

Application of Metabolomics to Diagnosis of Insulin Resistance

Michael V. Milburn and Kay A. Lawton

Metabolon Inc., Durham, North Carolina 27713; email: mmilburn@metabolon.com, klawton@metabolon.com

Annu. Rev. Med. 2013. 64:291–305

The *Annual Review of Medicine* is online at med.annualreviews.org

This article's doi:

10.1146/annurev-med-061511-134747

Copyright © 2013 by Annual Reviews.
All rights reserved

Keywords

biomarkers, biochemistry, diabetes, prognostics, diagnostics

Abstract

Metabolomics, the global interrogation of the biochemical components in a biological sample, has become an important complement to genomics and proteomics to aid in the understanding of pathophysiology. Major advantages of metabolomics are the size of the metabolome relative to the genome or proteome and the fact that it provides a view of the existing biochemical phenotype. As such, metabolomics is fast becoming an important discovery tool for new diagnostic and prognostic biomarkers. Although many methods exist for performing metabolomics, relatively few have led to successful development of new diagnostic tests. This review will aid the reader in understanding various metabolomic methods and their applications, as well as some of their inherent advantages and disadvantages. In addition, we present one example of the application of metabolomics to the identification of new fasting blood biomarkers for the diagnosis and monitoring of insulin resistance.

INTRODUCTION TO GLOBAL METABOLOMICS

Over the past decade, “omics” technologies have become key to discovering novel biomarkers. Biomarkers can be nearly anything that distinguishes one individual from another. They can be based on a diagnostic test (e.g., cholesterol measurements), physical characteristic (e.g., body mass index), genetics (e.g., single nucleotide polymorphisms), or other distinguishing characteristics (e.g., diet). Discovery technologies such as transcriptomics (profiling mRNA), proteomics (profiling proteins), and others (e.g., genome wide association studies) start with an idea of the biomarker type needed to determine a condition, disease, or health risk. Unfortunately, general screening methods for biomarker discovery have been challenging, with few success stories. Biomarker validation is a relatively new and evolving concept. The challenges are mathematical, technological, and/or statistical (insufficient number of well-characterized samples). One technology with promise to overcome these issues is metabolomics, which profiles biochemicals and metabolites. This review introduces applications of metabolomics in the diagnosis of insulin resistance.

The goal of metabolomics is to measure all of the small molecules (i.e., biochemicals, metabolites) in a sample. As such, it is grounded in biochemistry and provides a snapshot of the biochemical or metabolic status of the subject under investigation. The development of metabolomics as a tool to leverage the large body of available biochemical knowledge has been a major challenge. Most laboratories have focused on targeted metabolomic analysis, specializing in the measurement of 20–100 different metabolites which, most often, are within a common class of compounds. For example, a number of companies and academic labs have developed methods for detecting lipid compounds. Although lipids represent only a subset of the total biologically relevant metabolites, these data have proven useful for biomarker discovery efforts (1).

Other groups have focused on methods to truly investigate all of the small molecules in samples. “Global” or “unbiased” metabolomics has been plagued by difficulties stemming from the diverse physical properties of small molecules. These properties can vary greatly, with significant differences in solubilities and with molecular weights ranging from 20 to >1,500 Da. It is therefore difficult to develop a single chromatography method to separate all of the compounds and even more difficult to analyze individual compounds without chromatographic separation. Further complications arise if studies are expected to be completed within a clinically useful turnaround time. These issues are currently being addressed through advanced multisystem approaches where the best separation and detection instrument technologies are developed to run in tandem. This approach allows for a comprehensive solution achieved by combining principles offered by various best-in-breed technologies. As this new technology develops and its use in biomarker detection studies increases, it is rapidly becoming clear that metabolomics will likely have a high impact in the diagnosis of disease, identification of drug targets, evaluation of the effects of drugs, and selection of patients most likely to respond to drug therapy (i.e., personalized medicine).

To develop a viable technology platform that harnesses the power of metabolomics requires engineering a system of chromatographic separation coupled to mass spectrometry (2–4). One such metabolomics analysis process operates in essentially four steps (**Figure 1**). Step one is extraction of the small molecules from the biological sample. Step two is the chromatography coupled with mass spectrometry and data collection. Step three is the automated and manual quality control (QC) analysis of the data. Step four is the statistical and biological interpretation of the data. Numerous methods of analyzing a wide range of very polar to nonpolar compounds from as little as 50 μ l of blood plasma have been published (5–7). For the case study



Figure 1

Data-generation steps of a global metabolomics method as applied to a biological sample. The three steps are (1) biochemical extraction, (2) multiple chromatography and mass spectrometry analysis, and (3) unbiased global informatics methods to reduce the raw machine data to the biochemicals in the sample and determine the relative concentration of each biochemical in each sample. In addition, a quality-control step is performed on the data prior to statistical analysis and biological interpretation.

presented in this review, the extracted samples were split into four aliquots for different chromatography and mass spectrometry (MS) platforms, two ultra-high-performance liquid chromatography (UHPLC) methods and one gas chromatography (GC) method, with one aliquot held in reserve (4). These three chromatography and MS systems complement one another in the range of biochemicals measured and provide an enhanced biochemical coverage of each sample (4). Approximately 70%–80% of the biochemicals are measured on more than one platform, with 30%–40% measured on all three platforms. For compounds observed on multiple platforms, the chromatography/analytical system with the best analytical characteristics (e.g., fewest interfering peaks or highest signal to noise) is generally used for the analysis of that compound. In general, the GC method provides better separation of molecules that tend to be more difficult to separate using a typical reverse-phase liquid chromatography (LC) method (e.g., carbohydrates).

Following acquisition of the raw data from the instruments, a suite of software methods (3, 8–14) automatically integrates each ion across retention time and then uses that ionic

information, which may include additional MS/MS fragmentation information and retention time, to identify the compound. After a compound is identified in a sample, one of the characteristic and stronger ions is used to determine a relative concentration of that compound in each sample. This chemo-centric approach assures that the compound will be represented only once in the subsequent statistical analysis. When the software has finished analyzing the samples, all of the data are loaded into a visual user interface that allows manual curation of the data. For QC purposes, a scientist visually inspects how well each compound was identified and verifies that only those compounds with the highest degree of confidence are included in the final data set used for statistical analysis (14).

After QC curation, a variety of statistical approaches can be applied to the final data set, including ANOVA, t-tests, Random Forest, Principal Component Analysis, etc. The goal of these statistical treatments is to identify the biochemicals that best represent the most significant changes in concentration between the groups in the study. One advantage of biochemistry is that multiple compounds in a particular biochemical pathway may often be significantly

altered, giving an even higher degree of confidence to the importance of that biochemical change. In this respect, it is important to point out that most statistical treatments assume independent variables when, in fact, we know that certain biochemicals are related to the same or similar pathways. Consequently, knowledge of the biochemical context can enable the identification of and increase confidence in the candidate biomarkers.

