TOWARDS MICROARRAY DIAGNOSIS OF INFECTION IN THE NEWBORN.

Claire Lindsay Smith

A thesis submitted for the degree of Doctor of Medicine

The University of Edinburgh
2010

Dedication

I dedicate this thesis to my husband John Thompson, my daughter Eilidh, my parents David and Linda Smith and my brother Kevin, with love and thanks.

Contents

Declaration

I have composed this thesis myself.

The experimental work on the umbilical cord study was done entirely by myself. The experimental work on the neonatal study was performed predominantly by myself as part of a research team; contributions from co-workers have been acknowledged in each section where appropriate. Others' contributions to the microarray sections have also been clearly acknowledged in each section. The sections of work where data has been previously published in the literature or the data has been used previously by others are clearly stated.

I hold the degrees of MBChB (1999) and BSc med sci honours first class (medical microbiology) (1996) from the University of Edinburgh.

I have not submitted this thesis in candidature for any other degree, postgraduate diploma or professional qualification.

Signature:

Acknowledgements

I would like to thank the following for their contributions to this study:

The parents and babies whose participation made the study possible.

Chiesi pharmaceuticals for their small project grant which funded a proportion of the umbilical cord blood studies.

GE Healthcare for the discount and training they provided on the use of the CodeLink™ Microarray System.

Ben Stenson for his supervision and advice on the clinical aspects of the study.

Peter Ghazal for his supervision and advice on the genomic aspects of the study.

Marie Craigon for her assistance and supervision during the early microarray stages and for running microarrays for the operator-balanced Codelink™ experiment. Also for helping prepare samples for the Illumina® arrays when I was unable to do so due to being in the latter stages of pregnancy.

Alan Ross for his patience in orientating me to the workings of the genomics laboratory, for carrying out RNA extractions and quality control on latter study samples and for running microarrays for the operator-balanced CodeLink™ experiment.

Petter Storm for running microarrays for the operator-balance CodeLink™ experiment.

Thorsten Forster and **Mizanur Khondoker** for their statistical advice and input.

Paul Dickinson for co-ordinating and contributing to the RNA quality control and preparation of samples for the Illumia® arrays in the latter stages of the study. Also for his advice on the genomic aspects of the study and for his help in generating some of the diagrams for the microarray section.

Paul Lacaze, who used my early study samples as the basis of his MSc study and in so doing increased my understanding of the process of using microarray data.

Allan Jackson and several other clinical colleagues for their help in recruitment of patients to the study and in taking study samples.

Al Ivens and **FIOS Genomics** for normalisation and statistical analysis of the microarray data.

Gen-Probe for running the Illumina® microarrays.

Abstract

Title

Towards Microarray Diagnosis of Infection in the Newborn

Background & Aims

Infection causes significant morbidity and mortality in newborn infants. Current methods for diagnosing infection are unreliable. It would be beneficial if a test could be found that could diagnose infection sensitively, accurately and reliably – particularly if more rapid and from smaller samples. Microarrays are a useful means of global analysis of gene expression. One particularly exciting application of microarrays could be in diagnosing infection by detecting alteration in host RNA phenotype in response to infection. I set out to investigate whether small volumes of neonatal blood could yield RNA of sufficient quality and quantity to carry out microarray analysis and to identify suitable methods of sample handling and RNA extraction. I then went on to determine if differences in gene expression profiles could be detected between infants with confirmed infection and a group of controls using microarray technology.

Methods

Umbilical cord blood was used to optimise blood collection tube, RNA extraction method and sample storage conditions. RNA quality and yield were assessed for each. RNA samples from neonatal blood taken from infants with confirmed infection and controls were then run on microarray: initially on CodeLink™ Whole Human Genome Microarray and later on Illumina® Human Whole-Genome Expression BeadChips. Normalised, validated microarray data was analysed to examine differences between control and infected samples. Functional annotation according to gene ontology and pathway analysis was performed.

Results

High quality RNA yields sufficient for microarray work were obtained. The optimum blood collection tube, RNA extraction method and sample handling conditions are

described. Results from the Codelink™ arrays are presented along with discussion of problems encountered using this platform. Results from 28 infected and 35 control samples run on the Illumina® platform are presented. 6221 features were significantly differentially expressed between infected and control groups (adjusted p-value < 0.001). 448 features had > 2-fold up-regulation and 341 features > 2-fold down-regulation ($p < 0.001$). Functional annotation and pathway analysis revealed a notable proportion of these are involved in immune functions.

Conclusions

Sufficient high quality RNA for microarray analysis can be obtained from small neonatal blood samples. Differential RNA expression profiles of host response can be detected between infected and non-infected infants. Such data may provide an alternative way for diagnosing infection in infants in the future. Large-scale studies are required to explore this further.

Word Count: 89776 (including tables and references)

Central Hypothesis

- Distinct patterns of host transcriptional response detected from neonatal whole blood may enable diagnosis of neonatal infection.

Secondary Hypotheses (Questions Arising from Central Hypothesis)

- It is possible to extract RNA of sufficient quality and quantity from neonatal blood to carry out successful microarray analysis.
- Host gene expression profiles differ between infected and non-infected infants.

Aims

To define a robust, reliable, reproducible method of RNA extraction from neonatal whole blood I aimed to:

- Determine if RNA of sufficient quality and quantity for microarray analysis can be extracted from clinically applicable volumes of umbilical cord blood & neonatal blood.
- Determine a suitable blood collection tube for RNA extraction from neonatal whole blood.
- Determine a suitably small blood volume to collect into the blood collection tube to be sufficient, provide accuracy and to be clinically applicable.
- Determine a suitable method for RNA extraction from neonatal whole blood.

- Determine an optimal time period between blood sampling and RNA extraction.
- Determine whether freezing of blood samples prior to RNA extraction was feasible.

To examine whether it is possible to detect differences in blood RNA expression profiles between infected and non-infected infants I set out to:

- Compare microarray results of infected and non-infected infants.

Ethical Consideration

A favourable ethical opinion was obtained from Lothian Local Research Ethics Committee for both the umbilical cord blood studies (LREC/2004/6/1) and the neonatal blood studies (05/S1103/3) presented here. Permission was also granted from the local NHS Research and Development Office for both phases of the project (R&D reference numbers: 2003/R/NE/12 and 04/S1103/3 respectively).

Chapter 1

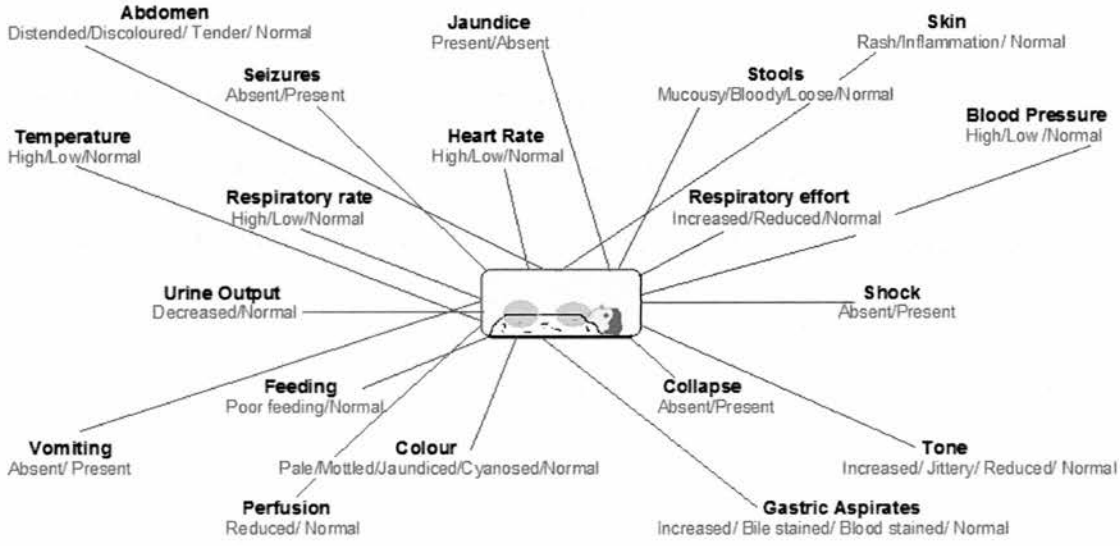
Background

1.1 The burden of neonatal infection

Despite advances in neonatal care, infection remains a significant cause of morbidity and mortality in newborn infants. The incidence of congenital infection is between 1 and 10 per 1000 live births (1). 65 % of extremely low birth weight infants develop presumed sepsis in the neonatal period (2). With mortality rates of 10-50 %, a 4-fold increase in cerebral palsy and increased risk of hearing, growth and neuro-developmental impairments as sequelae in this particularly vulnerable group, the costs are great (2). In Scotland, the overall neonatal mortality rate is 2.6 per 1000 live births, with the rate due to infection being 0.4 per 1000 live births (3). In other words 15 % of neonatal deaths in Scotland have infection identified as cause of death. Very low birth weight infants with late onset sepsis, especially those with gram negative or fungal infections, have prolonged hospital stays and are more likely to die than uninfected infants (4). Worldwide neonatal infection is a significant problem. Suspected infection is the commonest reason for admission to neonatal intensive care units in the USA (5). Infection is the commonest cause of neonatal mortality in South East Asia: overall mortality of 46.3 per 1000 live births with 30-40 % of these due to infection (6, 7). This makes it imperative that work should be done not only on prevention and treatment of infection but also on methods that could be used to diagnose infection more reliably and rapidly. It would be hugely beneficial if a test could be found that could diagnose infection sensitively, accurately and reliably – particularly if smaller samples could be used to achieve more rapid results. Such a test could help avoid unnecessary treatment in babies who are not infected and therefore reduce neonatal unit admissions, separation of mother

and baby and iatrogenic complications. Importantly, it would also allow more tailored anti-microbial therapy. Such progress would also be expected to reduce the financial burden of neonatal infection.

Figure 1.1: *Clinical Signs of Neonatal Infection*



1.2 Presentation of Neonatal Infection

Infection in neonates can present clinically in a multitude of ways. Clinical signs can range from very subtle and gradual to catastrophic with rapid progression. The source of infection is often difficult to pinpoint. It is therefore important that clinicians have a low threshold for considering infection, investigate expediently and treat whenever infection is suspected. Figure 1.1 displays possible signs of infection in the neonate and illustrates the variability in presentation for any given sign. This is not an exhaustive list. Current practice is to treat all babies with suspected sepsis with broad-spectrum antibiotics pending culture results. It is clear however that there is significant variation in practice among neonatologists when considering infants

who are suspected of having neonatal sepsis (5). If identification of a causative organism occurs at all, and it often does not, it is usually after antimicrobial therapy has already been started empirically. Clinical judgement still plays a large part in the diagnosis and management of suspected sepsis.

1.3 Definitions of Infection

Sepsis can be defined as a systemic illness caused by microbial invasion of normally sterile parts of the body (8). Consensus definitions for infection, sepsis, systemic inflammatory response syndrome (SIRS), severe sepsis and septic shock have been drawn up, e.g. International Pediatric Sepsis Consensus Conference 2002 and by The American College of Chest Physicians, the Society of Critical Care Medicine (9-11). These were designed to allow standardised discussions between groups and are rather long. The definition of infection I have chosen to use in this study is the summary definition from the International Sepsis Definitions Conference 2001. Infection was defined as "a pathological process caused by the invasion of normally sterile tissue or fluid or body cavity by pathogenic or potentially pathogenic microorganisms" (9).

1.4 Types of Neonatal Infection

When considering infection in infants less than 1 month of age, the most common causative bacteria may vary depending on the geographical location of the population. Common culprits include Group B Streptococcus, *Escherichia Coli* (and other gram negative enteric organisms), *Listeria monocytogenes, Streptococcus pneumoniae, Haemophilus influenzae* and *Staphylococcus aureus* (12, 13). Other important causative organisms worldwide include *Streptococcus pyogenes*,

Klebsiella species, Coagulase Negative Staphylococci, *Neisseria meningitidis*, *Enterococcus* species, *Pseudomonas aeruginosa, Enterobacter cloacae, Chlamydia trachomatis, Salmonella* species, *Citrobacter* species and *Neisseria gonorrhoeae (13-15).*

Neonatal sepsis is traditionally described as either early or late onset with congenital and vertically acquired infection most commonly presenting early (exceptions include Herpes Simplex Virus and late onset Group B Streptococcus) (15). Early onset sepsis is frequently defined as infection presenting within 72 hours of birth but definitions do vary from two to seven days postnatal life. Early onset sepsis (within 72 hours of birth) in very low birth weight infants in the USA is most commonly due to Group B Streptococcus, then *E.coli* and *H.influenzae* however Coagulase Negative Staphylococci have also been noted to have emerged as a cause of early onset sepsis (16). Although blood culture positive early onset infection is uncommon (1.9 % of very low birth weight infants), in a study by Stoll and colleagues, more than half of very low birth weight infants were considered to have clinical sepsis (16). Better means of diagnosis are therefore needed to identify those infants with infection and to enable safely stopping antibiotics in those without infection.

Late-onset sepsis (culture proven) occurs in 21 % of very low birth weigh infants surviving beyond three days (4). In these infants the majority of infections (70 %) are caused by gram positive organisms with Coagulase Negative Staphylococci accounting for 48 % of infections (4).

Group B Streptococcus or *Streptococcus agalactiae* is worth a particular mention as it can cause a particularly rapid deterioration if not treated expediently. It is an organism that is carried asymptomatically in the genital tract of a significant proportion of women and can colonise neonates asymptomatically or cause invasive disease up to around 90 days after birth (17). There is significant debate regarding the use of antenatal screening and prophylactic antibiotics and there is a wide variation in practice in this respect.

Infections due to organisms other than bacteria can also be important sources of morbidity and mortality in neonates. Viral illnesses can and do present in a similar manner to bacterial infections and although antibiotics will not be effective in such cases, the risk of severe adverse consequences should a bacterial infection be missed is too great to omit antibiotics unless viral infection is definitively confirmed. Enterovirus infections for example, are common in infants under 3 months of age (18). Herpes simplex virus infections are particularly important to detect due to serious potential sequelae. Fungal infections, most commonly with Candida species are also found, particularly in very premature and vulnerable sick infants who may have been exposed to broad-spectrum antibiotics.

1.5 Diagnosis of Neonatal Infection and Infection Markers

Investigation of infection almost universally involves obtaining a blood culture in addition to other blood tests that may include full blood count, coagulation screen, sugar, lactate, blood gas and C reactive protein. Completion of the "septic screen" would usually consist of a CSF sample being obtained by lumbar puncture, a urine

sample being obtained by clean catch or supra-pubic aspirate and a chest x-ray. In addition, swabs and samples of secretions may be taken if the clinical picture indicates. The problem with these investigations is that none are completely reliable and negative or normal results do not completely rule out infection. Each of these investigations and some additional tests and their drawbacks are discussed in turn below.

Blood cultures are still considered the "gold standard" for diagnosis of neonatal infection. However, they do not give a rapid result (generally taken as up to 48 hours incubation) and have poor sensitivity of 50-80 % at best (19, 20) although it has been argued that 36 hours of observation is sufficient to rule out infection in asymptomatic neonate with negative cultures (21). In addition, blood cultures require blood volumes that represent a significant proportion of an infant's circulating volume. The blood culture system in use in the neonatal unit in the Royal Infirmary of Edinburgh at present is the BacT/ALERT®PF. The manufacturers recommend 4 ml of blood be used in these bottles although they state that less blood can be used but that this would reduce the chance of organism detection (22). Our local guidelines recommend using 1.5 ml/kg of blood up to a maximum of 4ml (or 1.2 ml for infants weighing less than 1 kg) for blood culture. In order to maximise sensitivity and specificity of blood cultures attention should be paid to thorough skin preparation, culturing peripherally rather than (or in addition to) through existing access devices, sampling as early as possible when infection is suspected and paying attention to volume and number of cultures as appropriate (23). Although some would argue that multiple site culture gives optimal results, Sarkar and colleagues found that single

site cultures with volume of 1ml or more can detect sepsis with no loss of accuracy (24). Some researchers have suggested the use of umbilical cord blood instead of neonatal blood for culture to look for early onset infection (25, 26). This is an attractive option in some respects but would essentially be a screening tool, requiring that a large number of samples be taken and would be expensive. In addition, from experience, it is not always possible to obtain blood from umbilical cords. Use of blood culture results, as with other medical investigations, cannot be taken at face value or in isolation. An understanding of the limitations of the investigation and an evaluation of the whole clinical picture by the clinician are essential.

Gram-staining and microscopy of isolated organisms, while potentially giving an early indication of the organism responsible for an infection, can be misleading. This may be due to the subjectivity, training and experience of the technician, the quality of the staining or the uptake of the stain by the organisms (27). In addition, microscopy is only possible if the load of pathogen is great enough for visual detection (27). Furthermore, if antibiotics have been administered prior to obtaining clinical samples the likelihood of being able to detect organisms under the microscope may be reduced and the appearance of organisms seen may be altered.

Culture of body fluids remains the mainstay of pathogen detection. However, it is known to sometimes yield negative results in cases of bacterial infection. In some cases this can be attributed to antibiotics being administered prior to taking the clinical samples but false negative cultures can be obtained even without prior administration of antibiotics (27). Other factors leading to false negatives could

include use of the wrong culture medium, especially for more unusual or fastidious organisms and suboptimal sample handling, e.g. delay in reaching laboratory, drying out or handling at suboptimal temperature. In addition, there are some organisms that simply cannot be cultured in the laboratory and organisms that although cultured may not be successfully identified (27). In the past, gastric aspirate culture has been used in the investigation of neonatal infection. In general this has been found not to be a useful indicator of neonatal infection but it may indicate exposure to chorioamnionitis (20). A further drawback of working with live organisms in the laboratory is the potential threat to the health of laboratory workers (27).

Full blood counts can generate useful supporting information for infection. White cell and neutrophil counts may be elevated but this is not specific to infection and is seen with other inflammatory conditions. White cell, neutrophil and platelet counts can also be very low in neonatal infection, e.g. in overwhelming sepsis but again this is not specific to infection. Presence of toxic granulation and an elevated immature:total (I:T) neutrophil ratio may be considered to be more specific for infection, however these are of limited value in neonatal infection (1, 28). Blood film examination has been known to reveal bacteria within neutrophils in some cases of neonatal infection (29).

CRP is an acute phase protein produced in the liver approximately 6-8 hours following an inflammatory insult (20). It may not be elevated in early infection and the use of CRP in isolation to exclude infection could be dangerous as it does not have sufficient sensitivity or negative predictive value (20, 30). The positive

predictive value of an elevated CRP is also low and an elevated CRP is not specific to infection as it is also raised in conditions such as asphyxia (1, 20). Several groups feel that CRP may be more useful in monitoring response to treatment (1) or in aiding the decision to discontinue antibiotics (31). However, the sensitivity may not be sufficient to allow early discontinuation of antibiotics either (30). Gestational age is also a factor in whether CRP may rise in infection. A rise in CRP more likely with increasing gestation (32). In summary CRP of limited value in diagnostic tests in neonatal infection (28).

Antigen detection tests such latex agglutination tests can be used to identify organisms such as Group B streptococcus (20). Serological tests (antibody detection and microbial antigen detection) have limitations in that antibody responses may not be apparent if testing too early in the infection process and antigen detection may require a large pathogen load (27). Such tests are very specific and if a narrow differential diagnostic approach is not possible then such tests will be of little utility. Serological approaches may also be of limited use if the host is immuno-compromised in such a manner as to not mount an antibody response (27). ELISA-based tests to detect antibody response can require large amounts of time, sample volume and reagents (33). Newer tests however are trending towards using smaller sample volumes, e.g. D Bianchi and N Weinschenks' test to identify infant sepsis looking at leukocyte expression of cell surface antigens by labelled antibodies uses around 200 microlitres of blood (34). Multiplexed whole bacterial antigen microarrays are now emerging that allow automation of serodiagnosis, e.g. for culture negative endocarditis (35).

Until recently, testing for infection would mean looking at either single or small numbers of analytes. Simultaneous measurement of multiple analytes is now possible. Nucleic acid based methods and especially microarrays are discussed in detail later in this section but multi-analyte testing is not confined to the field of genomics. Recent advances in flow cytometry for example, allow simultaneous measurement of multiple markers from small blood volumes – careful selection of markers could aid in diagnostic decision making (36). As an example of this, Skogstrand and colleagues measured 25 inflammatory markers (23 cytokines, TREM-1 and CRP) simultaneously from blood spots (37).

There are a multitude of lab tests that are mentioned in the literature that have been used in the past but appear to have fallen out of favour, including nitroblue tetrazolium tests and acridine orange leucocyte cytospin test (38). Micro-ESR is more specific than CRP but less sensitive which, combined with risk of false positives with haemolysis and false negatives with disseminated intravascular coagulopathy, limits its usefulness (20). Other acute phase reactants such as orosomucoid and fibronectin are of limited value due to slower response to infection than CRP (20). Elastase-alpha-1-proteinase inhibitor rises rapidly in infection, is 100% sensitive but not particularly specific (20).

Despite an increase in the number of available potential markers of infection (many still in a research capacity) there is currently no single test or group of markers that will reliably diagnose infection enough to influence the initial decision to start or withhold antibiotics (36, 39). Several cytokines have been studied perhaps most

notably IL-6, IL-8, CD11b, CD64, TNFα, IL-1β and E-selectin (36, 40). IL-6 is elevated in infected infants giving high sensitivity in detection of infection, is increased earlier than CRP but is limited in usefulness due to short duration of elevation and the fact that it is also elevated in necrosis (1, 41). Another specific leukocyte marker that has been looked at as a marker of infection is CD11b – this looks promising with negative predictive value quoted as 100 %, positive predictive value of 99 %, sensitivity of 96 % and specificity of 100 % (42, 43). Neutrophil CD64 expression has been shown to be sensitive in predicting infection but with only moderate specificity of 81 % with Ng and colleagues suggesting it may allow discontinuation of antibiotics at 24 hours in infants with CD64 below cut-off who are clinically stable (44). TNF has an early peak after onset of infection (1). Cord blood TNFα is elevated in early onset sepsis with high sensitivity but is limited by low specificity (45). IL-1β is significantly decreased in neonates with sepsis (46). Several screening tests utilising various CRP and cytokines such as IL-1β, IL-6 and IL-8, TNFα and the soluble receptor of IL-2 have been described (42, 47-49). These invariably show that no individual test can reliably identify infected neonates and that combinations tend to be limited in either sensitivity or specificity (47, 48, 50). Some groups such as Franz and colleagues have suggested that in stable infants use of combinations of results can reduce the number of infants being commenced on antibiotics (49). However, in this example which looked at the combination of IL-8 and CRP, 14.5 % of infants with infection were not detected initially (49). Most would agree that such tests do not allow safe withholding of initial antibiotics but it may be that they could be used to help with early discontinuation of antibiotics where clinical condition permits (40, 41). Other notable markers of infection that

have been studied include procalcitonin and complement levels. Serum procalcitonin has been shown to rise rapidly in infants with sepsis and to have a better sensitivity than neutrophil count, CRP or IL-6 (50-55), but is still not reliable enough alone to be sole marker of neonatal sepsis (56). Procalcitonin is considered by some as a useful adjunct in sepsis diagnosis and serial measurements of procalcitonin have been described as useful in monitoring response to therapy (57). In terms of complement, although preterm infants have a relative deficiency of most complement components, they are still able to generate significant amounts of activation products of the complement cascade which temporally precede other lab indicators of infection such as white cell count and CRP (58). Specific complement related molecules that have been studied include C3a-desArg, C3bBbP, sC5b-9 (58), C3d which has been described as having a "reasonable" predictive value (20), C4d and Ba, with Ba performing better than CRP (1, 59). G-CSF , which is elevated in infected infants of all gestational ages has also been considered as a marker but has limited specificity (73 %) and poor positive predictive value (40 %) (60). ICAM-1, which is elevated in infection, has also been put forward as potentially having role in confirmation or prediction of infection (61). It has however also been noted to be elevated in RDS (62). Combinations of neutrophil CD64, procalcitonin and triggering receptor expressed on myeloid cells (TREM-1) are currently being investigated (63). Inter-alpha inhibitor proteins are independent of gestational age and significantly reduced in sepsis leading to the suggestion that their measurement may act as an adjuvant diagnostic marker (64). To re-iterate, none of these markers alone or in combination are reliable enough for confident diagnosis of infection.

Some groups have attempted to create reliable clinical scoring systems for neonatal sepsis or to use physiological parameters such as heart rate characteristics to add to laboratory information (65, 66). To date however none of these can confidently be relied on. In developing countries sepsis diagnosis using clinical algorithms is useful in areas where access to hospital care is limited (67). Some groups describe using close clinical observation rather than empirical antibiotics in selected infants, e.g. Chiu and colleagues observed febrile infants who were "low risk" according to criteria that included normal CRP, WBC and well appearance (68). Many neonatologists would be uncomfortable with such an approach. Clinical assessment however remains important in sepsis evaluation and should not be underestimated.

It is clear to me on reviewing the literature that currently available diagnostic tools and markers of infection are not ideal. This concurs with the view of Malik and colleagues who reviewed the literature and found few studies examining laboratory diagnosis of neonatal infection that were methodologically sound, with a few tests looking promising but with insufficient data to support their use as clinical tools (69) There is a real need to find an accurate and rapid means of diagnosing neonatal infection. It is my belief that microarray gene expression profiling is an important area to explore in this quest.

1.6 Host Response and Immunity

Host immune and inflammatory responses play a significant role in the pathophysiology of sepsis (8). As mentioned already the symptoms and signs of sepsis can range from very subtle to catastrophic. Factors that affect the symptoms

and signs at any given time can be considered to include: the virulence of the offending pathogen, the pathogen load, the route of entry (i.e. the site of infection), individual susceptibility of the host and the time in the course of infection (8). In the following sections I present current information on neonatal immunity and on host response to infection.

1.7 Neonatal Immunity

Neonates have generally been supposed to be immunologically immature. Reduced opportunity for trans-placental antibody acquisition is an important factor in increased risk of premature infants to infection (13). In addition, there are birth-weight related deficiencies in type-specific antibodies and complement opsonins, reducing efficiency of opsonisation (13). Sub-optimal cell mediated responses are also seen (70). Neonatal increased susceptibility to pathogens is due to a combination of immaturity of innate and adaptive immune responses with defects in phagocytic ability, and immaturity and dysregulation of T-cell mediated immunity. This does not mean that neonates are completely defenceless against infection and it does not mean that they fail to mount an immune response. It is increasingly recognised that neonates do mount a significant response to infection but that there are differences from the responses seen in adults. As in all aspects of examining immunity, outcome is a question of balance, e.g. fetal gut cells will mount a robust inflammatory response to lipopolysaccharide but this could be dangerous in the period shortly after birth where gut colonisation occurs (71). In this section I attempt to provide an overview of the neonatal immune system, looking at different aspects in turn and listing known deficiencies and differences from adult immunity.

Innate immunity: Skin and Mucosal Surfaces

When a baby is born, they move from a sterile environment to one that is full of antigens to challenge the immune system. The innate immune system plays a significant role in neonatal defence against infection. Physical barriers to microbial invasion are an important initial line of defence in the neonate and include the skin, respiratory tract, eye, gastrointestinal and urinary tracts. In preterm infants these barriers are more fragile, increasing their vulnerability to microbial invasion. In addition, in the setting of a neonatal unit such barriers are often breached through clinical necessity by the use of catheters and lines. As well as this barrier function, neonatal skin and mucosal surfaces provide defence against infection by their abundance of host defence proteins, lysozyme and lactoferrin, some of which increase with gestational age (71, 72). Vernix, which coats the skin at birth, also contains lysozyme, alpha defensins, ubiquitin, psoriasins and APPs – antimicrobial peptides and proteins (71).

Neutrophils and Phagocytic Function

Neonates are deficient in both quantitative and qualitative phagocytic cell responses. There is a relative inability to produce adequate numbers of neutrophils in response to infection due to insufficient numbers of myeloid progenitor cells, a decreased bone marrow neutrophil storage pool and high ratio of cells in the proliferative state (73). Neonatal neutrophils have impaired function and although this is not as yet completely characterised they show reduced chemotaxis, migration, opsonisation, phagocytic and intracellular bacterial killing ability, reduced oxidative responses and an inability to form extracellular traps that mediate extracellular killing of pathogens

(13, 74). Reactive oxygen species production tends to increase with birth weight and gestational age meaning that preterm infants are relatively more deficient in this respect (75). Neutrophil recruitment is also impaired due to low integrin and selectin proteins (71). In addition, neonatal neutrophils have impaired apoptosis and therefore are not expediently removed from inflamed tissues by monocytes trying to minimise injury (76).

Monocytes, Macrophages, Dendritic Cells, Antigen Presenting Cells

There are many papers in the literature looking at the function of adult monocytes compared to neonatal or cord blood lymphocytes. Satwani and colleagues. showed dysregulation of several cytokine and immunoregulatory genes in cord blood which may in part explain the relative immaturity of neonatal cell-mediated immunity (77). They showed reduced production of phagocytosis activating cytokines such as G-CSF, GM-CSF and M-CSF and of negative regulators of haematopoesis TGF-β1, MIP-1α, IL-8 (77). Cord blood monocytes have been also shown to have comparative down regulation of cytokine genes such as TGF-β, IL-8, IL-11, IL-12, IL-15, IL-18, IL-24, cytokine receptors such as IL-2Rα, IL-2Rβ, chemokines including RANTES, GCP-2, MIP-3β, surface antigens/molecules SLAM, IL-1R-antagonist, oncostatin M and lower expression of transcription factors such as NF-κB genes (78). There are also differences in expression of various phosphatases, kinases and cell regulatory genes (78). Neonatal monocytes also have reduced chemotactic and cytotoxic ability (13). Despite normal basal expression of Toll-like receptors (79), neonatal monocytes and macrophages have impaired responses to multiple TLR

ligands (80). In addition, TLR-induced monocyte TNFα release is inhibited by soluble factors found in neonatal blood (79).

Many studies go further and compare adult and neonatal monocyte responses to stimulation with bacterial antigens such as LPS. Significant differences in gene expression profiles are found (81). In neonates there is comparatively higher expression of IL-1β, IL-15, IL-15R, IL-17R, IL-2Rα, IL-2Rβ, IL-4R, IL-6, IL-8, IL-7R, G-CSF, GM-CSF and relatively lower expression of MHC class II molecules (78). Neonatal monocytes show normal basal expression of Toll-like receptors and membrane CD14 but LPS stimulation leads to relatively reduced release of Th1 polarising TNFα and IFNγ with preservation of Th2 polarising anti-inflammatory cytokines (79). Term neonates stimulated with specific bacterial antigens show a relatively down-regulated response compared to premature infants, i.e. premature infants may respond with uncontrolled inflammation to antigenic stimulation from organisms (82). Neonatal monocytes show impaired expression of IFNβ and IFN-inducible genes compared to adults when stimulated by LPS and also reduced IRF-3 and CREB-binding protein (IFN response dependent on this) (83). There are also differences within regulation of apoptosis (84).

Neonatal macrophages also show impaired type 1 responses to IFNγ and LPS (80).

When compared with mature dendritic cells, cord blood dendritic cells show relatively reduced expression of chemokines, cytokine receptors and cell surface molecules, and also lower expression of transcription factor, interferon regulatory

factors and structural regulatory genes (85). Delayed maturation of certain dendritic cells means limited IL-2 production (86). Cord blood dendritic cells also show relatively reduced TH1 response-related genes (87). Neonatal dendritic cells can be stimulated to express maturation surface molecules (towards specific immune response) but require combined stimulation with multiple pro-inflammatory signals (88).

Neonatal antigen presenting cells show functional changes, producing low levels of cytokines which could lead to defects of adaptive T cell response (89). They show impaired production of IFNα and IFNβ and reduced expression of TNF, IFNα, IFNγ, IL-12 and IL-1β (71). Neonatal antigen presenting cells and monocytes can however produce IL-6, IL-10, IL-23 in greater magnitude than adults when stimulated (71). Overall, antigen presenting cells show selective impairment of TH1 response following innate immune recognition (71).

Natural killer cell numbers present in cord blood increase with gestational age (90).

Toll Like Receptors

Toll like receptors (TLRs) play a key role in innate immunity, from pathogen recognition through to activation of adaptive immunity, and TLR polymorphisms can lead to increased susceptibility to infection (91). TLR 2 and TLR 4 receptors play a role in the gut and respiratory tract (71). There is conflicting data in the literature as to whether TLR 2 levels differ with age (71, 92). TLR4 expression and cytokine secretion following stimulation with LPS increases with gestational age (92).

Stimulation of neonatal blood with TLR activating stimuli is associated with a high IL-6/TNFα ratio and this is higher in the first week of life than in cord blood (93). TLR8 agonists are particularly proficient in activating co-stimulatory responses in neonatal antigen presenting cells (94). Cord blood TLR agonist induced production of Th1 polarising cytokines IL-12 and IFNα are generally impaired but for TLR3, TLR7 and TLR9 agonists increase to adult levels within a month whereas TLR4 shows slower maturation meaning low IL-12 and high IL-10 in response to LPS (95). Certain patterns (route, duration, load) of exposure to microbial TLR agonists in early life accelerate maturity of Th1 response and are protective against allergy – this is part of the hygiene hypothesis (71).

T cells

During fetal life, the immune system is necessarily skewed to avoid inflammatory responses that could precipitate delivery. The neonatal immune system is therefore biased against a pro-inflammatory Th1 response and towards a Th2 response leaving the neonate at risk of infection (71, 96). Antigen specific T cell responses can be mounted by neonates but CD4 T cell responses are often slower to mature and are biased towards a Th2 response (97). Exceptions to Th2 bias do exist: neonates can produce Th1 polarizing cytokines robustly in response to certain stimuli that can activate TLR pathways or complement pathways, e.g. Group B Streptococcus, BCG vaccine (71). In addition, Hassan and colleagues looked at the immune response to CMV in congenitally infected neonates and found a predominant Th1 response but with lower IFN-γ and higher IL-8 compared to adults (98). Omar and colleagues showed that neonates show a predominance of naïve T cells but are capable of

increasing memory T cell expression in response to infection (99). Mature memory T cells are higher in infected infants than controls even after recovery (99, 100). T regulatory and Th17 cell function are impaired at birth (101). Zhao and colleagues show that neonatal mice show much stronger inflammatory response to LPS stimulation than adults due to inadequate T cells with lack of regulation leading to a hyper-innate response and potentially increased morbidity and mortality (102). Gamma-delta T cells are important for immunoprotection at birth and could compensate for immaturity of alpha-beta cells (103). In addition they may have a role in rapid response to pathogen associated molecular patterns in the absence of antigen presenting cells (104). Gamma delta cells are relatively impaired in preterm infants in terms of TLR3 and TLR7 expression and suboptimal cytokine production (103).

Antibodies

As stated previously, there is reduced opportunity for trans-placental antibody acquisition in infants born prematurely. The number of immunoglobulin secreting cells is usually low at birth and rises with age. Infants with early onset sepsis have low levels of immunoglobulins in first 5 days but higher levels in week 2: elevated levels soon after birth may therefore be a surrogate marker for in-utero infection (105). T cell independent B cell antibody responses are impaired in infancy (97). T cell dependent antibody responses mature earlier but multiple immune encounters may be required to achieve the response of adults (97) .

Milk

Human milk can be thought of as an important part of the immune defences of the neonate. Without trying to cover all of the benefits of breast milk, it contains secretory immunoglobulin while cytokines, cytokine receptors, TLR agonists and antagonists, hormones, anti-inflammatory agents and nucleotides in milk modulate inflammation (106). Ingestible glycans found in milk act as prebiotics and inhibit pathogen binding (106). In addition, human milk has been found to have a capacity to modulate TLR-mediated responses specifically and differentially thereby influencing neonatal pathogen recognition (107).

Cytokines

Cytokines play a key role in the pathophysiology of sepsis. They are crucial in the response to infection but also contribute to tissue damage. A lot of the difference in cytokine production in neonates can be explained by the cellular characteristics described above. Neonatal cellular immune deficiency includes decreased production of IL-2/IL-23, IL-18, interferons, and other pro-inflammatory cytokines (80). Deficient IL-12 production may in part be compensated for by increased IL-27 production in neonatal dendritic cells in response to TLR agonists (IL-27 initiates Th1 responses in naïve T cells) (108). There is reduced IL-10 receptor expression on neonatal immune regulatory T lymphocytes and this, along with impaired IL-10, may play a role in the deficient neonatal anti-inflammatory response (109). From a study of dried blood spots at a few days of age: increased levels of IL-1β, IL-6, soluble IL-6r alpha, IL-8, matrix metalloproteinase and TGF-β1 and decreased levels of IL- 18, brain derived neurotrophic factor and CRP were associated with preterm birth (110).

Plasma concentrations of acute phase proteins change over the first few days of life : IL-1 and IL-6 rise in the first week and can trigger production of MBL, solCD14, CRP, LPS binding protein (71). IL-2 and IL-4 levels tend to be high and IFNγ low at birth and then trend towards normal over the first week of life (111). IL-6 acute phase response along with preserved a IL23-IL17 axis mobilises plasma antimicrobial peptides and proteins (71). IL-1β, IL-6, IL-8 have been seen to be elevated in fetal inflammatory response and may be indicators of in utero infection (112, 113). Mode of delivery influences the pattern of cytokines seen in neonatal blood at birth with caesarean section being associated with increased levels of IL-13 and IFNγ (114, 115).

Complement

Neonatal complement levels are low compared to adults: this means an impaired ability to halt bacterial replication and impaired adaptive responses (71). There is also reduced expression of complement receptor 3 (CD11/CD18) and L-selectin (71).

Maternal Immunity

Maternal immunity has an important part to play in fetal and neonatal immunity. Transplacental transfer of antibodies and transfer of antibodies in breast milk are well known as being important in protecting infants from infection. The protection provided will depend on the mother's own exposure to microorganisms. As an example of this, previous maternal exposure to varicella would be expected to lead to neonatal protection against varicella until the natural waning of maternal antibodies occurs. There is emerging evidence that other aspects of maternal exposure and

immunity also influence neonatal immunity. For example maternal farm exposure has been found to modulate neonatal immunity through regulatory T cells leading to reduction in the development of allergy (116). In addition, there is evidence that maternal leukocytes transferred from colostrum promote T cell immune development in calves (117).

Miscellaneous

There is reduced expression of many immune related pathways (TLR pathway, Jak-STAT pathway, cytokine-cytokine receptor interaction) in neonatal blood compared to maternal blood (118).

Preterm infants' susceptibility to Coagulase Negative Staphylococcal infection may in part be due to relative deficiency of serum proteins such as transplacental anti-coagulase negative staphylococci immunoglobulin and complement leading to suboptimal opsonisation and impaired bacterial killing (119). Neonates in general have a lower IL-6 response to Coagulase Negative Staphylococcus than adults and this may in part contribute to their vulnerability to this organism (82). Coagulase Negative Staphylococcus-induced production of pro-inflammatory cytokines IL-6 and TNFα in cord blood is dependent on gestational age, but of anti-inflammatory cytokines TGFβ and IL-10 is not (120).

1.8 Host Response to Infection

To discuss fully the host response to infection could fill volumes. The purpose of this section is rather to demonstrate the importance of host-pathogen interactions

and that microarrays have a place both in increasing knowledge in this area and in potentially using such knowledge to diagnose infection.

Put simply, when the host recognises presence of a pathogen, the initial response will be pro-inflammatory to try and eradicate the organism and later anti-inflammatory responses will work to restore the normal balance. Balance is important because either excessive inflammation or deficient anti-inflammatory response will result in pathology and can lead to shock or organ failure.

Many review articles are emerging that describe the promise of microarrays in monitoring host response and in monitoring bacterial gene expression in infection (121, 122). They also describe the potential for increased understanding of host-pathogen interactions and for discovering molecular signatures for infection (123). On the whole, there is a lot of optimism regarding the potential use of microarrays in microbial diagnostics (124).

Pathogens express molecules that activate the immune system by interacting with Toll-like receptors on immune cells (125). Different pathogens bind to different Toll-like receptors and Toll-like receptors are expressed differentially by various tissues, supporting the theory that the innate immune system is pathogen and tissue specific (125).

A large body of the work presented in the literature on host response to infection describes in vitro stimulation of specific cells with a particular bacteria, virus, yeast

or toxin, or of examination of immune cell response to stimulation (78, 126-145). Some of these immune responses may differ depending on strain of pathogen (135). There are now emerging views of a common response to infection characterised mainly by innate immune responses as well as pathogen specific host responses (125, 137, 145, 146). Some examples of this are listed here. Ren and colleagues compared host response in vitro to monocytes in response to *Coxiella burnetti* and *Chlamydia trachomatis* and found shared and pathogen-specific responses (147). Schnappinger and colleagues looked at macrophage responses to tuberculosis showing non-specific as well as pathogen specific response with tuberculosis inducing changes in TLR dependent and TLR independent transduction pathways (144). Das and colleagues looked at host gene expression profiles in monocytes in response to biological threat agents, e.g. anthrax using arrays (148). They found some expression patterns that were pathogen unique and others in response to multiple agents (148).

There are also papers emerging looking at in vivo host response to infection. Chinnaiyan and colleagues in a rat model looked at gene expression in several organs in response to sepsis following bowel perforation (149). They found a highly complex transcriptional response, with parts that were organ specific and others found in more than one organ, with pro-inflammatory genes often being balanced by genes with anti- inflammatory effects (149). Calvano and colleagues examined innate immune response to in vivo stimulation with endotoxin compared with healthy volunteers and found an initial pro-inflammatory phase and a subsequent anti-inflammatory phase (150). Reghunathan and colleagues compare monocytes from

the blood of SARS patients to those from healthy controls to examine differential gene expression and found that affected patients seemed to be mounting a mainly innate rather than specific immune response to virus (151). Whole blood whole genome expression profiling of children with septic shock by Wong and colleagues put forward the possibility that altered zinc homeostasis may be associated with poor outcome (152, 153). They found that although SIRS in children, sepsis and septic shock shared some common gene expression patterns, septic shock in particular displays down-regulation of genes involved in adaptive immunity and zinc related biology (154). Johnson and colleagues also found a unique gene expression profile in sepsis compared to uninfected SIRS (155). TLR and downstream signalling genes have been seen to be differentially expressed in sepsis compared with uninfected SIRS before phenotypic diagnosis using microarray (156). Gene expression profiling of children with uncomplicated Dengue showed a group host response profile of cytokines, IFN signalling, oxidative metabolism, protein ubiquination, and apoptosis whereas children with Dengue shock showed a more "benign response" (157). A gene signature for acute lung injury in adults with sepsis has been described (158).

