Proteomics and Biomarkers in Neonatology
Xuefeng B. Ling and Karl G. Sylvester
NeoReviews 2011;12:e585-e591
DOI: 10.1542/neo.12-10-e585

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://neoreviews.aappublications.org/cgi/content/full/neoreviews;12/10/e585
Proteomics and Biomarkers in Neonatology

Xuefeng B. Ling, PhD,* Karl G. Sylvester, MD*

Abstract

Proteomic technologies and disease-specific biomarkers are being increasingly explored across diverse fields of medicine. The care of the neonate is defined by both a unique patient population and acquired postnatal morbidities that are largely a function of failed adaptation to postnatal life. However, most current diagnostic clinical tests for the neonate suffer from poor sensitivity and specificity or simply rely on a morphologic description of end-organ damage. In this review, we discuss proteomic technologies for the discovery and translation of biomarkers to clinical use, emphasizing unique potential neonatal disease applications.

Objectives  After completing this article, readers should be able to:

1. Explain the role of proteomic technologies in helping to diagnose neonatal disorders.
2. Distinguish between primary (specific) and secondary (downstream) biomarkers.
3. Describe specific proteins that may act as primary biomarkers in the diagnosis of neonatal disorders such as intraventricular hemorrhage, chronic lung disease, and necrotizing enterocolitis.
4. Discuss the potential use of biomarkers in drug and other therapies.

Unmet Clinical Need in Neonatology

In the presence of significant improvements in neonatal critical care and a broadening array of increasingly sophisticated treatment modalities, molecular diagnostics involving proteomic, genomic, and metabolomic technologies have yet to provide parallel contributions in neonatal care. However, neonatal care likely could benefit dramatically from the application of emerging molecular technologies in the areas of improved disease-specific treatments and monitoring of response to therapy. Most major sources of newborn and, more specifically, preterm neonate morbidity are manifested by infection or exacerbations in tissue inflammation (eg, retinopathy of prematurity, chronic lung disease). Importantly, a current inability to define the differences between sterile inflammation and infection in the neonate leads to a generalized treatment approach that can have additional unintended treatment-associated morbidity. (1) Accordingly, investigations into the biochemical and molecular alterations associated with neonatal sepsis and inflammation are increasing. These efforts seek both a better basic biologic understanding of the unique aspects of the neonate’s response to infection and novel tools to define and monitor sepsis and inflammation in newborns better. These investigations stand to benefit from the maturation of newer technologies in genomics, proteomics, and companion diagnostic instruments as well as an increasing understanding of the immune response in infants. (2) In this review, we discuss current proteomic and biomarker-based technologies and their potential applications to neonatal disease. The application of proteomics for the discovery and translation to practice of disease-specific

*Division of Pediatric Surgery, Department of Surgery, Stanford University School of Medicine, Stanford, CA.

Abbreviations

CRP: C-reactive protein
GC: gas chromatography
IFABP: intestinal-specific fatty acid-binding protein
IL: interleukin
LC: liquid chromatography
m/z: mass-to-charge
MS/MS: tandem mass spectrometry
MS: mass spectometry
NEC: necrotizing enterocolitis
PAF: platelet-activating factor
PB: primary biomarker
SB: secondary biomarker
SIP: spontaneous intestinal perforation
TB: theragnostic biomarker
protein biomarkers may provide a solution to continuing unmet clinical challenges in neonatology.

Defining Proteomics for the Discovery of Biomarkers

The terms proteomics and biomarker tend to be commonly used in tandem. In simple terms, a biomarker can be defined as a molecular indicator of a specific biologic property, biochemical feature, or facet that can be used to measure the progress of a disease or the effects of disease treatment. Proteomics can be operationally defined as a field of study that is focused on the identification of proteins, peptides, or their interactions and posttranslational modifications. Clinical proteomics is currently conducted to detect or select biomarkers of disease. Common examples include prostate-specific antigen (PSA) for monitoring prostate disease and glycosylated hemoglobin for monitoring diabetic glycemic control. Investigational proteomics is much broader in scope and uses diverse detection platforms in an effort to identify differences in protein expression or posttranslational changes in protein structure (eg, phosphorylation, acetylation) across a broad array of analytes.