INTRODUCTION TO BIOMARKER DISCOVERY AND VALIDATION

Biomarker discovery typically involves, as a first step, the deployment of a discovery technology. Whether that technology is genomics/transcriptomics, proteomics, metabolomics, or something else, the process starts with an unmet need and an idea of the type of biomarkers that are required (single nucleotide polymorphisms, proteins, biochemicals, etc.). Clinical biomarker discovery studies usually involve (a) the identification of the appropriate study populations, using a “gold standard” test to distinguish individuals with the disease from the nondisease individuals; (b) the determination of the availability of appropriate sample types in sufficient numbers with the necessary meta-data for the desired study population; and, in many cases, (c) the development of clinical research agreements with academic and/or healthcare organizations.

For example, insulin resistance (IR) is a well-studied condition that can lead to a number of increased health risks, including diabetes, cardiovascular disease, and metabolic diseases. However, no simple fasting blood tests currently exist for IR or the assessment of the degree of IR. The current gold-standard test—namely, the hyperinsulinemic clamp (HI clamp)—is invasive, time-consuming, expensive, and rarely used outside of a clinical research setting. Currently available index tests have limitations in terms of time, invasiveness, complexity, and performance. New biomarkers that can be developed into a new diagnostic

test are needed to better manage this growing concern in humans and companion animals. If determining insulin sensitivity and identifying insulin-resistant subjects is the clinical test idea, then the next step is to perform a feasibility study to determine whether it is possible to separate the case from the control groups using the chosen technology and to identify candidate biomarkers that contribute to the separation. These proof-of-concept studies are typically simple in design and relatively small (<30 subjects per group).

As with many initial research studies, a biomarker feasibility study can produce mixed results even if well designed and sufficiently powered. Analysis of these initial studies can be confounded by the unforeseen impact of age, gender, or other differences (e.g., diet, medications, lifestyle) between the groups. A further important consideration for “omics-based” studies is that when a very large number of measurements (e.g., 1 million SNPs) are made on only a small number of subjects (e.g., 50 individuals per group), the probability of separating the groups by chance alone is enormous and often leads to false discovery. As a result, more than one feasibility study may be necessary to fully demonstrate the proof-of-concept and become sufficiently confident in the data to invest the resources necessary for the larger studies that are essential for biomarker validation. Each subsequent study design can be adjusted to take into account the confounding variables uncovered in prior studies.

Once it seems reasonable that the sample type for the biomarker is suitable (blood, urine, etc.), and the biomarker(s) from the feasibility study meets minimum performance criteria, then larger, well-powered discovery studies are warranted. These studies should focus on clinical utility and contain sufficient numbers of subjects to not only assess the reproducibility of the biomarker candidates uncovered in the feasibility study, and uncover additional biomarkers, but also provide an independent test set of subjects. That is, the clinical study population should be large enough to be divided into a “discovery validation set” (or “training set”) and a

“test set,” the latter of which is composed of subjects who have not been used to discover the biomarkers. A schematic of the diagnostic development from discovery study through validation and commercial formatting of the test is exemplified in **Figure 2**.

Another important consideration in biomarker discovery and validation is the demographic suitability of the subjects in the discovery population relative to those for whom the test will be applied. For instance, using only European populations for the discovery and development of the biomarkers may be a concern if the test will also be employed in Japan.

In addition to biomarker diagnostic performance validation, the laboratory assays used to measure the biomarkers must undergo analytical validation. Although analytical assay development and validation entail significantly less risk than developing the biomarker, these steps can be equally time consuming. Sufficient numbers of samples from the test population are necessary to confirm the laboratory performance of the analytical assays. In addition, the assay throughput requirements or the assay setting (e.g., hospital, point-of-care, central laboratory) often require specific formats for the biomarker measurement and may differ from the discovery assays. For example, gene chips are often used for the discovery analysis of genetic biomarkers, but to satisfy the throughput, cost, and quantitation requirements, other methods, such as PCR-based assays, have been developed.

ADVANTAGES OF METABOLOMICS FOR BIOMARKER DISCOVERY

The word “metabolomics” (or “metabonomics”) first appeared in journal articles in 2000. Only a few metabolomics papers were published that year, but in 2009 >1,300 published scientific papers reported metabolomics results. In fact, metabolomics publications are one of the fastest-growing areas of scientific publications in the past few years. Although

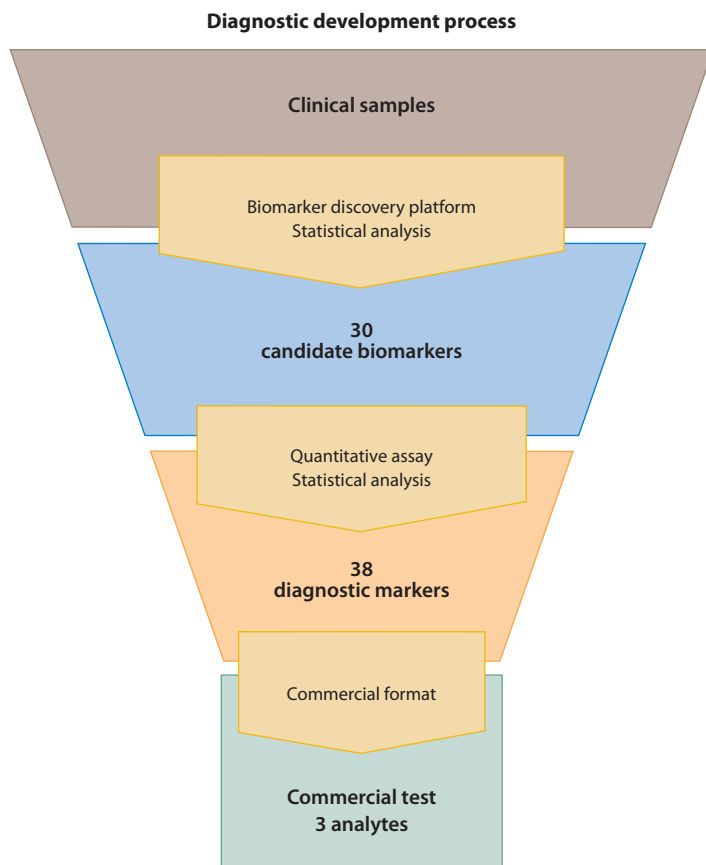


Figure 2

Process flow using “omics” technologies for biomarker discovery and validation. The initial work begins with decisions on the types of samples to be analyzed for the biomarker discovery study, number of subjects and samples required, and the relevant case and case controls that will be used. Many of these decisions are based on the types of technologies employed to analyze the samples. Other decisions depend more on how the ultimate biomarker or diagnostic test will be used clinically. After the samples are received, then the discovery technology can be employed and analyzed. The analysis of “omics” technologies is beyond the scope of this review but can involve complex multivariate analysis and methods to control for false discovery. Ultimately, the aim of the discovery study is to develop a shorter list of candidate biomarkers that can be further validated with independent assays and preferably independent cohorts of samples. Rarely is the discovery technology itself employed on an individual basis, although some technologies are moving in this direction. In most cases, the assay to validate a biomarker is a more specific quantitation of the analyte under strict clinical protocols. Ultimately, these assays will be further validated in the clinic to determine their clinical utility. Furthermore, many standard clinical measurements can be used in the performance evaluation of the new biomarkers; this explains the additional eight candidate markers in the figure, bringing the total biomarkers evaluated in this example to 38.

analytical chemists and biochemists have been identifying small molecules in biological samples since long before 2000, as a robust, nontargeted discovery tool the technology is new and rapidly evolving. In addition to the increasing number of publications in the past few years, several significant biomarker reports using metabolomics have been published that include supporting validation data for these discoveries (5, 15, 16).

