SECTION 1

BASIC METHODOLOGICAL STRATEGIES IN METABOLOMIC RESEARCH

1 Exploring the Human Metabolome by Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry

David S. Wishart

Although most of the focus in "omics" science over the past decade has been on sequencing the human genome (1) or annotating the human proteome (2), there is another equally important component of the human body that until more recently has been largely overlooked – the human metabolome. The human metabolome can be thought of as the complete collection of small molecule metabolites found in the human body. These small molecules include chemical entities such as peptides, amino acids, nucleic acids, carbohydrates, organic acids, vitamins, minerals, food additives, drugs, and just about any other chemical (with a molecular weight <1,500 Da) that can be used, ingested, or synthesized by humans.

Metabolites act as the bricks and mortar of our cells. They serve as the building blocks for all of our macromolecules, including proteins, RNA, DNA, carbohydrates, membranes, and all other biopolymers, that give our cells their structure and integrity. Metabolites also act as the fuel for all cellular processes, the buffers to help tolerate environmental insults, and the messengers for most intracellular and intercellular events. Together with the genome and the proteome, the human metabolome essentially defines who and what we are.

However, in contrast to the genome and proteome, the metabolome itself is not easily defined because the human metabolome is not solely dictated by our genes. Our environment (what we eat, breathe, drink) and our microflora (the bacteria that live in our intestinal tract) contribute to the metabolome. The human metabolome consists of a mix of both endogenous and exogenous compounds. Endogenous metabolites are small molecules synthesized by the enzymes encoded by our genome or our microfloral genomes, and exogenous metabolites are "foreign" or xenobiotic chemicals consumed as foods, drugs, or other additives. The fact that so many different chemicals from so many different sources can potentially appear in the human metabolome has made its characterization difficult. Nevertheless, several concerted efforts have been made to decipher the human metabolome – or, more appropriately, the human metabolome*s*. Beginning in 2005 and continuing to the present, the Human Metabolome Project (3) has been using a variety of high-throughput metabolomic studies in combination literature surveys to compile as much information about the "detectable" human metabolomes as possible. This information is released publicly through various Web-accessible databases.

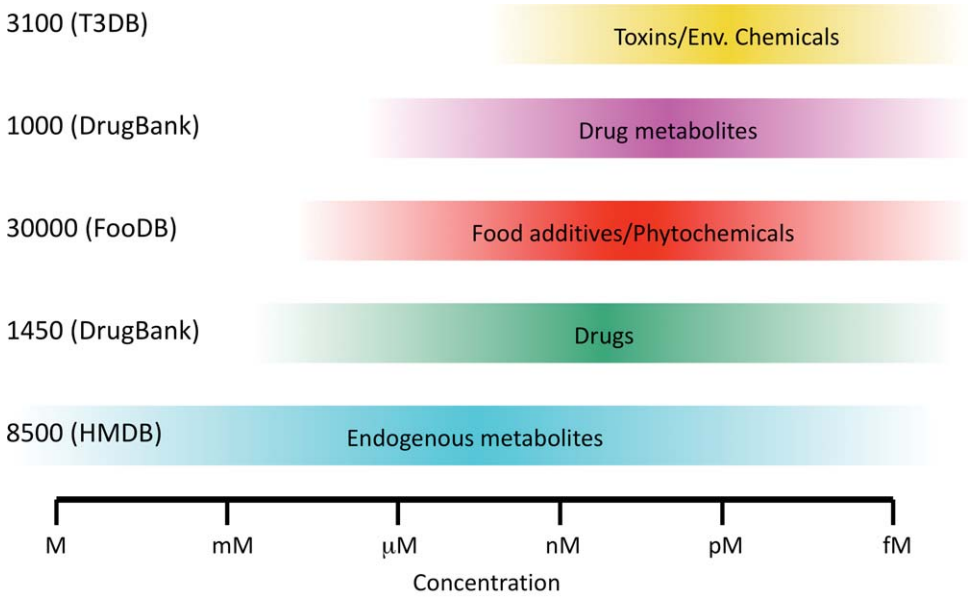

Figure 1.1. Schematic illustration of the different human metabolomes, including their concentration ranges (femtomolar to molar), their approximate size (number of compounds that have been identified or quantified or both), and the databases where this information is catalogued.