Investigational proteomics can be used to interrogate targeted proteins and their modifications or to index the entire proteome of a cell, system, or organism. Mass spectrometry (MS) is the central analytic technique used for most investigational proteomics (Fig. 1). In brief, MS involves the use of an ionizing source of energy to excite the molecular constituents of a peptide, producing ions of varying size (mass) and charge. The product ions are separated according to their mass-to-charge (m/z) ratio in an electromagnetic field. A detector processes the ion signals into mass spectra using quantitative methods that can calculate the abundance of each ion present. Taken together, current MS instruments consist of three modules: an ion source, a mass analyzer, and a detector. An example of a current advancement is the ability to vary the source of ionization (eg, electrospray ionizer) or method for ion mass analysis (eg, orbitrap). (3)

Various options in specific MS technique must take into account the specific needs of the analysis, with such variables including mass accuracy, resolving power, sensitivity, dynamic range, throughput, and quantification as well as the detection of protein modifications. The output of most MS interrogations can be both qualitative and quantitative. In addition, MS can be used simply to determine the molecular mass of a protein or to determine specific structural features such as amino acid sequence or posttranslational modifications. In the latter, complex multistage instruments and analyses are used and commonly referred to as tandem MS (MS/MS). In MS/MS experiments, after the initial mass of target or precursor ions is determined, specific ions are targeted for further fragmentation to produce product ions (Fig. 1). MS and quantitative techniques are currently used for clinical metabolomics to detect inborn errors of metabolism. Liquid or gas chromatography (LC or GC) is an analytic chemistry technique that is commonly used with MS to facilitate the physical separation of analytes (proteins and metabolites) inline and before MS-based ionization. Together, the combined resolving power of LC-MS for the detection of specific chemicals (eg, proteins or metabolites) in complex mixtures renders this platform ideally suited for the discovery of biomarkers.

If the objective of a proteomic analysis is to discover a biomarker or a biomarker panel and reduce the biomarker(s) to clinical utility, three distinct steps must be undertaken (Fig. 2). During biomarker discovery, the proteome is compared between two distinct diseases or stages of disease (class labels of disease). A candidate list of differentially expressed features is derived from an index of proteins or peptide m/z ratios, compiled using sophisticated bioinformatics techniques to distill the massive datasets that may distinguish the two diseases. The analysis either can be supervised, whereby disease class labels are applied before indexing and model building, or unsupervised, whereby class labels are revealed after patterns of distinguishing features have been
identified, as in a typical cluster analysis in microarray experiments. Next, candidate biomarkers must be verified or qualified on an independent and larger set of samples. Verification requires repeating the interrogation to establish the reproducibility of the entire experiment, ensure detection of meaningful biologic differences, and minimize the likelihood that a subsequent list of candidate biomarkers pursued in validation are spurious findings. During verification, the candidate list of biomarkers may evolve to accommodate better-performing features and remove more poorly functioning proteins. During the third and final step, biomarkers are validated on still larger sample sets using an alternative means of detection that differs from the “discovery platform.” Thus, the iterative process from discovery to validation requires increasing numbers of samples (guided by simulation analysis) and an evolution in detection modality. Ideally, all three steps produce both qualitative (present or absent) and quantitative (amount) results that may be additive in defining the disease or state of disease under study. Typically, because the objective is to derive clinically meaningful biomarker(s), immune-based testing using antibodies (eg, enzyme-linked immunosorbent assay, Western blotting) is the modality of choice for biomarker validation. The guiding principle throughout the biomarker discovery and validation process is to derive a panel that has maximum specificity (low false-positive rate) and sensitivity (low false-negative rate) for the disease under study. The clinical utility of biomarkers is normally reported as area under the curve represented on a receiver-operator curve or the plotting of sensitivity versus 1-specificity.

Current proteomic discovery investigations are considered high content due to the comprehensive nature of the interrogation and the massive complexity of the proteome of any given system that may encompass variability in number of distinct proteins, splicing variants, posttranslational modifications, and a concentration range that can span 10 orders of magnitude. (4) Thus, the number of protein(s) and their associated modifications that can be detected produces multidimensional data-handling challenges. The high cost and labor-intensive processing of samples frequently precludes high-throughput applications, thereby limiting the size of the discovery cohort.

The workflow for MS-based biomarker discovery involves several distinct steps. Initially, proteins from biologic samples are isolated and fractionated. Because many biologic samples (eg, plasma) obtained from individuals for discovery can be complex and dominated by several highly abundant proteins (eg, albumin or immunoglobulin G), an initial sample preparation step normally is undertaken that can either fractionate the total matrix according to specific biochemical properties (eg, size, charge, pH) and or remove dominant species. One objective of preprocessing is to reduce sample complexity to narrow the dynamic range subject to “discovery” as potential biomarkers. Specifically, when plasma or serum is being interrogated, most of the proteome (>90%) is comprised of albumin and immunoglobulin that is likely to obscure the search for low-abundance species from which most useful biomarkers are likely to be found. (4) In a bottom-up approach, proteins are subject to enzymatic proteolysis before MS. In a top-down analysis, naturally occurring peptides are subject to
MS profiling directly. The differentially expressed naturally existent peptides or proteins (features) can be subsequently determined through qualitative and quantitative MS and MSMS analysis. LC is normally paired with MS to provide multidimensional protein separation before MS. As an alternative, a targeted MS analysis of enriched or partially purified proteins can be performed following one- or two-dimensional gel electrophoresis of a complex mixture. Fractionation of biologic samples using gel electrophoresis is, in essence, based on size and charge, as would be the case when using LC.