Because metabolomics analysis is based on measuring the small molecules (i.e., biochemicals or metabolites of molecular weights <1,500 Da) in a biological sample, the interpretation of the resulting data is based on biochemistry. This provides a significant advantage. Relative to molecular biology and proteomics, biochemistry is a mature, highly developed field of science. In fact, many of the Nobel Prizes in Medicine before the 1960s were awarded in biochemistry. Today, we routinely depend on metabolite-based disease diagnosis. For instance, high glucose in urine was one of the earliest tests for diabetes, and cholesterol is used to measure the risk of heart disease. Metabolic panels of fatty acids, bile acids, sugars, creatine, creatinine, urea, etc. are routinely used clinically to assess organ function or risk for many diseases as well as for disease diagnosis. Clearly, the field of metabolomics is positioned to take advantage of this repository of biochemical pathway knowledge.

The major challenge for metabolomics has been to develop a technology that can extract, identify, and quantitate the entire spectrum of the small molecules (<1,500 Da) in any biological sample, unlike profiling DNA, RNA, or protein (see **Figure 3**). The exact number of small molecules in biological samples is a hotly debated subject. Some databases of metabolites list as many as 6,000 comprising the human metabolome. However, a deeper evaluation of these metabolites reveals that several thousand molecules can simply be grouped as different combinations of complex lipids or small peptides (17, 18). From a primary metabolism standpoint, and disregarding the combinations

of complex lipids or peptides as well as xenobiotic metabolites from drugs, diet, and the like, it is very likely that <3,000 human metabolites are significant for understanding metabolism and metabolic effects. Importantly, in any one sample matrix (i.e., blood, urine, tissue, etc.) there will always be dramatically fewer metabolites than the total number synthesized in the entire organism.

Most importantly, this number (<3,000 metabolites) is much smaller than the number of molecules involved in other “omics” technologies, such as genomics (>30,000 genes) or proteomics (>100,000 proteins). This difference may represent a significant advantage for metabolomics in biomarker discovery. A smaller number of total observed measurements for any individual allows the application of more robust statistical testing methodologies and results in fewer false discoveries. False discovery (19) represents a severe limitation for generally profiling genes and proteins. The underlying math is simple. As the total number of observations per individual subject increases, the likelihood of separating groups of individuals purely by chance increases significantly. For instance, in a study of two groups of 100 subjects, the likelihood of separating the groups by random measurements is significantly higher if one is measuring 100,000 variables per individual than if one measures 1,000 variables per individual.

Another important advantage of metabolomics, especially for developing tests that rely on noninvasive types of samples, is that essentially any type of sample can be analyzed. Metabolites can be routinely measured in urine, feces, sweat, saliva, blood, tissue, etc., which can often be challenging sources for obtaining genetic or protein information. A number of papers have been published recently describing the use of metabolomics assessment of more creative sample types for disease (20–22).

Many biological effects of drugs and disease result from the overall health of an individual, as well as his or her environment, lifestyle, and

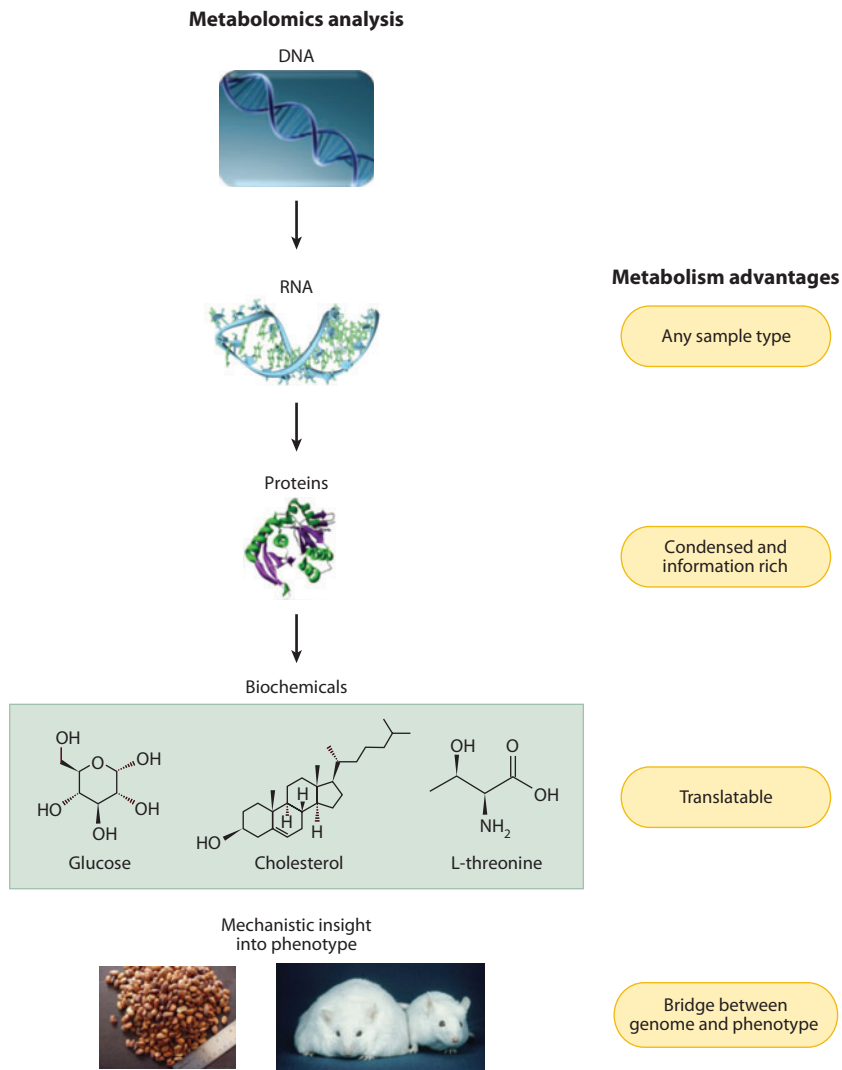


Figure 3

Progression of genes to metabolites and how each step fits into the “omics” technologies. In contrast to >30,000 genes and >100,000 possible transcripts and proteins, metabolites represent a far smaller total number, totaling <3,000. The consequence is that a discovery study in metabolomics is likely to be fraught with far fewer false discovery effects than these other technologies. In addition, nearly any type of sample will contain biochemicals (sweat, saliva, urine, feces, etc.). Perhaps the most striking difference is how closely the measurements of metabolites reflect the current phenotype of the test subject as compared to genomics and proteomics. This is likely why metabolomics results can be much more practical at solving metabolism and mechanism problems.

diet. Although genetics can play an important part in predisposing an individual to drug side effects or disease, the biochemistry of an individual is likely a more informative measurement of the individual’s current state

and condition. Combining genetic predisposition with the environmental and health status measurements that can be achieved with metabolomics will likely be a very important biomarker discovery method of the future.

