Biomarkers in Neonatology: The Next Generation of Tests

Pak C. Ng    Hugh S. Lam

Department of Paediatrics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, SAR, China

Key Words

Biomarkers · Newborns · Nucleic acids · Organ-specific biomarker

Abstract

Over the past two decades, neonatal clinicians have commonly used host response biomarkers to diagnose and assess the severity of systemic infection. Most of these biomarkers, such as acute-phase proteins or cytokines, are non-specific immunomodulating mediators of the inflammatory cascade. With advances in biochemical/genetic research, it is anticipated that future biomarkers will be ‘organ and/or disease specific’. There is also the quest for discovery of ‘novel’ biomarkers to assist diagnosis and prognosis of neonatal diseases using powerful mass-screening techniques, e.g. the next-generation sequencing, proteomics and arrays. This article aims to introduce the concept of the next generation of biomarkers to practising neonatal clinicians, and, hopefully, to integrate basic science research into day-to-day clinical practice in the future.
not usually confined to one specific organ, most investigators focus on up- or down-regulation of immunomodulating proteins in the inflammatory cascade as surrogate biomarkers for identifying infections or infection-related conditions such as NEC [1, 2]. Although these proteins, e.g., interleukins and chemokines [3, 4, 6], are not specific to a particular disease or infection, fortunately they are reasonably accurate in positively identifying neonatal infections as, apart from sepsis, there are relatively few inflammatory conditions that will substantially influence their circulating levels in term and preterm infants. For example, connective tissue and rheumatoid diseases, which often affect the immunological pathway and inflammatory cascade, do not present at such an early stage of life. However, tissue inflammation and necrosis induced by surgery can easily mask the biochemical signals of underlying sepsis. Whilst this approach can successfully identify genuine septicaemia in the majority of suspected cases, these ‘non-specific’ biomarkers fail to differentiate sepsis from NEC [5], which carries different management strategies and long-term outcomes. Similarly, an increase in concentrations of pro-inflammatory cytokines in tracheal aspirate specimens of ventilated preterm infants merely signifies inflammation in the respiratory tract [11]. It does not provide precise information on the type of cell damaged or the exact location of injury. In addition, the conventional approach to validate biomarkers in most current studies is to employ the ‘candidate approach’ of selecting known proteins/mediators and subject them to assessment in a case-control or cohort study to work out the diagnostic utilities (i.e. sensitivity and specificity for case-control studies, plus positive- and negative-predictive values for cohort studies) [12]. This approach is efficient but limiting, and suffers from the major drawback of evaluating only archived mediators. Novel biomarkers could not be discovered with such a study design. Thus, in recent years, investigators have fine-tuned their objectives towards the discovery of novel biomarkers [13], and using new biochemical technologies for more precise identification of specific diseases or specific organ pathologies.

**Disease- or Organ-Specific Biomarkers**

**Proteins**

Specific Protein Biomarkers

The identification of circulating proteins specific for organ injury or disease has become the prime objective in biomarker research. A typical example is the use of troponin for the diagnosis of acute myocardial infarction in adults [14]. Cardiac troponins are structural proteins bound to myofibrillar filaments of the cardiac muscle. An acute increase in the circulation following myocardial injury is due to a slow release of such proteins from damaged cardiac cells. The positive correlation between circulating troponin concentration and the magnitude of myocardial injury after heart surgery or cardiotoxic drug treatment renders this class of protein useful prognostic indicators of morbidity and mortality in the paediatric age group [15]. The same principle applies to term and preterm newborns for evaluating the extent of myocardial injury in newborns with perinatal depression and those suffering cardiopulmonary compromise secondary to respiratory distress syndrome. Recent studies suggested that circulating troponin T concentrations were significantly higher in infants with the most severe grade (Sarnat stage III) of neonatal encephalopathy compared with those with lower grades (Sarnat stages I and II) and healthy controls [16]. Similarly, significantly higher levels of troponin T were found in perinatally depressed infants with low ejection fraction, heart failure and those who did not survive [17, 18]. Preterm infants with respiratory distress syndrome, poor myocardial function and hypotension requiring inotropes also had elevated troponin T levels [19]. Troponin T is more cardioselective than conventional creatine kinase-MB and lactate dehydrogenase as a cardiac biomarker and undoubtedly a more precise indicator of cardiac injury.