Challenges in Biomarker Discovery and Clinical Application

Significant challenges arise in the search for clinically useful biomarkers because the biologic starting material is frequently highly complex and encompasses a wide dynamic range of proteins and peptides. Moreover, unlike in the search for cancer biomarkers, for many neonatal diseases the tissue of interest may not be available or able to be sampled for biomarker discovery or measurement. Thus, peripheral sampling of blood, urine, cerebrospinal fluid, and other bodily fluids must be used, which inherently narrows the spectrum of possible biomarker discoveries to proteins that are either circulating (in the cases of plasma) or filtered through the glomerulus when urine is the substrate.

In general, a proteomic interrogation can be either “unbiased” or employ the perhaps less complex “candidate approach.” In the unbiased approach, the search for biomarkers is not limited to specific known proteins and should not be constrained by technical limitations of the discovery platform. In an unbiased search via the shotgun approach, the protein sample is digested, producing numerous peptide fragments that are then subjected to tandem rounds of MS/MS. This approach can provide accurate quantification of identified peptides but presents a tremendous informatics challenge to decipher the large number of resulting mass spectra and determine protein identification. This process involves indexing and database searches to arrive at an accurate determination of significant protein differences. Conceptually, when considering the interrogation of either the proteome or peptidome of a specific body fluid, several practical issues must be considered and expectations reconciled.

In the candidate approach, the search for biomarkers is restricted to known proteins that are unique to the disease or host response to disease. A targeted approach using a platform in current use involves a cross-sectional examination of cytokines and chemokines that are reflective of the host immune response. The technology uses color-coded, bead-conjugated antibodies as a stationary phase assay to measure and report the simultaneous (multiplex) presence and quantity of up to 100 cytokine/chemokine proteins that have well-defined functions. Customized technology-based experiments can be constructed, and the platform can service hypothesis-driven studies of the cellular and humoral immune response in various diseases of clinical interest.

The critical care of newborns may be amenable to improved disease stratification (specific diagnosis or prognosis) through the identification and use of biomarkers. For example, current clinical practice for monitoring the host response to inflammation and infection includes examination of the total and differential white blood cell counts and C-reactive protein (CRP). As an alternative, if the intent is to identify proteins that may serve as a biomarker with specific clinical utility, the analysis may be confined to biomarkers that are highly specific or more stringently restricted to the organ of interest (eg, brain, gut, lung). One conceptually appealing framework for biomarker clinical utility is to divide biomarkers into primary biomarkers (PBs) (specific) and secondary biomarkers (SBs) (downstream) based on the biologic information that they report or reflect. A PB can be defined by an ability to identify and monitor a specific molecular or biochemical effect that is central to the pathology of interest. Specific examples of these might be proteins that have organ-specific expression in the brain for monitoring intraventricular hemorrhage, the lung for reporting on chronic lung disease, or the gut for the early detection of necrotizing enterocolitis (NEC). Alternatively, a PB may monitor the action of a drug such as a kinase inhibitor, and the phosphorylation status of a known target protein may be monitored by a pharmacodynamic biomarker. An SB would be used to examine and monitor downstream effects that occur as a result of disruption of the primary process of interest. Another useful term related to these concepts is the “theragnostic biomarker” (TB). A TB could be used to monitor the effect of a therapeutic intervention. For example, serial CRP measurement could be used as a TB or SB both to guide the potential need for therapy and to determine the neonate’s response to treatment in various infectious/inflammatory settings. However, to be truly useful, a PB or TB should reflect the central pathologic process that is being monitored. By these criteria, CRP monitoring, therefore, is a poor choice because as an acute-phase reactant, CRP is highly nonspecific.

