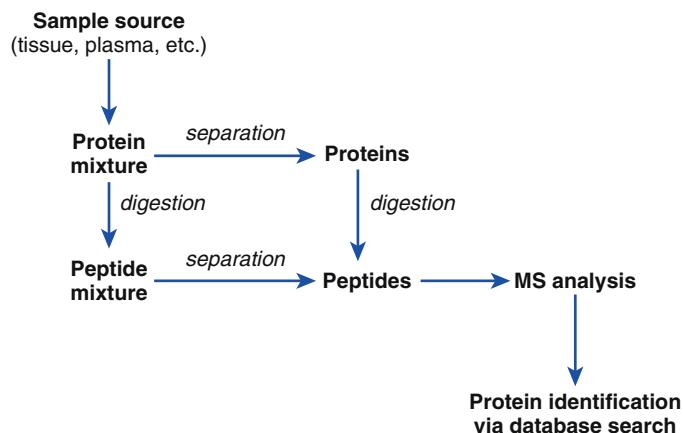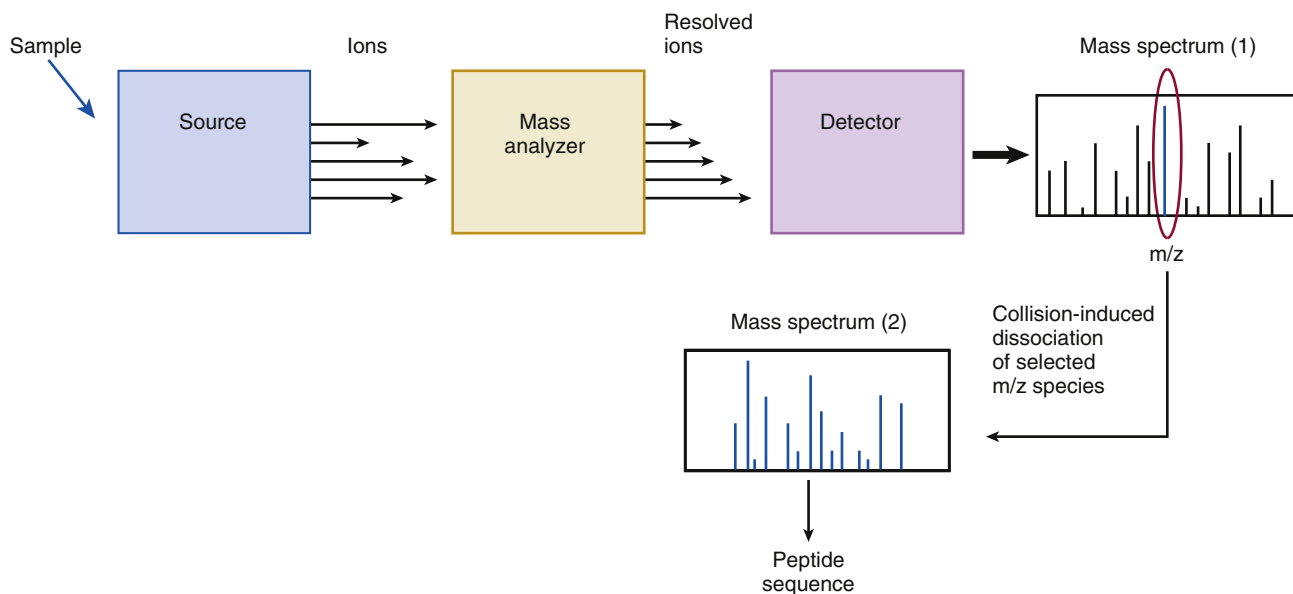

FIG. 8.4 Schematic of tandem mass spectrometry. *m/z*, mass-to-charge ratio.