Another example is the use of gastro-intestinal selective biomarkers for verifying enterocyte injury or cell death in preterm infants presenting with abdominal features, and differentiating life-threatening NEC from other benign neonatal abdominal conditions with similar clinical presentations such as gastro-intestinal dysmotility. Intestinal fatty acid-binding protein (I-FABP), a specific indicator of early enterocyte cell death [20, 21], is expressed by enterocytes in both small and large bowels and is liberated into the bloodstream as soon as the integrity of the enterocyte cell membrane is compromised. It is a small molecule and can readily pass through renal glomeruli. Recent studies found that the urinary I-FABP:creatinine concentration ratio was disproportionately elevated in infants who developed NEC or intestinal necrosis compared with those who did not [22, 23]. Thus, I-FABP is considered to be an organ-specific biomarker for intestinal mucosal injury and may potentially be developed as a screening test for NEC in newborns [22, 23].

**Brain-Specific Proteins.** Neurofilament triplet protein, glial fibrillary acid protein and S-100 were found to be 4- to 200-fold higher in infants with post-haemorrhagic hy-
hydrocephalus (PHH) than those without PPH. All these proteins were significantly higher in PHH infants who had parenchymal damage, died or were disabled at 1 year of age than in those who had PHH without parenchymal lesions, survived and were without disability at 12 months of age. Other proteins in the cerebrospinal fluid, such as transforming growth factor-β1, thrombopoietin or interferon-γ, are not brain specific, but their presence in the compartment is likely the result of physiological responses to intraventricular haemorrhage [24]. These proteins can reflect the inflammatory activity, extracellular matrix activity and the extent of cell death during the insult [24]. Thus, they have been used for the prediction of PHH and severity of brain injury [24]. Another example is the targeted assessment of different types of lung cell injury that can provide a distinctive blood biomarker profile, which enables respirologists to differentiate idiopathic pulmonary fibrosis from other causes of acute lung injury in adult patients [25].

Although investigators are progressing in the search of specific biomarkers for different diseases in neonatology, it is important: (1) to establish the normal range of the biomarker in term and preterm newborns, (2) to recognize the maturational and developmental differences in expression of biomarkers in adults, children and newborns, e.g. troponin I is not fully expressed in the neonatal myocardium until 9 months of age [26], (3) to recognize the differences in the disease spectrum between adults and newborns, e.g. acute myocardial infarction in adult patients versus cardiac ischaemia in perinatally depressed newborns, (4) to verify the influence of perinatal factors, such as mode of delivery, Apgar scores, gestational age and use of antenatal drugs on the expression of biomarkers, and (5) to recognize the limitations of ‘specific’ biomarkers, e.g. systemic hypotension or septicaemic shock, which compromises the hepatosplanchnic circulation, may potentially result in the release of gut barrier biomarkers into the bloodstream without causing NEC.

**Nucleic Acids**

The discovery of circulating nucleic acids in plasma has revolutionized biomarker research and initiated unbounded opportunities for diagnostic and prognostic applications in all branches of medicine. The spectrum includes tumour-derived DNA in cancer research, assessment of body organ injury in traumatology, monitoring the progress of organ transplantation and non-invasive prenatal diagnosis using maternal blood samples [27].

**Deoxyribonucleic Acid**

**Cell-Free Plasma DNA**

The greatest advantage of using circulating cell-free fetal DNA in maternal plasma is to provide direct genetic information of the fetus in a non-invasive manner. Cell-free fetal DNA constitutes about 3–6% of total DNA in the maternal plasma. Demonstration of the Y chromosome-specific sequence in maternal plasma can accurately determine the fetal gender and is vital for the prenatal diagnosis of X-linked diseases, haemophilia [28]. Accurate genotyping of fetal Rhesus blood group in Rhesus-negative mothers has reduced unnecessary use of immunoglobulin prophylaxis and invasive amniocentesis in high-risk pregnancies [29]. This type of screening test has already been translated into routine clinical practice in some countries. Also, positive identification of the DNA sequence of autosomal dominant diseases, such as Huntington’s chorea [30] and myotonic dystrophy [31], accurately indicates inheritance of the disease in the fetus, though a negative survey does not rule out the possibility of inheritance if the mother is a carrier. Further, an elevation of fetal DNA concentration in maternal blood may precede the onset of symptoms in pre-eclampsia [32, 33] and preterm labour [34], presumably due to increased fetal cellular trafficking or decreased fetal DNA clearance in pregnancies with complications [35, 36]. Thus, recent advances in the evolution of circulating cell-free DNA as molecular biomarkers have opened up a new avenue for future assessment of inherited diseases in utero.