Biomarkers and Specific Examples of Clinical Utility in Neonatology

Inflammatory mediators play a distinct role in the host’s response to infection. To date, most molecular studies of neonatal infection have confined investigations to largely known or “candidate” markers of infection, including
chemokines, cytokines, white blood cell surface antigens, and acute-phase reactants. (6) Taken together, these markers of infection have been used to monitor the host response to infection or inflammation, but they have no role or utility in identifying the presence or source of infection. As such, biomarkers that reflect the host response to inflammation and or infection could be best considered as SBs. For the specific detection of infectious organisms, including both bacteria and viruses, clinicians rely on blood culture or increasingly, polymerase chain reaction-based detection methods. However, each of these methods, and in particular culture-based methods, has significant inherent limitations, particularly in newborns. These include low sensitivity; prolonged lag time for the provision of useful results; and the need for blood sampling that is hard to obtain either in sufficient volume, serially (for longitudinal monitoring), or from different sources (eg, peripheral heelstick × 2, catheter, cerebrospinal fluid). Therefore, in addition to being sensitive and specific, optimized biomarkers of disease should ideally be readily obtained via minimally invasive methods that are amenable to serial measurement at low cost and with quick turnaround time.

To illustrate the potential for biomarker application to neonatal care, consider the clinical conundrum presented by neonatal sepsis, NEC, and sterile inflammation. The presentation of infection/sepsis in newborns can be very nonspecific, inconspicuous, and obscure in the early stages, often resembling exacerbations of other leading sterile inflammatory neonatal morbidities such as chronic lung disease. The unique biology of the neonate tends to compound these problems because of the heterogeneous response that the naive newborn immune system manifests in response to inciting agents. (2) Similarly, currently used clinical criteria, radiographic findings, and laboratory values for diagnosing NEC are only specific after disease onset. Moreover, the clinical and biochemical abnormalities in the earlier stages of NEC (Bell stages I and II), are common among the population at risk (sick neonates) and approximate a systemic inflammatory response that is indistinct from sepsis. The pathognomonic radiographic finding of pneumatosis intestinalis suffers from low sensitivity and vulnerability toward diminished specificity due to low concordance among interpreting radiologists. (5) The ability to identify biomarkers of disease before disease onset or in heralding the onset of progressive disease as early as possible is ideal. If prospective biomarkers that predict or forecast fulminant disease could be identified, new opportunities for possible earlier intervention or to prevent the progression or halt the onset of NEC, for example, could be undertaken. Additional opportunities include the need for specific delineation between early NEC and sepsis as well as differentiating sepsis and sterile inflammation. The identification of biomarkers that could define these entities on a molecular scale could lead to more specific treatment regimens and a possible ability to guide therapy, including the need for antibiotics and duration of antibiotic use.

**Recent and Current Markers of Disease**

Given the unique biology and clinical challenges presented by NEC, considerable effort has been expended by numerous investigators in the search for both PBs and SBs. CRP is commonly used as an SB by clinical neonatologists during the evaluation and screening for NEC and sepsis. CRP concentrations have been found to be elevated in Bell stages II and III NEC, and an association with progression to perforated NEC has been reported when CRP values remain elevated following the initiation of medical management (eg, antibiotics, discontinuation of enteral feedings). (8)(9) In practice, therefore, CRP is being used as both an SB and a TB. In this regard, CRP is a sensitive but highly nonspecific indicator of NEC, thus failing as a strong clinical biomarker because it cannot provide a specific diagnosis and guide therapy in early cases.

Another example of an SB or host response candidate for NEC is platelet-activating factor (PAF). (9)(10)(11) PAF is a potent proinflammatory phospholipid produced by platelets, leukocytes, and endothelial cells as a mediator of inflammatory pathophysiology, including sepsis. Several studies have reported good sensitivity and specificity for PAF in diagnosing NEC. (9)(10)(11) However, these studies each had differing control groups and examined different stages of NEC. Moreover, despite being first introduced in 1990 as a possible mediator of NEC, no large validation trials have determined that PAF is a useful biomarker.

Examples of published candidate primary biomarkers of NEC include the proteins intestinal- and liver-specific fatty acid-binding proteins, fecal calprotectin, and Claudin-3. Intestinal-specific fatty acid-binding protein (IFABP) has been reported to be a reliable plasma and urinary marker for enterocyte injury. (12)(13) Claudin-3 is a cellular tight junction protein that has been detected in the blood and urine of patients who have inflammatory bowel disease. Calprotectin is a peptide that is released from neutrophils upon inflammatory activation in the gut and is detectable in both the feces and plasma. Both calprotectin and Claudin-3 have been evaluated in very low-birthweight infants as potential specific biomarkers of NEC. (13)(14) Although these peptides show some promise as PBs, as reflected by favorable disease likelihood ratios, the number of cases examined is small, and almost all of the studies were conducted in patients in whom NEC was suspected. Similarly,
IFABP has shown some promise as a marker of disease or severity of disease in a few studies involving infants in whom disease was suspected. (13)