THE NEED FOR SIMPLE INSULIN RESISTANCE BIOMARKERS

Insulin resistance (IR) is a well-established risk factor for type 2 diabetes (T2D) and cardiovascular disease (CVD) progression (23–29). IR and compensatory hyperinsulinemia are often associated with obesity. When coupled with beta-cell dysfunction, IR is a major pathophysiological determinant of hyperglycemia (impaired fasting glycemia and impaired glucose tolerance) and T2D (30, 31). Conditions associated with high CVD risk, such as hypertension, dyslipidemia, and atherosclerosis, have also been associated with IR (30–33).

Although IR plays a central role in the development of numerous diseases, it is not readily detectable using many of the clinical measurements for prediabetic conditions. IR develops prior to the onset of hyperglycemia and is associated with increased production of insulin. Over decades, the ability of the cell to respond to insulin decreases, and the subject becomes resistant to the action of insulin. Eventually the beta cells of the pancreas cannot produce sufficient insulin to compensate for the decreased insulin sensitivity. The beta cells begin to lose function and apoptosis is triggered. Beta-cell function may be decreased as much as 80% in prediabetic subjects. As beta-cell function decreases, the production of insulin decreases, resulting in lower insulin levels and high glucose levels in diabetic subjects. Vascular damage is associated with the increase in IR and the development of T2D.

Unfortunately, traditional clinical fasting plasma measurements do not assess IR directly. The gold standard for assessing IR is the HI clamp. This involves glucose and insulin infusions and requires insertion of two catheters into the patient, who must remain immobilized in the clinic for up to six hours (34). As a result, the procedure is typically performed only in a research setting, and the associated cost and time constraints severely limit its clinical utility. Fasting insulin and derived indices (HOMA, QUICKI) have been used to assess IR (35), but owing to the complicated calculations

required, these have not been widely adopted in routine clinical practice. The identification of novel biomarkers to detect subjects at risk of IR and to stratify the risk of progression to T2D and/or CVD in subjects with IR, in order to implement effective strategies for prevention as well as to monitor treatment response, remains an unmet need. Furthermore, a test for IR could be particularly useful in identifying patients who have significant IR levels despite normal A1C levels.

Biomarker Discovery for IR Using a Global Metabolomics Analysis

Pilot studies had demonstrated the feasibility of identifying IR biomarkers, so biomarker discovery and validation studies were designed. The first step was to identify a sufficiently large cohort of nondiabetic subjects in whom the HI clamp had been performed and for whom fasting blood plasma samples were available for metabolomic analysis. The RISC (Relationship of Insulin Sensitivity to Cardiovascular Risk) cohort, subjects of a large, observational longitudinal study initiated to address how IR may contribute to progression to T2D and CVD, met these criteria. The cohort included a spectrum of insulin sensitivity as determined by the HI clamp (36, 37). The study design for the discovery and validation of the IR biomarkers with the RISC cohort is illustrated in **Figure 4**.

As previously reported, for biomarker discovery, fasting blood plasma samples collected at baseline from 399 subjects from the RISC study, all of whom were nondiabetic and clinically healthy, were analyzed using metabolomics (38). Roughly half the subjects were male and half were female; all were matched to age and body mass index (BMI). Each subject had his or her insulin sensitivity assessed using the HI clamp, which resulted in a measured insulin-mediated glucose disposal rate, M_{FFM} (expressed as $\mu\text{mol} \times \text{min}^{-1} \text{kg FFM}^{-1}$, where FFM refers to fat-free mass). The measured M_{FFM} values ranged from high to low and included a spectrum of insulin

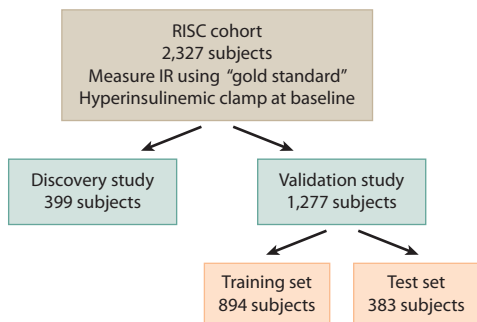


Figure 4

Experimental design schema illustrating the identification of a sufficiently large cohort, in this case the Relationship between Insulin Sensitivity and Cardiovascular Disease (RISC) Study, on which the hyperinsulinemic clamp test has been performed to serve as the biomarker discovery and validation cohort. The large number of subjects is partitioned into a smaller discovery study comprising 399 subjects and an independent validation study comprising a training set and a test set of subjects. IR, insulin resistance.

sensitivity from very insulin sensitive (high M_{FFM}) to insulin resistant (low M_{FFM}). Subjects with $<45 \mu\text{mol} \times \text{min}^{-1} \text{kg FFM}^{-1}$ belonged to the bottom tertile of the cohort analyzed and were defined as being insulin resistant whereas the top two tertiles ($M_{FFM} >45 \mu\text{mol} \times \text{min}^{-1} \text{kg FFM}^{-1}$) were defined as being insulin sensitive (38).

A total of 471 metabolites were measured in the blood plasma samples collected from the 399 subjects and analyzed for IR biomarkers (38). The authors assessed the ability to classify subjects as insulin sensitive or insulin resistant, with Random Forest analysis performed using the entirety of the screening data (i.e., 471 compounds). **Figure 5** shows the resulting Importance Plot, which ranks the classifiers by their contribution to the separation of the subjects into groups. This result did not change when normalizing the M value for kilograms of body weight rather than kilograms of FFM. A separate univariate correlation analysis of the screening data identified the correlation of the metabolites to the glucose disposal rate.

From these analyses, a panel of 30 candidate biomarkers was identified for further assay

development and evaluation (38). The candidate markers from this analysis are summarized in **Table 1**. For biomarker success, it was very important that two completely independent statistical treatments, in this case Random Forest and correlation analysis, resulted in the same top-performing metabolites and further confirmed the likely importance of these biochemicals as biomarkers for IR.