There are several studies that examine facets of common host response to pathogens. Host pathogen recognition receptors, e.g. CD14, TLRs, Nod1/2, scavenger receptors, CD18, CD55, CD11b, TREM-1, CXCR4 can trigger a common immune response (146). Cell wall components such as CpG DNA, flagellin and peptidoglycan interact with TLRs (e.g. TLRs 2,5,6,9), and CD14 may also be involved to produce a common response (125). Jenner and colleagues (159) defined a common host

response of 511 genes from 77 different host-pathogen interactions (from 32 studies) that included pro and anti -inflammatory mediators, chemokines, interferon-regulated genes, transcriptional regulators, signal transducers, and NF κB pathway signalling. Manger and colleagues eloquently describe host-pathogen interactions as "diverse, choreographed and regulated" (160). When looking closely at pathogen specific molecular patterns of LPS lipid A, lipotechoic acid and peptidoglycan, they found common responses of innate and acute phase responses as well as unique responses which were dependent among other things on timing and place of interaction (160). Host-host variability needs to be taken into account and more in vivo work carried out before diagnostic signatures can be confidently found (160).

Early host immune and inflammatory responses may differ greatly for infections caused by gram positive and gram negative organisms (134). Feezor and colleagues showed that comparison between cytokine profiles in patients with gram positive and gram negative sepsis show some commonality but also display distinct cytokine and gene expression patterns (134). Toll like receptors have been shown to differentially recognise and signal in response to gram negative and gram positive related antigens (161, 162). Zeyton and colleagues describe distinct cytokine and chemokine induction depending on stimulation with LPS (TLR4), peptidoglycan (TLR2) or dsRNA (TLR3) of antigen presenting cells (163).

Gram positive organisms cause sepsis by stimulating immune cells with cell wall components (lipotechoic acid, peptidoglycans) through TLR-2 or by toxin production acting as superantigens (125). Toxins bind to MHC 2 molecules and Vβ Chains of T

cell receptors leading to pro-inflammatory cytokine production (125). Recognition of bacterial components by cell receptors, e.g. toll like receptor 2 starts a NF-κB mediated response leading to secretion of IL1 and TNF (146). IL6 and IL8 are also involved in an early common response to gram positive organisms (146). A later host reaction to gram positive organisms has been shown to be triggered by inter-cellar pathogen activity or by metabolites secreted by the pathogen, leading to either elimination of the bacteria or of survival of the pathogen (146).

Lipopolysaccharide found on the outer membrane of gram-negative bacteria is key in inducing gram-negative sepsis. Epithelial cells exfoliate in an attempt to avoid colonisation when bacteria attach at mucosal surfaces (146). Interaction of LPS through TLR-4, with involvement of membrane proteins CD14 and MD-2, plays a role in inducing a specific response (125, 164).

It is also important to look at expression profiling of non-bacterial infection, e.g. candida (141) and of non-infectious pathologies. There are, for example, groups who have been investigating the host response to injury (165). Oberholzer and colleagues show that IL-18 levels are higher when comparing septic patients with those with severe injury (166). This is important if distinction between different types of infection and between infection and other pathologies is to be made. Chung and colleagues showed that sepsis induces changes in mouse leukocyte genes that can be used to distinguish between sepsis and systemic inflammation (167). Expression of lymphocyte mRNA for chemokines are different in inflammation due

to perinatal infection to those in perinatal asphyxia, with IL-8 elevated in infection and MCP-1 increased in asphyxia (168).

Pathogens will cause differential gene expression depending on site and time of sampling with common and unique responses. In addition, there is large variation in clinical manifestation and response to identical infecting organisms due to effects of genetic variation in both host and pathogen: these variations may affect outcome as well as susceptibility (169). HLA associations with various infections are well known (169). Investigations into genetic susceptibility to infection are taking place and show that, as well as immune system related genes, mitochondrial DNA may play a role in differences in survival (170). Polymorphisms in IL-10, IL-6 and CD14 may affect risk of sepsis in very low birthweight infants (171). There is now emerging research to look at whether there are predispositions to specific symptomatic infections that are not directly linked to the immune response, e.g. Kerr and colleagues looked at single-nucleotide polymorphisms associated with symptomatic infection and differential gene expression in healthy people with previous infection with human parvovirus B19 implicating the cytoskeleton, integrin signalling and oncosuppression in B19 pathogenesis (172). Inherited or acquired mutations of genes involved in the innate immune response may affect susceptibility to infection, e.g. polymorphisms in TNFα, lymphotoxinα and IL-1 receptor agonists and variation in the MHC complex on chromosome 6 and HLA genes (125, 173, 174). Genetic/genomic factors may also play a role in variations in the outcome of sepsis (175, 176). Balding and colleagues have shown that genetic variability in the IL-6, IL-10 and IL-1RN gene are associated with poor outcome of

meningococcal disease (177). Specific genetic variants have been shown to be important for particular infections, e.g. Roy and colleagues showed that a C reactive protein polymorphism is associated with susceptibility to invasive pneumococcal infection (178). Deficiency of mannose-binding lectin (component of innate immunity) is associated with increased risk of neonatal sepsis (179, 180). Genetic variation in the IL-6 gene may also be associated with increased risk of sepsis in preterm infants (181).

Infection leads to physiological changes in host and pathogen, and gene expression changes occur in both (182). Host expression changes can help elucidate host response and may also allow classification of organisms based on host response (182). Microarrays are increasingly giving more insight into host-pathogen interaction. Diehn and colleagues performed comparative analysis of published datasets of human gene expression in response to two respiratory pathogens, *Bordatella pertussis* and *Pseudomonas aeruginosa*, looking at overlap in up- and down- regulated genes and found a large set of co-induced and co-repressed genes (182).

Experimental conditions are important when considering gene expression profiling of infection (182). Consistency and reproducibility are important so it is important to understand any limitations and to adhere to standard operating protocols (183). Choice of cell type to study is important and if mixed cell types are used care should be taken not to extrapolate results inappropriately (184). Microarray related technical differences such as different platforms, alternative splicing, different

methods of producing labelled cDNA, use of total RNA or mRNA, and amplification steps can influence results (182). In addition infection-related factors such as site of infection, strain of infection and timing of sampling in the course of infection may be important (182). Responses to non-infectious stimuli must also be considered when attributing expression patterns to infection (182). It is also necessary to compare host response to different species and stimuli before attributing any response to a single pathogen (182).

1.9 Microarray Technology

Microarrays allow the simultaneous measurement of expression of tens of thousands of genes. While there was initially the impression of a great deal of scepticism and caution over the potential clinical applications of microarray technology there was also hopefulness that it may fulfil a significant role in the clinical setting and at the bedside (185-187) including within clinical paediatrics (188). In recent years research and clinical applications have emerged in numerous medical fields and this is reflected in the increase in the volume of literature describing such applications. There are ever increasing numbers of review articles and commentaries describing potential and actual applications of microarrays in healthcare and as diagnostic tools (189-199) as well as in paediatric and neonatal research (200, 201). Some examples of such applications are: in the field of oncology in classification and prognosis of tumours (202, 203), in the fields of toxicological, environmental and dental practice looking at the effects of various chemicals and substances on human tissues (204-207), in blood screening for the purposes of transfusion (196), in neurology to investigate migraine (208), examination of the effects of valproate in paediatric

epilepsy (209) and in identifying surrogate markers of neurological disease such as neurofibromatosis 1 and Tourette syndrome (210). Further applications include diagnosis of chromosomal abnormalities in children (211), genetic analysis of variation in gene expression (212), examination of gene expression in paediatric heart disease (213), investigation of the pathogenesis of neonatal hyperbilirubinaemia at neuronal level (214) in dermatological research (215) and research into inflammatory disease-related genes in rheumatoid arthritis, SLE and asthma (216-218). There are also papers describing expression profiling in Huntington's disease (219) and in experimental bronchopulmonary dysplasia induced by oxidative stress (220). In paediatrics, applications include classification in oncology, pathogen detection, detection of genetic mutations including in newborn screening and prediction of adverse drug reactions (188). There is also ongoing research to try and identify potential targets for therapy in children with diabetes (221). Genome-wide association studies attempting to identify causal genes of a number of diseases are now being described, e.g. asthma, inflammatory bowel disease (222-224). It is very encouraging to see research such as that by Tang and colleagues who found that blood genomic patterns in rats were different after stroke, seizures, hypoglycaemia and hypoxia (225). The ability to differentiate between different pathologies using blood gene expression profiling is clearly necessary when considering diagnostic tests.

As early as 1999 it was suggested that DNA microarrays could be used for neonatal screening of genetic conditions such as sickle cell disease, alpha-1-antitrypsin

deficiency, MELAS and Factor V Leiden from newborn blood cards (DNA) (226). Use of microarrays for newborn screening is still of significant interest (227).

Microarray technology can be used both to detect pathogens and the host response to pathogens. Interaction between pathogen and host consists of both host response to pathogen and pathogen response to host.

As my research has progressed, particularly in the past couple of years, there has been increasing discussion in the literature of the benefits of high-throughput array-type pathogen detection technologies. As well as the diagnostic applications of organism detection there are also the benefits gained from increased knowledge of virulence and resistance characteristics (228-232). The shortcomings of such technology are also described in terms of lack of knowledge of nucleic acid in the normal host, the difficulties posed by small organism load in small samples, cost, lack of expertise among microbiologists and the need for new methods of validation (233). These concerns are not likely to be insurmountable. Such microarrays do not determine pathogen viability however (234) and this point is key to the potential importance of considering the host response to infection in conjunction with pathogen detection. In addition, molecular assays can detect asymptomatic carriage of pathogens, e.g. Kumar and colleagues found this for *S.aureus* and *S. pneumoniae* in secretions while examining detection of organisms causing community acquired pneumonia in adults (235). Increasing data on genome sequences of important pathogenic organisms are emerging providing important insights into specific genes which may allow identification of organisms and greater understanding of

heterogeneity between strains of a particular organism, e.g. *Streptococcus agalactiae* (236). Whole-genome sequencing has been completed for many pathogens including *Borrelia burgdorferi, Campylobacter jejuni, Chlamydia pneumoniae, Chlamydia trachomatis, Haemophilus influenzae, Helicobacter pylori, Mycobacterium tuberculosis, Ureaplasma urealyticum* (237, 238) and is in progress for many others (238). Microbial genome sequencing will allow enhanced study of host pathogen interaction, allow identification of sequences to use for microbial identification in diagnosis and for markers of resistance, and may allow prediction of severity of disease (239).

Protein microarrays of microbial antigens have been developed to detect the presence or absence of specific antibodies, e.g. to ToRCH antigens (*Toxoplasma gondii*, rubella virus, Cytomegalovirus and HSV1&2 (33). Multiplex detection of multiple pathogens from clinical samples by bacterial protein microarray is also emerging (240).

PCR can be used to amplify small amounts of microbial RNA or DNA that would have been impossible to detect using conventional methods. Multiplex real-time PCR has been investigated in the diagnosis of late onset neonatal sepsis (241). PCR-based assays can be extremely sensitive - to the extreme of being able to amplify a single copy of a microbial gene fragment and detect it (27, 242). PCR based assays can also be very specific by using primers to unique segments of DNA (27). PCR based investigations have the advantage of speed but limitations of false positives from DNA contamination of samples or the laboratory process, and false negatives if

there are inhibitors present within the sample or if the volume of sample examined is too small (27, 243, 244). The sample volume issue is relevant for neonatal blood sampling. Other drawbacks could be considered to include cost and the importance of careful sample collection and handling (243, 244). Use of gene sequencing for pathogen detection is restricted to those transcripts for which homologous probes are available (245). Also, any mismatch between the primer used and target sequence will decrease the efficacy of hybridisation: this is an issue for variant and evolving viruses (243). Development of panels of diagnostic PCR-based assays to diagnose meningitis or pneumonia for example have emerged as well as more specific organism detection, e.g. of Group B Streptococcus or Enterovirus (246, 247). There are ever increasing examples of such tests including Kim and colleagues describing rapid detection of twelve respiratory viruses using PCR (248) and of microarray "Virochip" detection of respiratory viruses from nasopharyngeal aspirates (249). There is much interest within the field of blood transfusion for the detection of blood-borne viruses such as HIV and Hepatitis C (250) and within military spheres for rapid pathogen detection (148, 251). There are several groups who have been developing rapid and portable tests for microbial DNA or RNA and many also test antibiotic resistance. An example of rapid diagnostic testing for neonatal bacterial infection is described by Shang and colleagues in their paper describing PCR testing for bacterial 16S rRNA (252). This paper is pretty convincing of the utility of such testing, presenting data showing higher sensitivity and specificity than blood culture (252). Shang's group have focussed on the use of bacterial DNA detection in the realms of neonatal septicaemia (253) and used 0.5-1ml of blood or CSF. Jordan and colleagues compared a 16S rDNA PCR assay to blood culture in early onset neonatal

sepsis – this showed a specificity of 97.5 % and negative predictive value of 99.2 % but it failed to detect a significant number of culture-proven cases (254). Caution is therefore necessary when using molecular amplification based testing in neonates. There are several other papers describing microarrays being developed for use in pathogen detection/screening particularly using PCR amplification of 16S rRNA (255, 256). Most bacterial species have a unique 16S rRNA gene sequence allowing identification of isolated organisms and classification of novel agents (27). Other groups have used 16S rDNA detection (257) and microarray using detection of sequences of 23S ribosomal DNA to detect organisms such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii* in clinical specimens including blood, stool, pus, CSF and urine (258-260). 16S rRNA may also be used to help in the identification of pathogens that are uncultured (261). Yoo and colleagues developed a diagnostic DNA microarray for 39 pathogenic bacteria using 23S ribosomal DNA and 16S-23S rDNA regions as targets for pathogen detection (262). Microarray assays based on specific gene sequences and using other conserved bacterial gene regions, e.g. topoisomer genes gyrB and parE (263, 264), can be used to rapidly detect and identify specific microorganisms such as *Mycobacterium* species (265), *Staphylococcus aureus, Escherichia coli* and *Pseudomonas aeruginosa* (266). More complete bacterial genomic information is emerging all the time. With time, more reference sequences, standardisation of techniques aiding comparison and increased understanding of bacterial and host responses under different conditions are expected (245). Multi-pathogen arrays looking at viruses or combinations of bacteria, viruses and yeasts are also emerging (267-274). Such methods can be seen

to be more rapid, accurate and cost-effective compared to culture-based methods (275, 276).

As well as detection of pathogens, microarrays are being increasingly used to examine interactions between host and pathogen (277) and in gaining greater understanding of the immune system (223, 224). Fjaerli and colleagues compared whole blood mRNA expression in 5 infants hospitalised with respiratory syncytial virus (RSV) bronchiolitis with 5 controls using microarray and QRT-PCR and found that around half of the thirty most differentially expressed genes were involved in immunological processes (278). Kawada and colleagues looked at peripheral blood gene expression profiles of children with influenza and compared acute and convalescent profiles (279). They showed 200 up-regulated and 20 down-regulated genes in the acute compared with the convalescent phase – immune response genes were over-represented in the up-regulated genes and interferon-regulated genes were strongly up-regulated in the acute phase (279). Thach and colleagues used microarrays to classify army recruits with respiratory illness into febrile vs. non febrile, healthy vs. convalescent and febrile with adenovirus vs. febrile without adenovirus using host signature transcripts (280). Ardura and colleagues examined the host response of mononuclear cells from children with invasive *S.aureus* infection and found distinctive expression profiles with over-expression of innate and under-expression of adaptive immunity in invasive disease with decreased memory CD4 and CD8 T cells and increased numbers of monocytes (281). Tang and colleagues examined expression of peripheral blood mononuclear cells to identify differences between infection and non-infectious SIRS and between Gram positive

and Gram negative sepsis (282). They found signatures to differentiate sepsis and SIRS but not between Gram positive and Gram negative infection (282). Tang's group had previously examined differences in gene expression between sepsis and control patients describing a common non specific septic response (283) and had examined neutrophil gene expression in sepsis generating a set of signature genes (284). Examination of gene expression during recovery from sepsis is also being undertaken (285). Investigation of survivors of septic shock compared to non survivors shows over-expression of innate immune cytokines, chemokine receptors and TLR pathways, i.e. restoration of inflammatory and immune response is a key step in survival following septic shock (286).

Microarrays, therefore, should lead to greater understanding of host-pathogen interactions, microbial pathogenesis, innate immunity and gene expression during various stages and severity of infection (287-292). Although this is not without problems, for example rapid changes in expression with time, diversity of pathogens, heterogeneity of human samples and complex data analysis, there is work ongoing to overcome these issues (290). Increasing human genome sequence knowledge will also facilitate progress in understanding of human host interaction with pathogens (293).

Each microbe offers a distinct combination of pathogen-associated molecular patterns that interact with specific receptors on immune cells. The host is likely to discriminate between different classes of organisms and some host genes may be implicated in pathogenesis of microbes (293). This notion was backed by Ramilo

and colleagues in 2007 when they examined gene expression profiles from 131 paediatric patients with influenza A, *E.coli, S.aureus* and *S.pneumoniae* and found different signatures for each pathogen (294). Microarrays are well-placed to be a vital tool in exploring this exciting area of research which has the potential to revolutionise infection diagnostics.

1.10 Expression Profiling

Peripheral blood cells express a large proportion of the genes in the human genome (295). Changes in gene expression may occur in response to small changes in tissues and organs due to disease, with the continuous interaction between blood and tissue allowing blood cells to act as biosensors for such changes and thus have diagnostic potential (295). Cotten and colleagues in their paper "Genomic analyses: a neonatology perspective" describe examples of candidate gene studies in premature infants including for RDS and infectious diseases (181).

Gene expression in healthy people varies. Peripheral blood is a mixture of cell types. Expression differences may reflect age, gender, race, drugs, timing of sample collection or variation in the relative proportion of cell types present rather than pathology (194, 296, 297). Study of peripheral blood mononuclear cells from healthy volunteers showed variation related to time of collection in less than 1 % of genes with none noted in more than 1 individual (of 3) (296). Identifying variation in gene expression in health is important in order to recognise and understand patterns associated with pathology such as infection (297). Variation can reflect both physiological and inter-individual variation (297). Variation in gene expression

between healthy people is smaller than in blood from patients with bacterial infection or cancer, i.e. disruptions due to pathology cause greater than normal background variation thus supporting whole genome assessment for disease signatures (297).

As well as variation arising from pathology and inter-individual variation, technical factors such as sample collection, sample handling, method of nucleic acid extraction and sample storage can contribute to differences in expression profile (298, 299). Timing of sample collection within the course of a pathological process, time to RNA extraction and storage times also need to be considered carefully. RNA is particularly prone to degradation and RNA expression profiles can change if the RNA is not stabilised. If specific transcripts are to be studied it is important to determine if the methods being used are suitable for the study question (300). In addition, comparability of gene expression results between laboratories is affected by variables such as microarray platform, labelling and hybridisation protocols and approach to data analysis (301). Results have been found to be most comparable when GO nodes of biological theme are used rather than gene-by gene comparison (301). All of these factors are important when planning studies of gene expression. In summary, the whole process of gene expression profiling from sampling, transport, RNA extraction, labelling, detection and data analysis will affect the quality, reliability and reproducibility of the final result. Standardisation of the entire process is therefore necessary (194). Optimisation of the process from sample collection to RNA processing forms a significant proportion of the work contained in this thesis.

1.11 Foreword to Thesis

The prospect of using microarrays to investigate host response to infection in neonatal blood is exciting. Looking at the host response has several theoretical advantages over pathogen detection. Traditional culture-based pathogen detection requires that there is a certain load of viable organisms. Molecular methods of detection can detect tiny amounts of microbial material and do not require organisms to be viable. This raises the possibility of detection of tiny amounts of contaminant and also detection of organism where the host has already rendered it harmless. Given the important contribution that the host response makes to the pathogenesis of sepsis, it would provide valuable information to examine the host response to a microbial pathogen. Each of us likely experiences numerous host-pathogen encounters on a daily basis but only becomes ill relatively infrequently. Theoretically it could be possible to have a bacteraemia with a low load of pathogen that could be rapidly eradicated by the host response without causing discernable symptoms of sepsis. Microarray pathogen detection methods could detect these organisms but without the host response information it would be impossible to tell/predict the clinical relevance of the pathogen-host interaction. Ultimately, combining both pathogen detection and examination of host response is likely to yield the most useful and reliable information, and I believe that this will have an important role to play in infection diagnostics in the future.

When considering the aim of achieving a diagnostic tool for infection using microarray technology one could be easily discouraged by the apparently overwhelming number of obstacles to be considered. Potential obstacles to success

include variations introduced by differences in sampling technique, sample handling, RNA extraction, labelling, hybridisation and the microarray platform chosen. Biological differences depending on stage of infection, route of entry or site of infection, strain of infection and load of infection may also affect the results achieved. I felt strongly that this should not be a deterrent to working towards this aim. In fact, the main aims of this study were to examine many of these factors and to work to minimise their impact.

For this study, I set out to investigate and produce a standardised procedure for sampling neonatal blood in order to minimise technical variation. The original scope of this study was to determine whether it was possible to use RNA obtained from neonatal blood in microarray experiments, thus facilitating future microarray studies using neonatal blood derived RNA. In actual fact, after achieving these aims, due to promising results and the reducing costs of microarrays, I have been privileged to be able to continue further in this study as part of a research group to examine comparisons between control and infected groups. Some of this further data is presented in this thesis.

When I embarked on this research, there was no data in the literature regarding expression profiling of infection in children and more particularly not in the neonatal population. As well as the technical difficulties of collecting blood from neonates and the absolute necessity of any test being carried out on small samples due to the relatively small circulating volume of such patients, the fact that neonates have relatively immature immune systems and may be encountering infection for the first

time, made the prospect of studying neonatal infection profiles a challenging but exciting prospect. Over the time that I have been working on this project, there have been publications starting to emerge that are looking at gene expression profiling of a variety of conditions within the field of paediatrics. In addition there have been increasing numbers of publications using microarray technology to examine the host response to infection. It is both encouraging and exciting to see the increasing knowledge in this area. It reflects increasing acceptance of microarray use and of expression profiling as valid and useful tools. Although it would be potentially possible to create a diagnostic tool using the data from a single research group, the use of concurring data from a number of groups would create a more robust tool. I believe that the data presented in this thesis brings a useful contribution to this blossoming field of research.

Chapter 2

Methods

This chapter contains the methods that were used for the blood sampling and experimental work that is discussed in the chapters that follow. This chapter also contains discussion as to the rationale behind the choice of methods used. This chapter is split into the following sections:

- Methods used for blood sampling
- Methods used for RNA extraction
- Methods used for RNA quality assessment
- Methods used for labelling and hybridisation of microarrays

2.1 Methods Used for Blood Sampling

All blood samples were taken with written, informed, parental consent.

Umbilical Cord Blood Sampling

As discussed in chapter 3, umbilical cord blood was used a surrogate for neonatal blood for the initial work in this study. Umbilical cord blood sampling was performed as soon as possible after delivery of the placenta and in all except one case sampling was within 15 minutes of delivery of the baby (the exception was sampled at 20 minutes). Study samples were taken only after any samples that were required clinically had been obtained. The cord blood sampling method described here is similar to that described for established stem cell gathering techniques (302). Locally, clinical cord blood samples are taken routinely from every delivery. In all cases where consent for the study had been granted, I took both the clinical and the study samples. Gloves, apron and goggles were worn at all times. The clinical samples were taken using a needle from the umbilical cord as near to the cord clamp, i.e. as far from the placenta as possible. The umbilical cord was then cleaned with

filter-sterile phosphate buffered saline solution (Gibco®) and sterile swabs. Sterile scissors were then used to cut the cord nearer the placenta than the site of clinical blood sampling. The end of the cord still attached to the placenta was cleaned further with sterile swabs soaked in phosphate buffered saline and the umbilical vein catheterised with a sterile nasogastric tube (NG) (size 5) that was advanced until good blood flow was obtained. Blood samples were drawn off into sterile syringes connected to the end of the NG tube. The blood was then distributed immediately into tubes - either dripped directly into neonatal blood collection bottles or injected directly into a PAXgene™ Blood RNA tube through a sterile needle. Each sample bottle was inverted several times after sampling.

Samples were taken directly to the laboratory in sealed plastic containers. Those samples collected into PAXgene™ Blood RNA tubes were transported at room temperature. All other samples were transported on ice.

Neonatal Blood Sampling

The volume of blood sampled for the study in each case was between 0.5 – 1 ml. These values were chosen in order to display the natural variation in sampling volume that occurs clinically where precision measuring is not possible and at the same time reflecting the need to keep sample volumes as small as possible in this population. The circulating blood volume of a newborn is 75-80 ml per kg body weight. In general, for consistency, I aimed for samples to be around 0.6 – 0.7 ml in volume.

No extra venepuncture was performed for the sake of the study; extra blood was taken at the same time as clinical sampling only. If, after the required clinical samples were obtained, the needle or cannula did not bleed back or was dislodged, it was not re-sited for the sake of the study.

All samples for blood culture were taken according to the neonatal unit policy as much as possible. The skin was cleaned with iodine antiseptic as per unit policy, however instead of wiping the iodine antiseptic off with an alcohol swab, a sterile swab soaked in sterile saline was used to wipe the antiseptic off and the skin dried with a sterile swab. Blood culture samples were taken where possible from a closed system (syringe attached to cannula). The study sample was then aspirated from the hub of the cannula/ needle using a needle and syringe. The study sample was injected immediately into the PAXgene™ blood RNA tube. For safety, the PAXgene™ tube was in a test-tube rack to avoid risk of needle-stick injury. The PAXgene™ blood RNA tube was then inverted several times and transferred either directly to the lab or put straight into a -20 °C freezer. All samples were transferred to the laboratory in sealed plastic containers. Those that had been frozen in the neonatal unit were transferred on ice and immediately placed in a -20 °C freezer in the laboratory to ensure that defrosting did not occur.

2.2 Methods of RNA Extraction from Neonatal Whole Blood

Literature Review & Planning

At the time of study design (autumn 2003 to spring 2004), I performed searches of the literature to determine up-to-date information on methods of RNA extraction from human whole blood. The searches revealed that there were a variety of methods used to extract RNA from human whole blood but that there were no studies detailing extractions from small volumes, (as would necessary in using neonatal blood samples) that sought the yields necessary for microarray analysis.

Examination of the literature revealed that many groups opted for isolation of peripheral blood mononuclear cells (PBMCs) (303, 304) using methods such as Ficoll-Hipaque gradients / Percoll centrifugation when extracting RNA from human blood (305-312) or alternative methods such as using gelatine (313). From the outset it was decided that I would not to go down this route for several reasons. Firstly we wanted to find a method that utilised whole blood as this would simplify the journey of samples from the clinical area to the laboratory. Secondly we did not wish to restrict our examination of RNA expression profile to monocytes. Thirdly, it would avoid the additional time involved in cell separation and culture and fourthly we were conscious of cost. In addition, we made the decision not to opt for methods that would require amplification of sample RNA not only because of the extra time involved to incorporate such steps but also because this would be an additional source of variation.

Several papers used a guanidine-isothiocyanate-phenol-chloroform method of extraction (209, 305, 311, 314-319). Studies directly comparing three of these methods showed that Trizol-LS gave consistently better results (320). I therefore chose TRIzol®LS as my example of this method to try out. Other methods described, included on-column methods either alone or as clean-up (209, 307, 316). After reviewing the manufacturer's information on a few alternatives, I chose the QIAamp® RNA mini kit as our example of an on-column method as it looked as though it should be suitable for use with neonatal blood.

Filter paper dried blood spot techniques of sample collection were of particular interest with a neonatal population in mind. The literature I reviewed on blood spots was generally describing their use with forensic specimens for DNA or PCR analysis or were from mouse blood (321, 322). For example, blood stains can have RNA extracted using a guanidine-isothiocyanate-phenol-chloroform method and yield 170-260 ng RNA from a 50 μl stain (323). There have been reports of blood spots from Guthrie cards being used to successfully isolate mRNA using RNeasy, however these were pooled samples that went on to have nested PCR carried out (324). More recently Haak and colleagues have shown that RNA can be isolated and amplified from stored blood spot cards that can be analysed by quantitative microarray and qPCR (325). Another recent paper describes that DNA can be obtained from blood cards to allow whole genome microarray analysis for genetic testing (326). The method of filter-paper based RNA extraction that I chose was the Whatman® system. This decision was made on availability and cost at the time of planning.

PAXgene™ Blood RNA system is marketed as a technology that:

"consolidates and integrates the key steps of whole blood collection, nucleic acid stabilization and RNA purification... minimizing the unpredictability associated with RNA processing...enhanced accuracy of intracellular RNA analysis" (PreAnalytix website: www.preanalytix,com).

According to PreAnalytix, PAXgene™ tubes contain a proprietary reagent that immediately stabilizes intracellular RNA. This system looks at whole blood which, unlike if using separated peripheral blood mononuclear cells, will look at neutrophil expression. There were several encouraging papers comparing PAXgene™ RNA extraction to other methods with some going on to run RNA on array expression studies (327-330). This made the PAXgene™ system a very attractive option that was therefore included in this study.

After selecting one example of each method of extraction, a protocol for use on neonatal blood samples was drawn up. This was not always straightforward as the techniques had not been described for use with neonatal blood which generally has a higher white cell content than adult blood. There therefore followed several personal communications between myself and the companies involved to enable me to optimise protocols where possible – these are referenced at the appropriate places in this section.

Towards the end of my methodology experiments I came across MagaZorb®, a magnetic bead based method of RNA extraction which had just come on the market. As this was a completely different method to the others already examined, I went on to investigate it further.

Since my comparison of the methods described in this study, there have been several new RNA extraction kits on the market but these were not examined in retrospect as my goal of finding a method which would consistently extract high quality RNA from small neonatal blood samples of sufficient quantity to carry out microarray experiments had already been achieved.

Methods

In order to ensure absolute consistency and eliminate variation, checklists were compiled for each method of RNA extraction used. The checklists and the following formal descriptions of the methods used were compiled following the manufacturers' instructions and the general laboratory practice protocols for the Centre of Genomic Technology and Informatics (now Division of Pathway Medicine) with a few modifications. The modifications are discussed in detail in italics beneath each method. For all laboratory work described in this work, disposable gloves were worn both as personal protection and also to prevent RNase contamination of samples. Sterile disposable plastic ware was used at all times and the pipettes used were reserved solely for this work.

TRIzol®LS Extraction (Invitrogen™, Life Technologies)

1. 0.5 ml of RNase-free water was added to 0.5 ml of blood.
2. 3 ml of TRIzol®LS were then added and the cells lysed by repetitive pipetting.
3. The TRIzol®LS/Blood/water mix was then left to sit at room temperature for 5 minutes.

4. 800 microlitres of chloroform were then added, the samples capped securely and shaken vigorously by hand for 15 seconds.
5. The samples were then sat at room temperature for 2-3 minutes.
6. Samples were centrifuged at 4,000 rpm in a table top Eppendorf 5810R centrifuge for 1 hour at 4 °C.
7. The colourless aqueous phase was carefully removed with a pipette and transferred to fresh tubes: 0.5 ml per tube. (The volume of the aqueous phase is ~ 70 % of the volume of TRIzol®LS reagent used for homogenisation, i.e. around 2.1 ml).
8. The tubes that had been removed last sometimes needed a further period of centrifugation at 12,000 rpm for 10 minutes and the aqueous layer transferred to a new tube. This and all subsequent centrifugation steps were carried out in a Heraeus Biofuge fresco.
9. The RNA was precipitated by mixing with 0.5 ml isopropyl alcohol per tube.
10. The tubes were then incubated at room temperature for 10 minutes.
11. Centrifugation at 12,000 rpm for 10 minutes at 4 °C was carried out and the RNA precipitate was seen to form a gel-like pellet on the side of the tubes.
12. The supernatant was removed and discarded.
13. The RNA pellet was washed once with 75 % ethanol: 1 ml per tube.
14. Each tube was vortexed & centrifuged at 8,800 rpm for 5 minutes at 4 °C.
15. Some samples required a further 5-10 minutes centrifugation.
16. Most of the supernatant was removed by pipetting.
17. The RNA pellet was air-dried without letting the pellet dry out completely.

18. The RNA was dissolved in 100 microlitres of RNAse-free water by passing the solution a few times through a pipette tip and incubating for 10 minutes at 55 – 60 °C.

Notes on TRIzol®LS *method:*

- *Invitrogen recommended that the blood be diluted 1:1 with water (step 1 above) in order to optimise results (personal communication and* TRIzol®LS *Reagent information form no. 18057N).*
- *The centrifugation speed of 4,000 rpm on the Eppendorf Centrifuge 5810R in step 6 above was set after Invitrogen advised using a table-top centrifuge for 30-60 minutes (personal communication), i.e. longer time at lower force. This was necessary as I did not have access to a centrifuge that would have both achieved the recommended 15 minutes at 12,000 x g and taken tubes that could withstand* TRIzol®LS *and chloroform.*

QIAamp® RNA Mini Protocol for RNA Cleanup (Qiagen)

1. 350 microlitres of Buffer RLT was added to the 100 microlitres of dissolved RNA and the sample was mixed by pipetting.
2. 250 microlitres of 100 % ethanol were added to the lysate and mixed by pipetting.
3. 700 microlitres of sample was pipetted into a spin column sitting in a 2 ml collection tube and centrifuged for 15 seconds at 9,200 rpm.

4. Flow through and collection tube were discarded.
5. Using a new collection tube, any remaining sample was pipetted into the spin column and centrifuged at 9,200 rpm for 15 minutes.
6. The flow through and collection tube were discarded and the column placed into a new collection tube.
7. 350 microlitres of Buffer RW1 were pipetted into the column and centrifuged at 9,200 rpm for 15 seconds to wash.
8. Flow through was discarded.
9. Separately, 10 microlitres of DNase 1 stock solution (per sample) were added to 70 microlitres (per sample) Buffer RDD and mixed gently by tapping the tube.
10. 80 microlitres of DNase 1 incubation mix from step 9 was pipetted directly onto the column membrane.
11. The column was left at room temperature for 15 minutes.
12. 350 microlitres of Buffer RW1 were pipetted into the column and centrifuged at 9,200 rpm for 15 seconds.
13. Flow through was discarded and the column placed in a new collection tube.
14. 500 microlitres of Buffer RPE were added to the column and centrifuged for 15 seconds at 9,200 rpm.
15. Flow through and collection tube were discarded and a new collection tube used.
16. A further 500 microlitres of Buffer RPE were added to the column and centrifuged at full speed for 3 minutes.

17. The collection tube and flow through were discarded and the column placed in a new collection tube.
18. The column was centrifuged for 1 minute at full speed.
19. The column was transferred to a new 1.5 ml collection tube and 50 microlitres of RNase-free water was pipetted directly onto the column membrane.
20. After 1 minute, the column was centrifuged at 9,200 rpm for 1 minute to elute.
21. Keeping the same collection tube a further 50 ml of RNase-free water was pipetted onto the membrane.
22. After 1 minute the sample was centrifuged at 9,200 rpm for 1 minute.
23. The column was centrifuged for a further minute at 9,200 rpm.
24. The eluate was transferred into a new tube at stored at -80 °C.

Notes on the QIAamp® RNA Mini Protocol for RNA Cleanup:

- *The method was taken from the QIAamp® RNA Blood Mini Handbook – version January 1999, p24).*
- *An on-column DNase step was incorporated (steps 9-13 above).*
- *The optional extra 1 minute centrifugation step (steps 17-18 above) was incorporated.*
- *All centrifugation steps were performed using a* Heraeus Biofuge.

QIAamp® Blood RNA Mini Kit Extraction (Qiagen)

1. 0.5 ml of cord blood (in EDTA) was mixed with 2.5 ml of Buffer EL. In order to allow sufficient volume for efficient mixing 2 aliquots of 0.25 ml of blood were each mixed with 1.25 ml of Buffer EL for each sample.
2. Samples were incubated on ice for 20 minutes, being vortexed briefly twice during this period.
3. The samples were then centrifuged at 2,000 rpm for 10 minutes at 4 °C.
4. The supernatant was removed and discarded.
5. 0.5 ml of Buffer EL was added to each cell pellet and the cells re-suspended by vortexing briefly.
6. The tubes were then incubated for an additional 10 minutes on ice.
7. The tubes were centrifuged for 10 minutes at 4 °C at 2,000 rpm.
8. The supernatant was removed and discarded.
9. 300 microlitres of Buffer RLT were added to each tube and mixed by pipetting.
10. The lysate from each pair of tubes was combined and pipetted directly into a QIAshredder spin column,
11. The column was centrifuged at maximum speed for 2 minutes.
12. The shredder column was discarded and the homogenized lysate retained.
13. 600 microlitres of 70 % ethanol were added to the homogenized lysate and mixed by repetitive pipetting.
14. 700 microlitres of the sample were pipetted into a new QIAamp spin column sitting in a 2 ml collection tube.
15. The column was centrifuged at 9,200 rpm for 15 seconds.

16. The flow-through and collection tube were discarded and the column placed in a new collection tube.
17. Any remaining sample was pipetted into the QIAamp spin column and centrifuged at 9,200 rpm for 15 seconds.
18. The flow-through and collection tube were discarded and the column placed in a new collection tube.
19. 350 microlitres of Buffer RW1 were pipetted into the column and centrifuged at 9,200 rpm for 15 seconds to wash.
20. Flow through was discarded.
21. Separately, 10 microlitres of DNase 1 stock solution (per sample) were added to 70 microlitres (per sample) Buffer RDD and mixed gently by tapping the tube. The tube was centrifuged for up to 2 seconds at 2,000 rpm to collect droplets from the lid.
22. 80 microlitres of DNase 1 incubation mix from step 21 was pipetted directly onto the column membrane.
23. The column was left at room temperature for 15 minutes.
24. 350 microlitres of Buffer RW1 were pipetted into the column and centrifuged at 9,200 rpm for 15 seconds.
25. The flow-through and collection tube were discarded and the column placed in a new collection tube.
26. 500 microlitres of Buffer RPE were added the column and the column centrifuged for 15 seconds at 9,200 rpm.
27. The flow-through and collection tube were discarded and the column placed in a new collection tube.

28. A further 500 microlitres of Buffer RPE were added to the column and the column centrifuged at full speed for 3 minutes.
29. The flow-through and collection tube were discarded and the column placed in a new collection tube.
30. The column was centrifuged at full speed for 1 minute.
31. The column was transferred into a new 1.5 ml collection tube and 50 microlitres of RNase-free water were pipetted directly onto the QIAamp membrane.
32. After 1 minute, the column was centrifuged at 9,200 rpm for 1 minute to elute.
33. Using the same collection tube, a further 50 microlitres of RNase-free water were pipetted onto the column membrane.
34. After 1 minute, the column was centrifuged for 1 minute at 9,200 rpm.
35. The column was centrifuged for a further 1 minute at 9,200 rpm.
36. The eluate was transferred into a new tube and was stored at –80 °C.

Notes on the QIAamp® RNA Mini Protocol

- *The method was taken from the QIAamp® RNA Mini Protocol for Isolation of Total Cellular RNA from Whole Human Blood found on pages 13 – 16 of the QIAamp® RNA Blood Mini Handbook, version January 1999.*
- *Each sample was split into two aliquots that were processed in parallel until merged again in step 10 above. This was so that the sample did not exceed ¾ of the volume of the tube to allow efficient mixing as recommended.*

- *The additional time needed to allow the suspension to become translucent was incorporated into the incubation time in step 2 above.*
- *An additional 10 minute incubation on ice was incorporated (step 6 above) as recommended in the trouble shooting section on page 25 of the QIAamp® RNA Blood Mini Handbook, version January 1999. This was incorporated because the pellet obtained after step 3 above was red, indicating incomplete erythrocyte lysis.*
- *A total of 600 microlitres (300 microlitres to each aliquot of sample) of Buffer RLT was added to each sample in step 9 above. The larger amount of Buffer RLT was used in expectation that the cord blood would be rich in leukocytes.*
- *An on column DNase digestion was incorporated: steps 19-23 above.*
- *The optional centrifugation step after addition of Buffer RPE (step 30) was incorporated.*
- *All centrifugation steps were performed using a* Heraeus Biofuge.
- *This method can only be performed on fresh blood and cannot be used on frozen blood samples.*
- *The amounts used above were specifically worked out for 0.5 ml of blood. A single QIAamp spin column can process 1.5 ml of adult blood or a maximum of 1 x 10^7 leucocytes: greater amounts would mean that the leucocytes would not be fully lysed. The number of leucocytes in neonates would be expected to be in the range of 9 - 30 x 10^6 per ml, i.e. 0.9 –3.0 x 10^7 per ml. Considering this, and after discussion with a Qiagen*

representative, I felt that it was reasonable to use 0.5 ml of cord blood in the above protocol.

PAXgene™ Blood RNA System (PreAnalytix: Qiagen, BD)

All centrifugation was carried out at room temperature.

1. Each sample was incubated at room temperature for a minimum of 2 hours (4 hours after removal from freezer if frozen).
2. Each PAXgene tube was centrifuged for 10 minutes on a table-top centrifuge set to 4,000 rpm using a swing-out rotor (Eppendorf 5810R centrifuge).
3. The supernatant was decanted off and the rim of the tube dried with a clean tissue.
4. 5 ml of RNase-free water was added to the tube and the tube closed.
5. The pellet was then re-suspended by thorough vortexing and the sample centrifuged at 4,000 rpm in the Eppendorf 5810R centrifuge for 10 minutes.
6. The entire supernatant was then decanted and the rim dried with a clean tissue.
7. The pellet was re-suspended in 360 microlitres of Buffer BR1 by vortexing.
8. The sample was then pipetted into a fresh 2 ml microcentrifuge tube.
9. 300 microlitres of Buffer BR2 was added to the tube.
10. 40 microlitres of proteinase K was then added and the sample mixed by vortexing.