Nearly all of these studies have used a candidate approach to biomarker qualification, have been conducted in single centers that had small numbers of infants, and lacked the inclusion of powerful control infant cohorts who had inflammatory illness that was not NEC (eg, sepsis). One additional recent study used an unbiased MS-based approach to biomarker discovery and reported a composite NEC-sepsis score for very low-birthweight infants with proapolipoprotein CII and serum amyloid A. (15) The investigators employed several clinical parameters and novel biomarker discoveries to define a subset of infants who had NEC and sepsis together, but they made no attempt at distinguishing the two disease states.

Further potentially complicating the generalized classification of being either a PB or SB, all potential biomarkers, including those that have found some clinical utility in neonatology such as CRP, have additional known and unknown roles that potentially confound their clinical utility. For example, CRP is produced by hepatocytes in response to an increase in interleukin (IL)-6 that occurs as an early T-cell response to bacterial infection. (6) The combined early peak and short half-life of IL-6 that diminishes substantially within 24 hours of treatment and secondary response of CRP that peaks 24 hours after the onset of infection argues for development of a panel or combination of markers to provide the greatest sensitivity, specificity, and negative predictive value of the disease or its process in evolution.

As an alternative approach, consider the related clinical entities NEC and spontaneous intestinal perforation (SIP) that frequently are obscured by their similar presentation of enteric perforation. The pathophysiologic of both NEC and SIP involves a degree of intestinal injury, but it is self-limited and without significant systemic inflammatory response in SIP and involves progressive inflammation and more widespread intestinal necrosis in NEC. These differences suggest the opportunity for combining organ-specific (PB) and host response differences (SB) to provide an ensemble or integrated method for differentiating these disease entities. An ensemble approach that employs both clinical parameters and protein biomarkers to diagnose or risk-stratify infants may provide the most informative and powerful approach. More recent deeper understanding of the basic biology and the possibility of developing newer integrated markers (ensemble approach) hold promise for the development of biomarker panels that would assist in deciding whether to start or withhold antibiotic treatment. An example of this type of decision support through the use of biomarkers was provided by a group that measured CRP in combination with plasma IL-8 concentrations to guide the need for antibiotic use in the setting of possible early-onset or perinatal sepsis. In this study, the authors reported a significant reduction in the use of antibiotics through specific measurement of IL-8. (16)

Summary and Conclusions
A survey of the most prominent morbidities and sources of mortality in the neonatal intensive care unit suggests that most, if not all, are amenable to biomarker development. Infection, chronic lung disease, intraventricular hemorrhage, retinopathy of prematurity, NEC, and their consequences are all either caused by or result from inflammation and organ dysfunction associated with developmental stage-related frailty. Currently, combinations of various imaging modalities and empiric treatment strategies are used to track morphologic changes that are frequently reactive or due to a process that is progressive and has progressed. Stated differently, current modalities do not provide timely information that can allow for interventions that might prevent continued deterioration in end-organ function. This leads to a reactionary treatment approach that seeks to limit further injury and possibly reverse current findings and disruptions. Increased understanding of neonatal disease initiation and progression on a molecular level offers the opportunity both to detect disease and monitor its progression more specifically. One very appealing utility of biomarkers in the future is to define potential subject cohorts for clinical trials and as potential surrogate endpoints. Biomarkers also should play a larger and pivotal role in drug development or other therapeutics by defining the molecular endpoints or measures of desired effect. The interest in and use of biomarkers in neonatology appears poised to expand and offer potential for significant benefit to these fragile patients.

References
1. Cotten CM, Taylor S, Stoll B, et al; NICHD Neonatal Research Network. Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis...
12. Proteomics is defined as a field of study that is focused on the identification of proteins, peptides, or their interactions and posttranslational modifications. Clinical proteomics is currently being applied for detection of biomarkers of disease. Of the following, the most commonly used biomarker in neonatal-perinatal medicine is:

A. Glycosylated hemoglobin.
B. C-reactive protein.
C. Interleukin-8.
D. Platelet-activating factor.
E. Surfactant protein B.

13. Investigational proteomics uses diverse detection platforms to identify differences in protein expression or posttranslational changes in protein structure across a broad array of analytes. Of the following, the central analytic technique used most often for protein biomarker discovery is:

A. Gel electrophoresis.
B. Immune-based antibody testing.
C. Polymerase chain reaction.
D. Liquid chromatography–mass spectrometry.
E. Microarray analysis.