Since the initial analyses were based on data obtained from the nontargeted biochemical profiling technology, targeted isotopic dilution assays were developed for the panel of the 30 best candidate biomarkers to provide absolute quantitative results (38). These analyses showed a high correlation between the results from the screening platform and the quantitative targeted assay. Further, whether measured by the screening platform or the targeted isotopic dilution assay, the biomarker α -HB was consistently higher, and the biomarkers linoleoyl-lysoglycerophosphocholine (L-LPC) and oleate were lower, in insulin-resistant subjects than in insulin-sensitive subjects ($p < 0.0001$ for both the screening and targeted data). Thus, the targeted assay data validated the screening data in the test cohort.

Developing an IR Biomarker Panel into a Diagnostic Test

After the development and analytical validation of quantitative assays for the selected biomarker candidates, the next steps in developing the biomarkers into a clinically useful diagnostic test include (a) biomarker selection and algorithm development, (b) validation in an independent set of clinically relevant samples, and (c) demonstration of the clinical utility of the test.

The 30 biomarkers identified by Random Forest and Lasso Regression analyses as important to build models to predict IR (defined by the bottom tertile of M values in the cohort) are listed in **Table 1** (38). These cross-validated biomarkers were used to classify individuals as insulin resistant (bottom tertile M_{FFM}) or normally insulin sensitive (top two-thirds

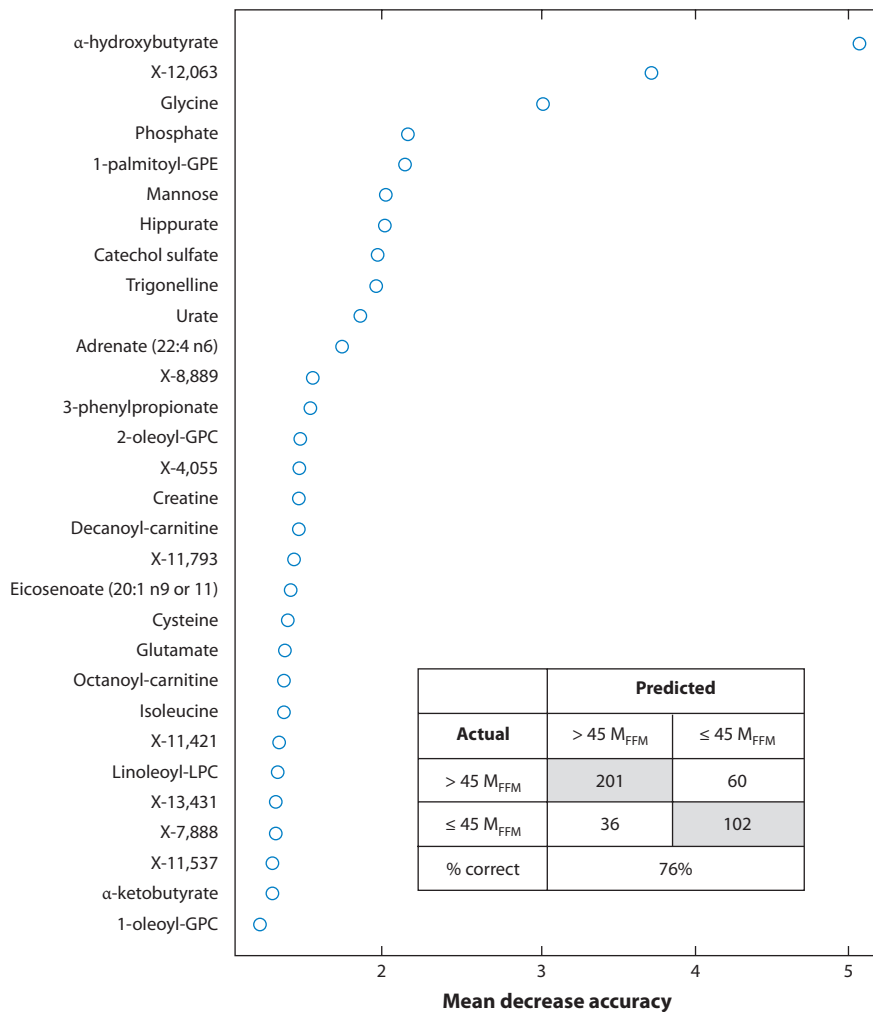


Figure 5

The results of Random Forest analysis of the previously reported (38) discovery study cohort. The Importance Plot shows the top-ranked biomarkers identified in the initial study. The inset table shows the results of the classification of subjects as being insulin sensitive (>45 M_{FFM}) or insulin resistant (≤45 M_{FFM}) using the biomarkers. The prediction accuracy of the separation of the top two-thirds, which are the insulin-sensitive subjects, from the bottom tertile, which represents the insulin-resistant subjects, is ~75%.

M_{FFM}). Samples from 90% of the subjects were used to build the model, and samples from the remaining 10% were used to test the predictive power of the model. These analyses identified biomarkers that are positively or negatively correlated with estimated M and ranked the biomarkers of most use to estimate M. These biomarkers are useful for determining insulin sensitivity and diagnosing IR. As shown in

Table 2, 164 subjects were classified as insulin resistant and 235 subjects were classified as insulin sensitive. These results indicate a sensitivity of ~85%, a specificity of ~91%, and an overall prediction accuracy of ~76% (38).

Next, an independent validation of the biomarkers was conducted using the targeted assays on fasting plasma samples from a

Table 1 Biomarker candidates

2-hydroxybutyrate	Linoleic acid	Stearate
3-hydroxybutyrate	Linolenic acid	Threonine
3-methyl-2-oxo-butyric acid	Margaric acid	Isovalerylcarnitine
3-phenylpropionate	Octanoyl carnitine	Linoleoyl-LPC
Catechol sulfate	Oleic acid	1,5-anhydroglucitol
Creatine	Oleoyl-LPC	Stearoyl-LPC
Decanoyl carnitine	Palmitate	1-palmitoyl-GPE
Docosatetraenoic acid	Palmitoleic acid	Octanoylcarnitine
Glutamic acid	Palmitoyl-LPC	alpha-ketobutyrate
Glycine	Serine	Cysteine

validation set of subjects. Samples from these subjects were also from the RISC cohort (36, 37) but had not been analyzed in the discovery study. The validation study comprised 1,277 fasting plasma samples obtained at baseline from the RISC participants, with 894 subject samples included in the training set and 383 subject samples in the test set (**Figure 4**). Continuous regression and logistic regression analyses were carried out to generate candidate diagnostic algorithms to estimate the M value. Shown in **Figure 6** is the spectrum of insulin sensitivity and the associated change in the selected biomarkers and comparisons with the measured M value, insulin, and BMI.

It is important in the development of any new test to benchmark the performance of the new analytes to previously used analytes. Part of the reason is to better understand the performance differences, both biologically and analytically. There may be simple analytical advantages for some biomarkers over others. BMI has been shown previously to be correlated with insulin, and using BMI or insulin with new analytes could improve the overall accuracy of IR assessment.