11. The sample was incubated for 10 minutes in a water bath at 56 °C with a brief vortex carried out after 5 minutes (cooling was avoided).
12. The sample was then centrifuged at maximum speed in a microcentrifuge for 10 minutes. This and all subsequent centrifuge steps were carried out using a Heraeus Biofuge.
13. The supernatant was transferred into a fresh 2 ml microcentrifuge tube using a pipette. A small amount of debris in the supernatant was accepted.
14. 350 microlitres of 100 % ethanol were added.
15. The sample was mixed by vortexing and centrifuged for less than 2 seconds at 2,000 rpm to remove droplets from the inside of the lid.
16. 700 microlitres of sample was added to a PAXgene column sitting in a 2 ml processing tube.
17. The column was then centrifuged at 9,200 rpm for 1 minute.
18. The flow-through and processing tube was discarded and the column placed in a new processing tube.
19. Any remaining sample was then pipetted into the column and centrifuged at 9,200 rpm for 1 minute.
20. The flow-through and processing tube was discarded and the column placed in a new processing tube.
21. 350 microlitres of Buffer BR3 was pipetted into the column and centrifuged for 1 minute at 9,200 rpm.
22. The flow-through and processing tube were discarded and the column placed in a new processing tube.

23. Separately, 10 microlitres of DNase 1 stock solution (per sample) was added to 70 microlitres (per sample) of Buffer RDD and mixed by tapping the wall of the tube. The tube was then centrifuged briefly to collect all droplets from the side/lid.
24. 80 microlitres of the DNase 1 incubation mixture from step 23 were pipetted directly onto the spin column membrane,
25. The column was then left on the benchtop for 15 minutes.
26. 350 microlitres of Buffer BR3 were pipetted into the column and the column centrifuged for 1 minute at 9,200 rpm.
27. The flow-through and processing tube were discarded and the column placed in a new processing tube.
28. 500 microlitres of Buffer BR4 were added to the column and the column centrifuged for 1 minute at 9,200 rpm.
29. The flow-through and processing tube were discarded and the column placed in a new processing tube.
30. A further 500 microlitres of Buffer BR4 were added to the column and the column centrifuged for 3 minutes at maximum speed.
31. The flow-through and processing tube were discarded and the column placed in a new processing tube.
32. The column was then centrifuged at full speed for 1 minute.
33. The flow-through and processing tube were discarded and the column placed in a 1.5 ml elution tube.
34. To elute, 40 microlitres of Buffer BR5 were pipetted directly onto the column membrane.

35. After 1 minute, the column was centrifuged at 9,200 rpm for 1 minute.
36. Using the same elution tube, a further 40 microlitres of Buffer BR5 were pipetted onto the column membrane.
37. After 1 minute, the column was centrifuged at 9,200 rpm for 1 minute.
38. The column was centrifuged at 9,200 rpm for a further 1 minute.
39. The elution tube was then incubated in a water bath set to 66 °C for 5 minutes.
40. The tube was then immediately transferred onto ice and chilled for at least 3 minutes.
41. The sample was then transferred into a new tube and stored at –80 °C.

Notes on the PAXgene™ Blood RNA Extraction:

- *The above method was taken from the PAXgene™ Blood RNA Kit Handbook, version April 2001.*
- *The centrifugation in step 12 above was increased to 10 minutes to ensure that the supernatant separated adequately – early experiments showed that the 3 minutes recommended was not sufficient.*
- *Steps 19 & 20 above were necessary as there was usually just over 700 microlitres of sample to be transferred onto the spin column.*
- *The optional on-column DNase digestion step was incorporated (steps 23-27 above) as described in Appendix A of the PAXgene™ Blood RNA Kit Handbook, April 2001.*
- *The optional additional centrifuge step (step 32 above) was incorporated. In order to maximise elution the columns were left for 1*

minute in steps 35 and 37 prior to centrifugation. In addition an additional 1 minute centrifuge step was added (step 38).

- *Note: Later versions of the PAXgene™ Blood RNA Kit Handbook (February 2004) suggested a 10 minute centrifuge step after the 1st waterbath incubation (as I had done anyway) or to use a QIAshredder spin column after the 1st waterbath incubation. I never used the QIAshredder as I had been achieving good results with the above protocol. Later versions also included the 2nd application of sample to the spin column.*

MagaZorb® RNA Mini-Prep Kit Extraction (Cortex Biochem™)

1. All reagents and samples were equilibrated to room temperature.
2. 20 microlitres of proteinase K solution was pipetted into the bottom of a 2 ml microcentrifuge tube.
3. 200 microlitres of whole blood (EDTA sample) was added to the tube and mixed gently by pipetting.
4. 200 microlitres of Lysis Buffer were added and the tube vortexed for 15 seconds.
5. The tube was incubated at 56 °C for 10 minutes.
6. 500 microlitres of Binding Buffer was added.
7. 20 microlitres of well-mixed MagaZorb® Reagent were added.
8. The tube was mixed gently and then incubated at room temperature using an end-over-end rotator.

9. The MagaZorb® RNA bound particles were sedimented using a magnetic rack.
10. The supernatant was aspirated and discarded.
11. The tube was then removed from the magnetic rack.
12. The particles were washed by adding 1 ml of Wash Buffer and inverting the tube several times.
13. The particles were sedimented using the magnetic rack.
14. The supernatant was aspirated and discarded.
15. Separately, a DNase 1 working solution was prepared by combining 20 microlitres (per sample) of DNase 1 with 80 microlitres (per sample) of DNase 1 Buffer.
16. The tube from step 14 was removed from the magnetic rack.
17. 4 microlitres of SUPERase IN was added to the tube and mixed gently.
18. 100 microlitres of the DNase 1 working solution from step 15 was added and mixed gently.
19. The tube was incubated at room temperature for 15 minutes while mixing in an end-over-end rotator.
20. The tube was removed from the end-over-end rotator and 1 ml of wash buffer was added and mixed by inverting the tube several times.
21. The tube was placed in the magnetic rack and the particles sedimented.
22. The supernatant was aspirated and discarded.
23. The tube was removed from the magnetic rack and a further 1 ml of Wash Buffer was added to the tube and the tube inverted several times.
24. The particles were sedimented using the magnetic rack.

25. The supernatant was aspirated and discarded.
26. 50 microlitres of RNase-free water was added to the tube.
27. The tube was placed in an end-over-end rotator and mixed for 10 minutes at room temperature.
28. The tube was placed into the magnetic rack and the particles sedimented.
29. Separately, a clean tube had 4 microlitres of SUPERase IN added.
30. The supernatant from step 28 was then transferred into the tube containing SUPERase IN, mixed gently and frozen at -80 °C.

Notes on the MagaZorb® RNA Mini-Prep Kit Extraction:

- *MagaZorb® RNA Mini-Prep Kit can only be used on fresh whole blood samples. Frozen samples cannot be used.*
- *The above method was taken from the MagaZorb® RNA Principle and Procedure Leaflet (2001-Rev3-04G08) and from the MagaZorb® Supplementary Protocol B (2004-RevA) supplied with the kit.*
- *The supplementary protocol B to remove DNA contamination was incorporated (steps 15- 22 above).*

Whatman FTA® Extraction Method

For the Whatman FTA® extraction method, 3M sodium acetate pH 5.2 was prepared in advance of the experiment. In addition, RNA processing buffer was made up fresh on each occasion and stored on ice. RNA processing buffer contained 10 mM Tris-

HCl pH 8.0, 0.1 mM EDTA, 800 U/ml of RNase Out, 200 µg/ml glycogen and 2 mM DTT achieved in the final volume.

- The samples on the Whatman FTA® cards were left to air dry at room temperature for at least 2 hours (minimum recommended is 1 hour).

For each Whatman FTA® experiment the following was carried out:

1. The sample disc was removed from the dried blood spot using the Harris punch and cutting mat as per manufacturer's instructions.
2. The disc was placed in a 0.2 ml RNase-free and DNAse free PCR tube.
3. 750 µl of RNA processing buffer was added to the tube and pipetted up and down twice.
4. The tube was capped and incubated on ice for 15 minutes – mixing by pipetting every 5 minutes.
5. The RNA was precipitated by adding 75 µl of 3M Sodium acetate pH 5.2 and 750 µl of ice cold 100 % isopropanol.
6. The sample was then incubated at –20 °C for 1 hour.
7. The tube was placed in a microcentrifuge tube and spun down at 12,000 x g for 5 minutes.
8. Supernatant was removed by pipetting.
9. The "pellet" was washed with 500 µl of ice cold 75 % ethanol and spun for 5 minutes at 12,000 x g.
10. *Additional steps in those samples with DNase treatment:*

- *A DNase working solution was prepared by combining 20 µl of DNase 1 with 80 µl of 10X DNase 1 Buffer*
- *The 75 % ethanol from the microcentrifuge tube containing the pellet was removed by pipetting.*
- *100 microlitres of DNase working solution was added to the tube containing the pellet and incubated at room temperature for 15 minutes.*
- *The DNase working solution was removed by pipetting.*
- *The 'pellet' was washed with a further 500 µl of ice cold 75 % ethanol and spun for 5 minutes at 12,000 x g.*

11. The supernatant was removed by pipetting and the "pellet" was air dried.
12. The pellet was re-suspended in 50 µl of TE^{-1} buffer.
13. Sample was stored at –80 °C until quality control checks could be carried out.

Notes on the Whatman FTA® Extraction Method:

- *The above method was taken from Whatman FTA Protocol BR01: Applying and Preparing Blood and Tissue/Cell Culture Samples on FTA® Cards for RNA Analysis. Reference was also made to: "Removing a Sample Disc from an FTA or CloneSaver card for Analysis" (protocol number BD09) and to "FTA® Protocols: Collect, Transport, Archive and Access Nucleic Acids... All at Room Temperature" by Whatman®. 2002. WB 120047.*
- *A few points in the protocol were noted to be inaccurate/vague, eg. the section describing the need to use the whole disc of cells to be processed for Northern blot stated that the wash step be "increased to 750 µl sterile*

water", then it goes on to describe the wash step being in RNA processing buffer. I sought clarification from the company directly by telephone and e-mail and have incorporated their advice into the method described above.

- *An additional DNase step was incorporated for some samples.*

Additional Information on RNA Extractions

1. DNase Steps.

For the on-column DNase steps described in the methods for the QIAamp® RNA Mini Protocol for RNA Cleanup, QIAamp® Blood RNA Mini Kit Extraction and the PAXgene™ Blood RNA System, I used the RNase-Free DNase Set from QIAgen. The DNase 1 solution was made up according to manufacturer's instructions:

- The lyophilised DNase 1 (1,500 Kunitz units) was dissolved in 550 microlitres of the RNase-free water provided in the kit.
- Mixing was performed by gently inverting the tube.
- The DNase 1 solution was then pipetted into aliquots of 20 microlitres and the aliquots stored at -20 °C until needed.

The DNase 1 solution in the MagaZorb® RNA Mini-Prep Kit Extraction was made using DNase 1 (2 U/µl) and the 10X DNase 1 Buffer from Ambion (Cat. No. 2222) as recommended in the MagaZorb® Supplemental Protocol B. These were also used in the Whatman FTA® extraction experiments.

2. Centrifuge speeds

For all RNA extractions it is important that the correct rotational speed/ g-force are used.

Relative centrifugal force (in g) = 1.12 x rotor radius (in cm) x $(rpm/1000)^2$

A. Heraeus Biofuge Fresco and Biofuge Pico

The rotor number for these centrifuges was #7500 3325 and the technical data is:

Max speed 13,000
Max rcf (at max speed) 16,060
Min speed 2,000
Min RCF 380
Max radius 8.5 cm
Min radius 5.9 cm
Angle 40^0

The manufacturer's instruction manual provided a graph to calculate the speed/g, I used this graph to calculate the following values which were essential to the study:

For TRIzol® LS:

- For the steps requiring centrifuge at less than 7,500 x g, the centrifuge was run at 8,800 rpm (7,500 x g = 9,000 rpm).
- For the steps requiring centrifuge at 12,000 x g, the centrifuge was run at 12,000 rpm (12,000 x g = 12,000 rpm).

For QIAamp®:

- For the steps requiring centrifuge at 400 x g, the centrifuge was run at 2,000 rpm (400 x g = 2,000 rpm)

- For the steps requiring centrifuge at over 8,000 x g, the centrifuge was run at 9,200 rpm (8,000 x g = 9,000 rpm)
- For the steps requiring centrifuge at 20,000 x g ("maximum speed"), the centrifuge was run at its maximum of 13,000 rpm (= 16,060 x g)

For PAXgene™:

- For the steps requiring centrifuge at 10,000 x g, the centrifuge was run at 10,000 rpm (10,000 x g = 10,000 rpm)
- For the steps requiring centrifuge at less than 1,000 x g, the centrifuge was run at 2,000 rpm (1,000 x g = 2,250 rpm)
- For the steps requiring centrifuge at over 8,000 x g, the centrifuge was run at 9,200 rpm (8,000 x g = 9,000 rpm)
- For the steps requiring centrifuge at 20,000 x g ("maximum speed"), the centrifuge was run at its maximum of 13,000 rpm (= 16,060 x g)
- For the experiment using aliquots of PAXgene™ fluid where the centrifuge was required to be at 3,000-5,000 x g, the centrifuge was set at 6,500 rpm (= 4000 x g)

For spinning the gel for the RNA 600 Nano Assay for the Agilent Bioanalyzer:

- Centrifuge was required to be set at 1,500 x g +/- 20 %. The centrifuge was set at 4,000 rpm (= 1500 x g).

B. Eppendorf Centrifuge 5810R

This table-top centrifuge has Rotor A-4-62.

Maximum speed 4,000

Maximum rcf (at max speed) 3,220

- For TRIzol LS method:

  12,000 x g for 15 minutes was recommended but I did not have access to a machine that was both large enough to take the necessary chemical-resistant tubes and able to achieve that force.

  After discussion with Invitrogen the Eppendorf Centrifuge was set at 4,000 rpm (3,161 x g, using 15 ml NUNC tube)

- For PAXgene™ method:

  3,000-4,000 x g was recommended.

  Centrifuge set at 4,000 rpm (3,072 x g, using PAXgene™ tube).

2.3 Assessment of RNA Quality

Quantitative Assessment of RNA Using Spectrophotometry

The spectrophotometer used for all of the umbilical cord blood RNA samples was the ThermoSpectronic Biomate 5 spectrophotometer and VISION 32 (bit Version 1.25). This was used for quantification and to obtain values to calculate A260:A280 ratio.

The initial neonatal blood RNA samples were also run on this spectrophotometer but later neonatal blood samples and any nucleic acid assessment during the course of microarray labelling and hybridisation was run on a ThermoSpectronic NanoDrop™1000 Spectrophotometer. All of the samples to be used with the Illumina® microarrays (chapter 5) that had initially been run on the Biomate were re-run on the NanoDrop™ to ensure consistency. The methods used for both spectrophotometers are detailed below.

Spectrophotometry Using the ThermoSpectronic Biomate 5

1. Unless otherwise stated in individual experimental methods,
   78 µl of 10 mM Tris-Cl pH 7.5 was pipetted into microcentrifuge tubes – one aliquot for each sample to be assessed.
2. 2 µl of each sample to be assessed was added to one of the tubes containing the 10 mM Tris-Cl pH 7.5.
3. Each tube was vortexed for several seconds to ensure thorough mixing.
4. The tubes were centrifuged briefly to collect any droplets from the lid (max 2 seconds).
5. The spectrophotometer was switched on and then the computer switched on.
6. The "remote operation" option was selected on the spectrophotometer.
7. The VISION software was opened on the computer and the following parameters were set for measurement:
   - Wavelengths selected at 260, 280
   - Absorbance selected

- Normal selected
- Time selected as 3 seconds
- Lamp selected as 324 nm (Blue+ 110-325 nm)
- Width selected as 2 nm
- Number of cycles selected as 3
- Time selected as Auto
- Number of samples was selected for the number of samples to be processed and the sample names inputted.
- The "one cuvette" option was selected

8. The cuvettes were cleaned thoroughly with RNaseZAP and then with RNase free water prior to use. Any residual fluid was removed carefully by pipetting.
9. 80 µl of 10 mM Tris-Cl pH 7.5 was added to each of two cuvettes and the cuvettes placed in the slot at the back right (black area to front) and the other in sample space 1 with the frosted end to the front.
10. With the lid of the spectrophotometer closed, the icon with the arrows pointing down was selected to blank the spectrophotometer.
11. The cuvette from sample space 1 was then removed and the Tris Cl removed by pipetting, ensuring that there was not residual fluid in the cuvette.
12. 80 µl of the first sample to be analysed was then pipetted into the cuvette and the cuvette placed back into slot 1.
13. With the lid of the spectrophotometer closed, the run icon (man running) was selected and the absorbances measured.

14. The sample was then removed from the cuvette by pipetting, the cuvette rinsed with 80 µl 10 mM Tris-Cl pH 7.5 and the Tris Cl removed by pipetting to ensure no residual fluid was left in the cuvette.
15. 80 µl of the next sample was then pipetted into the cuvette and the procedure repeated for this and subsequent samples.
16. At the end of each batch the data was saved.

As described above, the spectrophotometer was set to measure each sample in triplicate. If the values obtained were the same or similar within 0.002 then the results were considered valid. On occasion there was wider variation between the three results for each sample and these samples would be re-run, taking extra care to ensure adequate mixing prior to measurement.

A_{260}:A_{280} ratios were calculated using the A_{260} and A_{280} values obtained from the spectrophotometer. Quantity of RNA in each sample was calculated using the following equation:

A_{260} x dilution factor of sample x 44 = amount in ng/µl.

The dilution factor for each sample was 40 (2 µl sample in 78 µl Tris). 44 was the multiplier used as all of the samples had been diluted with 10 mM Tris Cl with pH 7.5. The total amount of RNA obtained in each case was calculated by multiplying the result of the above equation by the total amount of sample eluted. For the PAXgene™ samples this was 75 µl and for the TRIzol® LS or QIAamp® samples

this was 95 µl. This allowed for a residual 5 µl of dead space during elution using the spin columns (reference: personal communication with representative of Qiagen).

Spectrophotometry Using The ThermoSpectronic NanoDrop™1000 Spectrophotometer

1. The software was started up and the nucleic acid module selected and the RNA-40 settings selected.
2. Before each use, the sample pedestal of the spectrophotometer was wiped using a soft laboratory wipe.
3. The pedestal was then cleaned by pipetting a 2 µl aliquot of RNase-free water onto the pedestal, closing and opening the pedestal again and wiping it dry with a soft laboratory wipe.
4. To initialise the spectrophotometer, a further aliquot of water was loaded onto the pedestal, following the instructions on screen and pressing "OK" when instructed to do so. The pedestal was again wiped using a laboratory wipe.
5. Before making any sample measurements, a blank measurement would be made: 1 µl of RNase-free water was pipetted onto the pedestal, the pedestal closed and "make a blank measurement" selected on screen.
6. When the blank measurement was complete, both pedestals were wiped using a laboratory wipe.
7. A test measurement was made using water to check that the result was a spectrum with a relatively flat baseline and the pedestals wiped afterwards.

8. For each sample measurement, a 1 µl aliquot of sample was pipetted onto the pedestal, the pedestal arm closed and "measure" was selected on screen. The pedestal was observed carefully to ensure that a complete column of fluid was formed during the measurement: on the occasions where this did not occur, the measurement was repeated using a slightly bigger aliquot of sample.
9. For each measurement the following were recorded:
    - Absorbance at 260 nm
    - Absorbance at 280 nm
    - 260/280 ratio
    - Amount of RNA (ng/µl)
10. Following each sample the pedestal was wiped clean with a clean laboratory wipe, cleaned with an aliquot of water on the pedestal and then wiped dry with a laboratory wipe. The next sample would then be applied and the process repeated.

A260:A280 Ratio Target Values

The range of A260:A280 ratio that was considered desirable and indicative of good quality RNA was 1.8-2.2 (for samples processed in Tris Cl).

Qualitative Assessment of Total RNA Using Agilent 2100 Bioanalyzer (RNA 6000 Nano Assay)

RNA quality was assessed using both subjective assessment of Bioanalyzer traces and using RNA integrity number values that were generated by the Bioanalyzer software. The method of using the Bioanalyzer is detailed below. Notes on Assessment of the Bioanalyzer traces and on RIN values follow the method.

Before preparing the RNA 6000 Nanochip for analysis:

1. The chip priming station was set up, checking that the syringe was in place (new syringe used for each reagent kit), the base plate was in position C and the lever of the clip in the top position.
2. The bioanalyzer was set up and the Agilent 2100 Bioanalyzer software started.
3. The Bioanalyzer electrodes were decontaminated by filling the wells of an electrode cleaner with 350 µl RNaseZAP, placing the cleaner in the bioanalyzer and closing the lid for 1 minute. The electrode cleaner was then removed and another electrode cleaner filled with 350 µl of RNase-free water and then placed in the Bioanalyzer and the lid closed for 10 seconds. The electrode cleaner was then removed and the bioanalyzer lid left open for 10 seconds to allow the water on the electrodes to evaporate and then closed.
4. All of the reagents for preparation of the chips were allowed to equilibrate to room temperature for 30 minutes before use.

5. All of the RNA samples to be run on the chip and the RNA ladder were denatured for 2 minutes at 70 °C in an air incubator before use.

Chip preparation:

1. The gel was prepared by placing 550 µl of RNA 6000 Nano gel matrix (marked in the kit with a red cap) into a spin filter column and centrifuged at 4,000 rpm in the microfuge for 10 minutes. Aliquots of 65 µl of filtered gel were pipetted into 0.5 ml RNase-free microcentrifuge tubes. Any aliquots not used on that day were stored at 4 °C and used within one month.
2. The RNA 6000 Nano dye concentrate was vortexed for 10 seconds and then spun down in a microcentrifuge for around 2 seconds. The dye concentrate was protected from light at all times.
3. 1 µl of dye concentrate was pipetted into a 65 µl aliquot of the filtered gel from step 1. above and the tube capped and vortexed thoroughly, ensuring proper mixing of the gel and dye.
4. The dye-gel mix was centrifuged at room temperature for 10 minutes at maximum speed in a microcentrifuge.
5. A new RNA Nano chip was removed from its sealed bag and placed on the Chip Priming Station.
6. 9 µl of the gel-dye mix were pipetted into the bottom of the well marked with a white G in a black circle. Care was taken to avoid any particles from the bottom of the gel-dye tube and the pipette tip was inserted to the very bottom of the chip well.

7. The timer was set to 30 seconds, the plunger set to 1 ml and then the Chip Priming Station was closed ensuring that the latch lock clicked shut.
8. The plunger was then depressed until held by the syringe clip and held there for 30 seconds exactly.
9. When the 30 seconds was up, the plunger was released and after 5 – 10 seconds the plunger was slowly drawn back to the 1ml position.
10. The chip priming station was then opened.
11. 9 µl of gel-dye mix was pipetted into each of the chip wells marked with a black G on a grey background (2 wells).
12. 5 µl of RNA 6000 Nano Marker (green cap) was pipetted into the well marked with the ladder symbol and each of the 12 sample wells.
13. 1 µl of RNA 6000 Ladder was pipetted into the well marked with the ladder symbol.
14. 1 µl of each sample was pipetted into a sample well.
15. Any sample wells that did not have a sample added had an extra 1 µl of RNA 6000 Nano Marker added to ensure that the chip would run properly.
16. The chip was then placed in the chip adapter for the vortex mixer and vortexed for 1 minute at 2,400 rpm. The chip run in the Bioanalyzer was commenced within 5 minutes of this vortex step.

Running the chip in the Agilent 2100 Bioanalyzer:

1. The bioanalyzer lid was opened, the chip gently inserted and the lid closed slowly.
2. From the Assay Menu on the software screen, Eukaryote Total RNA Nano was selected.
3. The start button on the software screen was then clicked.
4. The sample information of the chip and each individual well was then entered.
5. After the assay was complete, the chip was removed from the bioanalyzer and disposed of.
6. To clean up, the Bioanalyzer electrodes were decontaminated by filling the wells of an electrode cleaner with 350 µl RNaseZAP, placing the cleaner in the bioanalyzer and closing the lid for 1 minute. The electrode cleaner was then removed and another electrode cleaner filled with 350 µl of RNase-free water and then placed in the Bioanalyzer and the lid closed for 10 seconds. The electrode cleaner was then removed and the bioanalyzer lid left open for 10 seconds to allow the water on the electrodes to evaporate and then closed.

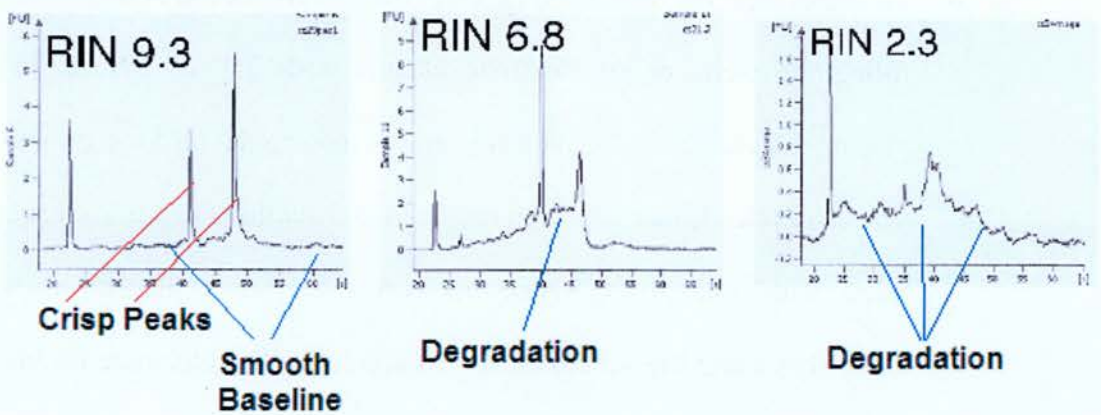
Notes on the RNA 6000 Nano Assay:

- *The above method was taken from the Agilent Technologies Reagent Kit Guide **RNA 6000 Nano Assay** Edition April 2003.*

Notes on Bioanalyzer Trace Assessment

A bioanalyzer trace is a graph of fluorescence against time. A good quality trace from intact RNA should have a marker peak and two crisp peaks corresponding to 18S and 28S fragments. There should also be very little "noise" along the baseline. The more "noise" there is the more degraded the RNA sample is. As examples of this, the graph on the left in figure 2.1 is from good quality intact RNA, the graph in the middle is of RNA showing some degradation and that on the right is from RNA showing a lot of degradation.

Figure 2.1: ***Examples of Bioanalyzer Traces for Intact, Partially Degraded and Very Degraded RNA samples.***



Notes on RNA Integrity Number (RIN)

RIN is a software generated number that gives an indication of RNA quality. The scale for RIN is 1 to 10. 1 would be poor quality, degraded RNA and 10 would be intact, high quality RNA. The higher the RIN the better. As an illustration of this, in figure 2.1 the sample on the left has RIN of 9.3 (good quality), the middle sample of 6.8 (intermediate quality) and the sample on the right of 2.3 (poor quality).

2.4 Methods Used For Labelling and Hybridisation of Microarrays.

Choice of Microarray Platform

At the time of study planning three microarray platforms were considered. These were microarrays produced by GE (CodeLink™ Whole Human Genome arrays), Affymetrix (GeneChip® Human Genome array) and Agilent. The Illumina® arrays used later in this thesis were not yet available.

The initial microarray work carried out in this thesis used Codelink™ Whole Human Genome Bioarrays. There were several reasons for this. Firstly, the platform was attractive because the amount of starting RNA was smaller than competing platforms at the outset of the microarray work in this study. Initial manufacturer's data recommended use of 0.2-2 μg of RNA, subsequently this recommended range was adjusted downwards to 0.1-1 μg. The small amount of starting material was a major factor because it meant that we could use some of our smaller neonatal samples on array.

The other main factors that influenced the choice of Microarrays used were:

- Timing. This was a platform that the Divison of Pathway Medicine was interested in looking into further at the time that my study reached the relevant stage of work.
- Cost. Microarray-based projects are costly. At the time of considering running my initial neonatal samples on array, each microarray would have

cost in the region of around £ 1,000 if Affymetrix arrays were used. This clearly would have limited the number of arrays run and the validity of any data obtained. At the time, CodeLink™ arrays were competitively priced compared to rival companies and would have enabled us to maximise the budget we had available at the time. In practice, the cost of microarrays has come down considerably during the course of the study.

When we encountered problems with the CodeLink™ system in the latter stages of the study, we then switched to Illumina® microarrays. The decision to use the Illumina® arrays was based on the lower amount of starting material required and cost. The Illumina® platform has been shown to provide good results from blood RNA extracted using the PAXgene™ system (331).

An overview of the features of these two microarray platforms is outlined below.

CodeLink™ Human Whole Genome Bioarray (as described on the manufacturer's website: www.appliedmicroarrays.com):

"CodeLink™ Human Whole Genome Bioarray is comprised of approximately 55,000 30-mer probes designed to conserved exons across the transcripts of targeted genes. These 55,000 probes represent well-annotated, full length and partial human gene sequences from major public databases. CodeLink™ Human Whole Genome Bioarrays target most of the known and predictive genes of the human genome as it is described today in the public domain. "

Illumina® Human Whole Genome Expression BeadChip (as described on the manufacturer's website: www.illumina.com):

"The HumanWG-6 v3.0 Expression BeadChip contains six arrays on a single BeadChip, each with >48,000 probes derived from human genes in the NCBI RefSeq and UniGene databases. Each array on the BeadChip provides genome-wide transcriptional coverage of well-characterized genes, gene candidates, and splice variants. The HumanWG-6 Expression BeadChips are arranged in an Array of Arrays format, providing multiple arrays on each slide. The arrays are separated from one another by a seal so that each array can be hybridized to a different sample; six samples can than be interrogated simultaneously. A minimal amount of total RNA (50-100 ng) is required for the single-round in-vitro transcription reaction."

The following section details the methods used for labelling, hybridisation and scanning of the CodeLink™ arrays. I have not included methods for the Ilumina® microarrays as this work was contracted out to an external laboratory.

Methods Used for CodeLink™ Microarrays

Described below are the CodeLink™ Microarray methods used for manual target preparation and bioarray hybridisation and detection. During the duration of the work on this project, the method for Manual Labelled cRNA Target Preparation was amended slightly by the manufacturers – latterly this was carried out using the CodeLink™ iExpress Assay Reagent Kit. The method for Bioarray Hybridisation and Detection remained the same. In practice, only a handful of arrays were run using the initial method, i.e. most of the work was carried out using the second method.

CodeLink™ Microarray Method (1)

(From Amersham Biosciences documents: 63005459 {080074-00/Rev. AA/2004-04} CodeLink Gene Expression System: Manual Labelled cRNA Target Preparation 63005460 {080075-00/Rev. AA/2004-04) CodeLink Gene Expression System: Single-Assay Bioarray Hybridization and Detection).

Day 1: Manual Target Preparation

1. Quality and quantity of the RNA was measured as described in section A.

2.1 A working solution of bacterial control mRNA was prepared as follows:

a) 5 µl of each of the 0.1 µg/µl bacterial control mRNAs (araB, entF, fixB, gnd, hisB and leuB) were mixed in a 2 ml tube by tapping and then centrifuged for 5 seconds at more than 10,000 x g in order to obtain a 16.7 ng/µl combined stock solution. This was divided into 3 µl aliquots and stored at -70 °C.

b) 3 µl of combined stock solution from a) above was mixed with 997 µl of nuclease-free water by pipetting. This achieved a 50.2 pg/µl combined spike dilution which was then centrifuged for 5 seconds at more than 10,000 x g.

c) The working solution of bacterial control mRNA was achieved by mixing 2 µl of the 50.2 pg/µl combined spike dilution from b) above with 998 µl of nuclease-free water.

2.2 Each RNA sample was prepared for manual target preparation by using the maximum quantity of mRNA that was possible to achieve the final volume of 12 µl in the following:

X µg RNA + X µl bacterial control solution + 1 µl T7 oligo dT primer + Y µl nuclease free water = 12

2.3 This mix was then incubated at 70 °C in a thermal cycler for 10 minutes followed immediately by 3 minutes on ice.

2.4 The tubes were centrifuged for 5 seconds to collect the samples at the base of the tubes and then returned to ice.

2.5 While keeping the tubes containing the 12 µl on ice, the following was added to each tube: 2 µl 10x first strand buffer, 4 µl 5 mM dNTP mix, 1 µl RNase inhibitor and 1 µl reverse transcriptase to give a final volume of 20 µl.

2.6 The tubes were incubated at 42 °C for 2 hours in an air incubator.

2.7 The tubes were centrifuged for 5 seconds to collect the samples at the bottom of the tubes.

3.1 To each 20 µl first strand cDNA synthesis reaction mix from step 2.7 the following was added: 63 µl nuclease-free water, 10 µl 10x second-strand buffer, 4 µl 5 mM dNTP mix, 2 µl DNA polymerase mix and 1 µl RNase H to give a final volume of 100 µl.

3.2 Mixing was performed by gently tapping the side of each tube and then each tube was centrifuged for 5 seconds at over 10,000 x g. The tubes were then incubated at 16 °C for 2 hours.

3.3 The tubes were centrifuged at maximum speed for 5 seconds to collect the samples at the bottom of each tube. The samples were mixed gently by

pipetting and then put on ice. On each occasion, step 4 proceeded directly after step 3.3.

4.1 500 µl of Buffer PB were added to each of the the tubes from step 3.3 and mixed gently by pipetting.

4.2 A QIAquick spin column for each sample was placed in a 2 ml collecting tube.

4.3 The cDNA Buffer PB mix was pipetted into the QIAquick spin column.

4.4 The spin column was centrifuged at more than 10,000 x g for 30-60 seconds.

4.5 The flow through was discarded. 700 µl of Buffer RPE was added to the columns and centrifuged at more than 10,000 x g for 30 seconds.

4.6 The flow though was discarded. The QIAquick spin column was placed in a new 2 ml collecting tube and centrifuged at more than 10,000 x g for 1 minute.

4.7 The QIAquick column was placed in a clean 1.5 ml microcentrifuge tube.

4.8 30 µl of Buffer EB were pipetted onto the centre of the column membrane, the column was left to stand at room temperature for 1 minute and then centrifuged at more than 10,000 x g for 1 minute. Using the same elution tube this step was repeated once to obtain 60 µl of eluate.

4.9 The cDNA was concentrated in a speedvac concentrator (medium heat) until a final volume of less than or equal to 9.5 µl was achieved for each tube.

4.10 For each sample, the volume was checked and the volume of each made up to 9.5 µl if necessary with nuclease-free water.

5.1 In a separate tube, the following were added, in the order listed to make up the IVT mixture: 4 µl of 10x T7 reaction buffer, 4 µl T7 ATP solution, 4µl T7

GTP solution, 4 µl T7 CTP solution, 3 µl T7 UTP solution, 7.5 µl 10 mM biotin-11-UTP and 4 µl 10x T7 enzyme mix to give a final volume of 30.5 µl.

5.2 The components of the IVT mixture were mixed by vortexing and then centrifuged at more than 10,000 x g for 5 seconds. The IVT reaction mixture (30.5 µl) was then added to each tube from step 4.10 and mixed by gentle pipetting.

5.3 The reaction was incubated in an air incubator for 14 hours at 37 °C.

Day 2: Manual Target Preparation continued.

6.1 Working solutions of Buffers RLT and RPE were prepared.

6.2 The IVT reaction tubes were centrifuged briefly to collect the sample at the bottom of the tube. The sample volume was then adjusted to 100 µl by adding 60 µl of nuclease-free water.

6.3 350 µl of Buffer RLT was added to the sample and mixed by pipetting.

6.4 250 µl of 100 % Ethanol was added to the sample and mixed by pipetting.

6.5 The sample (700 µl) was added to an RNeasy spin column in a 2 ml collection tube. The column was centrifuged for 15 seconds at more than 8,000 x g.

6.6 The RNeasy column was transferred to a new 2 ml collection tube. 500 µl of Buffer RPE were added to the column and the column centrifuged at more than 8,000 x g for 15 seconds. The flow through was discarded and this wash step repeated once.

6.7 The RNeasy column was placed in a new 2 ml collection tube and centrifuged at more than 8,000 x g for 2 minutes to dry the membrane.

6.8 The RNeasy column was transferred into a new 1.5 ml microcentrifuge tube. 50 µl of nuclease-free water was pipetted directly onto the column membrane.

6.9 The column was left at room temperature for 10 minutes and centrifuged at more than 8,000 x g for 1 minute.

6.10 Using the same elution tube, a further 50 µl of nuclease-free water were pipetted onto the column membrane.

6.11 The column was left at room temperature for 10 minutes and then centrifuged at more than 8,000 x g for 2 minutes.

6.12 The column was removed and the eluate mixed by tapping the side of the tube.

6.13 In each case the tube was moved directly into step 7.

7.1 cRNA quality and A_{260}:A_{280} ratio were measured **without** dilution using 1.5 µl of cRNA from step 7 in a nanodrop spectrophotometer.

7.2 Quantity and quality of cRNA were obtained using the nanodrop data and samples of sufficient quantity with A_{260}:A_{280} ratio of 1.8-2.1 were taken forward into the fragmentation and hybridisation stages.

Day 2 continued: Bioarray Hybridisation and Detection

1.1 For each bioarray to be loaded, 10 µg of cRNA from step 6.13 above was made up to a volume of 20 µl with nuclease-free water.

1.2 5 µl of 5x fragmentation buffer was added. Each tube was placed in a thermal cycler set to 94 °C and heated for 20 minutes using the heated lid feature.

1.3 The tubes were then cooled towards 0 °C in the thermal cycler for at least 5 minutes and as they did not reach 0 °C were placed on ice for 5 minutes.

2.1 The shaker incubator temperature was set to 37 °C and prepared for taking the 12 slide shaker-tray.

2.2 Each cRNA sample was transferred into a 1.5 ml microcentrifuge tube. For each bioarray to be processed a 260 µl hybridisation solution was prepared as follows: 78 µl of hybridisation buffer component A, 130 µl of hybridisation buffer component B, 27 µl of nuclease-free water were added to the 25 µl of fragmented cRNA (10 µg) from step 1.

2.3 The solution was vortexed for 5 seconds at maximum speed. The hybridisation mixture was incubated at 90 °C for 5 minutes to denature the cRNA.

2.4 The tubes were cooled on ice for at least 5 and a maximum of 30 minutes (Bioarrays were loaded within 30 minutes).

3.1 The 12 slide shaker-tray was placed on a level surface. The bioarrays were placed in the shaker-tray with the input-output ports face up.

3.2 The hybridisation reaction mixture was vortexed for 5 seconds at maximum speed. The tubes were centrifuged briefly to gather the liquid at the bottom of the tube and placed back on ice.

3.3 For each Bioarray chamber, 250 µl was drawn into a 1 ml wide-bore pipette tip. The pipette tip containing the hybridization solution was placed over an array input port and pressed in until the tip formed a seal with the chamber.

3.4 The entire sample was injected slowly into the Flexchamber without using the blowout feature of the pipettor. When the Flexchamber was full, pressure was maintained on the pipettor using the right hand, the left hand held down the bioarray (over the label) and the pipette was removed from the import port. Any

excess fluid around the port was aspirated with the pipette and any residual fluid blotted up with a lint-free wipe (without touching the port).

3.5 After all the Bioarrays were loaded, the ports were sealed using sealing strips and the port-sealing tool.

4.1 The 12 slide shaker-tray was fixed securely into the shaker incubator, ports facing up.

4.2 The shaker speed was set at 300 rpm and the slides were incubated for 21 hours at 37 °C.

4.3 In preparation for the next step, a large reagent reservoir was filled with 240 ml of filtered 0.75 x TNT Buffer. The reservoir was covered and incubated overnight at 46 °C.

Day 3: Bioarray Hybridisation and Detection Continued.

5.1 Each slot of the medium reagent reservoir was filled with 13 ml of filtered 0.75 x TNT Buffer. The Bioarray rack was placed in the reservoir and left at room temperature.

5.2 The 12 slide shaker-tray was removed from the shaker incubator and placed on a level surface.

5.3 The first Bioarray to be processed was placed into the Flex Chamber removal tool.

5.4 The Flex Chamber was removed by lifting the tab and slowly pulling it back at a 60 ° angle.

5.5 The Bioarray was place into a slot of the Bioarray rack in the medium reservoir containing the 0.75 X TNT Buffer. The Bioarray position tool (tooth side down) was used to ensure Bioarray slide was correctly positioned.

5.6 The Flex Chamber removal tool was rinsed with 0.75 x TNT Buffer after each Bioarray. The medium reagent reservoir was kept at room temperature until all of the Bioarrays had been processed.

5.7 Steps 5.3 – 5.6 were repeated for each Bioarray to be processed.

5.8 The Bioarray rack with the Bioarrays inside was transferred into the medium reagent reservoir containing the 0.75 x TNT Buffer that had been incubated at 46 °C overnight. The lid was replaced on the reservoir and the reservoir incubated at 46 °C for exactly 1 hour.

6.1 Each slot in the small reagent reservoir was filled with 3.5 ml of Cy5-Streptavidin working solution. The reservoir was covered with a black lid and left at room temperature.

6.2 The Bioarray rack was removed from the medium reservoir containing the 0.75 x TNT Buffer at 46 °C and placed in the small reagent reservoir containing the Cy5-Streptavidin working solution. The reservoir was covered with the black lid and incubated at room temperature for 30 minutes, covered from light with additional foil.

6.3 During the incubation period for step 6.2 I filled the large reagent reservoirs with 240 ml of 1 x TNT Buffer at room temperature.

6.4 After the 30 minute incubation period the Bioarray rack was removed from the reservoir containing the Cy5-streptavidin working solution and transferred to one of the large reservoirs containing 1 x TNT buffer. Following a gentle up-

and-down agitation the arrays were incubated in the reservoir for 5 minutes at room temperature covered from light.