These databases include the Human Metabolome Database (HMDB) (4), which covers endogenous human metabolites (including some very common food, drug, and microbial metabolites); DrugBank (5), which contains data on exogenous drugs and drug metabolites; Toxin and Toxin-Target Database (T3DB) (6), which covers pollutants, poisons, and environmental chemicals; and FooDB (7), which contains data on foods (i.e., phytochemicals) and food additives.

The approximate size of each detectable or potentially detectable human metabolome is shown in Figure 1.1. This figure also illustrates the concentration ranges typically reported for these compounds. What is not shown in Figure 1.1 is the size of the "theoretical" human metabolome. If all possible combinations of lipids, dipeptides and tripeptides, and disaccharides and trisaccharides were considered, the number of endogenous human metabolites could easily exceed 200,000 molecules (8,9). However, most of these theoretical metabolites have not been detected. Either they exist too transiently or they are at such low abundance that they cannot be seen with today's technologies.

Although all humans share pretty much the same endogenous metabolome, every human has a different exogenous metabolome. Furthermore, no single human (unless he or she regularly consumes all known drugs and all known foods and lives in a toxic chemical dump) has the full complement of known or detectable exogenous compounds in his or her body. If a large population is studied, many of these exogenous metabolites are observed – albeit at relatively low levels. A further source of variation in the human metabolome comes from the metabolites generated by the nearly 400 different microbial species that live in the human intestinal tract (10). In humans, the gut microflora weigh between 1 and 2 kg and constitute

a metabolically essential, albeit highly distributed, multicellular organ (10). Each human has his or her own unique gut microflora; these microbes (along with our diet) contribute significantly to our own unique metabolic phenotype, or "metabotype" (11).

The issue of exogenous versus endogenous metabolites is not the only complication associated with describing the human metabolome. Humans have more than 200 different cell types, several dozen different organs, and many highly compartmentalized biofluid systems. Each of these cell types, tissues, or organs is metabolically specialized in one way or another, often producing unique metabolites that are not found in other cells or organs. The same metabolic specialization is true for many biofluids as well. These biofluids include blood, milk, cerebrospinal fluid (CSF), bile, saliva, mucus, lung exudates, lacrimal secretions, semen, and lymph. As a result, specific cell, tissue, biofluid, and organ variations also make the human metabolome hard to define. Additionally, the human metabolome is hard to define because of the wide range of metabolite concentrations found in humans. These concentrations, which can range from picomolar levels (i.e., exogenous chemicals, certain hormones, and many signaling molecules) to near-molar concentrations (i.e., urea), are a function of diet, gender, time of day, age, health, and genetic background (12). Thus, the human metabolome is actually defined by when, where, and how it is measured.

This chapter is primarily concerned with describing how the human metabolome has been characterized using different analytical techniques. The chapter begins with a brief historical perspective to help frame the challenges and progress in identifying and quantifying human metabolites. This is followed by a description of the methods used to acquire and prepare human samples for metabolomic studies. The central focus of this chapter is the description and assessment of three different metabolomic methods commonly used to characterize metabolites in human biofluids and tissue extracts: 1) nuclear magnetic resonance (NMR) spectroscopy; 2) gas chromatography–mass spectrometry (GC-MS); and 3) liquid chromatography–mass spectrometry (LC-MS). The chapter concludes with a comparative assessment of these three technologies and a brief discussion of possible future directions for human metabolic profiling. Specific advantages of integrating MS with chromatography and NMR spectroscopy are described in Chapter 11.

1. Brief History of Human Metabolomics

Scientists and physicians have been attempting to characterize the human metabolome for thousands of years. In the fifth century BC, Hippocrates and Hermogenes believed that the color, taste, and smell of urine (i.e., urinary composition) were closely tied to the health of an individual. By the Middle Ages, physicians had created a field of medicine called "uroscopy," which included specialized "urine wheels" or urine charts that could associate particular colors of urine with twenty different disorders. The characterization of human biofluids eventually became more quantitative with the development of clinical chemistry in the mid-nineteenth century (13). Largely through the work and writings of numerous British scientists (e.g., Bostock and Bright), clinicians began to identify and quantify biofluid

constituents and associate them with various medical conditions. In the early twentieth century, Folin and Van Slyke helped to develop many of the colorimetric tests and early instrumentation used to quantify metabolites in blood and urine. Thanks to their efforts, metabolic profiling finally became a part of routine medical practice (14). Nowadays, blood and urine tests, which offer 5 to 50 different chemical readouts, are routinely performed by multicomponent clinical analyzers or by simple paper strip tests (15,16). These semiquantitative tests typically depend on colorimetric assays where specific reagents are added to a sample and reactions are monitored spectrophotometrically to identify or quantify a targeted metabolite.