Analytical/Clinical Validation

After the diagnostic biomarkers and algorithms are validated and the clinical utility is demonstrated, the diagnostic test itself must be analytically validated before it can be used clinically. **Figure 7** illustrates the prototype development and design control activities that precede clinical and analytical validation, or final product validation for an MS-based diagnostic test.

As demonstrated by this case study, a non-targeted metabolomics analysis has utility for the discovery of new biomarkers and further supports the increased application of this new technology for biomarker discovery. Using metabolomics, several novel IR biomarkers, including α -HB, oleate, and L-LPC, were identified and validated, and a specific diagnostic assay and algorithm for fasting blood samples were developed for the assessment of IR in nondiabetic individuals. Furthermore, these biomarkers and test, used alone and in combination with additional biomarkers identified in this work, are likely to provide clinically useful diagnostic tests for identifying and managing IR and its associated disorders, such as T2D, CVD and the like.

Table 2 Classification of insulin-resistant (IR) subjects by hyperinsulinemic (HI) clamp and by using IR biomarkers

	Number determined by HI clamp	Number predicted by biomarkers
Insulin resistant	138 (true positive)	164
Insulin sensitive	261 (true negative)	235

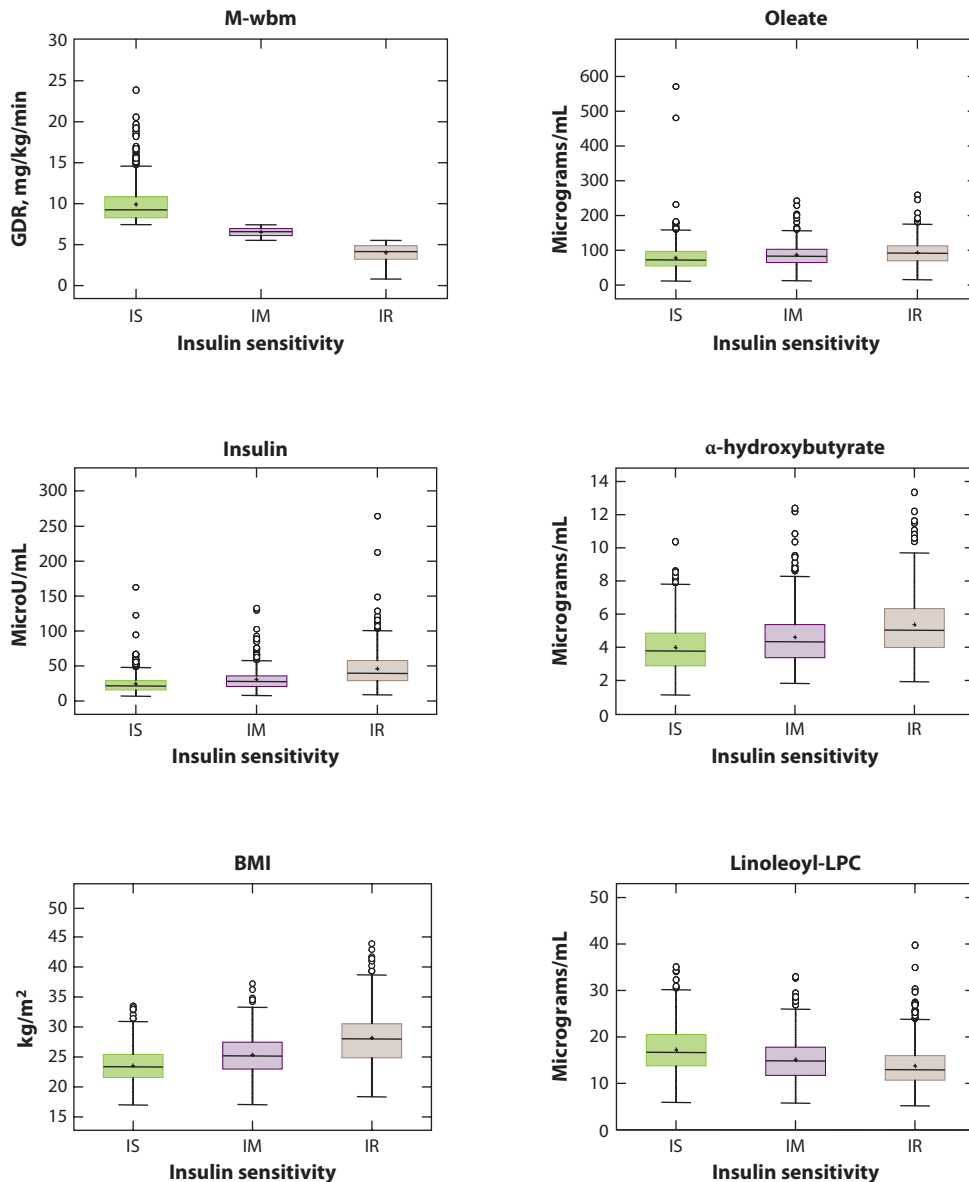


Figure 6

Association of insulin, body mass index (BMI), and insulin resistance (IR) biomarkers with various levels of insulin sensitivity. The box plots represent three categories of insulin sensitivity classification—insulin sensitive (IS) ($M\text{-wbm} > 7.5$ mg/kg/min), intermediate (IM) ($M\text{-wbm} > 5.6$ and < 7.5 mg/kg/min), and insulin resistant (IR) (< 5.6 mg/kg/min)—based on the gold-standard test of IR, the hyperinsulinemic clamp. GDR (glucose disposal rate): $M\text{-wbm}$ 5.6 mg/kg/min reflects the bottom tertile of insulin sensitivity in the entire baseline population used in training and test sets ($n = 1277$). In the box plots, the top and bottom of the box represent the 75th and 25th percentile, respectively; 75% of the data points lie within the boxed region. The top and bottom bars (“whiskers”) represent the entire spread of the data points, excluding “extreme” points, which are indicated with circles. The solid line in the box indicates the mean value and the + indicates the median value.