6.5 The Bioarray rack was removed from the first large reservoir and place in the second reservoir containing 1 x TNT buffer and gently agitated by moving up and down. The Bioarrays were incubated at ambient temperature for 5 minutes covered from light. This step was repeated with two additional large reservoirs containing 1 x TNT buffer to give 3rd and 4th washes.

6.6 During the third wash, a large reagent reservoir was thoroughly rinsed with distilled water and dried. It was then filled until completely full with 0.1 x SCC/0.05 % Tween 20 solution.

6.7 After the 4th wash step, the Bioarray rack was transferred into the large reagent reservoir containing the 0.1 x SCC/0.05 % Tween 20 solution. The slides were incubated at room temperature for 30 seconds while continually gently agitating up and down.

6.8 The Bioarray Rack was removed from the large reagent reservoir and the bottom edges of the slides were blotted on absorbent paper towel. The slides were removed from the Bioarray rack using the bioarray removal tool. The Bioarrays were placed in metal slide racks for the centrifuge and centrifuged at 2000 rpm with acceleration of 9, deceleration of 9 for 3 minutes. The centrifuge used was an Eppendorf 5810R.

6.9 The dry bioarrays were protected from light until scanned (scanned straight away).

6.10 The reservoirs were repeatedly rinsed with deionised water and inverted to dry.

6.11 The rack was washed thoroughly, rinsed with deionised water and air dried.

7.0 Scanning the CodeLink™ Bioarrays on an Agilent G2565 Microarray Scanner:

7.1 While wearing gloves, the cover of the Agilent slide holder was opened and the bioarray placed into the holder with the active bioarray surface facing up and towards the cover. The cover was locked by gently pressing it down with one thumb.

7.2 The slide holder was inserted into the carousel slot with the tapered end of the holder facing towards the centre of the carousel. This was repeated for each slide and the carousel cover was then closed.

7.3 The carousel was placed in the scanner.

7.4 The scanner was switched on and Agilent scan control was clicked.

7.5 The start and end slots were entered and edited for each slide.

7.6 The scan setting selected was a custom setting labelled Codelink with scan area 67.2 x 21.6 mm, dye channel red, red PMT 100 % and scan resolution of 5 μm. (The only exceptions to this were where the original few CodeLink slides were used as there was an issue with persistence of glue around the perimeter of the array. For some of these slides the scan area was reduced to 62.1 x 18.5 mm in order to allow the scanner to read the array. This included the entirety of the array and cut out only a margin of glass slide).

7.7 The output path was set to a designated folder in drive D.

7.8 The above values were checked and set.

7.9 Scanning was commenced by hitting "scan slots".

Codelink™ Microarray Method (2)

(From Amersham CodeLink™ iExpress Assay Reagent Kit. Product booklet: codes 67601000 AND).

1. The quality and quantity of total RNA starting material were assessed as described in section A.

2.1 Bacterial controls were prepared using the "alternative" bacterial mRNA controls dilution method for use in target synthesis: single concentration (found in appendix 10.2 of the iExpress Assay kit) as follows:

a) 5 µl of each of the 0.1 µg/µl bacterial control mRNAs (araB, entF, fixB, gnd, hisB and leuB) were mixed in a 2 ml tube by tapping and then centrifuged for 5 seconds at more than 10,000 x g in order to obtain a 16.7 ng/µl combined stock solution. This was divided into 3 µl aliquots and stored at -70 °C.

b) 3 µl of combined stock solution from a) above was mixed with 997 µl of nuclease-free water by pipetting. This achieved a 50.2 pg/µl combined spike dilution which was then centrifuged for 5 seconds at more than 10,000 x g.

c) The working solution of bacterial control mRNA was achieved by mixing 2 µl of the 50.2 pg/µl combined spike dilution from b) above with 998 µl of nuclease-free water.

Please note that this method of bacterial control preparation is identical to that in the original method so as to maintain consistency, i.e. the dynamic method described in the updated protocol was not used.

2.2 Each total RNA sample was prepared for target preparation as follows: by using the maximum quantity of mRNA that was possible to achieve the final volume of 12 µl in the following:

X µg RNA + X µl bacterial control solution + 1 µl T7 oligo dT primer + Y µl nuclease free water = 12

2.3 This mix was then incubated at 70 °C in a thermal cycler for 10 minutes followed immediately by 3 minutes on ice.

2.4 The tubes were centrifuged for 5 seconds to collect the samples at the base of the tubes and then returned to ice.

2.5 While keeping the tubes containing the 12 µl on ice, the following was added to each tube: 2 µl 10x first strand buffer, 4 µl 5 mM dNTP mix, 1 µl RNase inhibitor and 1 µl ArrayScript to give a final volume of 20 µl.

2.6 The tubes were centrifuged for 5 seconds at full speed to collect the sample in the bottom of the tube and then incubated at 42 °C for 2 hours in an air incubator.

2.7 The tubes were centrifuged for 5 seconds to collect the samples at the bottom of the tubes.

2.8 The tubes were placed on ice and second-strand cDNA synthesis was commenced without delay.

3.1 To each 20 µl first strand cDNA synthesis reaction mix from step 2.7 the following was added: 63 µl nuclease-free water, 10 µl 10x second-strand buffer, 4 µl 5 mM dNTP mix, 2 µl DNA polymerase mix and 1 µl RNase H to give a final volume of 100 µl.

3.2 Mixing was performed by gently tapping the side of each tube and then each tube was centrifuged for 5 seconds at over 10,000 x g. The tubes were then incubated at 16 °C for 2 hours.

3.3 The tubes were transferred to ice and taken directly forward to the purification of double-stranded cDNA steps.

4.0 The nuclease-free water was pre-heated to 50-55 °C and checks to ensure that the Wash Buffer had already been prepared were made.

4.1 250 µl of cDNA Binding Buffer was added to each tube of cDNA from step 3.3 above and mixed by gentle pipetting.

4.2 When each cDNA filter cartridge was firmly placed in a wash tube, the cDNA sample/cDNA Binding Buffer mixture from section 4.1 was pipetted into the centre of each cDNA filter cartridge. The cartridges were then sealed with the tube caps.

4.3 The spin columns were then centrifuged at 10,000 x g for 1 minute.

4.4 The flow-through was discarded and the cDNA filter cartridges replaced in wash tubes.

4.5 In order to wash, 500 µl of Wash Buffer was added to each cartridge, each cartridge sealed and then centrifuged at 10,000 x g for 1 minute.

4.6 The flow-through was discarded. The cDNA filter cartridges were placed in wash tubes, sealed and centrifuged at 10,000 x g for 1 minute to dry the column.

4.7 The cDNA filter cartridges were transferred to cDNA elution tubes.

4.8 To elute the cDNA, 12 µl of pre-heated nuclease-free water was pipetted onto the centre of each cDNA filter cartridge filter. The cartridges were sealed

and left at room temperature for 2 minutes before being centrifuged at 10,000 x g for 1.5 minutes.

4.9 The elution step in 4.8 was repeated with a further 12 µl of pre-heated nuclease-free water added to each cartridge filter. This was to generate 20 µl of eluate from each cartridge.

4.10 The eluate was carried directly to the IVT section.

5.1 The IVT mixture for each sample was made by adding the following components in the order listed: 12 µl biotin-NTP mix, 4 µl 10x T7 reaction buffer and 4 µl 10x T7 enzyme mix to give a final volume of 20 µl per reaction. (Where a master-mix was used for multiple reactions: the volume of each reagent was multiplied by the number of reactions plus an extra 5 %).

5.2 The components of the IVT mixture were mixed by gentle vortexing and the tube was centrifuged to collect the reagents at the bottom of the tube. The tube was then placed on ice.

5.3 Each cDNA sample from 4.10 was measured using a pipette to ensure that it was exactly 20 µl and, if not, was made up to 20 µl with nuclease-free water. 20 µl of IVT reaction mixture from step 5.2 was then pipetted into each of the tubes containing cDNA. Mixing was ensured by gentle pipetting followed by gently flicking the tube several times. The tubes were then centrifuged briefly at full speed to collect the reagents at the bottom of the tube.

5.4 The reaction was then incubated in a 37 °C air incubator for 14 hours.

6.0 The nuclease-free water was preheated to 50-60 °C for at least 10 minutes before RNA purification.

6.1 cRNA filter cartridges, one for each sample to be purified, were placed into cRNA collection tubes.

6.2 The IVT reaction volume from step 5.4 was then adjusted to 100 µl by adding 60 µl of room temperature nuclease-free water. This was mixed thoroughly by gentle vortexing. Each tube was then centrifuged briefly to collect the contents at the bottom of the tube.

6.3 350 µl of cRNA binding buffer was added to each sample.

6.4 Without delay, 250 µl of 100 % ethanol was added to each reaction sample and mixed well by pipetting.

6.5 Without delay, each sample (700 µl) was applied to a cRNA filter cartridge in a collection tube. The tubes were sealed and centrifuged for 1 minute at 10,000 x g. It was ensured that all mixture had passed through the filter.

6.6 The flow-through was discarded and the cRNA filter cartridges placed in collection tubes. 650 µl of wash buffer was added to each cartridge, sealed and centrifuged for 1 minute at 10,000 x g, ensuring that all the wash buffer had passed through the filter.

6.7 The flow-through was discarded and the filter cartridges placed in collection tubes. Each cartridge was re-sealed and centrifuged once more at 10,000 x g for 1 minute to dry the cartridges.

6.8 Each cRNA filter cartridge was transferred to a new cRNA collection tube and 100 µl of pre-heated nuclease-free water (50-60 °C) was pipetted directly onto the centre of each cRNA filter cartridge membrane (without touching the membrane).

6.9 The filter cartridges were sealed with the lids and incubated at room temperature for 2 minutes. The cRNA was then eluted by centrifugation at 10,000 x g for 1.5 minutes.

6.10 The columns were then removed. The collection tubes, containing the labelled cRNA were tapped gently to mix.

6.11 cRNA quality, concentration and purity was then assessed.

7.1 Concentration was measured using the A260 value obtained using a nanodrop spectrophotometer (undiluted). Although generally not necessary, were any sample found to have cRNA concentration <0.5 μg/μl, it would be concentrated in a SpeedVac concentrator before fragmentation.

7.2 Purity was measured using the A_{260}:A_{280} ratio obtained using a nanodrop spectrophotometer. Samples with A_{260}:A_{280} ratio of 1.8-2.1 were taken forward for hybridisation.

7.3 For some samples, size distribution was examined by running the labelled cRNA on an Agilent bioanalyzer.

Bioarray Hybridisation and Detection was carried out as described in method 1 above. Microarray scanning was carried out as described as in method 1 above.

Short Notes on Globin Reduction

Globin genes have been described as interfering with the discovery of biomarkers in whole blood (332). Globin mRNA transcripts have been estimated to constitute a significant proportion of whole blood mRNA. It is thought that these transcripts decrease the sensitivity of detection of the other mRNAs of interest when using microarray technology (333, 334). Indeed, some literature suggests that there is reduction in percent present cells and increased variability when looking at microarray data from whole blood mRNA compared to that extracted directly from white blood cells or compared to protocols incorporating globin reduction (333-335). Furthermore, globin reduction techniques have been put forward as ways to reduce this variability and to increase the sensitivity of detecting less abundant mRNAs (333-335).

We did not use globin reduction methods in our microarray studies for the following reasons:

- The globin reduction methods on the market were all targeted towards reduction of adult globin mRNA whereas the haemoglobin in neonatal blood is predominantly fetal haemoglobin. The effectiveness of globin reduction methods on neonatal samples was therefore in doubt.
- I investigated the use of a globin reduction method on one of the neonatal samples, comparing microarrays of the sample with and without globin reduction. This experiment did not provide evidence that globin reduction would be useful in neonatal samples. Rather, there was the suggestion that

omitting the globin reduction step may actually give superior results. I have provided the details of this experiment in appendix 1 for reference.

- The additional time and cost involved in incorporating a globin reduction step was felt not to be worthwhile given the doubts over the benefits in neonatal blood.
- It was felt that a globin reduction step could act as an additional source of variation between samples.

Chapter 3

Investigation of Sample Handling and RNA Extraction Methods

Background

When planning gene expression profiling experiments it is very important to pay due attention to the methods being used. It is known that sample handling conditions can effect gene expression and therefore minimisation of variation and establishment of sample handling protocols are paramount (194, 323, 326, 327). The overall aim of this study was to use neonatal blood in gene expressing profiling work. At the time of initial study planning and commencement of sampling there was no data in the literature regarding the use of neonatal whole blood in this respect. It was therefore necessary, not only to optimise RNA sampling and extraction technique but also to show that it was possible to extract sufficient amounts of RNA of high enough quality from small samples of neonatal blood to use for microarray work.

In order to achieve robust, reliable, reproducible results from gene expression profiling experiments, it is important to minimise any sources of technical variability. There is evidence that differences in blood collection and preparation can cause ex-vivo induction of cytokine mRNA (336). For example, anticoagulants can cause ex-vivo changes in cytokine production (329) and adherence to plastic blood collection tubes may induce pro-inflammatory cytokines (329). There can be also be changes in gene expression depending on storage temperature and duration of storage of whole blood – rapid lysis of whole blood or immediate RNA isolation after sampling is an important consideration (337, 338). The method of RNA isolation chosen is a critical source of variability (339). Standardisation of methodology for a study is therefore vital.

In order to achieve robust, high quality microarray data, the input RNA should be of the highest quality possible. High quality RNA has minimal degradation and minimal contamination with DNA or chemicals used in the extraction process. The work presented in this chapter was carried out in order to find a suitable method of RNA extraction from neonatal whole blood, aiming for the highest possible quality of RNA for use in microarray work. I also present work on optimising sample storage conditions, again with the aims of achieving consistency of method and of having the best quality of RNA possible for microarray work. A_{260}:A_{280} ratios, Bioanalyzer graph tracings and RIN values are used as measures of quality. A brief explanation of these can be found on pages 93 and 98 of chapter 2.

Yield of RNA is also examined for the experiments described in this chapter. The aim was find a method that consistently yielded sufficient RNA to carry out microarray work using the CodeLink™ arrays. If the minimum amount of RNA required for input into the CodeLink™ microarray protocol is taken at 200 ng (in a maximum of 10.8 µl), then the minimum yield required in 75 µl of eluted RNA is 1.39 µg.

Aims

The overall aim of this section was to optimise methodology for sample handling and RNA extraction from neonatal blood samples in order to obtain high quality RNA for microarray work. I therefore set out to:

- Determine an optimal blood collection tube at the point of clinical sampling to maximise the quality and quantity of RNA obtained. (3.1)
- Determine an optimal method of RNA extraction in order to maximise quality and quantity of RNA obtained. (3.2, 3.3, 3.9)
- Examine the volume of neonatal blood required in order to consistently obtain RNA of high quality and sufficient quantity to use in microarray experiments. (3.4)
- Determine optimum range of incubation period for neonatal blood in PAXgene™ Blood RNA Tubes at room temperature to maximise RNA quality and yield. (3.5, 3.6)
- Examine the feasibility of storing neonatal blood in PAXgene™ Blood RNA tubes at various temperatures and for various durations of time and to determine the effects on quality and quantity of RNA obtained. (3.7, 3.8)
- Assess a filter-paper based technique as a potential way of collecting blood for RNA extraction. (3.9)

All work in this section was carried out in its entirety by myself, Claire L Smith.

Methods

The Use of Umbilical Cord Blood

For the work in chapter 3, umbilical cord blood was used as a surrogate for neonatal blood. It would neither have been practical nor ethical to obtain the volumes of blood required for comparison from individual neonates in these studies. Umbilical cord blood is fetal blood and would logically be identical to neonatal blood in the initial stages of postnatal life. The method of umbilical cord blood sampling used throughout this chapter is described in chapter 2 (page 61).

Consent

Informed maternal consent for umbilical cord blood sampling was taken for each sample. In the initial studies samples were taken following elective caesarean sections, i.e. the mothers were not in labour. In these cases the prospective mothers were consented prior to delivery. Later studies used samples from normal, assisted and emergency deliveries. In these circumstances the mother was approached prior to delivery, i.e. during labour and verbal assent in principle for sampling the cord blood was obtained. The samples were collected and prepared as normal. These mothers were approached again after they had recovered from delivery and informed written consent was obtained at that time to confirm participation in the study. In the event of the mother declining consent after samples had already been taken, any such samples would have been destroyed. In fact consent was granted by all of these mothers. In all cases the women were given both verbal and written information regarding the study.

RNA Extraction

In each of the sections that follow, the methods used are stated. The methods themselves are described in chapter 2. Extracted RNA was stored at -80 °C until further analysis could be carried out in batches.

Assessment of RNA Quality and Quantity

For all RNA extracted in this chapter, RNA quantity was assessed by spectrophotometry and quality by running on the Agilent Bioanalyzer and examining A_{260}:A_{280} ratios. A_{260}:A_{280} ratios were calculated using the A_{260} and A_{280} values obtained from the spectrophotometer. The methods for these are described on pages 89 to 98 (chapter 2).

Statistics

In each section, statistical significance of differences between means for A_{260}:A_{280} ratio, yield of RNA and RIN were calculated using paired t-testing within the statistical software package SPSS version 12.

Results

The results of each experiment are laid out in the nine sections that follow. I have included a short background note and listed the methods used in each section so that this information is in close proximity to the results presented. A discussion of all of the findings in context of the literature and overall conclusions for this chapter follow these sections.

3.1: Optimisation of Blood Collection Tube.

Background

Choice of blood collection tube was an important consideration. Essentially, the requirement was for a collection tube that enabled extraction of consistently high quality RNA of sufficient quantity that would enable microarray experiments to be run. In addition, any tube needed to be easy to use and safe for use in a clinical setting. I chose to examine three different neonatal blood bottles that were regularly used on the neonatal unit – lithium heparin, EDTA and sodium citrate. Another factor that needed to be considered was that any in vitro alteration of RNA expression profile needed to be minimised, for example it is known that anticoagulant can lead to changes in RNA expression (323). I therefore included samples placed directly into TRIzol® LS reagent and samples injected directly into PAXgene™ Blood RNA tubes. TRIzol® LS reagent is commonly used in nucleic acid extractions in laboratory practise. The PAXgene™ Blood RNA tubes contain a proprietary agent which is described as stabilising the RNA at the point of injection into the tube.

Method Overview

For each of nine umbilical cords sampled, aliquots were injected immediately into one each of neonatal blood tubes containing Lithium Heparin and Sodium Citrate and into two tubes containing EDTA. All tubes were filled until full – to the fill line or as otherwise described by the manufacturer. Blood was also injected directly into a PAXgene™ blood RNA tube – the intention was to inject 2.5 ml on each occasion.

A further two aliquots of 0.25 ml of blood were each diluted with 0.25 ml of RNase-free water and each water/blood aliquot added directly into a micro-centrifuge tube containing 1.5 ml of TRIzol® LS reagent. With the exception of the PAXgene™ tubes, 0.5 ml of blood from each of the blood collection tubes was processed immediately using the TRIzol® LS extraction method followed by the QIAamp® on-column clean up method as described on pages 67 and 69. The PAXgene™ tubes were processed using the PAXgene™ blood RNA extraction kits after at least 2 hours incubation at room temperature (page 76).

Results

The raw data for each sample can be found in table 3.1.1 along with some explanatory notes in appendix 2. Figure 3.1.1 displays graphically the yield, A_{260}:A_{280} ratio and RIN values for each blood collection tube. Total yield is displayed on the purple line graph for each collection tube with an additional blue line displaying yield per 0.5ml of blood for the PAXgene™ tube. The A_{260}:A_{280} ratio is displayed as a red line graph with the y-axis on the right. The pastel bars represent the RIN values for each collection tube.

Figure 3.1.1: ***Yield, A260:A280 and RIN for each blood collection tube.***

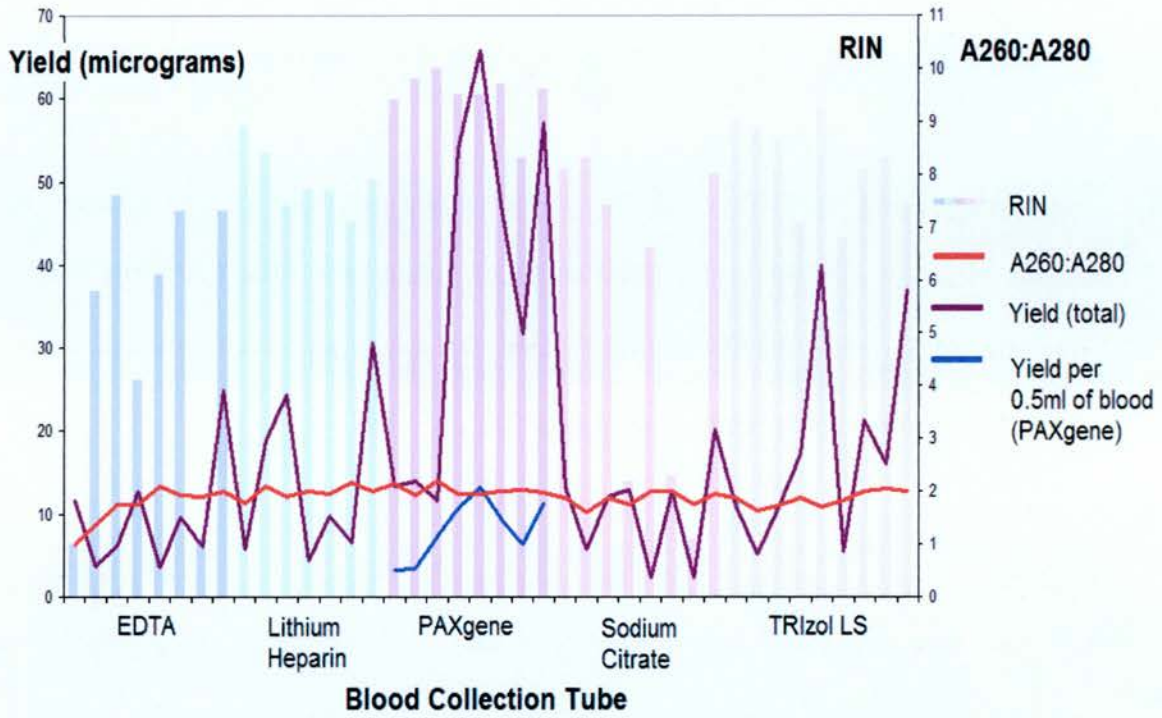


Table 3.1.2: ***Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each blood collection tube.***

Blood Tube	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA (μg) per 0.5 ml blood	Mean (range) RIN
EDTA	8	1.73 (1.01-2.10)	9.80 (3.51-24.91)	5.0 (1-7.6)
Li heparin	7	1.99 (1.79-2.17)	14.33 (4.35-30.60)	7.9 (7.1-8.9)
Na citrate	8	1.85 (1.62-2.0)	10.20 (2.34-20.23)	5.5 (1.1-8.3)
TRIzol®LS	9	1.86 (1.63-2.04)	18.21 (5.18-39.96)	8.2 (6.8-9.2)
PAXgene™	8	2.03 (1.93-2.2)	8.16 (3.33-13.17)	9.5 (8.3-10)

Table 3.1.2 shows the mean and range of values obtained with each collection tube for A_{260}:A_{280} ratio, yield of RNA obtained per 0.5 ml of blood and RIN values. From this table all of the mean A_{260}:A_{280} ratios are within the target range of 1.8-2.2 with PAXgene™ being the only method to have all ratios in the desired range. All

samples yielded sufficient RNA for microarray work with PAXgene™ having the lowest yield and TRIzol® LS the highest. PAXgene™ had the highest mean and best range of RIN values.

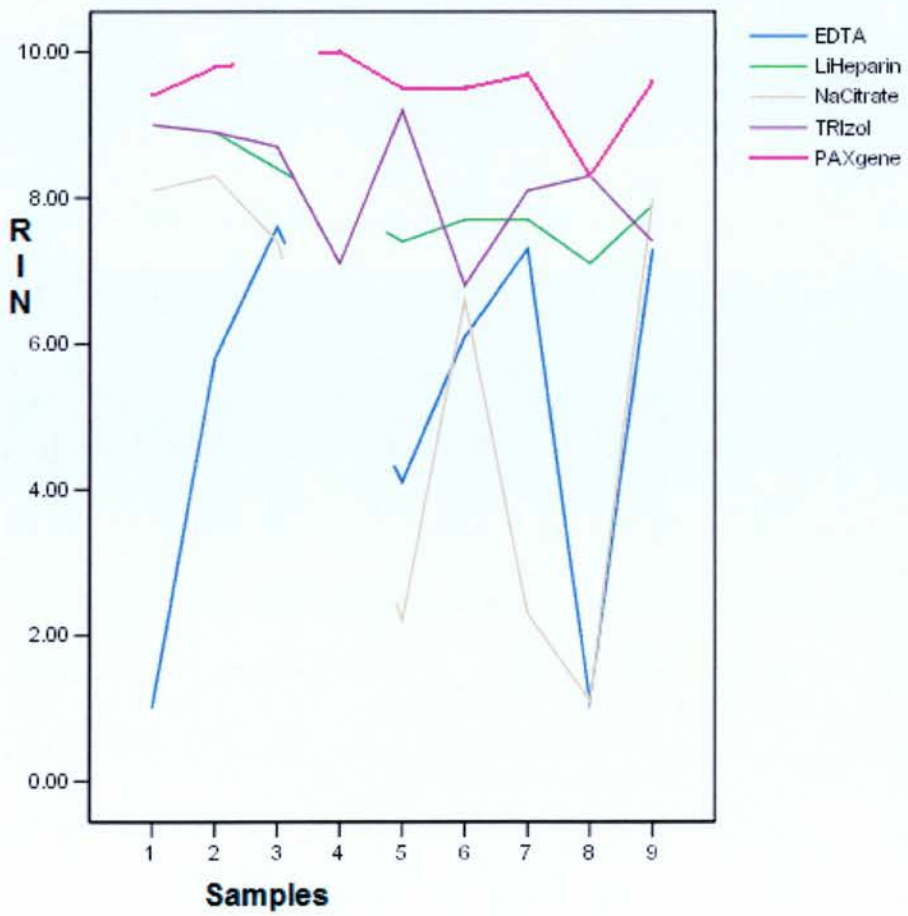
Table 3.1.3: ***Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of blood collection tubes (p values).***

	A260:A280 ratio	RNA per 0.5 ml (µg)	RIN
EDTA-Li Heparin	0.072	0.057	0.026
EDTA-Na Citrate	0.329	0.760	0.717
EDTA-TRIZol® LS	0.325	0.032	0.022
EDTA-PAXgene™	0.149	0.489	0.002
LiHep-Na Citrate	0.050	0.064	0.034
LiHep-TRIzol®LS	0.036	0.151	0.388
LiHep-PAXgene™	0.937	0.332	<0.001
Na Citrate-TRIzol®LS	0.903	0.063	0.046
Na Citrate-PAXgene™	0.032	0.571	0.010
TRIzol®LS-PAXgene™	0.025	0.035	0.011

Table 3.1.3 displays values of statistical significance when comparing means of each outcome studied for each pair of blood tubes. Taking p values of less than 0.05 as significant, the PAXgene™ tube gave statistically significantly higher A_{260}:A_{280} ratios than either TRIzol®LS or sodium citrate, and Lithium heparin tubes gave statistically higher ratios than TRIzol®LS. In terms of yield, TRIzol®LS gave the highest yield per 0.5 ml but this was only statistically significant when compared to EDTA or PAXgene™ tubes. When looking at the RIN values obtained, the PAXgene™ tubes gave statistically significantly higher RIN values than each of the other blood collection tubes examined. For the other collection tubes, the TRIzol®LS tube gave statistically higher RINs than EDTA or sodium citrate, and Lithium heparin tubes gave statistically significantly higher RINs than EDTA or sodium citrate. Figure 3.1.2. is a line sequence graph for the RIN values obtained

from each blood collection tube. The samples 1 to 9 are on the x-axis and the RIN values are on the y-axis. Each coloured line is a different blood collection tube. The PAXgene™ tube, for example, is represented by the yellow line and for this line the RIN can be seen to be between 8.3 and 10. The gap in the line is because there was no RIN value obtained for PAXgene for sample 3. Similar line sequence graphs are used in the other sections of this chapter. Bioanalyzer tracings of the RIN values can be found in table 3.1.4 (a and b) in appendix 2.

Figure 3.1.2: ***Graph of RIN values for each blood collection tube.***



Conclusion

Higher quality RNA is obtained from blood collected into the PAXgene™ blood RNA tube than blood collected into Lithium Heparin, Sodium Citrate, EDTA or TRIzol®LS tubes. All samples had mean A_{260}:A_{280} ratios in the desired range with only PAXgene™ having all ratios within the desired range. TRIzol®LS had the highest yield of RNA but all samples were of sufficient yield to use for microarray work. In order to obtain the best quality of RNA for microarray analysis the PAXgene™ RNA collection tube is therefore the best collection tube of those studied.

3.2: Optimisation of RNA Extraction Method Part 1.

Background

Once the most promising blood collection tube had been identified, the next step was to compare RNA extraction methods. The kits chosen were an example of a phenol-guanidine isothiocyanate based extraction (TRIzol®LS), an example of an on-column method (QIAamp® RNA Blood Mini Kit), an example of a proprietary RNA stabilisation agent and extraction kit (PAXgene™ Blood RNA Kit). In addition, the phenol-guanidine isothiocyanate extraction was combined with the on-column method's clean up protocol.

Methods

For each of seven umbilical cords, aliquots of blood were collected into EDTA tubes (at least three 1 ml tubes) and 2.5 ml of blood collected into a PAXgene™ blood RNA collection tube. For each sample, one each of a TRIzol®LS extraction, QIAamp® extraction, a TRIzol®LS with QIAamp® on-column clean-up and a PAXgene™ extraction was performed (pages 67 to 76).

Results

The raw data for each sample according to method of RNA extraction can be found in table 3.2.1 in appendix 2. This data is displayed graphically in figure 3.2.1, which displays yield, A_{260}:A_{280} and RIN for each extraction method. Total yield is displayed on the purple line graph for each extraction method with an additional blue line displaying yield per 0.5ml of blood for the PAXgene™ tube. The A_{260}:A_{280} ratio

is displayed as a red line graph with the y-axis on the right. The pastel bars represent the RIN values for each RNA extraction method.

Figure 3.2.1: ***RNA Yield, A260:A280 & RIN for Each Extraction Method.***

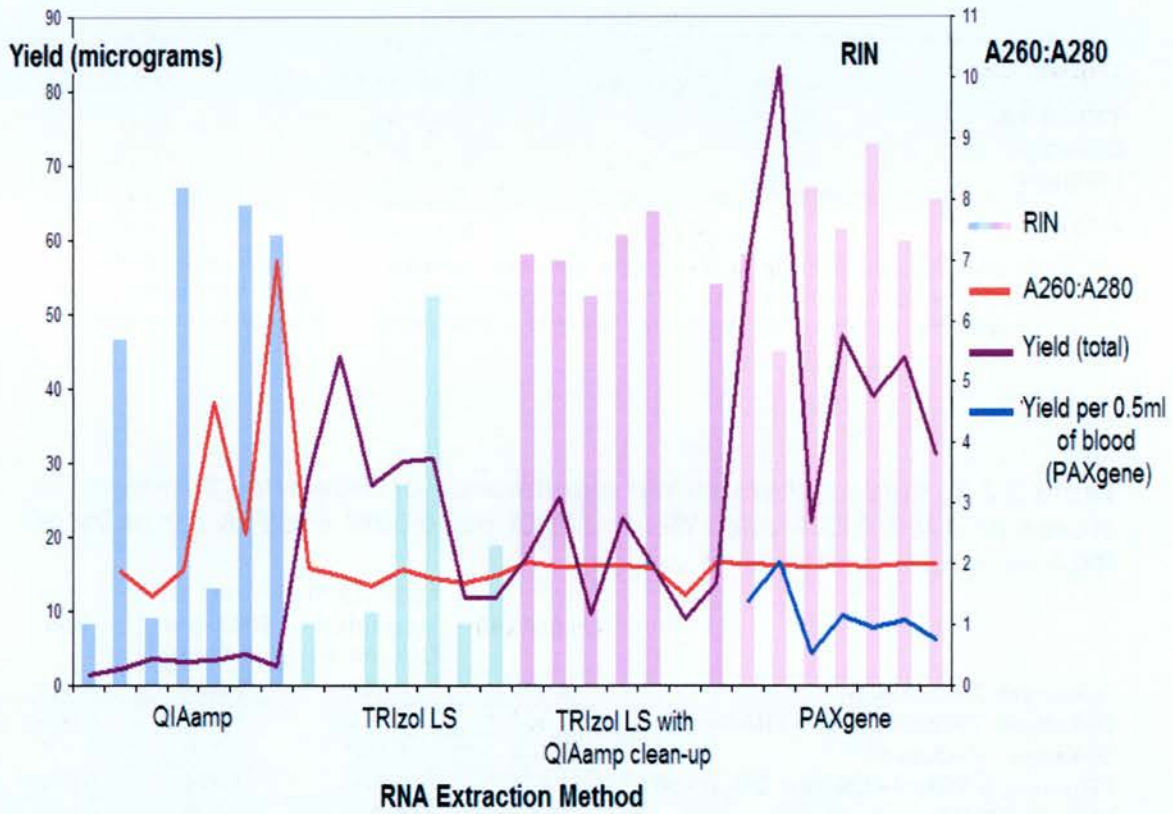


Table 3.2.2 displays the mean values for A_{260}:A_{280} ratio, RNA yield per 0.5 ml of blood and RIN for each method of RNA extraction. The means for both the original and the repeat A_{260}:A_{280} ratios are displayed for completeness (see note accompanying table 3.2.1 in appendix 2 for explanation).

Table 3.2.2: *Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each method of RNA extraction.*

RNA Extraction Method	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA (μg) per 0.5ml blood	Mean (range) RIN values
QIAamp®	7	2.77 (0-7) On repeat: 1.89 (1.82-2.0)	2.91 (1.34-4.18)	4.7 (1-8.2)
TRIzol® LS	7	1.80 (1.65-1.92)	26.35 (11.88-44.48)	2.2 (0-6.4)
TRIzol® & QIAamp® clean-up	7	1.91 (1.5-2.02)	16.29 (9.03-25.41)	6.0 (0-7.8)
PAXgene™	7	1.98 (1.95-2.01)	9.25 (4.41-16.66)	7.5 (5.5-8.9)

The *Analyst* article table (publication arising from this work) used the original spec values for the a260/a280 but repeat specs were done. The repeat ratios have now been added to the tables. The repeat values were the ones used to calculate the yield of RNA (verified). RIN values have also been checked.

Table 3.2.3: *Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of RNA extraction methods (p values).*

	A260:A280	A260:A280 (repeated QIAamp®)	Yield of RNA per 0.5 ml	RIN
QIAamp®-TRIzol®LS	0.312	0.028	0.002	0.316
QIAamp®-TRIzol®LS and QIAamp®	0.364	0.772	0.002	0.090
QIAamp®-PAXgene™	0.404	0.001	0.010	0.100
TRIzol®LS-TRIzol®LS and QIAamp®	0.101	0.101	0.014	0.006
TRIzol®LS-PAXgene™	0.002	0.002	0.003	0.001
TRIzol®LS and QIAamp®-PAXgene™	0.348	0.348	0.003	0.371

Table 3.2.3 displays the values of statistical significance when comparing difference of means of A_{260}:A_{280} ratio, yield of RNA per 0.5 ml and RIN for each pair of extraction methods. Looking at both Tables 3.2.2 and 3.2.3 it can be seen that in terms of yield the TRIzol®LS method had significantly greater yields of RNA per 0.5 ml of blood than the QIAamp® method, the TRIzol®LS method followed by the QIAamp®-on-column clean up and the PAXgene™ method. In addition the TRIzol®LS method followed by QIAamp®-on-column clean up had significantly

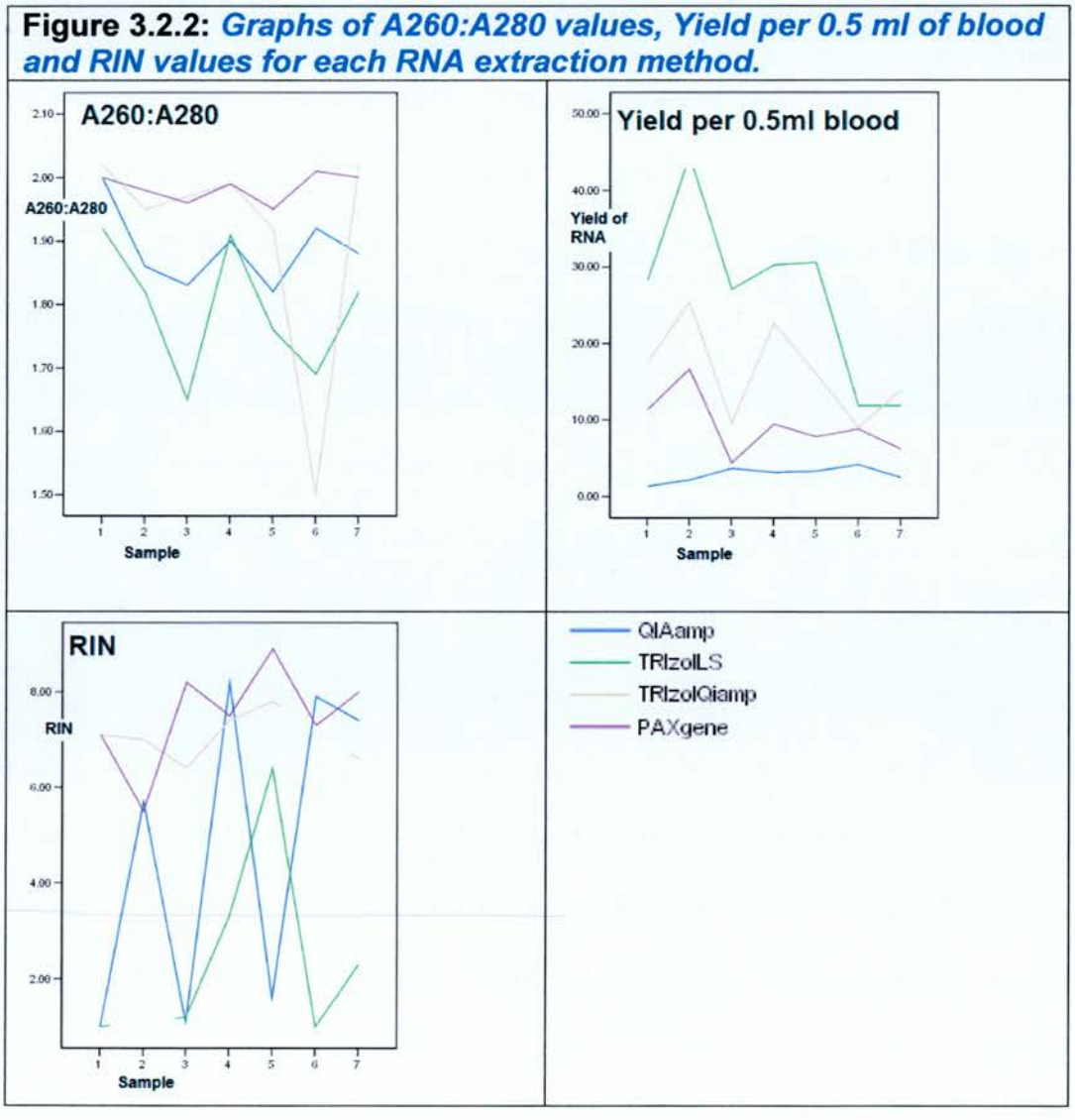
greater yields than the QIAamp® method alone and the PAXgene™ method. The PAXgene™ method yielded significantly greater RNA than the QIAamp® method. In terms of RIN values, although the PAXgene™ values were generally higher than those for each of the other methods, this difference was only statistically significant when comparing with the TRIzol®LS values. The only other statistically significant difference between groups with respect to RIN values was that TRIzol®LS with QIAamp®-on-column clean-up had higher values than those of TRIzol®LS alone. In terms of A_{260}:A_{280} ratios the only statistically significant differences were that the PAXgene™ values were higher (but not too high and therefore better) than those from TRIzol®LS and from QIAamp®, also QIAamp® gave statistically higher values than TRIzol®LS. Figure 3.2.2 gives a pictorial representation of the trends towards PAXgene™ having the best and most consistent A_{260}:A_{280} ratio, Trizol®LS having the greatest yield of RNA and PAXgene™ having the trend for highest RIN. Bioanalyzer graphs and RIN values for each sample can be found in table 3.2.4 in appendix 2.

Conclusions

From the results obtained in this section PAXgene™ extracted samples had the highest RIN values when compared to each of the other methods. Although this only reached statistical significance when compared with the TRIzol®LS method, I was satisfied that the trend towards better RIN values displayed was convincing enough to choose the PAXgene™ method as my method of choice for extracting RNA in future stages of the study. The consistency of the good A_{260}:A_{280} ratios obtained with PAXgene™ further consolidated this view. In terms of yield, TRIzol®LS was

clearly the best but I was happy that PAXgene™ consistently provided sufficient yield of RNA for use in microarray experiments. As the main priority was to maximise quality of RNA obtained, and as sufficient yield could be obtained, PAXgene™ was chosen to be taken forward to use in the next stages of the study as the optimum extraction method.

Figure 3.2.2: ***Graphs of A260:A280 values, Yield per 0.5 ml of blood and RIN values for each RNA extraction method.***



3.3. Optimisation of RNA Extraction Method Part 2 – PAXgene™ vs MAGAzorb®.

Background

After study 3.2 had been completed, I came across MAGAzorb®, an RNA extraction kit newly available in the UK that used a magnetic bead separation technology that did not require any centrifugation steps. As this was an entirely different method to those tried in 3.2, it was decided that this would be examined further. PAXgene™ blood RNA extraction was compared with MAGAzorb® RNA extraction.