By the 1970s, a new generation of clinical chemistry instrumentation permitted the identification of not just a single compound but a whole class of compounds. Gas chromatographic columns were coupled to mass spectrometers to create GC-MS systems that could detect organic acids from blood and urine. The birth of metabolomics (or metabolic profiling as it was called then) probably can be traced to a seminal article on GC-MS published in 1974 (17) that described the profiling of dozens of organic acids in human urine. Many other studies have since followed (18,19), and GC-MS continues to be the method of choice in organic acid profiling, especially for genetic disease testing and monitoring (20,21).

NMR spectroscopy has also been used to characterize human (and other mammalian) metabolites for many decades. In the 1970s, carbon-13 (^{13}C) isotope-tracer analysis was used in conjunction with NMR to decipher ethanol metabolism (22). The richness of information contained in high-resolution NMR spectra along with the ability to identify multiple metabolites at once soon made NMR a favorite tool for many other metabolism researchers. NMR-based studies of human urine and blood in the 1980s and early 1990s led to the identification of several putative biomarkers of cancer and coronary artery disease (23,24). These studies were complemented by the pioneering studies of Nicholson and colleagues (25,26), who used NMR spectra from urine to characterize inborn errors of metabolism (IEM) and drug toxicity. More systematic studies looking at the metabolic composition of CSF and blood followed (27,28). NMR has been included in hundreds of metabolomic studies since (29).

In the 1990s, tandem mass spectrometry (MS/MS) emerged as a powerful new approach for the nontargeted detection of human metabolites. MS/MS permits very rapid (1 to 2 minutes), sensitive (femtomole detection limits from dried blood spots), and, with appropriate internal standards, accurate quantification of dozens of metabolites. Because of these appealing features, MS/MS or direct injection mass spectrometry (DIMS) has become increasingly used in newborn screening programs in North America and Europe, with a particular focus on identifying amino acid, nucleic acid, and acylcarnitine markers for IEM (30). Other developments in metabolite profiling over the past decade include the introduction of capillary electrophoresis (CE) methods for more precise and rapid metabolite separation (31), the use of ultra-high-pressure liquid chromatography (UPLC) and two-dimensional high-performance liquid chromatography (HPLC) methods for improved compound partitioning (32,33), and the application of Fourier transform mass spectrometry (FT-MS) methods for large-scale metabolite screening (34).

Table 1.1 summarizes the advantages and disadvantages of the three major technologies (NMR, GC-MS, and LC-MS) used in modern human metabolomic

Table 1.1. *Comparison of nuclear magnetic resonance, gas chromatography–mass spectrometry, and liquid chromatography–mass spectrometry for metabolomics*

Technology	Advantages	Disadvantages
NMR spectroscopy	• Quantitative • Nondestructive • Fast (1 hr to collect data and analyze a sample) • Requires no derivatization • Requires no separation • Detects all organic classes • Allows ID of novel chemicals • Robust, mature technology • Can be used for metabolite imaging (MRSI) • Large body of software and databases for metabolite ID	• Not very sensitive (~5 μM) • Expensive instrumentation • Large instrument footprint • No detection of nonprotonated compounds, inorganic ions, and salts by ^{1}H NMR[a]
GC-MS	• Robust, mature technology • Relatively inexpensive • Quantitative (with calibration) • Small sample volume (50 μL) • Good sensitivity (100 nM) • Large body of software and databases for metabolite ID • Detects most organic and some inorganic molecules • Excellent separation reproducibility	• Sample not recoverable • Requires sample derivatization • Requires separation • Relatively slow (2–3 hr to collect data and analyze sample) • Cannot be used in imaging • Novel compound ID is difficult
LC-MS	• Superb sensitivity • Very flexible technology • Detects most organic and some inorganic molecules • Minimal sample size requirement (10 μL) • Can be used in metabolite imaging (MALDI) • Can be done without separation (direct injection) • Has potential for detecting largest portion of metabolome	• Sample not recoverable • Not very quantitative • Expensive instrumentation • Slow (hours to days to analyze a sample) • Poor separation resolution and reproducibility (vs. GC) • Less robust instrumentation than NMR or GC-MS • Limited body of software and databases for metabolite ID • Novel compound ID is difficult

[a] However, compounds with NMR-active nuclei other than protons can be detected and quantified by heteronuclear NMR spectroscopy (for ^{13}C and ^{31}P NMR *see* Chapters 18 to 22 and 24).