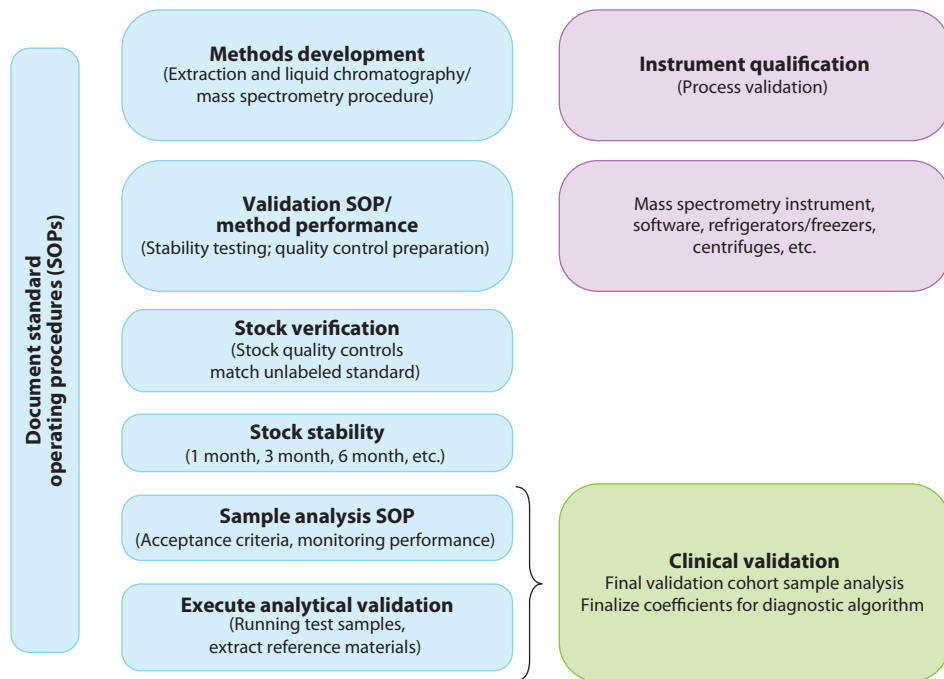


Figure 7

Prototype workflow describing the development and design control activities that are necessary for final validation of a multi-analyte diagnostic test. In this example, the test methods are based on liquid chromatography–mass spectrometry. A central requirement of design control is the development and documentation of standard operating procedures for each aspect of the diagnostic test.

CONCLUSIONS AND ACKNOWLEDGMENTS

Metabolomics offers a number of advantages that enable biomarker discovery and diagnostic test development. Profiling biochemicals in blood is relatively straightforward and is less likely than other “omics” technologies to be affected by false discovery. Profiling biochemicals facilitates understanding of metabolism and the effects of disease on metabolic phenotype, as well as drug alterations of metabolism in the liver, kidney, testes, and other organs, which also affect metabolic phenotype. Understanding the mechanisms of disease and drug effects will provide a wealth of potential new biomarkers.

Routine metabolomics could become the discovery method of choice, not only for diag-

nostic biomarker discoveries based on lifestyle and diet, but also for segregating biomarkers of population genetic differences in drug response and toxicity. One of the most exciting recent developments in metabolomics with potential to impact diagnostics is the new understanding of the penetrance of an individual’s genetics on metabolite levels in the blood (39, 40). The largest challenge to this growing technology is the early stage of the technology itself; few groups have developed fully validated global metabolomics methods, and many have very limited experience in using and understanding the data. It is hoped that as the importance of metabolomic studies is realized, a renaissance of biochemical learning will inspire a new breed of scientists who are as well versed in biochemistry as they are in molecular biology and genetics.

DISCLOSURE STATEMENT

M.V.M. and K.A.L. are employees of Metabolon, Inc.. The authors have no other 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 this article apart from those disclosed.

LITERATURE CITED

1. German JB, Gillies LA, Smilowitz JT, et al. 2007. Lipidomics and lipid profiling in metabolomics. *Curr. Opin. Lipidol.* 18:66–71
2. Ryals J, Lawton K, Stevens D, et al. 2007. Metabolon, Inc. *Pharmacogenomics* 8:863–66
3. Dehaven CD, Evans AM, Dai H, et al. 2010. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J. Cheminform.* 2:9
4. Evans AM, Dehaven CD, Barrett T, et al. 2009. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* 81:6656–67
5. Scalbert A, Brennan L, Fiehn O, et al. 2009. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 5:435–58
6. Römisch-Margl W, Prehn C, Bogumil R, et al. 2012. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 8:133–42
7. Lee Do Y, Bowen BP, Northen TR. 2010. Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging. *BioTechniques* 49:557–65
8. Bryan K, Brennan L, Cunningham P. 2008. MetaFIND: a feature analysis tool for metabolomics data. *BMC Bioinform.* 9:470
9. Bowen BP, Northen TR. 2010. Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* 21:1471–76
10. Dunn WB, Bailey NJ, Johnson HE. 2005. Measuring the metabolome: current analytical technologies. *Analyst* 130:606–25
11. Thielen B, Heinen S, Schomburg D. 2009. mSpecs: a software tool for the administration and editing of mass spectral libraries in the field of metabolomics. *BMC Bioinform.* 10:229
12. Xia J, Psychogios N, Young N, et al. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* 37:W652–60
13. Cloarec O, Dumas ME, Craig A, et al. 2005. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem.* 77:1282–89
14. DeHaven CD, Evans AM, Dai H, et al., eds. 2012. *Software techniques for enabling high-throughput analysis of metabolomic datasets, metabolomics.* doi: 10.5772/31277. <http://www.intechopen.com/books/metabolomics/software-techniques-for-enabling-high-throughput-analysis-on-metabolomic-datasets>
15. Brindle JT, Antti H, Holmes E, et al. 2002. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabolomics. *Nat. Med.* 8:1439–44
16. Sreekumar A, Poisson LM, Rajendiran TM, et al. 2009. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457:910–14
17. Kopka J, Schauer N, Krueger S, et al. 2005. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21:1635–38
18. Wishart DS, Tzur D, Knox C, et al. 2007. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35:D521–26
19. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57:289–300
20. Barnes VM, Teles R, Trivedi HM, et al. 2010. Assessment of the effects of dentifrice on periodontal disease biomarkers in gingival crevicular fluid. *J. Periodontol.* 81:1273–79
21. Ohta T, Masutomi N, Tsutsui N, et al. 2009. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol. Pathol.* 37:521–35