Method

The MAGAzorb® kit contained enough materials for 12 extractions. 6 umbilical cord samples were used for comparison. These were collected as described previously. For the first 2 samples, blood was collected into EDTA, lithium heparin, sodium citrate and fluoride oxalate tubes as well as the PAXgene™ tube. This was to gain an idea of optimal blood collection tube for the method. The 4 subsequent cord blood samples were injected into EDTA tubes and PAXgene™ tubes only. While appreciating that such a small sample number would not give statistical assurance of best method if they were all similar, I wished only to get an idea if this method could rival the PAXgene™ method. Had the PAXgene™ and MAGAzorb® methods been shown to be equally good, I would have optimised the MAGAzorb® method further and carried out further comparisons

The cord blood samples in their collection tubes were taken directly to the lab and the PAXgene samples were left to sit on the bench for 2 hours. The MAGAzorb® extractions were carried out straight away. The PAXgene™ and MAGAzorb® RNA extraction methods and methods for assessment of RNA quality and quantity used are described in chapter 2 (pages 76 and 80).

Results

Table 3.3.1: ***Raw values for samples according to RNA extraction method and collection tube.***

	Method/Tube	260/280 ratio	A260	Total RNA (μg)	RNA/ 0.5 ml (μg)	RIN
cs25	**MAGAzorb® EDTA**	1.56	0.042	3.99	9.98	1.8
cs25	**PAXgene™**	1.29	0.022	2.91	2.91	9.4
cs26	**MAGAzorb® EDTA**	1.77	0.046	4.37	10.93	1.4
cs26	**MAGAzorb® Li Heparin**	1.81	0.049	4.66	11.65	1.7
cs26	**MAGAzorb® Na Citrate**	2.04	0.049	4.66	11.65	1.1
cs26	**MAGAzorb® Fl Oxalate**	1.73	0.057	5.42	13.55	1
cs26	**PAXgene™**	1.9	0.019	2.51	2.51	8.5
cs27	**MAGAzorb® EDTA**	1.77	0.078	7.41	18.56	1
cs27	**MAGAzorb® Li Heparin**	2.08	0.05	4.75	11.88	2.1
cs27	**MAGAzorb® Na Citrate**	1.85	0.048	4.56	11.4	1.7
cs27	**MAGAzorb® Fl Oxalate**	1.93	0.081	7.7	19.25	1
cs27	**PAXgene™**	2.32	0.044	5.81	5.81	9
cs28	**MAGAzorb® EDTA**	1.66	0.058	5.51	13.78	1.9
cs28	**PAXgene™**	2	0.016	2.11	2.11	8.3
cs29	**MAGAzorb® EDTA**	1.8	0.079	7.51	18.78	2.1
cs29	**PAXgene™**	1.92	0.023	3.04	3.04	9.3
cs30	**MAGAzorb® EDTA**	1.85	0.074	7.03	17.58	2.9
cs30	**PAXgene™**	2.29	0.016	2.11	2.11	n/a
cs34	**MAGAzorb® EDTA**	1.53	0.092	8.74	21.85	2.3
cs34	**PAXgene™**	1.31	0.017	2.24	2.24	8.6

Table 3.3.1 displays the raw data for samples according to collection tube and RNA extraction method. Table 3.3.2 shows the mean and range of values obtained for A_{260}:A_{280} ratio, yield of RNA per 0.5ml of blood sampled and RIN values. For the

MAGAzorb® extractions, the "0.5 ml of blood" is derived rather than being an actual value as only 200 microlitres of blood was used for each extraction.

Table 3.3.2: ***Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each method of RNA extraction.***

Method/Tube	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA /0.5 ml blood (μg)	Mean (range) RIN
MAGAzorb® EDTA	7	1.71 (1.53-1.85)	15.92 (9.98-21.85)	1.9 (1-2.9)
MAGAzorb® Li Heparin	2	1.95 (1.81-2.08)	11.77 (11.65-11.88)	1.9 (1.7-2.1)
MAGAzorb® Na Citrate	2	2.06 (2.04-2.08)	11.53 (11.4-11.65)	1.4 (1.1-1.7)
MAGAzorb® Fl Oxalate	2	1.83 (1.73-1.93)	16.4 (13.55-19.25)	1 (1-1)
PAXgene™	7	1.86 (1.29-2.32)	2.96 (2.11-5.81)	7.6* (0-9.4)

***For the PAXgene tube that had RIN value of 0 after RNA extraction at 2 hours, another tube was processed at 24 hours and the RIN value was 9.4, if this was factored in instead of the 0: mean RIN= 8.9.**

From the data in table 3.3.2 and from the significance values displayed in table 3.3.3, there is no significant difference between the A_{260}:A_{280} ratios obtained with each extraction method. On looking at the data for RNA yield, the MAGAzorb® RNA yields from an EDTA collection tube were significantly higher than the yields using PAXgene™ system. Although not reaching statistical significance, all of the other collection tubes with MAGAzorb® extractions achieved higher yields of RNA than PAXgene™ per 0.5 ml of blood. When looking at the RIN values however it is clear that the MAGAzorb® samples were of much poorer quality than the PAXgene™ samples. This difference in quality was highly statistically significant when comparing PAXgene™ with the MAGAzorb® extracted samples collected in EDTA, Lithium Heparin or Sodium Citrate tubes. Although statistically the difference in RIN obtained using MAGAzorb® extracted samples collected in Fluoride Oxalate

was not significant (0.021), the results suggest that the RNA obtained using PAXgene™ was of better quality. Of note, one PAXgene™ sample did not give a reading for RIN on the Bioanalyzer - the graph is displayed in table 3.3.4 (appendix 2). For this same sample as part of another experiment, a further PAXgene™ tube was incubated at room temperature for 24 hours and gave a RIN value of 9.4. If this value was used in place of the unreadable RIN, the mean RIN for PAXgene™ would rise to 8.9 which would only serve to strengthen the evidence that PAXgene™ yields higher quality RNA.

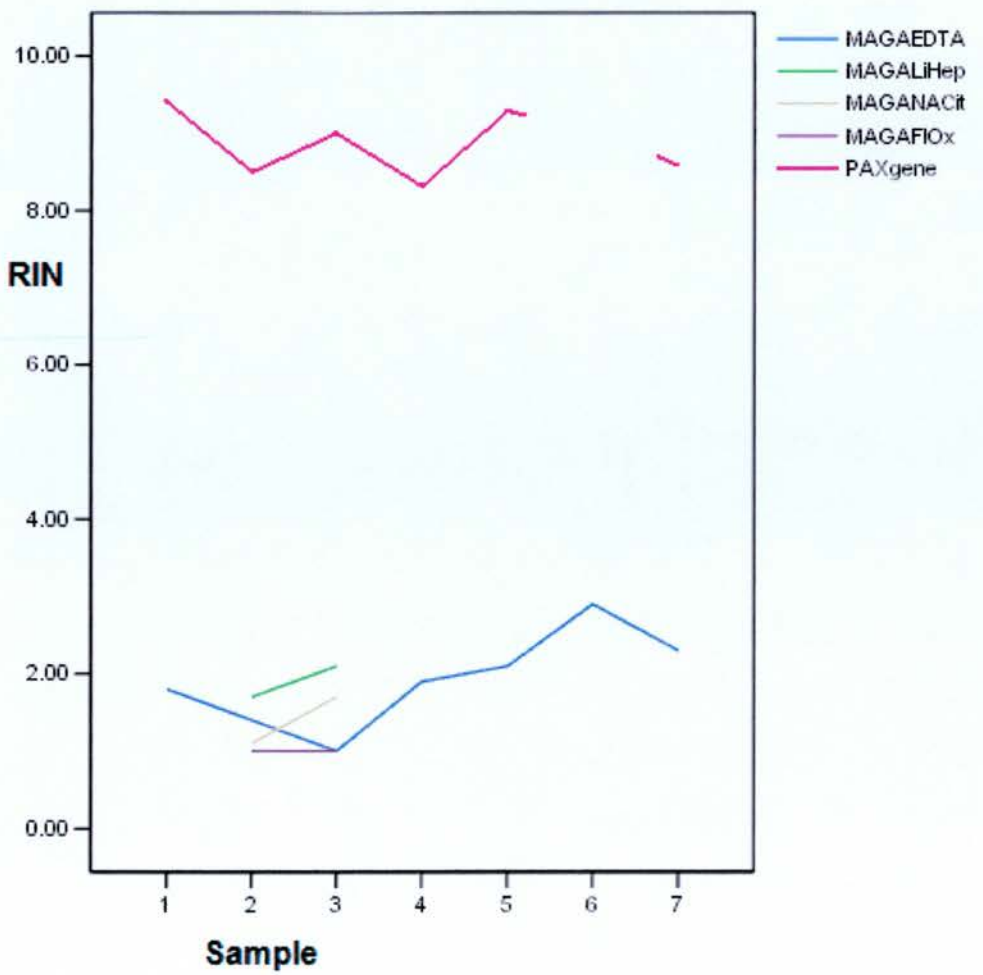
Table 3.3.3: ***Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for collection tube and RNA extraction method (p values).***

	A260:A280	Yield of RNA per 0.5 ml of blood	RIN
MAGAzorb® EDTA - MAGAzorb® Li Heparin	0.418	0.568	0.330
MAGAzorb® EDTA – MAGAzorb® Na Citrate	0.317	0.564	0.758
MAGAzorb® EDTA – MAGAzorb® FI Oxalate	0.656	0.336	0.500
MAGAzorb® EDTA - PAXgene™	0.239	<0.001	<0.001
MAGAzorb® Li Heparin – MAGAzorb® Na Citrate	1.000	0.500	0.126
MAGAzorb® Li Heparin – MAGAzorb® FI Oxalate	0.188	0.339	0.139
MAGAzorb® Li Heparin - PAXgene™	0.272	0.127	0.005
MAGAzorb® Na Citrate – MAGAzorb® FI Oxalate	0.661	0.349	0.410
MAGAzorb® Na Citrate – PAXgene™	0.684	0.151	0.004
MAGAzorb® FI Oxalate - PAXgene™	0.238	0.062	0.021

Given the low sample numbers examined in this study, statistical testing should be viewed with caution, the differences in RIN quality however are striking. Figure 3.3.1 illustrates the RIN values for each sample and method of extraction. These give clear visual representations of the higher quality of RNA obtained using the PAXgene™ method compared to the MAGAzorb® method when determining quality using a Bioanalyzer. Bioanalyzer traces and RIN values for each sample can be found in table 3.3.4 in appendix 2.

The MAGAzorb® extracted samples performed uniformly poorly in terms of RIN values, despite appearing to have acceptable A_{260}:A_{280} ratios and yields. It could be argued that something within the MAGAzorb® extracted samples could be interfering with the Bioanalyzer analysis. Bioanalyzer results are used to determine suitability for running individual samples on microarray. The consistently poor readings therefore ruled MAGAzorb® out as a viable RNA extraction method for further use in this study.

Figure 3.3.1: ***Graph of RIN values according to collection tube and extraction method.***



Conclusions

RNA extracted using MAGAzorb® gave higher yields but consistently poor quality RNA (as indicated by RINs) compared to PAXgene™. It was therefore decided to continue using PAXgene™ for RNA extraction for this study.

3.4: Volume of Blood Required

Most research utilising RNA from whole blood has been restricted to the use of adult blood samples and it is only recently that data on the use of neonatal blood has begun to emerge. There are differences that need to be considered in terms of composition of the blood, for example the white cell count of neonatal blood is generally considerably higher than in healthy adults. In addition, it is vital to look at the volume of sample required if any test using neonatal blood is ever going to be clinically practical. Validation studies carried out on behalf of PreAnalytiX generally used 3 PAXgene blood RNA tubes per adult patient in order to get sufficient RNA (337) to carry out microarray studies using the Affymetrix system. In other words 7.5 ml of blood per patient was required. The circulating blood volume of a term neonate is around 75-85 ml/kg – adult volumes of blood for samples would clearly not be practical and indeed would almost certainly be detrimental to the infants' well being. It was therefore necessary to investigate the use of smaller blood volumes using the PAXgene™ system.

PAXgene™ blood RNA tubes each contain a fixed volume of 6.9 ml RNA stabilisation agent (I originally obtained this information directly from PreAnalytiX as a personal communication; later it could also be found in the PAXgene™ Blood RNA Tube product circular). The tubes are designed for use with a particular blood collection system that leads to the tubes being filled with 2.5 ml of adult blood, giving a ratio of 2.76 ml of stabilisation agent per ml of blood. I set out to determine whether it was feasible to use smaller amounts of umbilical cord blood in each

PAXgene™ tube (0.5 ml) compared with 2.5 ml of umbilical cord blood. In addition, I pipetted aliquots of 0.69 ml of PAXgene™ RNA stabilisation agents from the PAXgene™ blood RNA tubes into two 2 ml microcentrifuge tubes and added 0.25 ml of umbilical cord blood to each at the time of sampling (these two aliquots were later merged). This was to explore the use of the same proportions of reagents as those intended for the PAXgene™ tubes filled with 2.5 ml of blood but using the smaller blood volume of 0.5 ml (reagents used in the subsequent extraction process were also proportionately reduced).

Methods

Umbilical cord blood was sampled from seven umbilical cords as described previously. For each sample, 2.5 ml was injected into a PAXgene™ blood RNA tube, 0.5 ml was injected into another PAXgene™ blood RNA tube and aliquots of 0.25 ml were added to two microcentrifuge tubes each containing an 0.69 ml aliquot of PAXgene™ blood RNA tube fluid. The samples were transported to the laboratory and incubated, upright, for at least 2 hours.

For each PAXgene™ blood RNA tube, RNA extraction was performed as described in chapter 2 (page 76). For the micro-centrifuge tubes containing RNA stabilisation agent, the protocol was varied slightly in order to keep reagent ratios comparative. The micro-centrifuge tubes were centrifuged in a Heraeus Biofuge Fresco set to 6,500 rpm for 10 minutes. The supernatant was then removed and 0.5 ml of RNAse free water added to each tube. The tube was then vortexed thoroughly and centrifuged for a further 10 minutes at 6,500 rpm. The entire supernatant was

removed and discarded and then 180 microlitres of Buffer BR1 was added to each tube. Mixing was achieved by repetitive pipetting followed by thorough vortexing. The contents of both micro-centrifuge tubes were then pipetted into a single fresh 2 ml micro-centrifuge tube. From the point of addition of 300 microlitres of Buffer BR2, the RNA extraction process continued as per the original protocol.

Note: From this study, it became apparent that 10 minutes of centrifugation after the 55 degree incubation step was not always sufficient to separate the debris adequately to allow complete removal of the supernatant. I found that increasing the duration of centrifugation was effective in leading to adequate separation. The length of centrifugation was increased for each sample until separation was seen to take place (the average time of centrifugation required was longer for the 2.5 ml samples in the PAXgene™ tubes than either of the 0.5 ml sampling methods). The operating instructions from the manufacturer have since been changed and now recommend that where a water bath or heating block have been used the sample should be vortexed for 30 seconds prior to centrifugation or alternatively to transfer the sample into a QIAshredder™ spin column, centrifuge for 3 minutes at maximum speed then to transfer the supernatant into a fresh microcentrifuge tube and proceed to the 10 minute centrifugation step as previously.

Results

From the raw data presented in table 3.4.1 and the mean values found in table 3.4.2 it can be seen that all of the tubes had consistently good A_{260}:A_{280} ratios and from the statistical significance values displayed in table 3.4.3 that there was no statistically

significant difference between tubes in this respect. In addition, it can be seen as expected that the greatest overall yields were obtained from the PAXgene™ tubes with 2.5 ml of blood. When the amount of RNA per 0.5 ml of blood is considered, the highest yields were obtained from 0.5 ml of blood in aliquots of PAXgene™ fluid, followed by 2.5 ml of blood in the PAXgene tube and were least for 0.5 ml of blood in the PAXgene™ tube.

Table 3.4.1: *Raw values for samples according to volume of blood sampled and volume of PAXgene™ RNA stabilisation agent.*

	Volume of blood / PAXgene™	260/280 ratio	A260	Total RNA (μg)	RNA/ 0.5 ml blood (μg)	*RIN*
cs10	**2.5 ml in PAXgene™ tube**	2	0.43	56.76	11.35	7.1
cs10	**0.5 ml in PAXgene™ tube**	2.06	0.035	4.62	4.62	8.8
cs10	**0.5 ml in aliquot of PAXgene™ fluid**	2.05	0.131	17.29	17.29	8.3
cs11	**2.5 ml in PAXgene™ tube**	1.98	0.631	83.29	16.66	5.5
cs11	**0.5 ml in PAXgene™ tube**	2	0.034	4.49	4.49	7.9
cs11	**0.5 ml in aliquot of PAXgene™ fluid**	2.07	0.149	19.67	19.67	6.8
cs12	**2.5 ml in PAXgene™ tube**	1.96	0.167	22.04	4.41	8.2
cs12	**0.5 ml in PAXgene™ tube**	1.81	0.038	5.02	5.02	9.4
cs12	**0.5 ml in aliquot of PAXgene™fluid**	1.88	0.062	8.18	8.18	8.6
cs13	**2.5 ml in PAXgene™ tube**	1.99	0.358	47.26	9.45	7.5
cs13	**0.5 ml in PAXgene™ tube**	2	0.07	9.24	9.24	9.4
cs13	**0.5 ml in aliquot of PAXgene™ fluid**	1.92	0.123	16.24	16.24	8.5
cs14	**2.5 ml in PAXgene™ tube**	1.95	0.296	39.08	7.82	8.9
cs14	**0.5 ml in PAXgene™ tube**	2.14	0.045	5.94	5.94	9.7
cs14	**0.5 ml in aliquot of PAXgene™ fluid**	2	0.112	14.78	14.78	9
cs15	**2.5 ml in PAXgene™ tube**	2.01	0.335	44.22	8.84	7.3
cs15	**0.5 ml in PAXgene™ tube**	1.83	0.044	5.81	5.81	9.3
cs15	**0.5 ml in aliquot of PAXgene™ fluid**	2.12	0.104	13.73	13.73	7.9
cs16	**2.5 ml in PAXgene™ tube**	2	0.236	31.15	6.23	8
cs16	**0.5 ml in PAXgene™ tube**	2.05	0.043	5.68	5.68	8.3
cs16	**0.5 ml in aliquot of PAXgene™ fluid**	2.11	0.078	10.30	10.30	8.3

The difference in yields obtained per 0.5 ml of blood in the PAXgene™ tubes with 2.5 ml of blood and those with 0.5 ml of blood were not statistically significant. The greater yields obtained using the aliquots were statistically significantly different when compared with either of the PAXgene™ tube conditions. Importantly, the yields obtained for the 0.5 ml of blood in the PAXgene™ tubes were still sufficient for use in microarray experiments with range of yield from 4.49 to 9.24 μg. The RIN values were seen to be best for the 0.5ml of blood in the PAXgene™ tube, followed by the 0.5 ml of blood in aliquots of PAXgene™ fluid and least good in the PAXgene™ tubes with 2.5 ml of blood in them, with mean RIN values of 9, 8.2 and 7.5 respectively. The mean RIN values obtained from the 0.5 ml of blood in PAXgene™ were statistically significantly better than either of the other 2 conditions.

Table 3.4.2: ***Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each blood volume/PAXgene™ RNA stabilisation agent volume.***

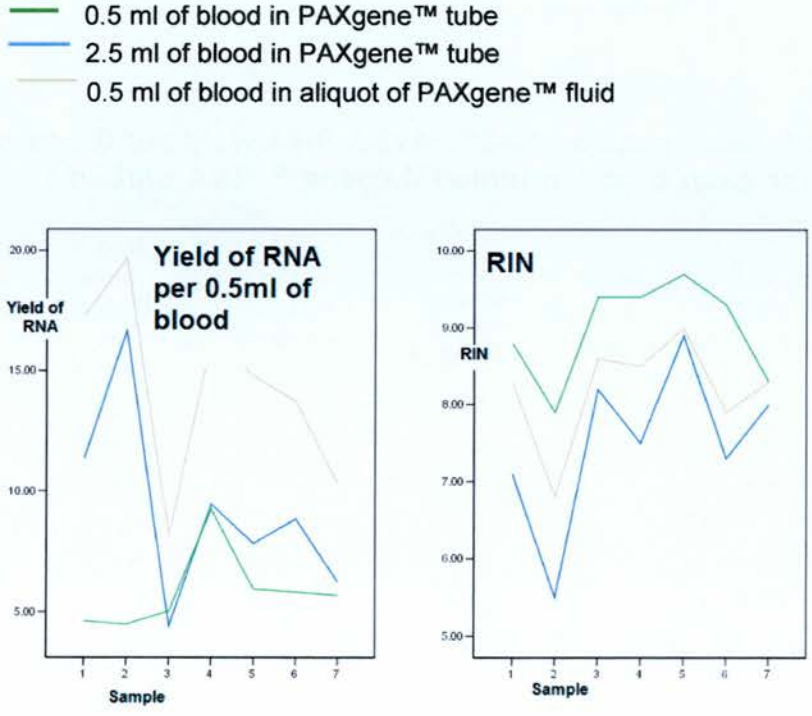
Volume Blood / PAXgene™ RNA stabilisation agent	Number of samples (n)	Mean (range) A260:A280	Mean (range) total RNA (μg)	Mean (range) RNA/ 0.5 ml blood (μg)	Mean (range) RIN
2.5 ml in PAXgene™ tube	7	1.98 (1.95-2.01)	46.26 (22.04-83.29)	9.25 (4.41-16.66)	7.5 (5.5-8.9)
0.5 ml in PAXgene™ tube	7	1.98 (1.81-2.14)	5.83 (4.49-9.24)	5.83 (4.49-9.24)	9.0 (7.9-9.7)
0.5 ml in aliquot of PAXgene™ fluid	7	2.02 (1.88-2.12)	14.31 (8.18-19.67)	14.31 (8.18-19.67)	8.2 (6.8-9.0)

Table 3.4.3: ***Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of blood volume/PAXgene™ RNA stabilisation agent volume (p values).***

	A260:A280	Yield of RNA per 0.5 ml blood	RIN
2.5 ml in PAXgene™ tube – 0.5 ml in PAXgene™ tube	1.000	0.094	0.002
2.5 ml in PAXgene™ tube – 0.5 ml in aliquot of PAXgene™ fluid	0.268	<0.001	0.008
0.5 ml in PAXgene™ tube – 0.5 ml in aliquot of PAXgene™ fluid	0.503	0.002	0.004

Graphs displaying yield per 0.5 ml of blood and RIN value for each sample are displayed in figure 3.4.1. Bioanalyzer traces and RIN values for each sample can be found in table 3.4.4 in appendix 2.

Figure 3.4.1: ***Graphs of Yield per 0.5 ml of blood and RIN values for each blood/PAXgene™ RNA stabilisation agent volume.***



Conclusions

All three combinations of blood volume/PAXgene™ fluid volume examined gave consistently good A_{260}:A_{280} ratio. All three combinations also consistently yielded sufficient RNA for reliable use in microarray experiments, however the greatest yield per 0.5 ml was obtained using the aliquot method. The RIN values were best with the 0.5 ml of blood injected directly into a PAXgene™ blood RNA tube. As the main aim was to optimise RNA quality while achieving sufficient yield and having a requirement for the smallest possible sample, these results indicated that the 0.5 ml of blood injected into the PAXgene™ was the best to take forward for use in future work. The additional advantage of this method over the aliquot method is that the PAXgene™ fluid remains completely isolated within the PAXgene™ tube.

3.5: Incubation time of PAXgene™ tube prior to RNA extraction. Part 1

Background

Investigation of optimum incubation period at room temperature was important both to determine the optimum time to maximise yield and quality of RNA obtained and also to help set parameters within which clinical sampling could be confidently carried out. For example, if samples could be incubated at room temperature for 24 or 72 hours this would mean that samples could be transferred to the lab from the clinical area without freezing and batched for extraction.

Method

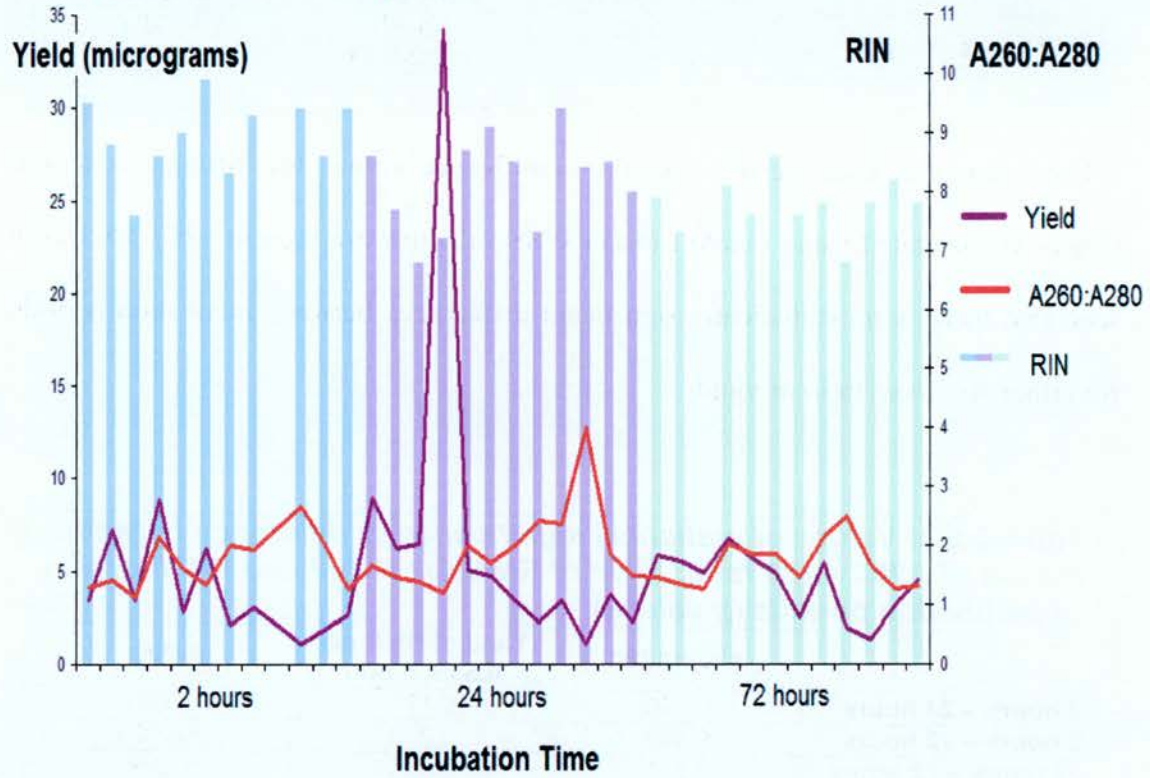
Blood was sampled from 12 umbilical cords. For each cord, three aliquots of 0.5 ml of blood were taken and each injected into a PAXgene™ Blood RNA Tube. The samples were then taken to the lab and incubated at room temperature. For each cord sampled one tube was incubated for 2 hours at room temperature, one for 24 hours at room temperature and the third for 72 hours at room temperature prior to RNA extraction. RNA extraction, RNA quantification and Bioanalyzer analysis were carried out as described in chapter 2.

Results

Raw data for each sample with incubation periods 2, 24 and 72 hours can be found in appendix 2, table 3.5.1. This data is displayed graphically in figure 3.5.1 with yield, A260:A280 and RIN displayed for each incubation period. Total yield is

displayed on the purple line graph for each incubation time. The A_{260}:A_{280} ratio is displayed as a red line graph with the y-axis on the right. The pastel bars represent the RIN values for each collection tube.

Figure 3.5.1: RNA Yield, A260:A280 Ratio and RIN for Each Incubation Time (2, 24 and 72 hours).



The mean values for A_{260}:A_{280} ratio, yield of RNA and RIN values for each incubation period are displayed in table 3.5.2.

Table 3.5.2: *Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for incubation of 2,24 or 72 hours prior to RNA extraction.*

Incubation Time	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA /0.5 ml blood (μg)	Mean (range) RIN values
2 hours	12 *(11 for RIN)	1.76 (1.13-2.67)	3.73 (1.06-8.84)	8.9* (7.6-9.9)
24 hours	12	1.97 (1.19-4.0)	6.81 (1.06-34.19)	8.2 (6.8-9.4)
72 hours	12 #(11 for RIN)	1.69 (1.26-2.5)	4.37 (1.32-6.73)	7.8# (6.8-8.6)

When examining these results and the significance values for differences in the means as calculated using paired t-tests (SPSS v12) that are seen in table 3.5.3, it is seen that there is no statistically significant differences between incubation periods for either A_{260}:A_{280} ratio or yield.

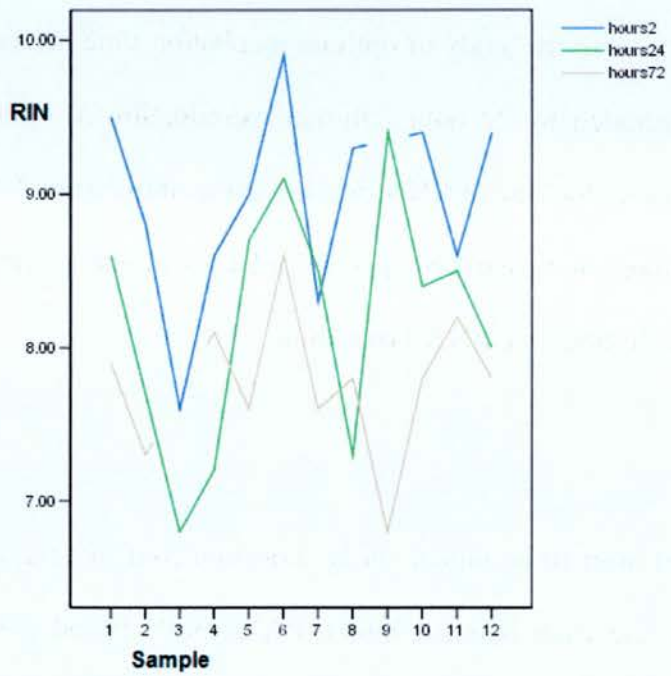
Table 3.5.3: *Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of incubation periods (p values).*

	A260:A280	Yield of RNA per 0.5 ml blood	RIN
2 hours – 24 hours	0.199	0.171	0.001
2 hours – 72 hours	0.592	0.218	<0.001
24 hours – 72 hours	0.221	0.316	0.076

When looking at the RIN values however, the samples incubated for 2 hours at room temperature have significantly higher values than either those incubated for 24 hours or those incubated for 72 hours. Although there was no statistically significant difference between RIN values incubated for 24 hours and those incubated for 72 hours, the mean RIN was lower at 72 hours than at 24 hours. This indicates that with continued incubation at room temperature, RNA degradation occurs by 24 hours and continues to degrade further by 72 hours. RIN values are displayed in figure

3.5.2 and Bioanalyzer traces for each sample can be found in appendix 2 (table 3.5.4).

Figure 3.5.2: ***Graph of RIN values for each incubation period***



Conclusion

On incubation at room temperature, RNA degradation appears to occur by 24 hours as evidenced by a significant reduction in RIN value. This indicated that further investigation into optimum duration of incubation for periods less than 24 hours was warranted.

3.6: Incubation time of PAXgene™ tube prior to RNA extraction. Part 2

Background

Following the first part of the study of optimal incubation time presented in 3.5 that suggested RNA degraded by 24 hours, further examination of optimal incubation time was carried out by looking at RNA quality and quantity after 2,4,6,8,10 and 12 hours. It was important to explore this in order to define parameters of RNA handling which would optimize RNA expression.

Method

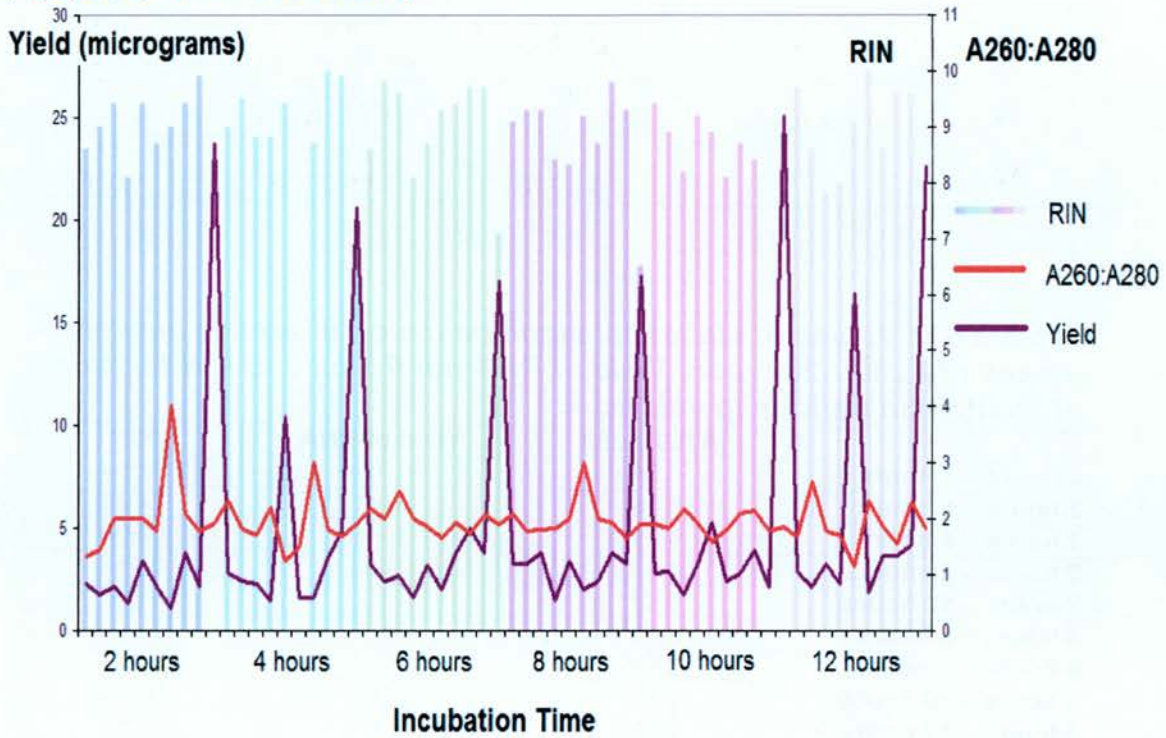
Blood was sampled from 10 umbilical cords. For each cord, six aliquots of 0.5 ml of blood were taken and each injected into a PAXgene™ Blood RNA Tube. The samples were then taken to the lab and incubated at room temperature. For each cord sampled one tube was incubated for 2 hours at room temperature, one for 4 hours at room temperature and one each at 6, 8, 10 and 12 hours at room temperature prior to RNA extraction. RNA extraction, quantification and quality assessment were carried out as described in chapter 2.

Results

The raw data for A_{260}:A_{280} ratio, RNA yield and RIN values for each tube are displayed in table 3.6.1 (a and b) which can be found in appendix 2. The raw data is displayed graphically in figure 3.6.1 for each incubation period. Total yield is displayed on the purple line graph for each incubation time. The A_{260}:A_{280} ratio is

displayed as a red line graph with the y-axis on the right. The pastel bars represent the RIN values for each collection tube. This data along with the mean values displayed in table 3.6.2 show similar A_{260}:A_{280} ratios and RNA yields for all of the tubes. In addition there is a suggestion of higher RIN values for those samples extracted after 2, 4 and 6 hours incubation (mean RIN of 9 – 9.1) compared with those extracted at 8, 10 or 12 hours incubation (mean RIN of 8.6 – 8.8).

Figure 3.6.1: ***RNAYield, A260:A280 and RIN for Each Incubation Time (2, 4, 6, 8, 10 and 12 hours).***



When the statistical significance of differences between the means are examined (table 3.6.3), it can be seen that there is no statistical significance between any of the incubation times for A_{260}:A_{280} ratios, yields or RIN values except for yield at 2 hours compared with 10 hours. All incubation conditions produced RNA of sufficient quantity and of high enough quality to run on microarray. Graphs of A_{260}:A_{280} ratio,

yield and RIN for each sample are displayed in figure 3.6.2. Bioanalyzer tracings and RIN values for each sample can be found in table 3.6.4 in appendix 2.

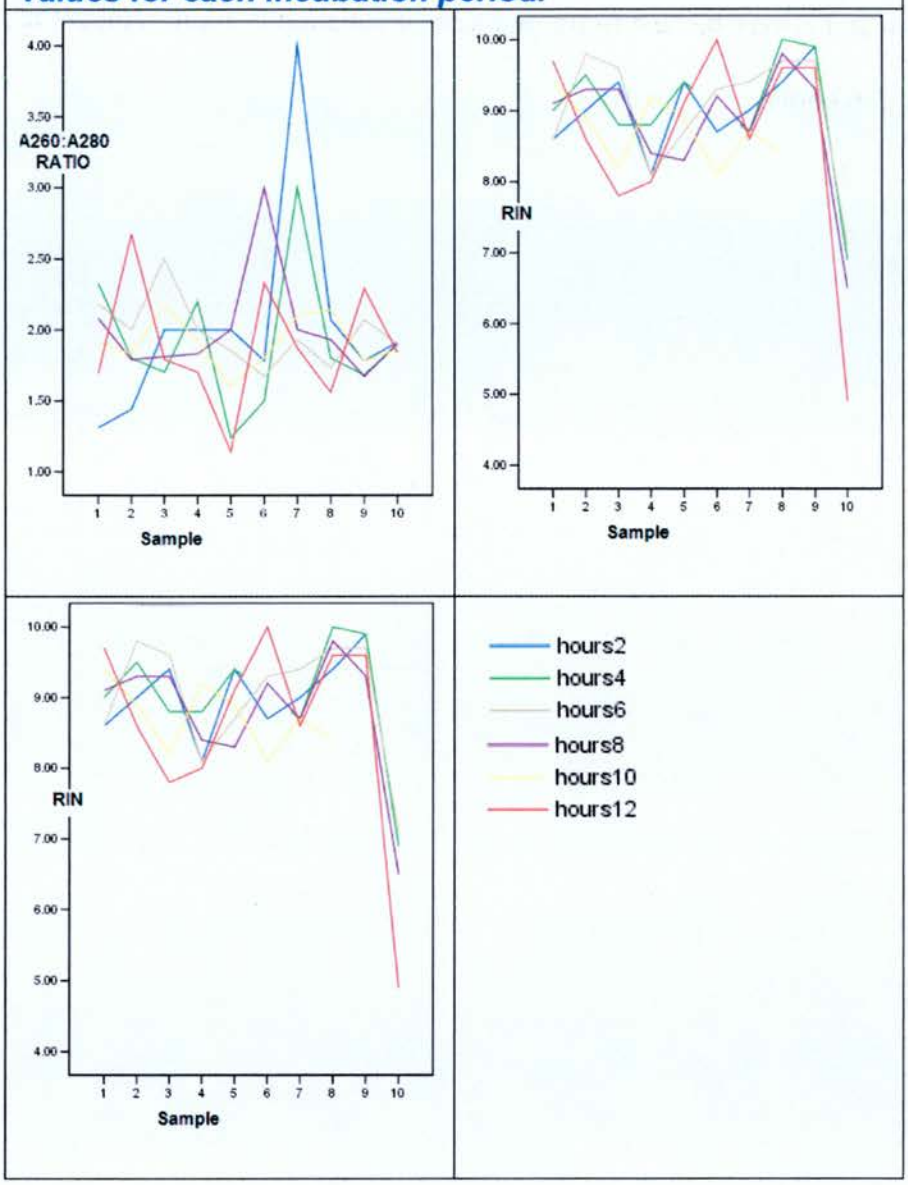
Table 3.6.2: *Mean values of A260:A280, RNA yield according to incubation period before RNA extraction.*

Incubation times (hours)	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA /0.5 ml blood (μg)	Mean (range) RIN
2	10 *(9 for RIN)	2.03 (1.31.-4.0)	4.37 (1.06-23.76)	9.1* (8.1-9.9)
4	10 #(9 for RIN)	1.92 (1.24-3.0)	5.15 (1.45-20.59)	9.0# (6.9-10)
6	10	1.98 (1.67-2.5)	4.46 (1.58-17.03)	9.0 (7.1-9.8)
8	10	2.00 (1.67-3.0)	4.41 (1.45-17.29)	8.8 (6.5-9.8)
10	10 $(8 for RIN)	1.91 (1.6-2.17)	5.23 (1.72-25.08)	8.7$ (8.1-9.4)
12	10	1.89 (1.14-2.67)	6.30 (1.85-22.57)	8.6 (4.9-10)

Table 3.6.3: *Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of incubation periods (p values).*

	A260:A280	Yield of RNA	RIN
2 hours – 4 hours	0.542	0.374	0.354
2 hours – 6 hours	0.860	0.910	0.328
2 hours – 8 hours	0.918	0.959	0.954
2 hours – 10 hours	0.598	0.011	0.458
2 hours – 12 hours	0.641	0.165	0.852
4 hours – 6 hours	0.687	0.457	0.863
4 hours – 8 hours	0.692	0.396	0.129
4 hours – 10 hours	0.977	0.923	0.227
4 hours – 12 hours	0.899	0.089	0.051
6 hours – 8 hours	0.917	0.842	0.130
6 hours – 10 hours	0.371	0.418	0.263
6 hours – 12 hours	0.560	0.220	0.258
8 hours – 10 hours	0.535	0.369	0.367
8 hours – 12 hours	0.519	0.197	0.487
10 hours – 12 hours	0.883	0.393	0.577

Figure 3.6.2: ***Graphs of A260:A280 Ratio, Yield and RIN values for each incubation period.***



Conclusions

There is no statistically significant difference either in quality or quantity of RNA obtained when umbilical cord blood in PAXgene™ Blood RNA tubes is incubated for 2, 4, 6, 8, 10 or 12 hours at room temperature prior to incubation. All of these incubation conditions produced RNA good enough for use in microarray

experiments. There was however a suggestion of slight reduction in RIN by 8 hours incubation and it may be best to try and keep incubation at room temperature within the range 2-6 hours.