Abbreviations: MRSI, magnetic resonance spectroscopic imaging; ID, identification; MALDI, matrix-assisted laser desorption/ionization.

studies. Figure 1.2 illustrates the relationship between the three different technologies and their overall sensitivity. NMR is typically capable of detecting 50 to 75 compounds in a given human biofluid, with a lower sensitivity limit of about 5 μM (29,35). Most of the compounds detected by NMR are intrinsically polar molecules, such as organic acids, sugars, amino acids, and small amines (Table 1.2). GC-MS is capable of detecting 50 to 100 compounds (depending on the biofluid), with a lower sensitivity limit of about 100 nM (36). GC-MS provides relatively

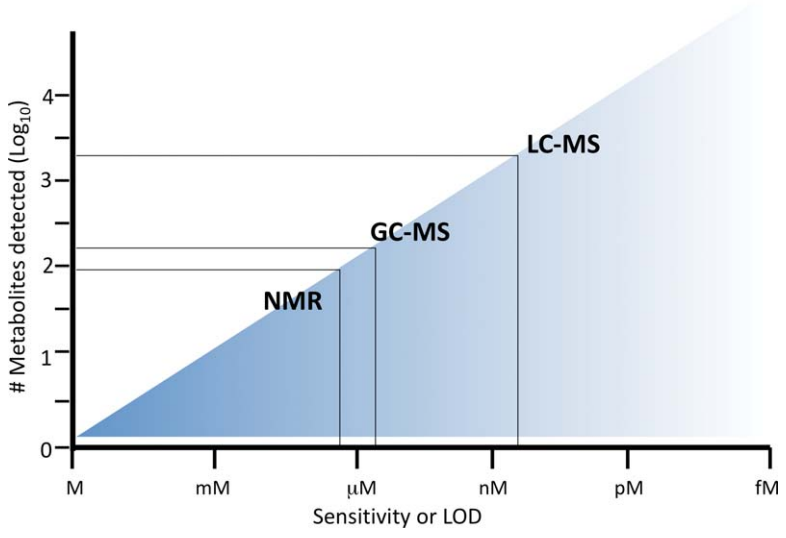

Figure 1.2. Graph illustrating the different sensitivities (limit of detection [LOD]) of NMR, GC-MS, and LC-MS for metabolite detection. The typical number of metabolites detected or identified or both by each technology is shown on the *y*-axis. NMR is the least sensitive method (limit of detection approximately 5 µM), and LC-MS is the most sensitive (limit of detection <1 nM).

broad metabolite coverage routinely detecting amino acids, sugars, organic acids, phosphorylated compounds, fatty acids, and cholesterol (Table 1.2). Because of the exquisite sensitivity of today's MS instruments, LC-MS methods can detect hundreds to thousands of "features" (37). However, the number of compounds that can be positively identified is typically much less (Table 1.3). LC-MS methods are particularly useful in targeted metabolomic studies of human serum lipids, in which up to 500 different lipids and fatty acids can be detected and quantified (38,39).

2. Sample Preparation for Human Metabolomic Studies

Key to any successful effort in all human metabolomic experiments is having a high-quality sample. The choice of the sample (e.g., fluid, tissue) is dictated by the questions being asked, the sensitivity of the instrument, and the kind of metabolites being studied. Metabolomic studies have been reported on extracted human tissues or biopsy specimens, fine-needle aspirates, dried blood spots, plasma or serum, urine, CSF, bile, seminal fluid, feces, saliva, and many other biofluids. Overall, most human metabolomic measurements are performed on biofluids, not tissues. The choice of fluids over tissues is based on the assumption that the chemicals found in most biofluids are largely reflective of the physiological state of the organ that produces, or is bathed in, that fluid. Hence, urine reflects the processes occurring in the kidney, bile reflects the processes occurring in the liver, and CSF reflects the processes occurring in the brain. The blood is a special biofluid because it potentially reflects all processes occurring in all organs. This can be both a blessing and