FUTURE DIRECTIONS IN BIOMARKER DISCOVERY

In addition to their use in neonatal screening, MS-based assays of small molecules are increasingly used in clinical chemistry. For example, robust clinical workflows measure vitamins, sex hormones, and drug levels. By contrast, high-throughput proteomic approaches that account for the complex protein constituents of human plasma are less mature, even for research purposes. Two emerging approaches to address limitations of present tools in analyzing the blood proteome incorporate new classes of affinity (e.g., binding) reagents coupled to high-throughput readouts based on established DNA technologies. One new technique uses aptamer-based technologies to selectively probe the plasma proteome. Aptamers are small RNA or single-stranded DNA nucleic acids that can bind with great specificity to targeted proteins and related cell targets. Aptamers can be incubated with plasma and, using standard bead immobilization techniques, ultimately separated into bound and unbound fractions. Once eluted, these bound aptamers (reflecting their accompanying protein targets) are hybridized to microarrays with complementary single-stranded DNA probes to quantify the specific fluorescent tags. A second high-throughput technique conjugates antibodies with nucleotide "bar codes" which can be quantified by polymerase chain reaction (PCR) or next-generation DNA sequencing. The DNA readout mitigates interferences that severely limit the use of multiple enzyme-linked immunosorbent assays (ELISAs) performed in the same plasma sample. A recent example of the aptamer technique includes an analysis in the Heart and Soul and HUNT-3 studies that measured 1130 proteins.[8] Of these, nine proteins were identified as being predictive of vascular risk, and a risk score derived from these nine proteins was able to separate high from low risk. These new high-throughput techniques are now enabling analyses in very large, well-phenotyped, human cohorts that hold unprecedented promise to illuminate new biomarkers and cardiovascular disease pathways. Of course, the generalizability of these findings requires further work to validate clinical utility in terms or early diagnosis or reclassification and, perhaps most importantly, whether the new approaches can identify novel therapeutic targets.

Identification of new biomarkers for cardiovascular disease depends on the complementary power of genetics, transcriptional profiling, proteomics, and metabolomics. The clinical usefulness of new biomarkers will require rigorous evaluation of their ability to improve the prediction of risk or to direct and monitor management in an individual, the ultimate goal of personalized medicine (see also Chapter 10). In addition to risk biomarkers, diagnostic biomarkers could help in making challenging acute diagnoses such as reversible myocardial ischemia, pulmonary embolism, and aortic dissection. The evolution of a clinical biomarker requires a long journey and an arduous transition from the research environment to clinical practice. Emerging technologies such as those described earlier have the potential to permit systematic assessment of variation in genes, RNA, proteins, and metabolites for identification of "uncorrelated" or "orthogonal" biomarkers that probably would not emerge with a focus on candidates from well-studied pathways.

REFERENCES

1. Stanford KI, Lynes MD, Takahashi H, et al. 12,13-diHOME: an exercise-induced lipokine that increases skeletal muscle fatty acid uptake. *Cell Metab.* 2018;27:1111–1120.e3.
2. Washburn J, Millington DS. Digital microfluidics in newborn screening for mucopolysaccharidoses: a progress report. *Int J Neonatal Screen.* 2020;6:78.
3. Lavelle A, Sokol H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nature Rev Gastroenterol Hepatol.* 2020;17:223–237.
4. Chen Z-Z, Gerszten Robert E. Metabolomics and proteomics in type 2 diabetes. *Circ Res.* 2020;126:1613–1627.
5. White PJ, Newgard CB. Branched-chain amino acids in disease. *Science.* 2019;363:582.
6. Lotta LA, Scott RA, Sharp SJ, et al. Genetic Predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a mendelian randomisation analysis. *PLoS.* 2016;13:e1002179.
7. Roberts A, Gu X, Buffa J, et al. Development of a gut microbe–targeted nonlethal therapeutic to inhibit thrombosis potential. *Nat Med.* 2018;24:1407–1417.
8. Ganz P, Heidecker B, Hveem K, et al. Development and validation of a protein-based risk score for cardiovascular Outcomes among patients with stable Coronary heart disease. *J Am Med Assoc.* 2016;315:2532–2541.