3.7: Storage of Samples For Up To One Week at -20 °C

Background

After I had shown that it was feasible to produce high quality microarray results from RNA extracted from small volumes of neonatal blood, it became important to consider ways to maximise our ability to study clinical samples. From a study view point it was important to minimise any variation between handling of samples and to ensure that conditions of transport and storage of samples were standardised. From a clinical point of view, samples could be obtained at any time of the day or night and it would often not be possible for the samples to have RNA extracted after the exact desired incubation at room temperature. In addition, if the study was to ever become multi-centred, the time of travel to the laboratory may exceed the desired incubation period if samples were to be kept at room temperature. Freezing samples at -20 °C immediately after sampling and transferring to the lab frozen and then defrosting and incubating at room temperature for a fixed period would make standardisation of technique much more robust. This however, would only be feasible if RNA quality and quantity were maintained for samples that were frozen and I therefore set out to investigate this.

Methods

Blood was sampled from six umbilical cords by the method previously described in chapter 2 (page 61). For each cord, aliquots of 0.5 ml of blood were injected directly into six PAXgene™ Blood RNA tubes. The samples were taken straight to the lab – in practice this took less than 10 minutes and would be a similar time to that required

to finish off any sampling procedure and label any samples. One of the six tubes was then left to incubate at room temperature and the other five were placed directly into a -20 °C freezer. The sample left at room temperature had RNA extraction performed after 2 hours. The next sample had RNA extraction commenced at 15 hours. All frozen samples were left at room temperature for 4.5 hours (allowing over 2 hours to defrost thoroughly and 2 hours incubation) after removal from the freezer. The third to sixth sample had RNA extraction commenced after 1 day, 3 days, 4 days and 7 days respectively: all were removed from the freezer and incubated at room temperature immediately prior to RNA extraction. These timings were chosen to investigate the possibility of samples being frozen overnight, over a weekend, long weekend or longer to facilitate lab extractions during the normal working week for potential future studies. RNA extractions were performed using the modified PAXgene™ protocol, RNA was quantified using the spectrophotometer and the samples run on Bioanalyzer as described in chapter 2.

Results

Raw data for this experiment is shown in table 3.7.1. Cord cs49 only yielded sufficient blood for five PAXgene™ tubes and the "3 day" incubation was therefore omitted for this sample.

Table 3.7.1: *Raw data for each sample according to whether frozen at -20 °C and duration of freezing.*

	Incubation conditions	Spec Reading	260/280 ratio	Total RNA (μg)	RIN
cs47	2hr RT only	0.013	1.86	1.72	1.4
cs47a	15 hr at -20 °C	0.054	2	7.13	9.6
cs47b	1 day at -20 °C	0.029	1.93	3.83	9.7
cs47c	3 days at -20 °C	0.034	1.89	4.49	9.3
cs47d	4 days at -20 °C	0.041	1.95	5.41	9.8
cs47e	7 days at -20 °C	0.027	1.93	3.56	9.2
cs48	2hr RT only	0.044	1.91	5.81	9.2
cs48a	15 hr at -20 °C	0.064	1.78	8.45	9.4
cs48b	1 day at -20 °C	0.068	1.89	8.98	9.4
cs48c	3 days at -20 °C	0.048	1.78	6.34	8.6
cs48d	4 days at -20 °C	0.069	1.92	9.11	10
cs48e	7 days at -20 °C	0.06	2	7.92	9.8
cs49	2hr RT only	0.029	1.93	3.83	7.9
cs49a	15 hr at -20 °C	0.05	2	6.6	9.3
cs49b	1 day at -20 °C	0.055	1.96	7.26	9.4
cs49c	3 days at -20 °C	.	.	.	.
cs49d	4 days at -20 °C	0.015	1.88	1.98	1.3
cs49e	7 days at -20 °C	0.027	1.8	3.56	8.4
cs50	2hr RT only	0.037	1.85	4.88	9.6
cs50a	15 hr at -20 °C	0.059	1.9	7.79	9.2
cs50b	1 day at -20 °C	0.044	1.91	5.81	8.3
cs50c	3 days at -20 °C	0.063	1.85	8.32	9.5
cs50d	4 days at -20 °C	0.048	1.78	6.34	9.6
cs50e	7 days at -20 °C	0.049	1.81	6.47	9.1
cs51	2hr RT only	0.031	1.55	4.09	9.4
cs51a	15 hr at -20 °C	0.099	1.94	13.07	9.5
cs51b	1 day at -20 °C	0.069	1.97	9.11	10
cs51c	3 days at -20 °C	0.029	1.71	3.83	9.8
cs51d	4 days at -20 °C	0.099	2.02	13.07	9.9
cs51e	7 days at -20 °C	0.038	1.65	5.02	9.9
cs52	2hr RT only	0.035	2.06	4.62	9.5
cs52a	15 hr at -20 °C	0.055	1.72	7.26	9.4
cs52b	1 day at -20 °C	0.05	1.32	6.6	9.7
cs52c	3 days at -20 °C	0.043	1.79	5.68	9.9
cs52d	4 days at -20 °C	0.05	1.79	6.6	9.8
cs52e	7 days at -20 °C	0.054	1.64	7.13	9.7

Table 3.7.2 displays the mean and range of values for A_{260}:A_{280}, yield of RNA and RIN values for each incubation condition. Paired t-tests were performed on the data using statistical software SPSS v12 and the results are displayed in table 3.7.3.

Table 3.7.2: *Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each duration of freezing at -20 °C.*

Incubation conditions	No. of samples (n)	Mean (range) A260:A280	Mean (range) RNA (μg) per 0.5 ml blood	Mean (range) RIN values
2 hours at room temp only	6	1.86 (1.55-2.06)	4.16 (1.72-5.81)	7.8 (1.4-9.6)
15 hours at -20 °C	6	1.89 (1.72-2.0)	8.38 (6.6-13.07)	9.4 (9.2-9.6)
1 day at -20 °C	6	1.83 (1.32-1.97)	6.93 (3.83-9.11)	9.4 (8.3-10)
3 days at -20 °C	5	1.80 (1.71-1.89)	5.73 (3.83-8.32)	9.4 (8.6-9.9)
4 days at -20 °C	6	1.89 (1.78-2.02)	7.09 (1.98-13.07)	8.4 (1.3-10)
7 days at -20 °C	6	1.81 (1.64-2.0)	5.61 (3.56-7.92)	9.4 (8.4-9.9)

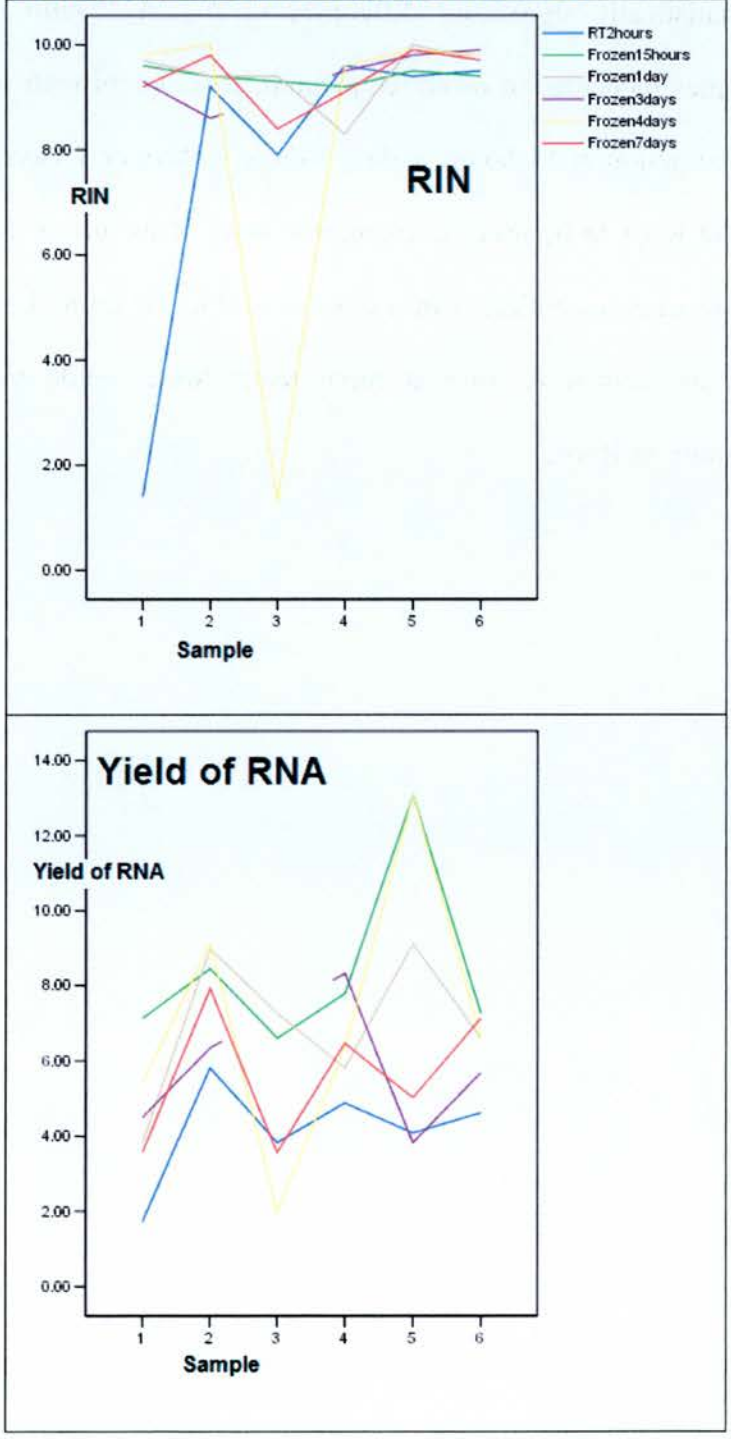
Table 3.7.3: *Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of duration of freezing (p values).*

	A260:A280	Yield of RNA	RIN
2 hours room temp -15 hours frozen	0.784	0.010	0.298
2 hours room temp- 1 day frozen	0.852	0.005	0.307
2 hours room temp – 3 days frozen	0.592	0.095	0.370
2 hours room temp – 4 days frozen	0.783	0.100	0.782
2 hours frozen – 7 days frozen	0.524	0.016	0.285
15 hours frozen – 1 day frozen	0.446	0.128	0.936
15 hours frozen –3 days frozen	0.277	0.142	1.000
15 hours frozen – 4 days frozen	1.000	0.147	0.507
15 hours frozen – 7 days frozen	0.281	0.068	0.825
1 day frozen – 3 days frozen	1.000	0.446	1.000
1 day frozen – 4 days frozen	0.521	0.906	0.509
1 day frozen – 7 days frozen	0.795	0.182	0.808
3 days frozen – 4 days frozen	0.250	0.275	0.210
3 days frozen – 7 days frozen	0.976	0.703	0.692
4 days frozen – 7 days frozen	0.253	0.342	0.476

From these tables it can be seen that there is no statistically significant difference in A_{260}:A_{280} ratio or RIN values for any of the incubation conditions. While not reaching statistical significance, the frozen samples did however seem to have higher RIN values than the samples incubated at room temperature for 2 hours. In addition, when considering the yield of RNA obtained for each condition, the amount of RNA yielded from samples incubated at room temperature for 2 hours was statistically

significantly lower than the frozen samples extracted after 15 hours (p = 0.01), 1 day (p = 0.005) and 7 days (p = 0.16). The RIN values are displayed in figure 3.7.1 along with RNA yields.

Figure 3.7.1: ***Graphs of yield of RNA and RIN values for each incubation condition***



Bioanalyzer tracings along with RIN values for each sample can be found in table 3.7.4 in appendix 2.

Conclusion

There is no statistically significant difference in A_{260}:A_{280} ratio or RIN when comparing samples incubated at room temperature for 2 hours with samples frozen prior to RNA extraction at 15 hours, 1 day, 3 days, 4 days or 7 days. All samples yielded sufficient RNA to be used in microarray work. This means that it would be feasible to freeze samples collected in a clinical setting for up to 1 week, enabling standardization of incubation time at room temperature, while maintaining the quality and quantity of RNA.

3.8: Examination of Storage Time and Storage Temperature

Background

If neonatal blood in PAXgene™ Blood RNA tubes could be frozen and stored for longer periods of time without loss of quality of RNA obtained and without compromising yield, this could have significant benefits for research involving RNA from neonatal blood. This could mean that samples could be stored and RNA samples extracted in batches meaning optimal use of laboratory time. In addition, if samples could be frozen and transported to the laboratory while frozen, they could be gathered at sites remote to the research laboratory. This could also allow samples to be gathered in multiple centres and transferred to the research laboratory. I therefore set out to study a cohort of samples frozen at -20 °C for 1 week and 2 years. In addition, I had frozen a small number of samples at -80 °C which were studied as part of this experiment.

Methods

Samples from eight umbilical cords were used in this section. All of the samples in this section were extracted using the PAXgene™ method described previously (page 76). For each umbilical cord, aliquots of 0.5 ml of cord blood were injected into 2 (cs17, cs18) or 3 (cs45, cs46,cs47, cs48, cs50 and cs52) PAXgene™ Blood RNA tubes and transported directly to the laboratory. For each sample, one aliquot was incubated at room temperature for 2 hours prior to having RNA extracted. For samples cs17 and cs18 and cs45 one aliquot was incubated by immediately placing in a -20 °C freezer and then transferring to a -80 °C freezer. Extraction of these samples

was carried out after transferring to a -20 °C freezer for 4 hours, then transferring to a fridge (4 °C) overnight and then incubating at room temperature for 4 hours. This was to allow freezing and thorough defrosting without damaging the tubes and including the minimum of two hours of incubation at room temperature after defrosting. Cs17 and cs18 were each extracted after 1 year at -80 °C and cs45 after being frozen for 1 week at -80 °C. An additional aliquot of cs45 was placed directly in a -20 °C freezer and left there until taken out for extraction after 2 years. For all of the other samples, two aliquots were placed directly in a -20 °C freezer with one aliquot extracted after 1 week and the other after 2 years. Frozen samples were extracted after being transferred onto the benchtop and incubated at room temperature for 4 hours 30 minutes to allow for thorough defrosting and to include the minimum 2 hours incubation after defrosting.

Results

Table 3.8.1 displays the raw values and table 3.8.2 displays the mean values for A_{260}:A_{280} ratio, RNA yield and RIN values for each incubation condition. When examining this data it can be seen that all of the incubation conditions yielded sufficient RNA for running on microarray. The RIN values on the whole were also good, with only one sample incubated at room temperature for 2 hours having a particularly poor RIN value of 1.4. This sample also had a lower yield and when looking at the Bioanalyzer graph displayed in table 3.8.4 a and b (appendix 2), the trace does not look degraded and the RIN value is probably more a reflection of low yield.

Table 3.8.1: *Raw data for each sample according to storage time and storage conditions.*

	Storage time	Storage temperature	A260:A280 Ratio	A260	RNA (μg)	RIN
cs17	**2hours**	**RT**	1.6	0.072	9.50	8.5
cs17y1	**1 year**	**-80 °C**	2.17	0.052	6.86	7.7
cs18	**2hours**	**RT**	1.5	0.039	5.15	9.5
cs18y1	**1year**	**-80 °C**	2.04	0.047	6.20	9.7
cs45	**2 hours**	**RT**	1.63	0.044	5.81	8.9
cs45w1	**1 week**	**-80 °C**	1.75	0.042	5.54	9.4
cs45y2	**2 years**	**-20 °C**	2.29	0.064	8.45	8.3
cs46	**2 hours**	**RT**	1.69	0.027	3.56	9.7
cs46w1	**1 week**	**-20 °C**	2.15	0.028	3.70	8.9
cs46y2	**2 years**	**-20 °C**	1.88	0.047	6.20	8.6
cs47	**2 hours**	**RT**	1.86	0.013	1.72	1.4
cs47w1	**1 week**	**-20 °C**	1.93	0.027	3.56	9.2
cs47y2	**2 years**	**-20 °C**	2	0.046	6.07	8.6
cs48	**2 hours**	**RT**	1.91	0.044	5.81	9.2
cs48w1	**1 week**	**-20 °C**	2	0.06	7.92	9.8
cs48y2	**2 years**	**-20 °C**	1.78	0.073	9.64	7.9
cs50	**2 hours**	**RT**	1.85	0.037	4.88	9.6
cs50w1	**1 week**	**-20 °C**	1.81	0.049	6.47	9.1
cs50y2	**2 years**	**-20 °C**	1.71	0.058	7.66	8.1
cs52	**2 hours**	**RT**	2.06	0.035	4.62	9.5
cs52w1	**1 week**	**-20 °C**	1.64	0.054	7.13	9.7
cs52y2	**2 years**	**-20 °C**	7	0.021	2.77	8.2

Table 3.8.2*: Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each storage condition.*

Incubation conditions	Number of samples (n)	Mean (range) A260:A280	Mean (range) RNA /0.5 ml blood (μg)	Mean (range) RIN
2 hours RT	8	1.76 (1.5-2.06)	5.13 (1.72-9.5)	8.3 (1.4-9.7)
1 week at -80 °C	1	1.75 (1.75-1.75)	5.54 (5.54-5.54)	9.4 (9.4-9.4)
1 year at -80 °C	2	2.11 (2.04-2.17)	6.53 (6.2-6.86)	8.7 (7.7-9.7)
1 week at -20 °C	5	1.91 (1.64-2.15)	5.76 (3.56-7.92)	9.3 (8.9-9.8)
2 years at -20 °C	6	2.78 (1.78-7.0)	6.80 (2.77-9.64)	8.3 (7.9-8.6)

The A_{260}:A_{280} ratios display a bigger spread of values than in previous experiments.

Some of the samples incubated at room temperature for 2 hours had a lower ratio

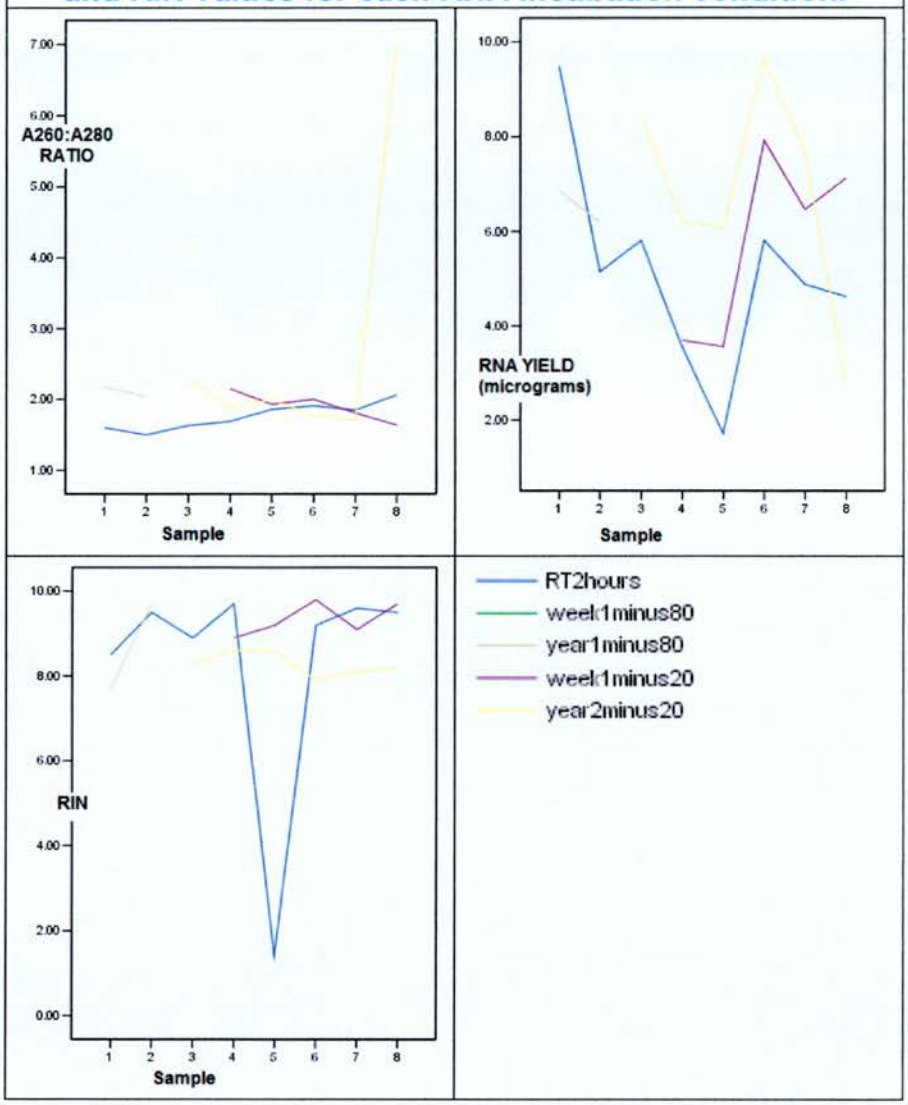
than would be expected given the values obtained in experiments 3.1 to 3.7. It is not clear why this is and may be a feature of biological variation between blood samples. In addition there is one sample that was incubated at -20 °C for 2 years (cs52y2) that was a significant outlier with a ratio of 7. It would therefore be more informative to examine the median rather than the mean for the ratio for the samples incubated at -20 °C for 2 years – this would be 1.94. When this is taken into consideration, the A_{260}:A_{280} ratios are found to be generally acceptable.

Table 3.8.3: ***Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of RNA storage conditions (p values).***

	A260:A280	Yield of RNA	RIN
2 hours at RT – 1 year at -80 °C	0.017	0.741	0.656
2 hours at RT – 1 week at -20 °C	0.832	0.015	0.424
2 hours at RT – 2 years at -20 °C	0.296	0.044	0.874
1 week at -20 °C – 2 years at -20 °C	0.428	0.611	0.022

Table 3.8.3 displays the significance values for comparison of the means of A_{260}:A_{280} ratio, RNA yield and RIN for each incubation condition. The only means that were statistically different for yield were those for incubation at -20 °C for 1 week and for 2 years when compared with 2 hours at room temperature, with the frozen samples yields being greater. There was a statistically significant difference in A_{260}:A_{280} ratio after incubation at room temperature for 2 hours compared with incubation at -80 °C for 1 year (1.76 compared with 2.11). Both of these means are well within the target range of 1.8 to 2.2. All of the mean RIN values were above 8 with the only statistically significant difference being for samples frozen at -20 °C for 1week compared to those at -20 °C for two years. Graphical representations of A_{260}:A_{280} ratio, RNA yield and RIN values for each sample can be found in figure 3.8.1. Bioanalyzer traces and RIN values for each sample can be found in table 3.8.4 (a and b) in appendix 2.

Figure 3.8.1: ***Graphs of A260:A280 values, Yield of RNA and RIN values for each RNA incubation condition.***



Conclusions

Neonatal blood samples in PAXgene™ Blood RNA tubes yield RNA of sufficient quantity and of consistently high quality to run on microarray whether incubated at room temperature for 2 hours, at -20 °C for 1 week, at -20 °C for 2 years or at -80 °C for 1 year. Any statistically significant differences between groups were not of any practical significance, in that all of the samples were suitable to run on microarray. Comparisons between groups in the main were not statistically significant.

This has positive implications for storage and batching of samples. This experiment does not provide any information on the preservation of RNA expression profiles, only on the quality and quantity of RNA obtained. Examination of preservation of RNA expression profiles would require multiple microarray or PCR- based experiments and was beyond the scope of funding for this project.

3.9: Assessment of a Filter- Paper Based Technique

Background

Neonatal venepuncture requires considerable skill and experience to be consistently successful. Neonatologists and parents of babies are generally familiar with the concept of neonatal blood samples being collected on filter paper. This is the method used for the routine newborn screening tests carried out at 5 days of age. Capillary blood samples taken by heel prick are sent off for screening for a series of treatable congenital conditions such as congenital hypothyroidism, phenylketonuria and cystic fibrosis. The general acceptance of this method of blood collection and the ease of the procedure means that almost any baby could have blood sampled in this manner. This made filter paper based collection of whole blood an attractive possibility for gathering RNA samples that could not be ignored. In addition, dried blood spots are already used in PCR based and DNA based examinations, for example in the field of forensic science (321-325).

Whatman FTA® cards are filter-paper type cards that are designed to facilitate nucleic acid isolation from whole blood samples. The cards are chemically treated in order to lyse cell membranes and denature proteins on contact. Nucleic acids remain stable in the card and are protected from microbiological, nuclease, oxidative and UV radiation actions. In addition the cards render infective pathogens inactive on contact. These factors make the cards very appealing for use with clinical samples where nucleic acid isolation is desired. The RNA protocol is described as being suitable for RT-PCR and northern blot analysis in the company literature.

My initial excitement regarding this method was tempered somewhat when I examined the on-line protocol from the company. The protocol, although at first glance fairly straightforward, contained a few inaccuracies and was vague on some points. For example, the section describing the need to use the whole disc of cells to be processed for Northern blot stated that the wash step be "increased to 750 µl sterile water", then it goes on to describe the wash step being in RNA processing buffer. I sought clarification from the company directly by telephone and e-mail and have incorporated their advice into the method described in chapter 2 (page 82).

Aims

To compare quality and quantity of RNA extracted using our preferred method of PAXgene™ blood RNA system with that extracted using Whatman FTA® cards and protocol.

Context

After making enquiries I was sent a trial sample pack of Whatman FTA® cards and punches. This, along with the aliquots of reagents required to make up the RNA processing buffer, meant that I had sufficient materials to carry out 7 experiments. The results of this study would be taken as an indication as to whether it would be worth carrying out further studies of this method.

Methods

Umbilical cord blood was obtained as described previously. From each cord blood sample, aliquots of 0.5 ml of blood were injected into PAXgene™ blood RNA bottles and in addition blood (~125 µl per 1 inch circle) was dropped onto the Whatman FTA® cards from a syringe. As per Whatman®'s advice, puddling of the blood on the cards was avoided and the blood was not rubbed or smeared on the card. The samples were then taken to the lab for processing.

PAXgene™ samples were processed according to the protocol described in chapter 2 after 2 and 24 hours (1st cord sample) and at 2 hours and at 1 week of incubation (2nd cord sample). The 2 and 24 hour incubations were all at room temperature. The sample incubated for 1 week (cs45pw1) was frozen to -80 °C and defrosted as described more fully in section 3.8, with 4 hours incubation at room temperature immediately before RNA extraction.

The samples on the Whatman FTA® cards were left to air dry at room temperature for at least 2 hours (minimum recommended is 1 hour). The Whatman® experiments carried out were:

- cs44W: Whatman® whole 1 inch blood spot
- cs44WX: Whatman® 1 x black (2 mm diameter) Harris punch disc of dried blood spot
- cs44WX2: Whatman® 2 x black Harris punch disc of dried blood spot
- cs44WD: Whatman® whole blood spot plus DNase step.
- cs44W24: Whatman® whole blood spot left 24 hours to dry.

- cs45WX2: Whatman® 2 x black Harris punch disc of dried blood spot
- cs45WX2D: Whatman® 2 x black Harris punch discs of dried blood spot plus DNase step.

The method used for RNA extraction from the Whatman® cards is described in chapter 2 (page 82).

Results

The raw results are displayed in table 3.9.1 and the mean values are displayed in table 3.9.2. Looking at this data, although none of the samples had perfect A_{260}:A_{280} ratios, all of the Whatman® samples had exceptionally poor ratios. The RINs obtained from the PAXgene™ tubes, with the exception of one were good as expected. The RIN values for the Whatman® samples were, on the other hand, uniformly poor at 1. Although these results make the results obtained for yield somewhat inconsequential, the mean yield of the PAXgene™ extracted samples was higher than the mean yield of Whatman® extracted samples. This is true whether looking at Whatman® when the whole disc was used, when 1-2 punches were used and whether or not a DNase step was included.

When the statistical significance values displayed in table 3.9.3 (calculated using paired t-test within the SPSS v 12 software program) are examined it is clear that none of the results achieved in this experiment reached statistical significance. This is due to the fact that only a limited number of Whatman® extractions could be carried out using the trial kit making the numbers examined too small to reach statistical significance.

Table 3.9.1: *Raw data for samples according to RNA extraction method and incubation conditions.*

Sample	Explanation	A260:A280 Ratio	A260	RNA Yield (μg)	RIN
cs44w	**whole disc Whatman®**	0.53	0.094	8.27	1
cs44wd	**whole disc Whatman® and DNase step**	0.38	0.039	3.43	1
cs44wx	**1 punch of Whatman®**	0.96	0.049	4.31	1
cs44w24	**whole disc Whatman® at RT for 24 hours**	1.7	0.034	2.99	1
cs44pax	**PAXgene™ at RT for 2 hours**	1.66	0.048	6.34	n/a
cs44pax24	**PAXgene™ at RT for 24 hours**	1.4	0.028	3.70	8.4
cs45wx	**1 punch of Whatman®**	1.43	0.02	1.76	1
cs45wx2	**2 punches of Whatman®**	1.2	0.036	3.17	1
cs45wx2d	**2 punches of Whatman® and DNAse step**	1	0.018	1.58	1
cs45pax	**PAXgene™ at RT for 2 hours**	1.63	0.044	5.81	8.9
cs45pw1	**PAXgene™ frozen for 1 week**	1.75	0.042	5.54	9.4

Table 3.9.2*: Mean values of A260:A280, RNA yield per 0.5 ml of blood and RIN for each RNA extraction method and incubation condition.*

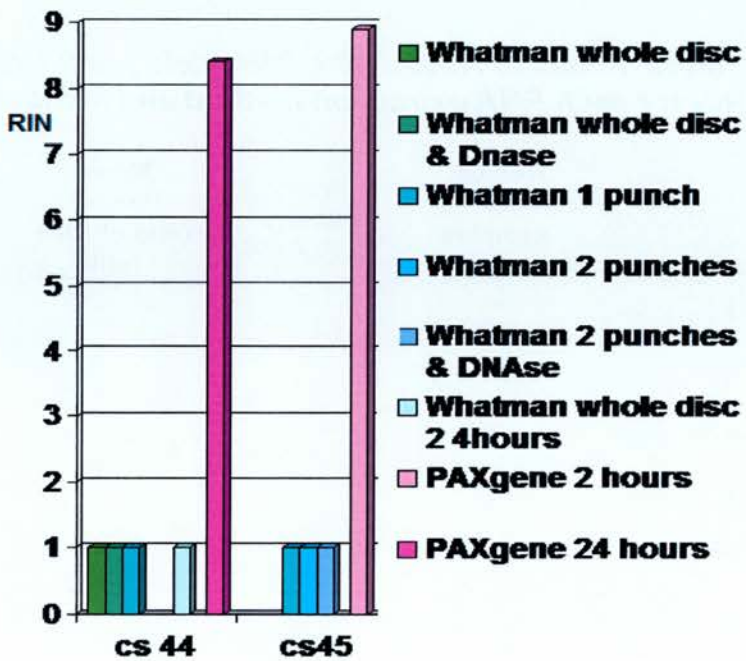
Method and incubation	Number of samples (n)	Mean (range) A260:A280	Mean (range) Yield of RNA (μg)	Mean (range) RIN
PAXgene™ 2 hours	2 (*1 for RIN)	1.65 (1.63-1.66)	6.08 (5.81-6.34)	8.9 (8.9-8.9)*
PAXgene™ any condition	4	1.61 (1.40-1.75)	5.35 (3.70-6.34)	8.9 (8.4-9.4)
Whatman® whole disc	3	0.87 (0.38-1.7)	4.90 (2.99-8.27)	1.0 (1.0-1.0)
Whatman® 1-2 punches	4	1.15 (0.96-1.43)	2.71 (1.58-4.31)	1.0 (1.0-1.0)
Whatman® with DNase step	2	0.69 (0.38-1.00)	2.51 (1.58-3.43)	1.0 (1.0-1.0)
Whatman® any condition	7	1.03 (0.38-1.70)	3.64 (1.58-8.27)	1.0 (1.0-1.0)

Table 3.9.3: ***Values of statistical significance of difference between means of A260:A280 ratio, Yield of RNA per 0.5 ml and RIN for pairs of RNA extraction conditions (p values).***

	A260:A280	Yield of RNA	RIN
PAXgene™ 2 hours – Whatman® DNase	0.209	0.116	N/A
PAXgene™ 2 hours – Whatman® 1 punch	0.323	0.204	N/A
PAXgene™ other – Whatman® with DNAse step	0.096	0.457	0.040
PAXgene™ other - Whatman® 1 punch	0.100	0.602	0.040
Whatman® with DNase step – Whatman® 1 punch	0.094	0.372	N/A

The raw data, the Bioanalyzer graph tracings (table 3.9.4, appendix 2) and the bar chart of RIN values for each extraction condition displayed in figure 3.9.1 are fairly convincing in showing that the quality of RNA obtained using the Whatman® method was uniformly poor.

Figure 3.9.1: ***Bar chart of RIN values for each RNA extraction condition.***



Conclusions

Although use of a filter-paper based technique would be very attractive for use in extracting RNA from neonatal blood, it is clear from this data that the Whatman® based technique was not a viable option for this study. In this experiment it did not yield any good quality RNA at all. The PAXgene™ method remains the method of choice for this work.

Chapter 3: Discussion and Summary of Conclusions

When I embarked upon this work there was no data in the literature regarding extraction of RNA from neonatal blood.

From the work presented in sections 3.1, 3.2 and 3.3 it was seen that the PAXgene™ Blood RNA tube and extraction technique were optimal for RNA extraction from umbilical cord blood from those examined. This concurs with studies using adult blood which showed PAXgene™ tubes do better than heparinised or EDTA tubes (298, 299) and that the PAXgene™ RNA extraction method is better than other methods (327-330). I acknowledge that since I carried out this work there have been other RNA extraction kits brought onto the market for which there are also favourable reports in the literature, e.g the Tempus system (340). My main objective was to find a good, reliable method of RNA extraction which yielded the best possible results at the time. I achieved this aim and as consistency of method is key in RNA work, I did not consider it necessary to consider these newer kits in retrospect.

Having chosen the PAXgene™ system for my study, I then moved on to examine the use of smaller blood volumes than the recommended 2.5 ml that the PAXgene™ tubes are designed for. This was novel work and my results show that high quality RNA of sufficient quantity to run on microarray can consistently be achieved by injecting 0.5 ml of blood directly into the PAXgene™ tubes. This was an important finding in the context of this study because it showed that the use of small neonatal

samples is possible using the PAXgene™ method. More recently, Carrol and colleagues have carried out similar work examining small volumes of blood in a paediatric population (341). They examined, in a similar manner to me, use of aliquots of PAXgene™ fluid with small volumes of paediatric blood (0.3 ml) and found adequate yields and quality for use in RT PCR work (341). My findings that direct injection of small volumes of blood into the PAXgene™ tubes yielded highest quality RNA led us to adopt this approach rather than use aliquots. This had the added advantage that we were able to keep the PAXgene™ fluid enclosed within the tubes while in the clinical area. The PAXgene™ system was the only blood RNA extraction method at the time to have FDA approval for clinical use. This was encouraging because it meant that the tubes were deemed to be safe to use in the clinical setting and this is an important consideration.

I then concentrated on determining optimal incubation times prior to RNA extraction and optimum blood tube storage. These are important parameters to consider in study design and my results were important in deciding how to proceed with the protocol for this study. I found that incubation at room temperature up to 12 hours did not lead to any loss of quality of RNA yielded but that by 24 hours and certainly by 72 hours there were signs of degradation. This is different to published data using adult blood which showed increased yield and quality after 24 hours incubation (342). This could be due to differences in adult compared to neonatal blood and highlighted the importance of studying neonatal blood before setting parameters for future work. My finding that blood samples that were frozen at -20 °C showed no difference in quality from those extracted after 2 hours incubation at room

temperature alone is supported by findings in the literature that adult blood frozen and extracted using PAXgene™ system gives better results than heparinised blood samples (299). In addition, Rainen and colleagues describe freezing as the best method of storage for PAXgene™ blood samples (298).

Given that filter-paper based techniques have been used in RNA work from forensic samples and blood spots (321-325), albeit with pooling of samples and amplification techniques, it was disappointing that the filter-paper based method examined for this study was found not to be suitable for our purposes. With more time and resources it may have been possible to find and refine a filter-paper based technique. Despite this, the work in this chapter identified a robust method of RNA extraction for future work with neonatal blood. I have summarised the findings of this chapter below:

- Of the collection tubes studied for use with umbilical cord blood, the PAXgene™ Blood RNA tube yielded the highest quality RNA and was of sufficient quantity for use in microarray work.
- The PAXgene™ Blood RNA extraction method studied was found to be an optimum method for maximising quality of RNA obtained from umbilical cord neonatal blood. It also consistently gave yields sufficient for microarray use.
- Sufficient quantity of RNA for microarray use can be consistently obtained from 0.5 ml of umbilical cord neonatal blood in the PAXgene™ Blood RNA system. Of the variations studied, injection of 0.5 ml of blood directly into the PAXgene™ tube yielded the highest quality RNA.

- Incubation of umbilical cord blood in PAXgene™ Blood RNA Tubes does not lead to any statistically significant change in quality of RNA obtained for incubation periods of 2 to 12 hours. By 24 hours however, there is statistically significant degradation of RNA quality and this becomes even more apparent by 72 hours.
- There is no statistically significant difference in quality or yield of RNA obtained when comparing umbilical cord blood in PAXgene™ Blood RNA tubes for those incubated at room temperature for 2 hours and for those frozen at -20 °C for 15 hours, 1 day, 3 days, 4 days or 7 days.
- There is no statistically significant difference in quality and no practical difference in yield of RNA obtained from umbilical cord blood in PAXgene™ Blood RNA tubes for those incubated at room temperature and for those frozen at -20 °C for 1 week, -20 °C for 2 years or at -80 °C for 1 year.
- The filter-paper based method of blood RNA extraction examined in this study was shown not to be feasible for use with umbilical cord blood in the conditions studied.

These findings enabled me to move forward to the next phase of the study and gather blood samples from a neonatal population, confident that the RNA obtained should be of high quality and sufficient quantity to use in microarray work. Chapter 4 describes the neonatal samples in detail.

Work from this chapter has subsequently been published in the Analyst:

Claire L Smith, Paul Dickinson, Thorsten Forster, Mizanur Khondoker, Marie Cragion, Alan Ross, Petter Storm, Stewart Burgess, Paul Lacaze, Benjamin J Stenson and Peter Ghazal. Quantitative assessment of human whole blood RNA as a potential biomarker for infectious disease. *Analyst* 2007, **132**, 1200-1209.

A copy of the paper can be found in Appendix 4: (reproduced by permission of The Royal Society of Chemistry) and can also be found at:

http://www.rsc.org/publishing/journals/AN/article.asp?doi=b707122c)

Chapter 4

Examination of Neonatal Blood Sampling

Aims

- To describe the sampling of neonatal blood samples taken for use in the remainder of this study.
- To examine the quality and quantity of RNA obtained from these clinical neonatal samples.
- To examine the effects of the following variables on the quality and quantity of RNA obtained from neonatal samples:
    - Whether samples were taken using a needle or cannula
    - Individual taking the blood sample
    - Whether or not samples were frozen
    - The duration of freezing
    - Time at room temperature prior to RNA extraction
    - Operator carrying out RNA extraction
    - White cell count of samples
    - Neutrophil count of samples

Of the 201 samples described in this chapter:

- 77 were consented for and taken by myself and 124 were taken by 38 clinical colleagues on my behalf (for 1 sample it is not documented who took it).
- 94 samples had the RNA extraction and assessment of quantity and quality of RNA carried out by me, 107 were carried out by two of my laboratory colleagues (Alan Ross and Paul Dickinson) on my behalf.
- All of the clinical data collection and all of the analyses and comparisons in this chapter were done by myself.

Background

Following the umbilical cord blood validation studies, I then moved on to study neonatal blood. The data obtained using cord blood suggested that it should be possible to obtain sufficient quantity and quality of RNA from clinically relevant small samples of blood from newborn infants. It was important however to confirm this in the process of gathering clinical neonatal samples for analysis. It also remained vital to eliminate or reduce as far as possible, any sources of variability in sampling, transport, storage and analysis.

Population

Eligible Babies

All infants who were admitted to the neonatal unit at Simpson Centre for Reproductive Health, or who were being reviewed by the neonatal team on either the postnatal wards or the neonatal unit and who, for clinical reasons, were having blood cultures taken through a needle or cannula were eligible for inclusion in the study. In addition, healthy infants who were having blood samples taken on the neonatal unit for other reasons were eligible as controls. Examples of such other reasons include screening of infants of mothers with thyroid disease, infants of rhesus negative mothers who missed having cord blood screening for blood group, screening for congenital adrenal hyperplasia or infants who were having blood taken as scheduled preterm clinical blood sampling.

Exclusions

In order to meet with laboratory regulations, samples that could be considered "high risk" were excluded. Infants were not included in the study if the mother was known to be positive for hepatitis B, HIV or hepatitis C viruses. In cases where the mother was known to have a history of drug misuse and had not had antenatal screening for blood borne viruses, the infants were also excluded. Other exclusion criteria were infants who did not require clinical blood samples and infants for whom extra blood sampling might be of particular risk, e.g. infants with an underlying disorder causing anaemia.

Consent

All samples were taken with informed parental consent prior to sampling. Parents of eligible infants were approached and given both verbal and written information regarding the study. The parents were given the opportunity to ask any questions regarding the study. If, after hearing the explanation and reading the information sheet, the parents were satisfied that they had sufficient time to make their decision and signed a consent form, the infant was considered recruited and blood sampling was attempted at the time of clinical blood sampling.

Methods

Neonatal Blood Sampling

The volume of blood sampled for the study in each case was between 0.5 – 1 ml. These values were chosen in order to display the natural variation in sampling volume that occurs clinically where precision measuring is not possible and at the same time reflecting the need to keep sample volumes as small as possible in this population. The circulating blood volume of a newborn is 75-80 ml per kg body weight. In general, for consistency, I aimed for samples to be around 0.6 – 0.7 ml in volume.

No extra venepuncture was performed for the sake of the study; extra blood was taken at the same time as clinical sampling only. If, after the required clinical samples were obtained, the needle or cannula did not bleed back or was dislodged, it was not re-sited for the sake of the study.

All samples for blood culture were taken according to the neonatal unit policy as much as possible. The sampling technique is described in detail in chapter 2 (page 62).