Table 1.2. *Compounds typically found in human serum by nuclear magnetic resonance and gas chromatography–mass spectrometry metabolomic studies*

NMR	GC-MS	
2-Hydroxybutyric acid	2-Aminobutyric acid	Malonic acid
2-Oxoisovaleric acid	2-Hydroxyisobutyric acid	Maltose
3-Hydroxybutyric acid	2-Methylbutanoic acid	Mead acid
Acetaminophen	3-Hydroxybutyric acid	Methionine
Acetic acid	4-Hydroxybutyric acid	Methylmaleic acid
Acetoacetic acid	Acetaminophen	Methylmalonic acid
Acetone	Adrenic acid	*Myo*-inositol
Alanine	Alanine	Myristic acid
Arginine	Aminomalonic acid	*N*-Acetyl-glycine
Asparagine	Arachidonic acid	*N*-Acetyl-L-lysine
Aspartate	Asparagine	Nicotinic acid
Betaine	Aspartic acid	Nonadeca-10(Z)-enoic acid
Carnitine	Benzoic acid	Oleic acid
Choline	Bovinic acid	Ornithine
Citric acid	Cervonic acid	Oxalic acid
Creatine	Cholesterol	Palmitelaidic acid
Creatinine	Citric acid	Palmitic acid
Cysteine	Clupanodonic acid	Palmitoleic acid
Cystine	Cysteine	Pentadecanoic acid
Ethanol	Cystine	Phenylalanine
Formic acid	D-Fructose	Phosphoric acid
Glucose	D-Galactopyranose	Proline
Glutamate	D-Galactose	Pyroglutamic acid
Glutamine	Decanoic acid	Ribitol
Glycerol	Dihomo-γ-linolenic acid	Salicylic acid
Glycine	Dodecanoic acid	Serine
Histidine	Eicosanoic acid	Stearic acid
Hypoxanthine	Eicosenoic acid	Stearidonic acid
Isobutyric acid	Erythronic acid	Succinic acid
Isoleucine	Fumaric acid	Tartaric acid
Isopropanol	Glucitol	Tetradecanoic acid
Lactic acid	Gluconic acid	Threonine
Leucine	Glucopyranose	Timnodonic acid
Lysine	Glucose	Tryptophan
Malonic acid	Glutamic acid	Tyrosine
Methanol	Glutamine	Urea
Methionine	Glyceric acid	Uric acid
Methylmalonic acid	Glycerol	Vaccenic acid
Ornithine	Glycine	Valine
Phenylalanine	Heptadecanoic acid	Xylitol
Proline	Histidine	α Linolenic acid
Propylene glycol	Hydroxyproline	γ-Linolenic acid
Pyruvic acid	Isobutyric acid	
Serine	Isoleucine	
Threonine	Lactic acid	
Tryptophan	Lauric acid	
Tyrosine	Leucine	
Urea	Linoleic acid	
Valine	Linolenic acid	
Xanthine	Lysine	

Table 1.3. *Compounds that can typically be found (and quantified) in human serum by reverse phase liquid chromatography–tandem mass spectrometry methods*