Our research team has also used Gram-positive and Gram-negative organism-specific nucleic acids for identifying and differentiating these two types of septicaemia in preterm infants [10]. None of the sepsis episodes were classified into the wrong Gram-specific category. More importantly, despite negative blood cultures in 5 infants suffering from intra-abdominal sepsis, the quantitative polymerase chain reaction (PCR) test could identify the Gram-specific category of the causative organisms in blood. Thus, the test is highly specific and a positive result is able to demonstrate bacterial bloodstream infection with high precision [10]. A recent study on the broad-range 16S gene using real-time PCR also demonstrated that 7 infants with positive PCR results and negative blood cultures had bacterial sepsis [37].

**Metagenomics**

Another fast-developing technique is the use of metagenomics for molecular identification of a vast array of micro-organisms. For example, high-throughput 16S rRNA sequencing has been used to obtain specific bacterial se-
quence signatures in stool samples for evaluating the status of the gut micro-environment of microbial colonisation. This vital information can theoretically improve the ability to diagnose and treat gut microbial-mediated conditions [38]. A recent study suggested that a significant change in intestinal luminal microbiota with upsurge of Proteobacteria and decrease in Firmicutes was observed in preterm infants who subsequently developed NEC [39]. The recognition of specific bacterial sequence signatures can lead to better understanding of the aetiology of NEC, and this technique may potentially transform into an efficient diagnostic tool or a prophylactic measure. Hopefully, neonatologists in the future may screen for foreign, e.g. viral, bacterial and fungal nucleic acids in blood samples of septicaemic babies and obtain accurate information for the identification of the causative organism plus antimicrobial sensitivity profile without the need for conventional medium culture or serological test.

Next-Generation Sequencing Technique

The development of next-generation sequencing (NGS), has substantially improved the efficiency and sensitivity of nucleic acid sequencing and detection in plasma [40]. This breakthrough in technology using maternal plasma has recently been translated into the future non-invasive screening test for detection of trisomy 21 fetuses. The principle is to show the presence of an increased amount of chromosome 21 sequences in maternal blood. The difficulty of measuring a minute increment in chromosome 21 DNA concentration has been overcome by NGS [41], which can accurately identify and quantify millions of DNA fragments in the blood sample within a short time period (days) [42]. In a multicentre study of 753 high-risk pregnancies using a designated protocol (2-Plex Protocol), Down’s fetuses were detected with extremely high accuracy and diagnostic utility (100% sensitivity, 97.9% specificity, 96.9% positive and 100% negative predictive value) [43]. It is anticipated that only a very small percentage of all pregnant women (estimated 0.1%) will in the future require referral for chorionic villus sampling and amniocentesis for genetic confirmation [43]. The same technique may be applied for detecting other trisomies, such as trisomy 18 or 13. It is not difficult to foresee that in the near future, perinatologists will be able to perform a non-invasive genetic and mutation scans for the whole fetal genome, and maternal plasma nucleic acid sequencing will become an integral part of routine prenatal screening for genetic diseases [43]. The same principle can apply for RNA sequencing and is termed ‘transcriptome analysis’.

Ribonucleic Acid

Messenger RNA

Analysis of cell-free plasma RNA offers another dimension for the development of disease-related biomarkers. The messenger RNA (mRNA) expression profile unique to a disease or tissue may be exploited as a specific diagnostic tool. The detection of circulating RNA offers a number of advantages over circulating DNA [44]. First, if an RNA transcript unique to a specific organ/tissue is chosen, this method is usually more applicable to the disease of that organ because production of mRNA depends on gene expression, which varies depending on the cell type. Secondly, quantitatively more RNA than DNA will be released from the injured cells. This is due to the fact that many copies of a RNA transcript are present in each cell, whereas only a single diploid genome equivalent of DNA is contained in one cell. In one study, liver-derived albumin (ALB) mRNA was evaluated as a diagnostic biomarker for detecting liver pathologies [45]. The results suggested that plasma ALB mRNA was significantly increased in patients with hepatocellular carcinoma, cirrhosis and active chronic hepatitis B compared with healthy subjects. Overall, plasma ALB mRNA was significantly increased (cutoff 835 copies/ml) in 73.8% of patients with liver pathologies whereas only 21.5% had abnormal alanine aminotransferase levels. Similarly, 91.4% patients with hepatocellular carcinoma had elevated plasma ALB mRNA concentrations, whereas only 48.6% had abnormal serum $\alpha$-fetoprotein levels [45]. Much evidence has now accumulated to suggest that the release of mRNA into plasma is the consequence of cell death. Thus, plasma ALB mRNA may be considered to be a more sensitive biomarker for diagnosing liver cell death than the conventional liver function test or $\alpha$-fetoprotein. In addition, the investigators have demonstrated that a proportion of the supposedly liver-specific ALB mRNA detected in whole blood was in fact derived from haematopoietic cells due to illegitimate gene transcription, and that only ALB mRNA detected in plasma was liver specific [45]. A similar principle of using mRNA can, therefore, apply to investigate liver disease, e.g. parenteral nutrition-associated cholestasis or other organ-specific injuries in neonatal patients.