From late 2006 onwards, clinical colleagues began sharing in the consent and sampling of neonatal samples. As the project was progressing we felt it was important that we looked at ways to expand the project and potentially recruit other centres. We therefore began routinely freezing the PAXgene™ blood RNA tubes at

-20 °C immediately after sampling and when ready to extract the RNA, the samples were taken out of the freezer and left at room temperature for 4 hours prior to RNA extraction.

Review of Clinical Data

Clinical data was gathered for each recruited infant and recorded on a standardised data sheet. Data was gathered from the infants' paper hospital records (including medical, nursing and prescription charts), the infants' computerised hospital records (Badger system), the infants' computerised laboratory results (Apex system) and maternal paper hospital maternity records.

RNA Extraction

RNA extraction was carried out according to the protocol in chapter 2 (page 76) using the PAXgene™ blood RNA tubes and reagent kits. For the earlier samples that were not frozen, the PAXgene™ blood RNA tubes were left upright at room temperature until transported to the laboratory in a sealed plastic container and the RNA extraction process commenced. In order to ensure that the samples were given at least 2 hours in the stabilisation agent, the first centrifugation was aimed to take place between two and a half and four and a half hours after sampling to reflect timing consistency within plausible clinical parameters. In practice this was achieved for the majority of samples but a handful of samples were left longer than this before processing (longest gap 16 hours 15 minutes) – almost exclusively on days where there were two or more babies recruited meaning that a choice had to be made between commencing RNA extraction on the first sample within the target

timeframe and missing an eligible study subject or collecting both samples and being outwith the target time frame for the earlier sample.

After 2006 when the samples were frozen at -20 °C immediately following sampling, the samples were taken in batches to the lab, on ice to prevent thawing and transferred to a -20 °C freezer in the laboratory until RNA extraction could take place. The RNA was extracted, usually taking batches of 6 samples which, when taken from the freezer, were left at room temperature for 4-4.5 hours prior to commencing the first centrifugation step. All of the samples except three were extracted before 5 hours 20 minutes at room temperature. Three samples were left to defrost overnight which was not following the protocol as had been intended.

Assessment of RNA quality and quantity

Extracted RNA was frozen at -80 °C until assessment of RNA quality and quantity could be carried out in batches. RNA quantity was assessed using spectrophotometry – either using a ThermoSpectronic Biomate 5 spectrophotometer and VISION 32 (bit Version 1.25) software or using a Thermo Scientific NanoDrop™ Spectrophotometer with software v3.5.2 as described in chapter 2 (pages 89 and 92). RNA quality was assessed using A_{260}:A_{280} ratios obtained from the spectrophotometry and from both RIN values and subjective assessment of the tracings obtained using an Agilent 2100 Bioanalyzer (RNA 6000 Nano Assay) as described in chapter 2 (pages 93 to 98).

Statistical Testing

For sections 4.2 to 4.9, differences between groups were examined by using SPSS v12 statistics software to carry out Kruskal-Wallis testing. ANOVA using the same software was carried out for any conditions that appeared to show statistical significance on Kruskal-Wallis testing.

During my discussion of this section I refer to statistical significance and practical significance. Statistical significance was defined as p value <0.05. The best possible quality and higher yields of RNA are always desirable. Practically however, RNA will be suitable for use in microarray work as long as it is of sufficient yield and high enough quality. Sufficient quality was felt to be indicated by RIN values of 8 and above and A_{260}:A_{280} ratios within the range 1.8 to 2.2. If the minimum amount of RNA required for input into the CodeLink™ microarray protocol is taken at 200 ng (in a maximum of 10.8 µl), then the minimum yield required in 75 µl of eluted RNA is 1.39 µg. The yield I have taken as being sufficient for use on array was therefore 1.39 µg. Provided that the samples met these conditions there would be no practical consequences for differences in actual values – they would all be able to be run on array, i.e. any differences in value would not be of practical significance.

Chapter 4: Results

4.1: Overall Yield and Quality of RNA

Table 4.1.1a: *Raw RNA Yield and Quality Data for Neonatal Blood Samples*

Sample	RIN	A_{260}:A_{280} Ratio	RNA yield (µg)	Sample Volume (ml)	RNA/0.5 ml of sample (µg)
csb1	10	1.95	9.77	0.8	6.11
csb2	9.2	1.90	7.79	0.8	4.87
csb3	1	2.20	1.45	0.5	1.45
csb4	9.1	1.84	4.62	0.6	3.85
csb5	10	1.80	4.75	0.7	3.39
csb6	10	1.86	5.15	0.6	4.29
csb7	10	1.71	6.34	0.5	6.34
csb8	9.5	2.19	4.62	0.6	3.85
csb9	1.1	2.80	1.85	0.6	1.54
csb10	9.8	2.03	10.16	ns	-
csb11	6.6	1.76	13.73	ns	-
csb12	9.2	1.80	4.75	0.5	4.75
csb13	10	1.72	8.18	0.75	5.46
csb14	9.9	1.76	6.73	0.5	6.73
csb15	6.6	1.81	14.78	0.8	9.24
csb16	9.5	1.64	4.75	0.8	2.97
csb17	8.5	1.73	3.43	0.6	2.86
csb18	8.3	1.95	9.77	0.7	6.98
csb19	9.5	1.98	14.39	0.5	14.39
csb20	8.6	1.70	2.24	0.5	2.24
csb21	9.5	1.50	5.54	0.8	3.47
csb22	8.8	1.85	3.17	0.6	2.64
csb23	9.1	2.02	14.39	0.75	9.59
csb24	9.2	1.98	14.39	0.6	11.99
csb25	9.6	2.04	6.73	0.6	5.61
csb26	8.7	2.05	15.18	0.7	10.84
csb27	9.1	2.25	3.56	0.7	2.55
csb28	9.4	2.06	4.88	0.6	4.07
csb29	8.2	1.33	7.39	0.65	5.69
csb30	8.4	1.91	5.81	0.7	4.15
csb31	na	4.50	1.19	0.65	0.91
csb32	6.8	2.50	3.30	0.7	2.36
csb33	9.9	2.47	4.88	0.6	4.07
csb34	10	2.27	7.79	0.6	6.49
csb35	10	2.15	9.37	0.6	7.81
csb36	10	2.31	3.96	0.65	3.05
csb37	9.1	1.40	1.84	0.6	1.53
csb38	9.7	1.98	11.75	0.6	9.79
csb39	9.8	1.94	15.54	0.65	11.95
csb40	7.2	1.71	3.17	0.65	2.44
csb41	8.9	1.58	2.51	0.6	2.09
csb42	8.5	1.80	2.38	0.65	1.83
csb43	9.4	1.24	2.77	0.6	2.31
csb44	9.9	1.84	9.24	0.65	7.11

Table 4.1.1b: *Raw RNA Yield and Quality Data for Neonatal Blood Samples*

Sample	RIN	A_{260}:A_{280} Ratio	RNA yield (µg)	Sample Volume (ml)	RNA/0.5 ml of sample (µg)
csb45	4.8	1.83	18.61	0.75	12.41
csb46	8.8	1.53	3.04	0.65	2.34
csb47	8.9	1.68	4.88	0.6	4.07
csb48	1.7	1.27	2.51	0.45	2.79
csb49	9	1.65	4.36	0.55	3.96
csb50	8.4	1.79	9.24	0.6	7.70
csb51	9.2	2.00	2.64	ns	-
csb52	na	2.00	1.32	ns	-
csb53	9.4	2.01	21.78	0.8	13.61
csb54	8.2	2.00	9.24	0.7	6.60
csb55	9.4	2.00	7.13	0.75	4.75
csb56	na	Error	0.26	0.5	0.26
csb57	7.4	2.00	5.81	0.75	3.87
csb58	9.3	2.00	2.11	0.7	1.51
csb59	9.5	2.08	3.30	2	0.83
csb60	9.4	1.94	16.63	1	8.32
csb61	9.3	1.47	3.70	0.5	3.70
csb62	9.4	1.73	9.11	0.9	5.06
csb63	9.1	1.77	7.00	1	3.50
csb64	9.2	1.96	13.20	1	6.60
csb65	9.3	1.80	5.94	1	2.97
csb66	4.2	1.77	3.04	0.6	2.53
csb67	9.4	1.61	7.00	1	3.50
csb68	9.4	1.83	20.06	1	10.03
csb69	9.6	1.55	6.34	0.8	3.96
csb70	8.8	1.70	6.73	1	3.37
csb71	7.2	1.71	9.50	ns	-
csb72	9.6	1.63	8.58	ns	-
csb73	9.6	1.94	4.36	1	2.18
csb74	8.3	1.99	43.69	1	21.85
csb75	8.7	1.91	5.81	0.7	4.15
csb76	9	2.01	30.23	0.5	30.23
csb77	2.4	1.93	3.56	0.5	3.56
csb78	1	1.00	4.29	ns	-
csb79	9.3	1.87	5.68	1	2.84
csb80	9.8	1.33	24.16	0.5	24.16
csb81	8.8	1.81	4.79	0.9	2.66
csb82	na	1.13	2.81	0.75	1.87
csb83	na	2.45	100.06	0.7	71.47
csb84	na	2.14	309.94	0.6	258.28
csb86	8.8	1.29	2.97	1	1.49
csb87	10	1.83	15.68	1	7.84
csb88	na	2.36	173.05	0.9	96.14
csb89	9.6	1.90	20.96	1	10.48
csb91	9.4	2.00	7.95	1	3.98
csb92	1.1	1.86	0.21	0.5	0.21
csb93	na	1.87	0.56	0.6	0.47
csb94	9.3	1.78	1.94	0.6	1.62
csb95	8.6	2.06	10.29	0.8	6.43
csb96	8.6	2.04	3.92	0.7	2.80

Table 4.1.1c: *Raw RNA Yield and Quality Data for Neonatal Blood Samples*

Sample	RIN	A_{260}:A_{280} Ratio	RNA yield (µg)	Sample Volume (ml)	RNA/0.5 ml of sample (µg)
csb97	7.9	2.01	3.03	0.7	2.16
csb98	2.2	1.77	0.46	0.5	0.46
csb99	na	2.63	0.98	0.6	0.82
csb100	8.7	2.61	0.74	ns	-
csb101	9.1	2.08	5.98	0.5	5.98
csb102	9.1	2.05	7.06	ns	-
csb103	9.1	2.06	4.50	1	2.25
csb104	2.1	2.56	0.84	0.75	0.56
csb105	na	2.14	5.06	0.7	3.61
csb106	na	1.72	0.4	0.5	0.40
csb107	5	1.92	4.43	0.75	2.95
csb108	5.9	1.98	2.60	0.5	2.60
csb109	8.9	2.21	5.34	0.6	4.45
csb110	8.4	1.99	10.31	0.5	10.31
csb111	2.5	2.33	1.1	0.5	1.10
csb112	7.1	2.03	2.7	0.6	2.25
csb113	9.1	2.04	8.88	0.8	5.55
csb114	7.9	2.02	5.38	1	2.69
csb115	7.8	2.19	2.37	1	1.19
csb116	9.7	2.09	7.46	0.8	4.66
csb117	na	2.87	0.84	0.7	0.60
csb118	na	2.14	10.69	0.5	10.69
csb119	na	2.09	3.96	0.5	3.96
csb120	na	3.15	0.45	0.5	0.45
csb121	9.4	2.03	25.3	0.8	15.81
csb122	9.1	2.06	8.17	0.8	5.11
csb124	8.9	2.27	1.86	0.6	1.55
csb125	na	2.12	13.83	0.75	9.22
csb126	8.8	2.00	4.84	0.8	3.03
csb127	na	3.38	0.59	0.5	0.59
csb128	8.8	2.08	32.73	1	16.37
csb129	na	2.12	10.57	1	5.29
csb130	9.6	2.23	4.57	0.5	4.57
csb131	4.7	2.05	1.80	0.6	1.50
csb132	9.5	1.94	3.15	0.8	1.97
csb133	8.8	2.02	6.55	1	3.28
csb134	8.9	1.89	3.84	0.6	3.20
csb137	9.3	2.00	11.14	1	5.57
csb138	9.3	1.95	4.50	0.7	3.21
csb139	9.1	2.11	12.73	0.5	12.73
csb140	9.2	2.00	21.96	1	10.98
csb141	7.6	2.76	1.55	0.6	1.29
csb142	8.9	1.93	5.89	1	2.95
csb143	9.6	2.10	14.94	ns	-
csb145	na	1.66	0.78	0.5	0.78
csb146	na	1.87	1.33	0.5	1.33
csb147	8.9	2.01	8.67	0.6	7.23
csb148	8.7	2.28	4.39	1	2.20
csb149	9.6	2.88	1.57	1	0.79
csb150	8.9	2.35	5.93	0.75	3.95

Table 4.1.1d: *Raw RNA Yield and Quality Data for Neonatal Blood Samples*

Sample	RIN	A_{260}:A_{280} Ratio	RNA yield (µg)	Sample Volume (ml)	RNA/0.5 ml of sample (µg)
csb151	na	3.08	0.93	0.7	0.66
csb152	8.7	2.34	4.41	0.5	4.41
csb153	na	1.70	0.17	0.5	0.17
csb154	na	0.26	0.07	0.5	0.07
csb155	9.2	2.21	7.39	0.7	5.28
csb156	8.9	2.10	7.52	0.6	6.27
csb157	9.4	2.56	1.83	0.4	2.29
csb158	9.1	2.15	4.83	0.6	4.03
csb159	9.4	2.19	7.34	0.6	6.12
csb160	9.3	2.21	4.75	0.6	3.96
csb161	9.1	2.40	6.08	1	3.04
csb162	8.5	2.75	2.38	0.6	1.98
csb163	8.7	2.25	4.52	1	2.26
csb164	na	3.19	1.52	0.5	1.52
csb165	9.2	2.29	9.15	0.7	6.54
csb166	na	0.33	0.07	0.5	0.07
csb167	8.8	2.67	1.84	0.9	1.02
csb168	na	1.09	0.49	ns	-
csb169	3.3	2.14	10.48	0.8	6.55
csb170	9.1	2.08	27.35	1	13.68
csb171	7.9	2.44	2.77	1	1.39
csb172	8.7	2.05	1.28	1	0.64
csb173	9.2	2.09	22.14	1	11.07
csb174	9.2	2.08	5.40	0.5	5.40
csb175	7.8	2.37	2.45	0.8	1.53
csb176	8.8	2.06	7.05	1	3.53
csb177	9.6	2.09	16.61	1	8.31
csb178	8.5	2.20	3.30	1	1.65
csb179	9.2	2.31	4.48	0.8	2.80
csb180	9	2.11	14.54	1	7.27
csb181	9.4	2.16	11.79	1	5.90
csb182	9.4	2.38	2.95	0.6	2.46
csb183	9.6	2.12	17.12	0.6	14.27
csb184	9.7	2.15	11.24	0.5	11.24
csb185	9.3	2.05	30.32	ns	-
csb186	9.6	2.08	10.12	0.6	8.43
csb187	9.4	2.13	8.33	1	4.17
csb188	9.3	2.10	16.48	0.8	10.30
csb189	9.4	2.10	12.29	0.5	12.29
csb190	9.9	2.15	9.15	1	4.58
csb191	6.5	2.42	3.34	0.7	2.39
csb192	8.5	2.14	6.55	1	3.28
csb193	9.4	2.14	8.18	0.5	8.18
csb194	8	2.33	3.00	ns	-
csb195	7.4	2.05	15.55	0.7	11.11
csb196	9.2	2.08	23.55	ns	-
csb197	9.4	2.07	11.81	1	5.91
csb198	9.6	2.91	1.55	0.5	1.55
csb199	na	6.47	0.58	1	0.29
csb200	9.3	2.19	5.86	1	2.93

Table 4.1.1e: *Raw RNA Yield and Quality Data for Neonatal Blood Samples*					
Sample	**RIN**	**A_{260}:A_{280} Ratio**	**RNA yield (µg)**	**Sample Volume (ml)**	**RNA/0.5 ml of sample (µg)**
csb201	Na	5.46	0.61	0.5	0.61
csb202	8.8	2.17	6.23	0.8	3.89
csb203	7.2	2.40	2.14	0.7	1.53
csb204	9.3	2.07	23.82	1	11.91
csb205	8.4	2.17	3.46	0.8	2.16
csb206	8	2.38	2.88	0.8	1.80
csb207	9.1	2.15	6.08	0.8	3.80

The raw data of the 201 neonatal blood samples obtained are displayed in table 4.1.1 (a to e). Samples are numbered csb1-csb207. There are 6 numbers that do not have a sample – csb85, csb90, csb123, csb135, csb136, csb144. Four of these tubes were removed from the sampling box but never found – presumed discarded without sampling. Two of the tubes had blood samples taken but unfortunately the accompanying data sheet was not filled in and therefore these samples could not be identified and therefore could not be used. Immediate efforts were taken to avoid such a situation being repeated.

Of the 173 samples assigned a RIN value, RIN ranged from 1 to 10. The mean RIN was 8.5 and the median 9.1 (table 4.1.2). Of the 28 samples who were not assigned a RIN value (marked as "na" in the RIN column in table 4.1.1), subjective examination of the Bioanalyzer traces was carried out. The Bioanalzyer traces can be found in table 4.1.3 (a - j) in appendix 3. This examination revealed that 5 of the traces were too small on which to make any judgement (csb120, csb146, csb154, csb166, csb168), 2 traces were small but looked poor quality (csb117, csb127) and 21 traces were small but were suggestive of good quality with two clear peaks (csb31, csb52,

csb56, csb82, csb83, csb84, csb88, csb93, csb99, csb105, csb106, csb118, csb119, csb125, csb129, csb145, csb151, csb153, csb164, csb199, csb201).

200 of the samples had A_{260}:A_{280} ratio values. The remaining sample had a measured A_{280} value of 0 so calculation of a ratio was not possible. The A_{260}:A_{280} ratio values ranged from 0.26 to 6.47 (table 4.1.2) with a mean of 2.08 and median of 2.04.

Table 4.1.2: *Overall Quality and Yield of RNA Obtained from Neonatal Blood Samples*

	RIN	A_{260}:A_{280} Ratio	Yield (µg)	Yield per 0.5 ml of blood (µg)
Mean	8.5	2.08	10.13	7.07
Median	9.1	2.04	5.38	3.7
Range	1-10 (28 "na")	0.26-6.47 (1 na)	0.07-309.94	0.07-258.14 (12 did not state vol., n=189)
Percentage of samples sufficient for microarray:			**88.1%**	

The actual RNA yield obtained from the samples ranged from 0.07 µg to 309.94 µg with a mean of 10.13 µg and a median of 5.38 µg. 177 of the 201 samples (88.1 %) yielded sufficient RNA to take forward for microarray analysis.

4.2: Yield and Quality of RNA: Sampling With Needle or Cannula.

The vacuum based blood sampling system recommended for use with the PAXgene™ is unsuitable for use in neonates as the vacuum applied to small neonatal veins would cause the vein to collapse. At the time of this study there was no data in the literature regarding alternative sampling methods. Blood sampling in neonates is usually carried out by slowly aspirating blood into a syringe from a needle or cannula or by dripping blood directly into a blood tube. The dripping method is not suitable for PAXgene™ tubes due to the reagents contained in it being harmful. It was therefore important both to confirm that sampling into a syringe and then injecting into the PAXgene™ tubes yielded RNA of sufficient quantity and quality, and to identify if RNA quality or quantity was different depending on whether needle or cannula was used for sampling. Sampling is sometimes carried out through heparinised lines but this was generally avoided where possible for this study,

Of the 201 samples in the study 56 were sampled from a cannula, 132 from a needle, 6 from another non-heparinised line (5 from a peripheral arterial line at the time of insertion and one from a non-heparinised long-line) and 2 from heparinised lines. For 5 samples it was not stated which means of sampling was used.

Table 4.2.1 displays the mean, median and range of values for RIN, A_{260}:A_{280} ratio and yield for each method of sampling. All modes of sampling are presented in this table for completeness but I will comment only on needle and cannula. The mean RIN for samples taken with a needle was 8.1 (median 9.1) and for those taken from a

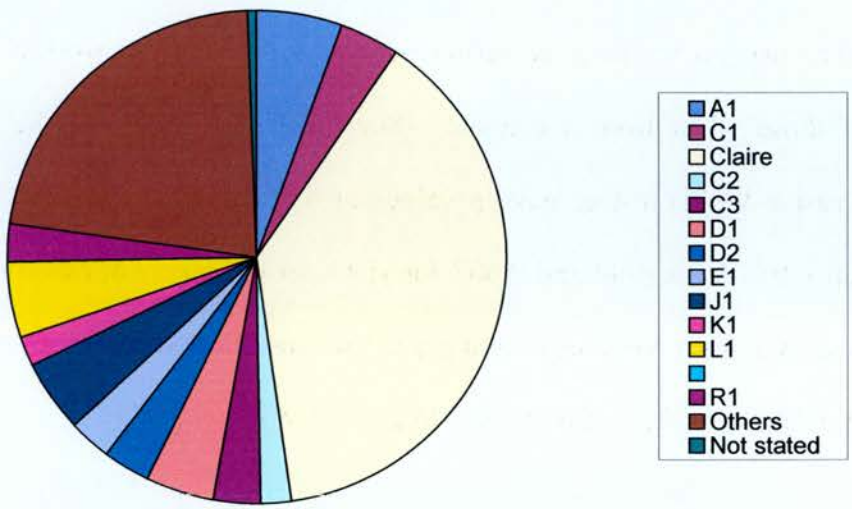
cannula was 8.7 (median 9.1). The mean (and median) values for A_{260}:A_{280} for samples taken by needle and those by cannula were well within the target range of 1.8-2.2 with values of 1.92(1.94) and 2.17(2.08) respectively. 89.3 % of samples taken using a needle gave sufficient yield for microarray work compared to 87.9 % of those taken from a cannula. Statistical comparison of the two groups using Kruskal-Wallis testing gave p values of 0.752 for RIN values, 0.006 for A_{260}:A_{280} ratio, 0.006 for yield and 0.007 for yield per 0.5 ml. On carrying out ANOVA for A_{260}:A_{280} ratio, yield and yield per 0.5 ml statistical significance did not persist with significance values of 0.062, 0.346 and 0.418 respectively.

Table 4.2.1: ***Mean, Median and Range of RIN, A260:A280 Ratio and Yield of RNA for Each Mode of Sampling.***

		Cannula (n=132)	Needle (n=56)	Other line (non-heparinised) (n=6)	Heparinised Line (n=2)
RIN	**Mean**	8.7	8.1	7.8	9.3
	Median	9.1	9.1	9.0	9.3
	Range	2.1-10	1-10	1.1-9.8	8.5-10
A_{260}:A_{280} Ratio	**Mean**	2.17	1.92	2.00	1.99
	Median	2.08	1.94	1.91	1.99
	Range	0.26-6.47	0.33-2.91	1.33-2.75	1.83-2.14
Yield (µg)	**Mean**	12.59	5.00	7.72	11.12
	Median	6.03	3.74	3.40	11.12
	Range	0.07-309.94	0.07-23.82	0.21-24.16	6.55-15.68
	% Sufficient Yield for Microarray	87.9 %	89.3 %	83.3 %	100 %
Yield per 0.5ml blood (µg)	**Mean**	8.78	3.15	6.50	5.56
	Median	4.07	2.73	3.20	5.56
	Range	0.07-258.28	0.07-11.91	0.21-24.16	3.28-7.84

4.3: Yield and Quality of RNA: Clinician Taking the Sample.

Figure 4.3.1: ***Proportion of samples taken by each clinician***



In order for any investigation to be useful clinically it is important that sampling can be done effectively by clinicians at the point of clinical need. It was therefore important to examine the quality and quantity of RNA obtained from sampling by multiple clinicians. Overall there were 200 samples taken by 39 clinicians (1 sample was excluded as sampler not stated). Figure 4.3.1 displays the proportion of samples taken by each clinician. The mean, median and range of volume sampled, RIN, A_{260}:A_{280} ratio and yield according to clinician sampling are displayed in table 4.3.1. Any clinician who took less than 4 samples was included in the "other" row.

Looking at these results, with the exception of sampler R1, the RNA obtained was generally of good quality as determined by RIN, with the lowest mean RIN being 7.8 and lowest median being 8.8. Mean and median A_{260}:A_{280} ratios were generally within the acceptable range of 1.8 to 2.2 with a few outliers.

Table 4.3.1: *RIN, A260:A280 Ratio and Yield of RNA According To Clinician Sampling*

Clinician	Volume per sample (ml) Mean Median (Range)	RIN Mean Median (Range)	A260:A280 Ratio Mean Median (Range)	RNA Yield (µg) Mean Median (Range)	% of samples with yield sufficient for microarray
Claire (n=77)	0.65 0.6 (0.35-1.0) 3ns	8.5 9.2 (1-10) 3na	1.96 2.00 (0.33-2.8)	7.08 5.15 (0.07-23.82)	94.8 %
A1 (n=11)	0.73 0.7 (0.6-1.0) 2ns	8.9 9.1 (7.8-9.5) 4na	2.31 2.19 (1.94-3.08)	41.87 3.15 (0.74-309.94)	72.7 %
C1 (n=8)	0.83 0.9 (0.5-1.0) 1ns	8.9 8.8 (7.9-9.6) 1na	1.77 1.82 (1.29-2.04)	6.18 4.36 (0.4-20.06)	87.5 %
C2 (n=4)	1.0 1.0 (1.0-1.0) 1ns	8.6 8.8 (7.2-9.6)	1.85 1.85 (1.71-1.99)	20.03 15.23 (5.94-43.69)	100 %
C3 (n=6)	0.88 1.0 (0.6-1.0)	9.1 8.9 (8.8-9.3) 1na	2.07 2.05 (2.0-2.21)	12.84 9.62 (6.55-32.73)	100 %
D1 (n=9)	0.78 0.85 (0.5-1.0) 1ns	9.0 9.2 (7.2-9.7)	2.14 2.1 (2.05-2.40)	11.43 11.24 (2.14-30.32)	100 %
D2 (n=6)	0.66 0.6 (0.5-1.0)	7.8 8.9 (4.7-9.8) 1na	1.92 1.96 (1.33-2.35)	6.57 3.48 (0.56-24.16)	83.3 %
E1 (n=6)	0.77 0.80 (0.6-0.9)	9.1 9.0 (8.8-9.7) 2na	2.14 2.12 (2.0-2.36)	34.11 6.4 (4.84-173.05)	100 %
J1 (n=9)	0.75 0.75 (0.5-1.0) 1ns	9.2 9.3 (8-10) 2na	2.52 2.13 (1.47-5.46)	11.06 8.33 (0.61-30.23)	88.9 %
K1 (n=4)	0.69 0.63 (0.5-1.0)	8.7 8.7 (8.7) 3na	2.41 2.2 (1.87-3.38)	5.04 2.86 (0.59-13.83)	50 %
L1 (n=10)	0.91 0.85 (0.5-2.0)	8.0 9.2 (2.2-9.5)	1.88 1.85 (1.73-2.08)	6.55 6.41 (0.46-16.63)	90 %
R1 (n=5)	0.74 0.75 (0.5-1.0)	4.4 3.8 (2.1-7.9) 1na	2.34 2.33 (1.92-2.87)	2.52 1.1 (0.84-5.38)	40 %
Other (n=45: 27 samplers, each with <4 samples)	0.74 0.7 (0.5-1.0) 2ns	8.6 9.1 (2.4-9.6) 10na	2.22 2.08 (0.26-6.47) 1na	6.12 4.5 (0.45-21.96)	82.2 %

There was considerable variance in the yield obtained. Two clinicians (K1 and R1) had only 50 % and 40 % of their samples respectively yielding sufficient RNA for microarray. This is notably different to the samples from all of the other clinicians which yielded sufficient RNA 72.7 to 100 % of the time.

Using Kruskal-Wallis testing to examine differences between samplers, there was no significant difference in RIN between samplers with significance value of 0.306, there was a significance value of 0.036 for yield and a significance value of less than 0.001 for A_{260}:A_{280} ratio. On carrying out ANOVA statistical significance was not seen to persist for A_{260}:A_{280} ratio with significance value of 0.064 but the significance value for yield was 0.011.

4.4: Yield and Quality of RNA: Samples Not Frozen or Frozen.

Table 4.4.1: *RIN, A260:A280 Ratio and Yield of RNA According to Samples Being Not Frozen or Frozen*

	RIN Mean Median (Range)	A260:A280 Ratio Mean Median (Range)	RNA Yield (µg) Mean Median (Range)	% of samples with yield sufficient for microarray
Not Frozen (n= 60)	8.6 9.2 (1-10) 3na	1.94 1.94 (1.24-4.5) 1 na	6.88 5.02 (0.26-412.86)	95 %
Frozen (n=141)	8.4 9.1 (1-10) 25na	2.13 2.08 (0.26-6.47)	68.01 5.98 (0.07-853.60)	85.1 %

Comparison of yield and quality of RNA depending on whether a sample was frozen or not prior to extraction was carried out. For both groups the mean and median values for A_{260}:A_{280} were well within the optimal range of 1.8-2.2. The mean and median values of RIN were also good for both groups with the frozen group having mean RIN of 8.4 and median of 9.1 compared to 8.6 and 9.2 for the not frozen group. Although the mean and median values of yield for the frozen samples looked to be greater, the proportion of samples yielding sufficient RNA for microarray was greater for the samples not frozen. When differences between the groups were examined using Kruskal-Wallis testing, there was no significant difference either in RIN with significance value of 0.079 or in yield with a significance value of 0.773. Comparison of mean A_{260}:A_{280} ratio between groups revealed significance value of less than 0.001 when comparing groups. On carrying out ANOVA for A_{260}:A_{280}, the statistical significance was found to be 0.036.

4.5: Yield and Quality of RNA: Duration of Freezing.

Table 4.5.1: *RIN, A260:A280 Ratio and Yield of RNA According to Duration of Freezing*

	RIN Mean Median (Range)	A260:A280 Ratio Mean Median (Range)	RNA Yield (µg) Mean Median (Range)	% of samples with yield sufficient for microarray
0-7 days (n=23)	7.65 9.5 (1-9.9) 3na	1.97 1.96 (1.00-2.91)	9.0 5.68 (0.56-43.69)	91.7 %
8-14 days (n=24)	9.1 9.3 (7.2-10) 3na	2.03 2.03 (1.63-3.15)	10.38 8.38 (0.40-30.32)	91.7 %
15-21 days (n=17)	9.2 9.3 (8.6-9.8) 4na	1.93 2.06 (1.13-3.38)	9.37 6.34 (0.59-32.73)	88.2 %
22-28 days (n=5)	9 9 (8.5-9.5) 1na	1.87 2.06 (1.09-2.2)	3.99 3.30 (0.49-7.05)	80 %
29-35 days (n=12)	8.6 9 (7.1-9.2) 3na	2.18 2.13 (1.66-2.61)	40.02 5.74 (0.74-309.94)	83.3 %
>5-6 weeks (n=4)	6.3 8.4 (1.1-9.3) 1na	2.12 2.12 (1.86-2.36)	50.14 13.64 (0.21-173.05)	75 %
>6-7 weeks (n=1)	9.3 9.3 (9.3)	2.19 2.19 (2.19)	5.86 5.86 (5.86)	100 %
>7-8 weeks (n=2)	5.6 5.6 (3.3-7.9)	2.08 2.08 (2.02-2.14)	7.93 7.93 (5.38-10.48)	100 %
>8-9 weeks (n=7)	7.6 8.8 (2.5-9.6) 1na	2.88 2.33 (2.05-6.47)	4.25 1.84 (0.58-16.61)	57.1 %
>9-10weeks (n=7)	9 9.2 (8.5-9.4) 2na	2.94 2.56 (2.07-5.46)	4.55 2.38 (0.61-11.81)	85.7 %
>10-11weeks (n=7)	6.9 8 (2.1-9.4)	2.21 2.19 (1.92-2.56)	4.54 4.43 (0.84-7.52)	85.7 %
>11-12 weeks (n=3)	8.3 8.5 (7.8-9.1)	2.22 2.15 (2.14-2.37)	4.61 4.83 (2.45-6.55)	100 %
>12 -13 weeks (n=4)	9 9 (8.7-9.2) 2na	1.63 1.95 (0.26-2.34)	3.01 2.29 (0.07-7.39)	50 %
>3-6 months (n=11)	8.9 8.9 (7.8-9.7) 5na	2.19 2.19 (0.33-3.08)	3.51 2.37 (0.07-8.67)	63.6 %
>6 months (longest 41 weeks 2 days) (n=13)	8.6 8.9 (4.7-9.6)	2.13 2.06 (1.93-2.76)	9.17 5.89 (1.55-25.3)	100 %

Of the samples that had been frozen, more specific comparison of yield and quality of RNA obtained depending on duration of freezing that was then carried out. 140 samples are examined in this section. It was not possible to calculate the duration of the remaining one sample from the sampling information that was documented.

Table 4.5.1 displays the mean, median and range of RIN, A_{260}:A_{280} ratio and yield of RNA for each period of freezing. For most durations of freezing both the median and mean values for RIN were good at over 8. Four durations of freezing (0-7 day, 5-6 weeks, 8-9 weeks and 10-11 weeks) had lower mean RIN values (6.3-7.65) but median RIN values of over 8. The samples frozen for between 7 and 8 weeks had poorer RIN values with both mean and median values of 5.6. With the exceptions of samples frozen for between 8 and 10 weeks which had slightly high mean and median A_{260}:A_{280} ratios and the 12-13 week group which had a low mean ratio at 1.63, the other groups showed acceptable A_{260}:A_{280} ratios (target range 1.8-2.2). Yield varied between groups with the percentage of samples sufficient for microarray ranging from 50-100 %. The lowest percentage of sufficient yields occurred after 12-13 weeks freezing and the highest after 6-8 week, 11-12 weeks and over 6 months.

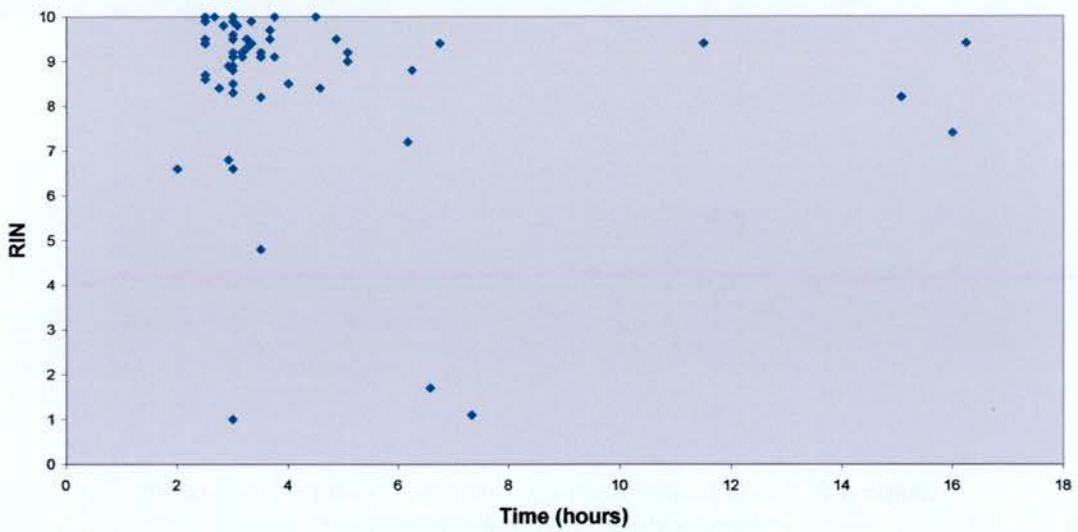
Kruskal-Wallis testing between groups showed no significant difference in RIN with significance value of 0.107 and no significant difference between groups with respect to yield with significance value of 0.107. The significance value for difference between groups in terms of A_{260}:A_{280} ratio was 0.004. When ANOVA was carried

out to look at differences between the groups in terms of A_{260}:A_{280} ratio the significance value was 0.007.

4.6: Yield and Quality of RNA: Time at Room Temperature Prior to RNA Extraction.

Yield and quality of RNA were examined depending on time of incubation at room temperature prior to RNA extraction. Samples were looked at separately depending on whether they were frozen or not.

Graph 4.6.1: ***RIN against incubation time at room temperature for samples not frozen***

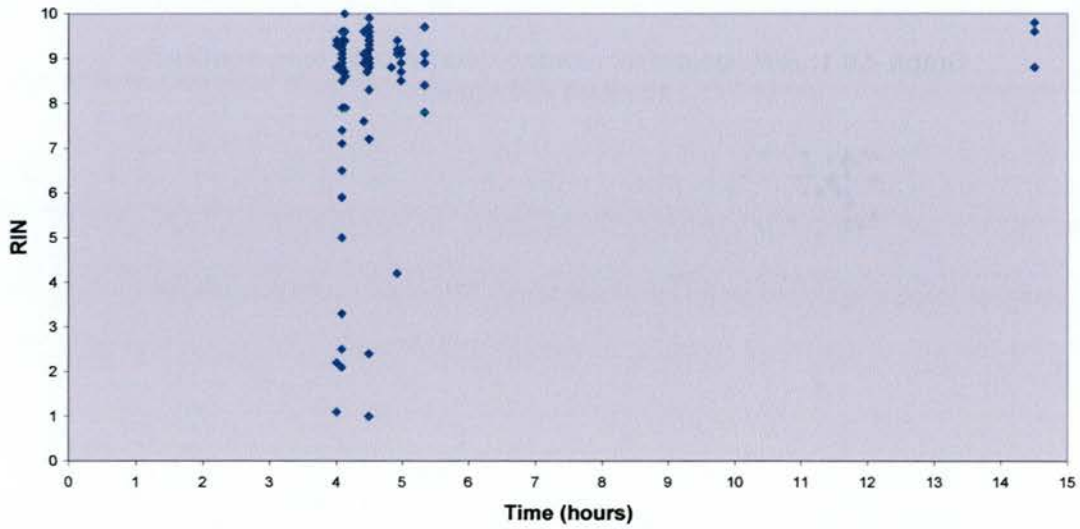


On examination of graphs 4.6.1 and 4.6.2, plotting RIN against incubation time for samples that were not frozen and were frozen respectively, there is no clear pattern relating incubation time to RIN value. In both graphs, it can be seen that the majority of samples have good RIN values and that those at lower RIN do not seem to occur at a particular duration of incubation.

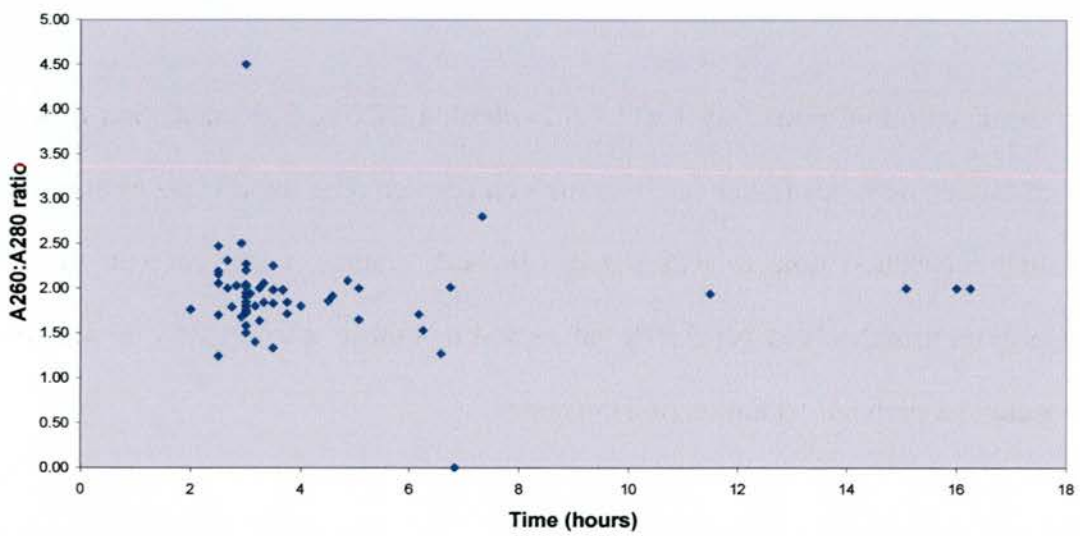
Graphs 4.6.3 and 4.6.4 plot A_{260}:A_{280} ratio against incubation time at room temperature prior to RNA extraction for samples that were not frozen and frozen respectively. There is no clear pattern relating A_{260}:A_{280} ratio to incubation time. Of

note, there is a greater spread of A_{260}:A_{280} ratio values for samples that were frozen but this does not seem to be related to any particular incubation periods.