11-HETE	6-trans-LTB4	lysoPC a C18:1	Phosphatidylcholine aa C38:5
11,12-DiHETrE	8-HETE	lysoPC a C18:2	Phosphatidylcholine aa C38:6
11,12,15 TriHETrE	8,15-DiHETE	lysoPC a C20:3	Phosphatidylcholine aa C40:1
11(12)-EpETrE	8,9-DiHETrE	lysoPC a C20:4	Phosphatidylcholine aa C40:2
12-HEPE	8(9)-EpETrE	lysoPC a C24:0	Phosphatidylcholine aa C40:3
12-HETE	9-HETE	lysoPC a C28:0	Phosphatidylcholine aa C40:4
12-HpETE	9-HODE	lysoPC a C28:1	Phosphatidylcholine aa C40:5
12-KETE	9-HOTE	Methionine	Phosphatidylcholine aa C40:6
12,13-DiHODE	9-HpODE	Nonaylcarnitine	Phosphatidylcholine aa C42:0
12,13-DiHOME	9-KODE	Octadecadienylcarnitine	Phosphatidylcholine aa C42:1
12(13)-EpODE	9,10-DiHODE	Octadecanoylcarnitine	Phosphatidylcholine aa C42:2
12(13)-EpOME	9,10-DiHOME	Octadecenoylcarnitine	Phosphatidylcholine aa C42:4
12(13)Ep-9-KODE	9,10,13-TriHOME	Octanoylcarnitine	Phosphatidylcholine aa C42:5
13-HODE	9,12,13-TriHOME	Octenoylcarnitine	Phosphatidylcholine aa C42:6
13-HOTE	9(10)-EpODE	Ornithine	Proline
13-HpODE	9(10)-EpOME	Prostaglandin B_2	Propionyl-L-carnitine
13-KODE	Acetyl-L-carnitine	Prostaglandin D_2	Resolvin D_1
14,15-DiHETE	Arginine	Prostaglandin E_1	Resolvin E_1
14,15-DiHETrE	Butyryl-L-carnitine	Prostaglandin E_2	Serine
14(15)-EpETE	Decadienylcarnitine	Prostaglandin E_3	Sphingomyelin (OH) C14:1
14(15)-EpETrE	Decanoylcarnitine	Prostaglandin $F_{2\alpha}$	Sphingomyelin (OH) C16:1
15-deoxy PGJ_2	Decenoylcarnitine	Prostaglandin J_2	Sphingomyelin (OH) C22:1
15-HEPE	δ-2-Prostaglandin J_2	Phenylalanine	Sphingomyelin (OH) C22:2
15-HETE	DL-carnitine	Phosphatidylcholine aa C28:1	Sphingomyelin (OH) C24:1
15-HETrE	Dodecanoylcarnitine	Phosphatidylcholine aa C30:0	Sphingomyelin C16:0
15-HpETE	Glutaconyl-L-carnitine	Phosphatidylcholine aa C30:2	Sphingomyelin C16:1
15-KETE	Glutamine	Phosphatidylcholine aa C32:0	Sphingomyelin C18:0
15,16-DiHODE	Glycine	Phosphatidylcholine aa C32:2	Sphingomyelin C18:1
15(16)-EpODE	Hepoxilin A_3	Phosphatidylcholine aa C32:3	Sphingomyelin C20:2
16(17)-EpDoPE	Hexadecadienylcarnitine	Phosphatidylcholine aa C34:1	Sphingomyelin C22:3
17-HDoHE	Hexadecanoylcarnitine	Phosphatidylcholine aa C34:2	Sphingomyelin C24:0
17,18-DiHETE	Hexadecenoylcarnitine	Phosphatidylcholine aa C34:3	Sphingomyelin C24:1
17(18)-EpETE	Histidine	Phosphatidylcholine aa C34:4	Sphingomyelin C26:0
19,20-DiHDoPE	Hydroxybutyrylcarnitine	Phosphatidylcholine aa C36:0	Sphingomyelin C26:1
19(20)-EpDoPE	Isoleucine/Leucine	Phosphatidylcholine aa C36:1	Tetradecadienylcarnitine
20-carboxy-LTB_4	Lipoxin A_4	Phosphatidylcholine aa C36:2	Tetradecanoylcarnitine
20-HETE	Leukotriene B_4	Phosphatidylcholine aa C36:3	Tetradecenoylcarnitine
20-hydroxy-LTB_4	Leukotriene B_5	Phosphatidylcholine aa C36:4	Threonine
5-HEPE	Leukotriene E_4	Phosphatidylcholine aa C36:5	Tiglyl-L-carnitine
5-HETE	lysoPC a C14:0	Phosphatidylcholine aa C36:6	Tryptophan
5-KETE	lysoPC a C16:0	Phosphatidylcholine aa C38:0	Thromboxane B_2
5,15-DiHETE	lysoPC a C16:1	Phosphatidylcholine aa C38:1	Tyrosine
5,6-DiHETrE	lysoPC a C17:0	Phosphatidylcholine aa C38:3	Valeryl-L-carnitine
6-keto-PGF1α	lysoPC a C18:0	Phosphatidylcholine aa C38:4	Valine

Note. This list is highly dependent on the separation methods and equipment used.

a curse because metabolite perturbations in the blood, although easily detectable, cannot be easily traced to a specific organ or a specific cause.

In metabolomics, the choice of biofluids over tissues is also dictated by the fact that fluids are far easier to acquire, process, and analyze with NMR, MS, or HPLC