MicroRNA

The microRNA (miRNA) molecules that circulate in human plasma are another recent discovery. This is a class of small (about 22 nucleotides) non-coding RNA fragments that inhibit gene expression, thereby halting the translation machinery that produces functional pro-
miRNA) have been identified with specific organ dysfunction and diseases. Consequently, miRNA has been used as a diagnostic biomarker for detecting sepsis [47], drug-induced liver injury [48], myocardial injury [49], and malignancies [50]. Investigators have also identified specific miRNAs for differentiating different diseases within an organ. For example, heart-associated miR-208b and miR-499 were increased 1,600- and 100-fold, respectively, in acute myocardial infarction due to cardiomyocyte injury [49]. This specific profile of miRNA expression may enable clinicians to differentiate patients with acute myocardial infarction from those with other cardiac conditions, such as viral myocarditis, acute heart failure and diastolic dysfunction. Also, in a separate study, intestinal region-specific miRNAs (e.g., terminal ileum- and colon-specific miRNA) have been identified [51]. The findings suggested that different miRNAs were differentially expressed in different diseases and different sites of the bowel, including ulcerative colitis, Crohn’s colitis and Crohn’s ileitis [51]. Thus, we may soon be able to locate precisely the site of active disease and to differentiate closely immunologically related conditions, such as Crohn’s disease and ulcerative colitis, with blood biomarkers. In principle, similar techniques may also apply to newborns with NEC, thus assisting neonatologists and paediatric surgeons to locate the exact region of the diseased bowel.

Although there are advantages in using nucleic acids as diagnostic or prognostic biomarkers, there are also limitations, including: (1) a large volume of blood may be required for quantifying nucleic acids of low abundance, (2) cell-free mRNA has a relatively short half-life in the circulation as plasma is rich in nucleases, (3) expression levels of mRNA in the tissue may not necessarily correlate with the serum concentration, and (4) tissue-specific mRNA may be difficult to identify as many tissue mRNAs are also produced by circulating or infiltrating inflammatory cells.

Searching for ‘Novel’ Biomarkers – the ‘Hypothesis-Free’ Approach

The ultimate aim of biomarker research is the discovery of new and previously unreported biomarkers for clinical use. In 2010, our research team has successfully used the mass spectrometry-based proteomic profiling technology to identify known and novel host-response signature proteins, serum amyloid A and proapolipoprotein CII, respectively, for the early diagnosis of neonatal infection and NEC [13]. Using a mathematical formula to calculate the ApoSAA score, we could confidently withhold or stop antimicrobial treatment within 24 h in 61% of true-negative sepsis/NEC cases [13]. In this study, a very stringent experimental design was adopted to minimize the effect of bias and the chance of identifying false-positive candidate biomarkers. This study design is based on the fundamental principles of novel biomarkers discovery and can be considered as one of the models to which similar future studies may refer.

In the first part of the study, the profiling process consisted of two distinctive phases: the biomarker discovery phase and the independent validation phase. First, in the discovery phase, a 3-step approach was adopted: (1) An initial case-control comparison of plasma proteomic profiles between disease subjects and control subjects was used to identify the desired differential proteomic peaks. (2) These peaks would be considered as potential biomarkers only if their levels demonstrated a reversal pattern upon disease recovery in the longitudinal follow-up with a separate group of patients (this separate group included both disease and control cases). This essential step was to avoid selecting proteomic peaks caused by systematic bias. (3) The plasma concentrations of the identified proteins were then subjected to logistic regression analysis and a diagnostic equation was constructed. Secondly, in the validation phase, other independent groups of disease and control infants were evaluated for validating the diagnostic equation to determine their areas under the receiver operating characteristic curves. Thereafter, in the second part of the study, the selected biomarkers and equations were subjected to the final validation process using a prospective cohort of patients to determine the ultimate diagnostic utilities of the test [13]. Thus, only with such an elaborate study design could new and robust biomarkers be discovered for clinical use.

Conclusion

Advances in molecular diagnostic technologies, in particular gene sequencing and other powerful mass screening techniques, such as proteomics and arrays, have greatly opened up new opportunities for discovering new biomarkers for diagnostic and prognostic applications. A comprehensive and robust approach for biomarker discovery has been outlined in this review. The next generation of biomarkers, whether they are circulat-
NG/LAM

Disclosure Statement

The authors declare that there is no conflict of interest.

References