22. Wetmore DR, Joseloff E, Pilewski J, et al. 2010. Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. *J. Biol. Chem.* 285:30516–22
23. Ginsberg H, Olefsky JM, Reaven GM. 1974. Further evidence that insulin resistance exists in patients with chemical diabetes. *Diabetes* 23:674–78
24. Harris MI. 1995. Epidemiologic studies on the pathogenesis of non-insulin-dependent diabetes mellitus (NIDDM). *Clin. Invest. Med.* 18:231–39
25. Lyssenko V, Jonsson A, Almgren P, et al. 2008. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* 359:2220–32
26. Reaven GM. 1998. Insulin resistance and human disease: a short history. *J. Basic Clin. Physiol. Pharmacol.* 9:387–406
27. Reaven GM, Olefsky JM. 1978. The role of insulin resistance in the pathogenesis of diabetes mellitus. *Adv. Metab. Disord.* 9:313–31
28. Shen SW, Reaven GM, Farquhar JW. 1970. Comparison of impedance to insulin-mediated glucose uptake in normal subjects and in subjects with latent diabetes. *J. Clin. Invest.* 49:2151–60
29. Warram JH, Martin BC, Krolewski AS, et al. 1990. Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann. Int. Med.* 113:909–15
30. Abdul-Ghani MA, DeFronzo RA. 2009. Pathophysiology of prediabetes. *Curr. Diabetes Rep.* 9:193–99
31. DeFronzo RA. 2009. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* 58:773–95
32. Reaven GM. 1999. Insulin resistance: a chicken that has come to roost. *Ann. NY Acad. Sci.* 892:45–57
33. Reaven GM, Chen YI, Coulston AM, et al. 1983. Insulin secretion and action in noninsulin-dependent diabetes mellitus. Is insulin resistance secondary to hypoinsulinemia? *Am. J. Med.* 75:85–93
34. DeFronzo RA, Tobin JD, Andres R. 1979. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am. J. Physiol.* 237:E214–23
35. Muniyappa R, Lee S, Chen H, et al. 2008. Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. *Am. J. Physiol. Endocrinol. Metab.* 294:E15–26
36. Hills SA, Balkau B, Coppack SW, et al. 2004. The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and objectives. *Diabetologia* 47:566–70
37. Ferrannini E, Balkau B, Coppack SW, et al. 2007. Insulin resistance, insulin response, and obesity as indicators of metabolic risk. *J. Clin. Endocrinol. Metab.* 92:2885–92
38. Gall WE, Beebe K, Lawton KA, et al. 2010. alpha-Hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One* 5:e10883
39. Gieger C, Geistlinger L, Altmaier E, et al. 2008. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 4:e1000282
40. Suhre K, Shin SY, Petersen AK, et al. 2011. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477:54–60



Contents

Abiraterone and Novel Antiandrogens: Overcoming Castration Resistance in Prostate Cancer <i>R. Ferraldeschi, C. Pezaro, V. Karavasilis, and J. de Bono</i>	1
Antibody-Drug Conjugates in Cancer Therapy <i>Eric L. Sievers and Peter D. Senter</i>	15
Circulating Tumor Cells: From Bench to Bedside <i>Marija Balic, Anthony Williams, Henry Lin, Ram Datar, and Richard J. Cote</i>	31
Cytokines, Obesity, and Cancer: New Insights on Mechanisms Linking Obesity to Cancer Risk and Progression <i>Candace A. Gilbert and Joyce M. Slingerland</i>	45
Glioblastoma: Molecular Analysis and Clinical Implications <i>Jason T. Huse, Eric Holland, and Lisa M. DeAngelis</i>	59
Harnessing the Power of the Immune System to Target Cancer <i>Gregory Lizée, Willem W. Overwijk, Laszlo Radvanyi, Jianjun Gao, Padmanee Sharma, and Patrick Hwu</i>	71
Human Papillomavirus Vaccines Six Years After Approval <i>Alan R. Shaw</i>	91
Reduced-Intensity Hematopoietic Stem Cell Transplants for Malignancies: Harnessing the Graft-Versus-Tumor Effect <i>Saar Gill and David L. Porter</i>	101
The Need for Lymph Node Dissection in Nonmetastatic Breast Cancer <i>Catherine Pesce and Monica Morrow</i>	119
The Role of Anti-Inflammatory Drugs in Colorectal Cancer <i>Dingzhi Wang and Raymond N. DuBois</i>	131
The Human Microbiome: From Symbiosis to Pathogenesis <i>Emiley A. Elloe-Fadrosh and David A. Rasko</i>	145
The Rotavirus Saga Revisited <i>Alan R. Shaw</i>	165

Staphylococcal Infections: Mechanisms of Biofilm Maturation and Detachment as Critical Determinants of Pathogenicity <i>Michael Otto</i>	175
Toward a Universal Influenza Virus Vaccine: Prospects and Challenges <i>Natalie Pica and Peter Palese</i>	189
Host Genetics of HIV Acquisition and Viral Control <i>Patrick R. Shea, Kevin V. Shianna, Mary Carrington, and David B. Goldstein</i>	203
Systemic and Topical Drugs for the Prevention of HIV Infection: Antiretroviral Pre-exposure Prophylaxis <i>Jared Baeten and Connie Celum</i>	219
Hyperaldosteronism as a Common Cause of Resistant Hypertension <i>David A. Calhoun</i>	233
Mechanisms of Premature Atherosclerosis in Rheumatoid Arthritis and Lupus <i>J. Michelle Kahlenberg and Mariana J. Kaplan</i>	249
Molecular Mechanisms in Progressive Idiopathic Pulmonary Fibrosis <i>Mark P. Steele and David A. Schwartz</i>	265
Reprogrammed Cells for Disease Modeling and Regenerative Medicine <i>Anne B.C. Cherry and George Q. Daley</i>	277
Application of Metabolomics to Diagnosis of Insulin Resistance <i>Michael V. Milburn and Kay A. Lawton</i>	291
Defective Complement Inhibitory Function Predisposes to Renal Disease <i>Anuja Java, John Atkinson, and Jane Salmon</i>	307
New Therapies for Gout <i>Daria B. Crittenden and Michael H. Pillinger</i>	325
Pathogenesis of Immunoglobulin A Nephropathy: Recent Insight from Genetic Studies <i>Krzysztof Kiryluk, Jan Novak, and Ali G. Gharavi</i>	339
Podocyte Biology and Pathogenesis of Kidney Disease <i>Jochen Reiser and Sanja Sever</i>	357
Toward the Treatment and Prevention of Alzheimer's Disease: Rational Strategies and Recent Progress <i>Sam Gandy and Steven T. DeKosky</i>	367
Psychiatry's Integration with Medicine: The Role of DSM-5 <i>David J. Kupfer, Emily A. Kuhl, Lawson Wulsin</i>	385

Update on Typical and Atypical Antipsychotic Drugs <i>Herbert Y. Meltzer</i>	393
Ataluren as an Agent for Therapeutic Nonsense Suppression <i>Stuart W. Peltz, Manal Morsy, Ellen M. Welch, and Allan Jacobson</i>	407
Treating the Developing Brain: Implications from Human Imaging and Mouse Genetics <i>B. J. Casey, Siobhan S. Pattwell, Charles E. Glatt, and Francis S. Lee</i>	427
Genetic Basis of Intellectual Disability <i>Jay W. Ellison, Jill A. Rosenfeld, and Lisa G. Shaffer</i>	441
Sickle Cell Disease, Vasculopathy, and Therapeutics <i>Adetola A. Kassim and Michael R. DeBaun</i>	451
Duty-Hour Limits and Patient Care and Resident Outcomes: Can High-Quality Studies Offer Insight into Complex Relationships? <i>Ingrid Philibert, Thomas Nasca, Timothy Brigham, and Jane Shapiro</i>	467
Quality Measurement in Healthcare <i>Eliot J. Lazar, Peter Fleischut, and Brian K. Regan</i>	485

Indexes

Cumulative Index of Contributing Authors, Volumes 60–64	497
Article Titles, Volumes 60–64	501

Errata

An online log of corrections to *Annual Review of Medicine* articles may be found at
<http://med.annualreviews.org/errata.shtml>