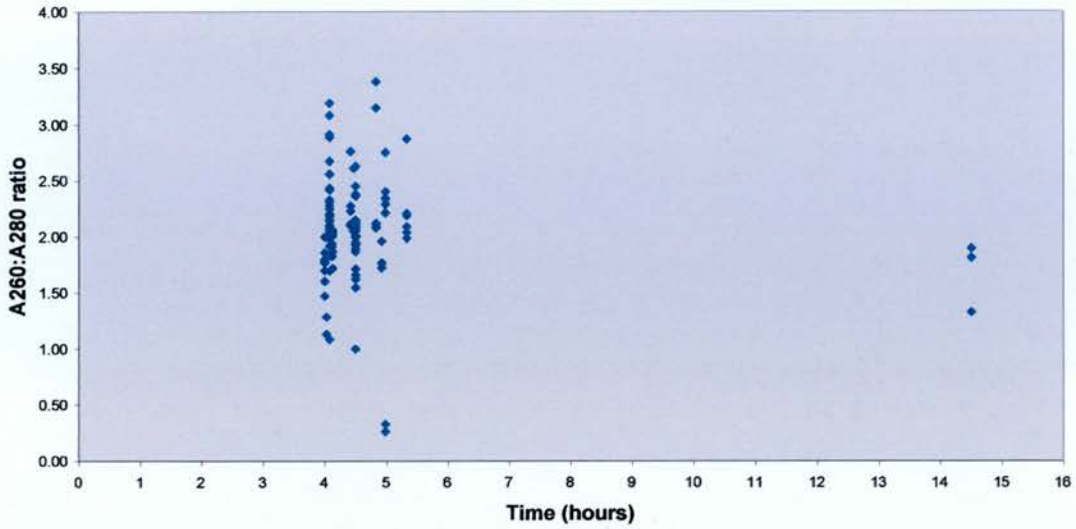
Graph 4.6.2: ***RIN against incubation time at room temperature of samples that were frozen***



Graph 4.6.3: ***A260:A280 ratio against incubtation time at room temperature of samples not frozen***

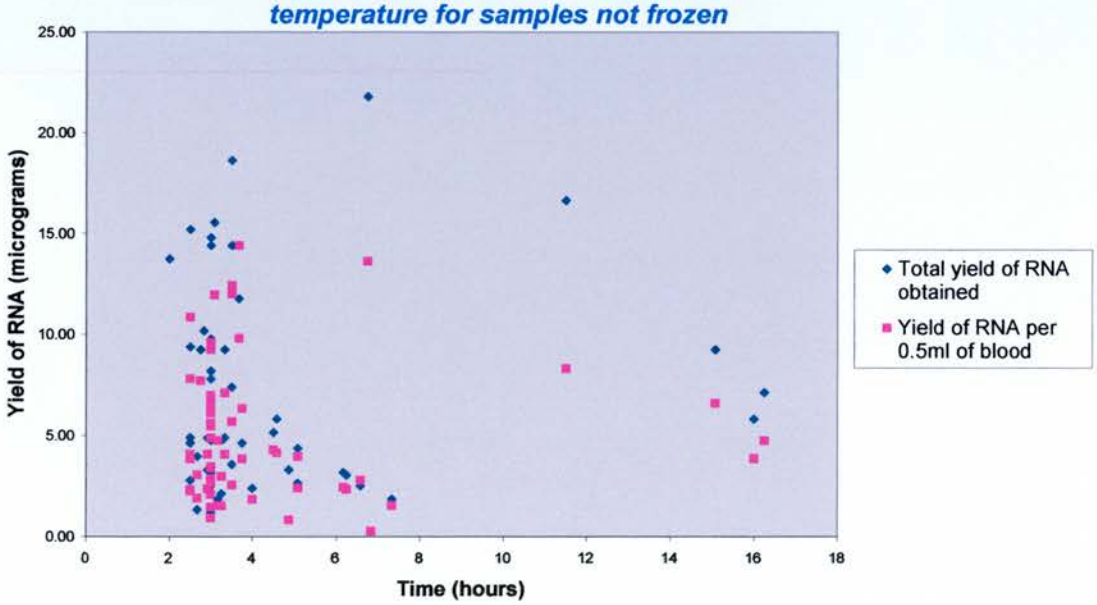


Graph 4.6.4: ***A260:A280 ratio against incubation time at room temperature for samples that were frozen***

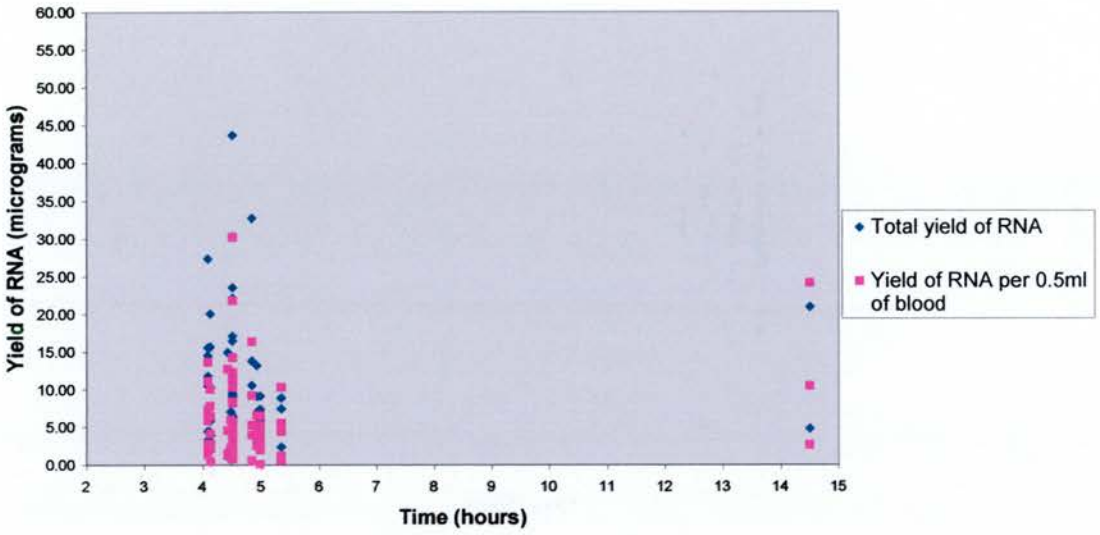


Looking at graphs 4.6.5 and 4.6.6 which plot total yield (blue diamond) and yield per 0.5 ml of blood (pink square) against incubation time at room temperature for samples that were not frozen and that were frozen respectively, there is not any clear pattern of yield in relation to duration of incubation.

Graph 4.6.5: ***Yield of RNA against incubation time at room temperature for samples not frozen***

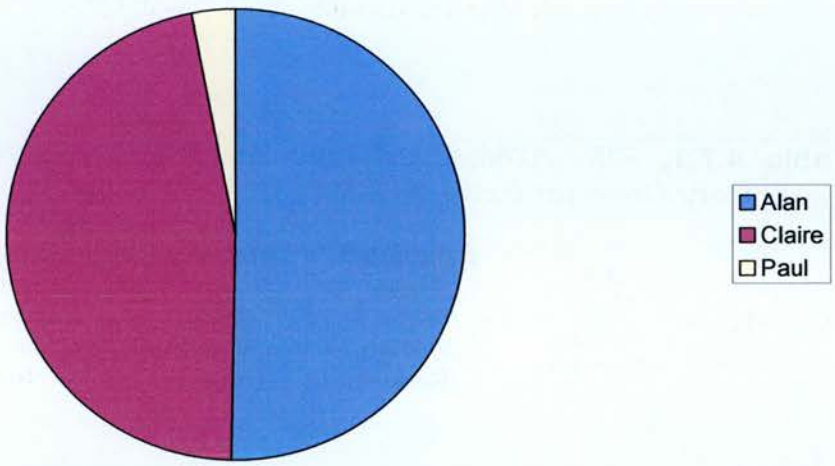


Graph 4.6.6: ***Yield of RNA against incubation time for samples that were frozen***



4.7: Yield and Quality of RNA: Operator Carrying Out RNA Extraction.

Figure 4.7.1: ***Proportion of RNA Extractions by Each Laboratory Operator***



Consistency of RNA extraction method is vital and it is also important for any method to be reproducible between laboratory workers. Yield and quality of RNA obtained from three different laboratory workers (myself, Alan Ross and Paul Dickinson) was therefore examined (Fig 4.7.1). The PAXgene™ Blood RNA extraction protocol on page 76 was always used. The mean (and median) RIN values for each operator were similar at 8.3 (9.1) for Claire, 8.5 (9.1) for Alan and 8.8 (8.9) for Paul. All three operators had mean and median A260:A280 ratios within the desired range of 1.8 to 2.2. 89.4 % of samples extracted by Claire gave sufficient yields for microarray compared to 87.1 % of those extracted by Alan and 83.3 % of those extracted by Paul.

On Kruskal-Wallis testing, there was no significant difference in RIN between operators with significance value of 0.881. Neither was there significant difference between operators with respect to overall yield with significance value of 0.961 nor

with respect to yield per 0.5 ml of blood with significance value of 0.69. The significance value for difference between operators in terms of A_{260}:A_{280} ratio was 0.002. When ANOVA was carried out to look at differences between the operators in terms of A_{260}:A_{280} ratio the significance value was 0.225.

Table 4.7.1: ***RIN, A260:A280 Ratio and Yield of RNA According to Laboratory Operator Extracting RNA***

Operator	RIN Mean Median (Range)	A260:A280 Ratio Mean Median (Range)	RNA Yield (µg) Mean Median (Range)	% of samples with yield sufficient for microarray	RNA Yield (µg) per 0.5 ml blood Mean Median (Range)
Claire (n=94)	8.3 9.1 (1-10) 16na	2.00 1.98 (1.13- 4.5) 1na	12.53 4.88 (0.26-309.94)	89.4 %	9.21 3.85 (0.26-258.28) 4ns
Alan (n= 101)	8.5 9.1 (1-10) 11na	2.14 2.09 (0.26-6.47)	8.15 5.86 (0.07-43.69)	87.1 %	5.15 3.53 (0.07-30.23) 8ns
Paul (n=6)	8.8 8.9 (7.8-9.7) 1na	2.23 2.14 (1.99-2.87)	5.87 6.4 (0.84-10.31)	83.3 %	4.46 4.56 (0.60-10.31)

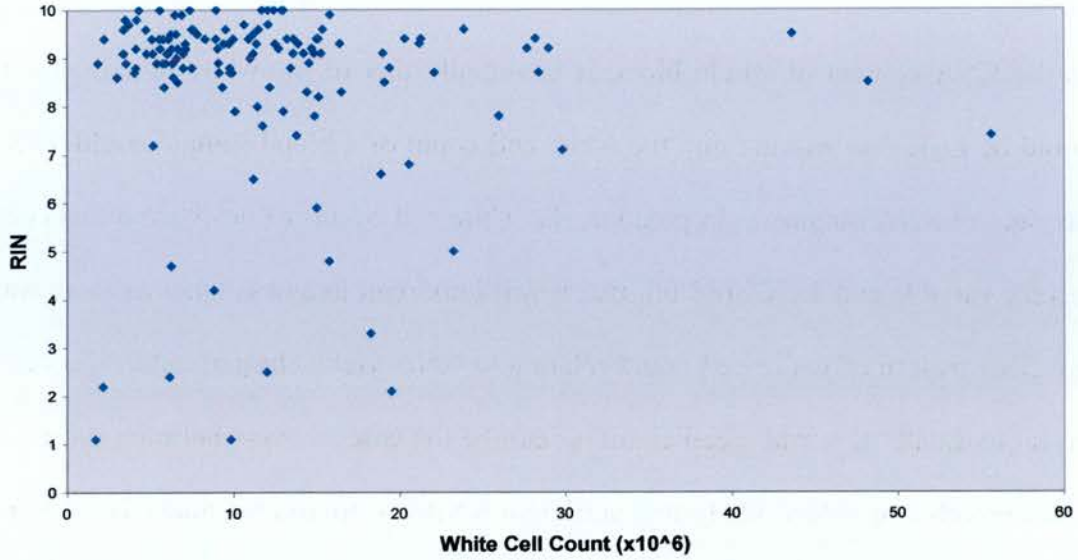
4.8: Yield and Quality of RNA: White Cell Count of Samples

As the RNA content of whole blood is essentially that of the white blood cells, it would be logical to assume that the white cell count of a blood sample could affect the yield of RNA obtained. In practice, the white cell counts of newborn infants can be very variable and therefore I felt that it was important to assess whether there was any clear pattern of white cell count relating to RNA yield. In particular, if a baby has an unusually low white cell count, as can be the case in overwhelming sepsis, is it still possible or indeed likely that sufficient RNA for microarray analysis could be obtained?

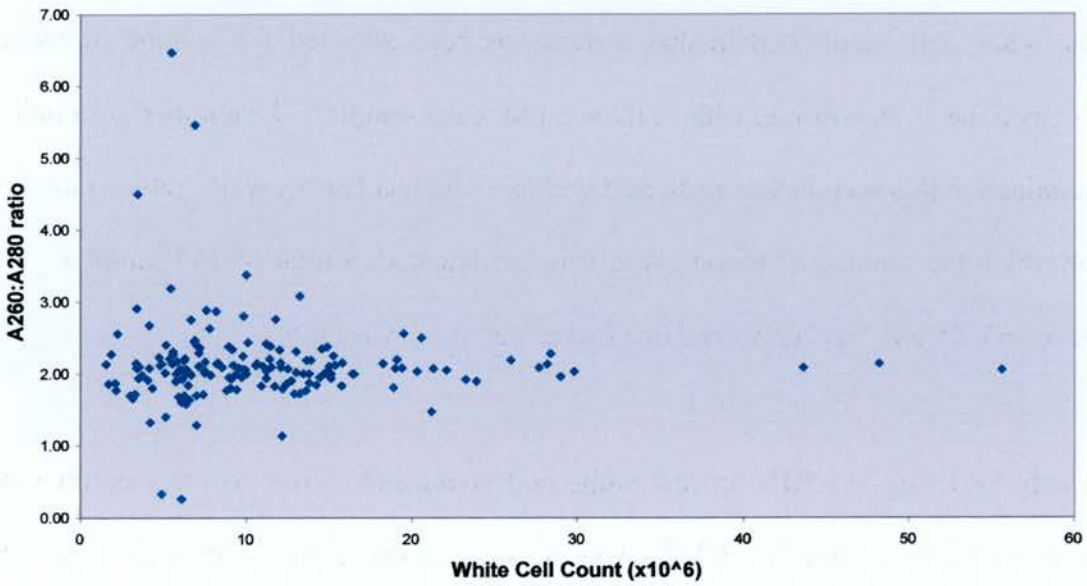
The white cell count used in this section has been adjusted for volume of blood obtained, i.e. is an absolute white cell count for each sample. The number of samples examined in this section was restricted to those who had known white cell counts and for which the volume of blood taken was documented, a total of 163 samples. Of these 163, 25 had "na" RINs and one had a "na" A_{260}:A_{280} ratio.

Graph 4.8.1 displays RIN against white cell count and shows no clear correlation between the two. Graph 4.8.2 displays A_{260}:A_{280} ratio against white cell count. It shows that the majority of samples are within the desirable range of 1.8 to 2.2.

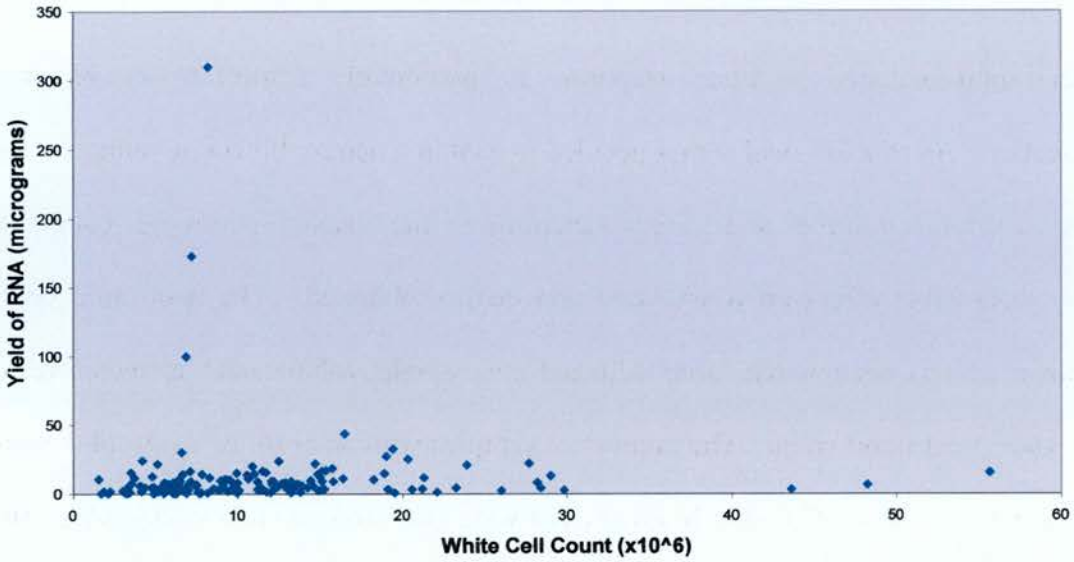
Graph 4.8.1: ***RIN against white cell count***



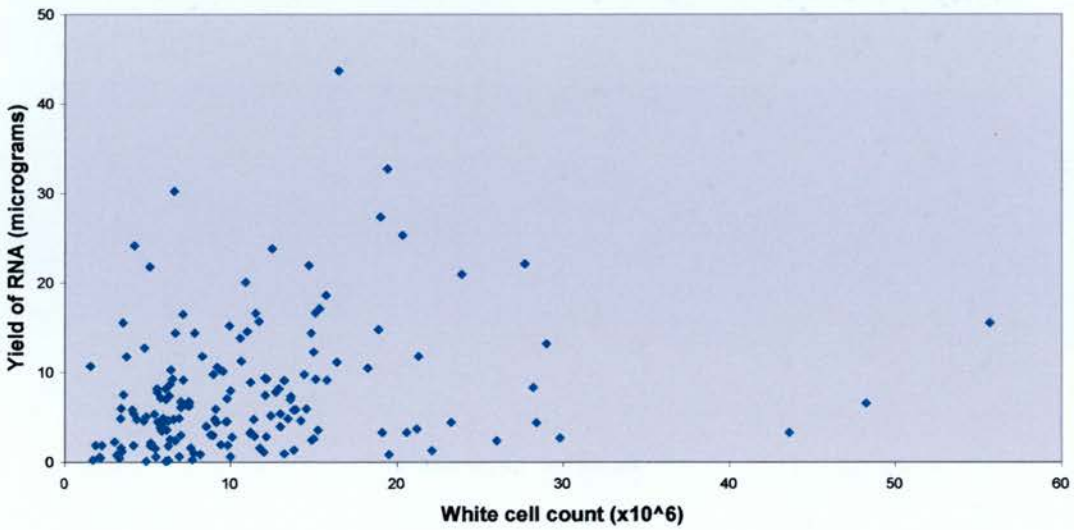
Graph 4.8.2: ***A260:A280 ratio against white cell count***



Graph 4.8.3: ***Yield of RNA against white cell count***



Graph 4.8.4: ***Yield of RNA against white cell count (magnified for yield under 50 micrograms)***

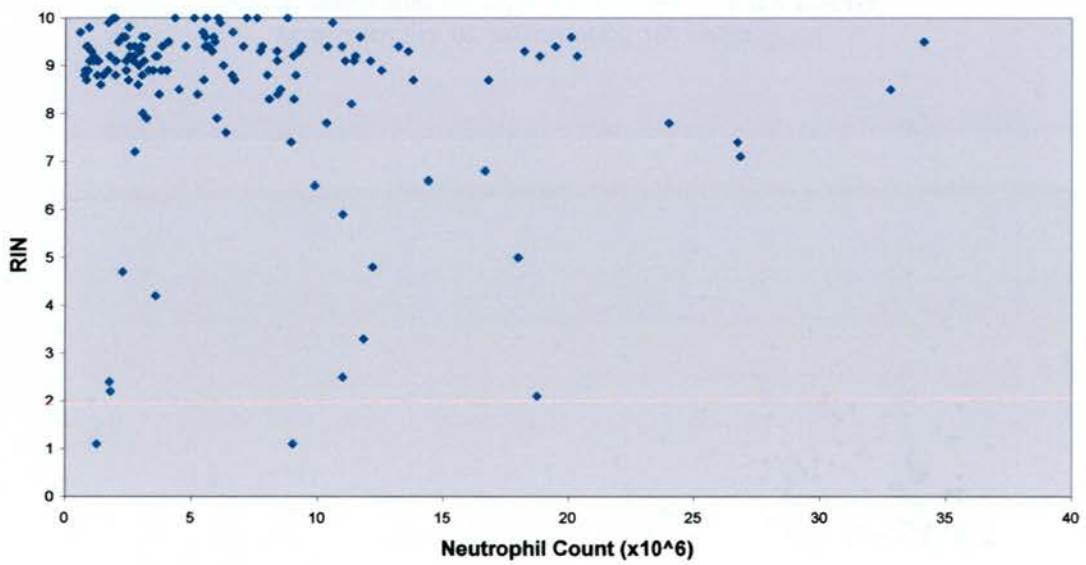


Graphs 4.8.3 and 4.8.4 plot yield of RNA against white cell count. Both graphs display the same data but with 4.8.4 magnifying the yield values of under 50 µg. In other words this includes all except three of the values from 4.8.3.

4.9: Yield and Quality of RNA: Neutrophil Count of Samples.

Neutrophil-mediated immune response is particularly important in bacterial infection. As an additional step, I decided to examine neutrophil count values for the neonatal blood samples and try and determine if the absolute neutrophil count has any discernible effect on RNA yield and quality obtained. The neutrophil count shown in this section has been adjusted for sample volume and is therefore an absolute neutrophil count. The number of samples with an absolute neutrophil count was 163. Of these, 25 had RIN values that were "na" and one had an A_{260}:A_{280} ratio that was "na".

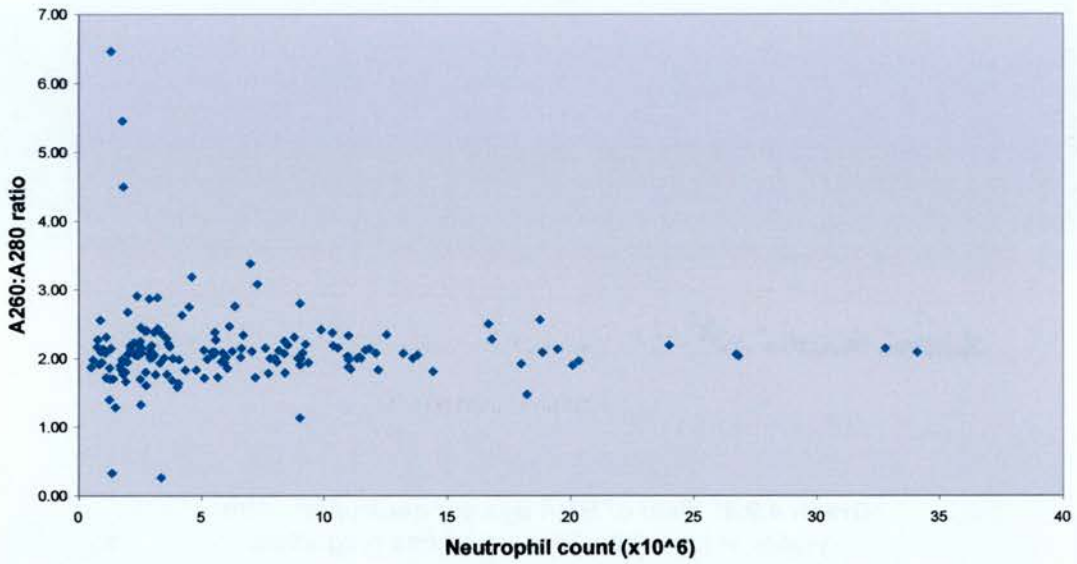
Graph 4.9.1: ***RIN against neutrophil count***



Graph 4.9.1 plots RIN against neutrophil count and does not show a clear relationship between the two factors. Lower RIN values are found over a wide range

of neutrophil counts, not just those that were low. Most samples with low neutrophil counts had good RIN values.

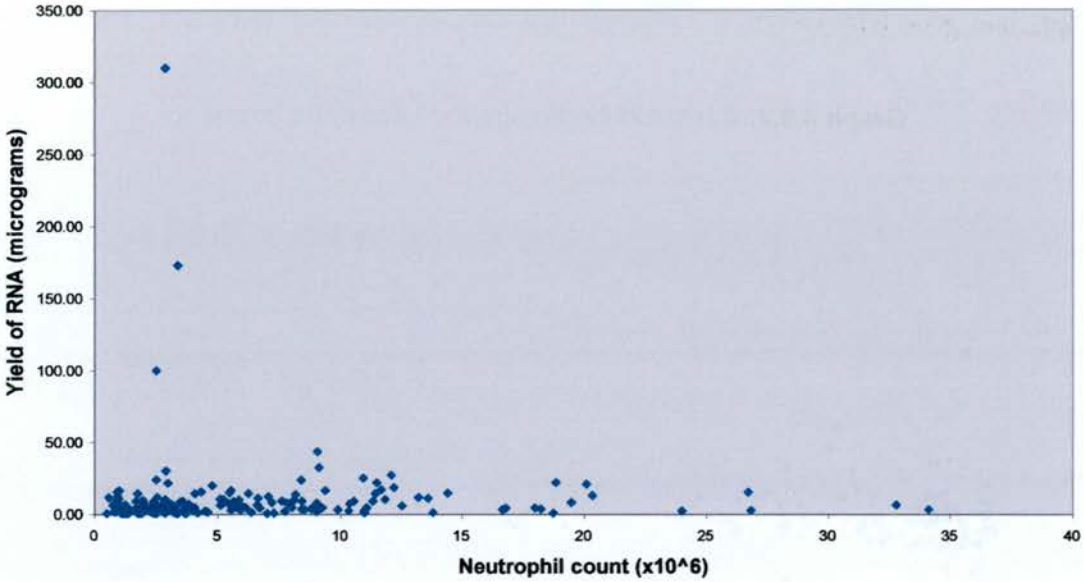
Graph 4.9.2: ***A260:A280 ratio against neutrophil count***



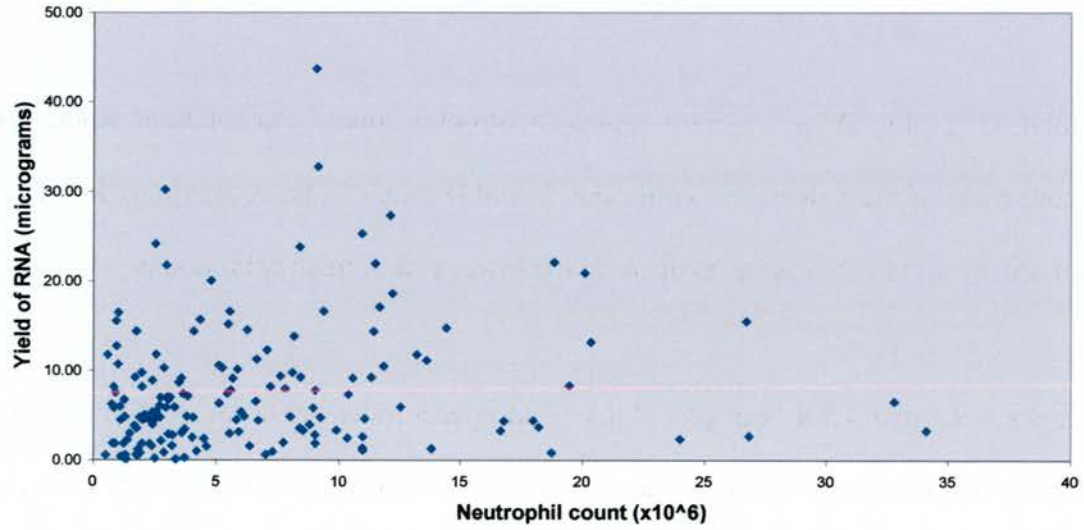
Graph 4.9.2 plots A_{260}:A_{280} ratio against neutrophil count. As for total white cell count, most of the values are within the desirable range. The A_{260}:A_{280} ratio values that are most extreme appear to occur at the lower end of neutrophil count.

Graphs 4.9.3 and 4.9.4 both plot RNA yield against neutrophil count. Graph 4.9.4 displays identical data points to Graph 4.9.3 but magnifies yield of less than 50 µg. This excludes only 3 of the data points found in graph 4.9.3. There is a suggestion that higher neutrophil counts may trend towards higher yields of RNA but there are many samples where this was not the case.

Graph 4.9.3: ***Yield of RNA against neutrophil count***



Graph 4.9.4: ***Yield of RNA against neutrophil count (yields of less than 50 micrograms magnified)***



Chapter 4: Discussion

As discussed in chapter 3, at the time of this work, there was no data in the literature regarding the use of neonatal whole blood for use in microarray work. Neonatal blood composition differs from that of adults, particularly with respect to white cell count and there are well established reference ranges for the newborn (343). In addition, neutrophil counts tend to be lower in very low birth weight infants and separate reference values have been defined (344). These differences in cell composition along with the absolute requirement of small sample volumes meant that validation work was essential. From the work in chapter 3 using umbilical cord blood I could be confident that use of small volumes of neonatal blood for microarray work would be feasible and therefore neonatal blood sampling commenced. It was however, still important to take stock and examine the yield and quality of RNA obtained from the neonatal samples. Although umbilical cord blood is a good surrogate for neonatal blood, the samples used in chapter 3 were all taken from term or near-term births. The neonatal population being studied for the microarray work includes a large proportion of preterm and growth restricted infants who may have lower neutrophil counts at birth (344). In addition, the sampling from the neonatal population was not restricted to the time of delivery so age-related changes in blood composition may have an affect on yields of RNA obtained.

It was therefore reassuring to confirm in 4.1 that high quality RNA of sufficient quantity for microarray use was obtained in the majority of cases sampled. Examination of tables 4.1.1 and 4.1.2 revealed that RIN values ranged from 1 to 10 with the median and mean values good, at 8.5 and 9.1 respectively. It was also

reassuring that on examination of the bioanalyzer traces (appendix 3) for the samples with no RIN, many of the traces looked to have two crisp peaks but only very small. In other words, it appears likely that if many of the "na" bioanalyzer traces were repeated with concentrated samples that a higher RIN would be obtained. In addition, subjective examination of 9 further samples with RIN of 2.5 and under, showed that two samples (csb3, csb78: both with RIN of 1) were too small to make any further comment on and the other seven (csb9, csb48, csb77, csb92, csb98, csb104, csb111) were small but with two crisp peaks, suggesting that if more concentrated may have had higher RIN values. The mean and median A_{260}:A_{280} ratios were also both within the target range at 2.08 and 2.04 respectively. There were a few notable outliers of A_{260}:A_{280} values however and on closer examination, the majority of these were also very small samples that were insufficient to run on array. It is possible that these smallest samples could be generating aberrant ratios when measured on the spectrophotometer because small changes in measurement of either value could make a large difference to the ratio. Overall, I was very pleased with the quality of RNA obtained. The yield of RNA obtained ranged from 0.07 to 309.94 μg with 177 of the 201 samples (88.1 %) being sufficient to run on microarray. Overall, although this is a high percentage, I was disappointed with this level of success with the yield as this would mean that 1 in 10 samples taken was not sufficient. For interest, I calculated the overall yield of RNA per 0.5 ml of blood sampled in each case in order to allow comparison with the yields obtained from the umbilical cord blood studies in chapter 2. From the clinical samples 163/189 samples (86.2 %) yielded sufficient RNA from 0.5ml for microarray analysis. This compares with 97.1 % of the cord blood samples (0.5ml) from chapter 3.

I suspect that this difference may be due to over-estimation of the volume of blood sampled in many clinical cases and that a proportion of the samples were less than 0.5 ml. Unfortunately I did not weigh the samples to verify sample volume prior to RNA extraction and if I were to do the study again I would weigh each sample to verify the volume of blood in each tube before RNA extraction. Alternatively, the wider range of yields obtained may be a reflection of the greater diversity of the clinical group. The samples in chapter 3 were taken from umbilical cords at deliveries by elective caesarean section. In general this group would be expected to be around term in gestation and would be considered a "low risk" group in terms of ill-health for the baby. In addition, all of these samples were taken in an identical manner by one person (myself) and RNA was extracted by myself. The samples in this chapter are taken from a more heterogeneous group in terms of gestation, age, clinical condition at time of sampling, person sampling, method of sampling and person extracting RNA. It is likely that both underestimation and increased heterogeneity of the population play a part in the increased proportion of insufficient samples. For future sampling, the importance of taking adequate volumes of blood will be stressed in order to eliminate any avoidable causes of inadequate yields of RNA. Of note, the calculations used to determine if the yields were sufficient for microarray use were based on the required starting material for CodeLink™ microarrays. We have subsequently moved onto using Illumina® microarrays which require 50-100ng starting material. This means that 93.5% of the clinical samples would have sufficient yield if 100ng was used and 97.5% if 50ng was used. While this is markedly better, there is still room for improvement and my comments above remain pertinent.

From the results presented in 4.2 it can be seen that similar high levels of high quality RNA are obtained when sampling using a needle or cannula and that the percentage of samples yielding sufficient RNA are also similar. Statistical testing did not show any consistently statistically significant difference in the yield or quality of RNA obtained. This means that either method is suitable for blood collection. This means that current clinical practice with respect to blood sampling at the time of suspected sepsis would not need to change.

When the clinician taking the samples was examined in 4.3, it was pleasing that for the majority of clinicians the RNA quality was good. The notable exception to this was clinician R1. On statistical testing there was no persisting statistically significant difference in terms of RNA quality. There was a wide range of RNA yields seen however. This did not seem to show any consistent pattern when compared with the documented volume of blood sampled. As alluded to above I have concerns over the accuracy of documentation of volume of blood sampled and that some volumes may be over-estimates. There is wide variation in the percentage of samples sufficient for microarray between samplers – K1 and R1 are particularly notable in being low in this respect. In addition, statistical testing revealed that differences in yield according to clinician sampling were statistically significant. It may be postulated that samplers K1 and R1 in particular may have over-estimated their sample volumes. For sampler R1, the low yields and low quality may both be consequences of low sample volume. Alternatively, the low quality may have been due to non-adherence to sampling protocol. For future sampling, more stringent education on sampling technique, along with demonstration and supervision may

help improve the yields obtained. In addition, increasing the minimum volume of blood aimed for from 0.5 to 0.75 ml may help.

When examining samples depending on whether they were frozen or not (4.4) there was no statistically significant difference in either yield of RNA obtained or in RIN values. Statistical testing did reveal a statistically significant difference between the groups for A_{260}:A_{280} ratio with frozen samples tending to have higher values. The mean ratio for frozen samples was 2.13 and that for samples that were not frozen was 1.94. As these values are both well within the target range of 1.8 to 2.2, this difference is not of any practical significance. Both groups yielded high quality RNA with a high percentage sufficient for microarray work. This is important as it confirms the findings from the umbilical cord blood work in chapter 2. Freezing of samples allows more efficient sample processing. Samples taken off site could also be included in future studies if they were transported frozen which is very positive.

Duration of freezing (4.5) had no practical or statistically significant effect on RIN or yield of RNA obtained. On statistical testing there was found to be a statistically significant difference between groups in terms of A_{260}:A_{280} ratio, however this difference is unlikely to be of any practical significance as all of the groups other than those frozen for 8-10 weeks and 12-13 weeks had mean and median ratios within the target range of 1.8 to 2.2. Intuitively, it would seem unlikely that samples would worsen in terms of high A_{260}:A_{280} ratio if frozen for 8-10 or 12-13 weeks and then improve with further freezing. Overall, this data is encouraging and supports freezing of samples for up to several months. It would be prudent to keep

monitoring the A_{260}:A_{280} ratios however, and if any significant differences persist to investigate this further.

When the effect of period of incubation at room temperature prior to RNA extraction was looked at (4.6), no clear pattern relating time to yield, RIN or A_{260}:A_{280} ratio was seen. This was true both for samples that had been frozen and those that had not. As almost all of the samples were incubated at room temperature for 2.5 to 5 hours for fresh samples and 4 to 5 hours for the frozen samples, these observations confirm what was expected from the umbilical cord blood results in chapter 2.

The operator carrying out the RNA extraction (4.7) did not have any statistically significant effect on RIN or yield of RNA obtained. Although on Kruskal-Wallis testing there was a statistically significant difference in A_{260}:A_{280} ratio depending on operator, this statistical significance did not persist on ANOVA. For all three operators good levels of RIN and sufficient yields in a high percentage of samples were obtained. This was encouraging but was what was expected as all three operators were strictly following the same protocol. This may be taken as evidence that adherence to protocol for RNA extraction leads to consistency of results.

In section 4.8 I examined the quality and yield of RNA with respect to the white cell count of the samples. There was no clear correlation seen between RIN and white cell count (graph 4.8.1). It could be seen that most samples with lower white cell counts had high RIN values and that the poorer RIN values were spread over a range of white cell counts including both low counts and counts within the "normal range".

The few samples with extremely high white cell counts all had RIN values of over 7 but the number of samples is not sufficient to confidently make any statement regarding trend in this respect. The majority of samples had A_{260}:A_{280} ratios within the desirable range. Interestingly, the samples that were the greatest outliers in terms of A_{260}:A_{280} ratios all had lower white cell counts (graph 4.8.2). The majority of samples with lower white cell counts were, however, within the desired range. When the yield of RNA was considered there may be a suggestion of a trend towards increased white cell count yielding more RNA but there are many samples where this was not the case. In other words, yield of RNA may be higher for some but not all samples with higher white cell count. When neutrophil count was examined in section 4.9 there was no clear-cut relationship seen between neutrophil count and RIN values, A_{260}:A_{280} ratio or RNA yield. It is worth noting that some of the poorest A_{260}:A_{280} ratios were from some of the lowest neutrophil counts but most samples with low neutrophil count had good A_{260}:A_{280} ratios. Yield of RNA may be higher for some but not all samples with higher neutrophil count. These findings are of practical importance when considering clinical neonatal infection. Some of the sickest infants will have low white cell and neutrophil counts. It is therefore very encouraging that some of the highest yields of RNA were obtained from samples with low counts and that most samples with low counts had high RIN values. This means that the sickest babies with the lowest white cell counts can be included in these studies and can therefore benefit from any useful results obtained.

These results are novel findings as there are no similar reports specific to neonatal whole blood in the literature at the time of writing.

Chapter 4: Conclusions

- It is possible to obtain sufficient high quality RNA from neonatal blood samples in order to carry out microarray experiments.
- The clinician taking the sample did not seem to have any significant effect on the quality of the RNA obtained but may affect the yield. This may be due to variation in sampling technique or over-estimation of the blood volume being sampled. More stringent training on sampling and the importance of minimum blood volume may help with this.
- There was no statistically significant or practical difference in yield or quality of RNA obtained when comparing:
    - Sampling using needle or cannula
    - Samples being frozen or not prior to RNA extraction
    - Incubation time at room temperature prior to RNA extraction (for either frozen or non-frozen samples)
    - Laboratory operator carrying out RNA extraction
- Duration of freezing prior to RNA extraction does not affect RIN values or yield of RNA but may affect the A_{260}:A_{280} ratios achieved. This is not likely to be of any practical significance.
- There is no clear-cut relationship between either white cell count or neutrophil count and RIN values, A_{260}:A_{280} ratio or RNA yield.

Overall, the work in this chapter confirmed that neonatal whole blood is a suitable source of high quality RNA for microarray work. These findings meant that studies

into neonatal RNA expression profiling could be planned with confidence and we were happy to proceed onto the microarray work presented in the next chapter.

Chapter 5

Microarray Investigation of Host RNA Expression Profiles in Infected and Control Infants

Foreword

As work progressed on this study it became clear that the results looked promising. I have therefore been fortunate enough to be able to take advantage of the opportunity to carry out further array experiments. When I set out on this project, the aim was to show that it was feasible to run human neonatal RNA obtained from whole blood by using in the region of 6-10 microarrays. In this chapter, this aim is surpassed as I present larger, more detailed microarray studies examining differences in gene expression profiles between infected and control neonates.

Chapter 5.1: Investigation of Neonatal Infection Using CodeLink™ Microarrays.

Aim

To examine differences in RNA expression profiles between groups of infected and control and possibly infected infants within a statistically meaningful and balanced experimental design.

Experimental Design

This experiment was designed so that analysis should include at least 27 in each of infected, control and possible infection groups. From work carried out within the Division of Pathway Medicine (345) (*Analyst* paper, Appendix 4) using some of my non-infected neonatal samples the number projected for each group to achieve 90 % power for at least 90 % of genes at significance level of $\alpha = 0.001$ was 100. This was an early projection and as further results were analysed it became apparent that robust results were being achieved from smaller group sizes. 27 was chosen as the group size as this was felt to be achievable in the time frame being considered and allowed 3 operators for each group. The experiment was designed to ensure that arrays were run so that the samples for analysis could be operator-balanced in order to avoid this as a potential source of variation.

27 infected samples were run; 9 each by myself, Marie Craigon and Alan Ross

27 control samples were run: 9 each by myself, Marie Craigon and Alan Ross

30 possible infection samples were run: 10 each by myself and Marie Craigon and the other 10 were run by either Alan Ross (3) or Petter Storm (7).

Acknowledgements

RNA extractions were run by myself (55) and by Alan Ross (29). RNA quality control was carried out by myself and Alan Ross.

28 microarrays were run by myself, 28 by Marie Craigon and 28 by either Alan Ross or Petter Storm as detailed above.

Microarray data was sent to Al Ivens of FIOS Genomics for data quality control, data filtering and normalisation, statistical comparisons and functional analysis.

Infants studied

Details of eligibility to the study, exclusion criteria and consent were the same as those described in chapter 4 (page 188).

Infected samples were defined as samples that had positive microbiological evidence of infection from a usually sterile body fluid, i.e. from blood, cerebrospinal fluid, peritoneal fluid or fluid from an abscess. Cases of necrotising enterocolitis were only considered to be infected if there was a positive culture result from a usually sterile body site. All of these samples were taken from infants at the time of clinical suspicion of infection.

Control samples were taken from infants who were having clinical blood samples and for whom there was no clinical suspicion of infection. Specifically, samples

were taken from a group of term infants who were having their thyroid function tested in view of maternal thyroid disease, from a group of term infants in whom a diagnosis of congenital adrenal hyperplasia (CAH) was suspected, and from a group of preterm infants who were having routine clinical biochemistry/haematology monitored while clinically well. In order for the control samples to be eligible for inclusion, the thyroid screening or CAH screening had to be negative, the full blood count must not have shown any indicators of infection, and the infant must have remained clinically well. Had there been any doubt on any of these points, the sample would not have been used as a control on array. A third group of infants were examined. These were infants who had study blood samples taken because they were suspected of having sepsis but who subsequently did not have any positive microbiological culture results obtained from a usually sterile body site. These infants are referred to as the "possible infection group".

Each sample was from a different infant, i.e. no infant had more than one sample examined. All samples were taken with written, informed, parental consent. Ethics committee approval had been obtained for this study. Blood sampling and storage prior to RNA extraction was carried out as described in detail in chapter 2 (page 62).

RNA extraction and Quality Control

RNA extraction was carried out using the modified PAXgene™ protocol found on page 76 and assessment of quality and quantity of RNA was performed as described on pages 88 to 98.

Labelling, Hybridisation and Microarray Analysis

Labelling, hybridisation and microarray scanning of samples was carried out as described on pages 99 to 117.

Data Filtering, Normalisation and Analysis

Microarray data was supplied to FIOS genomics. Normalisation of the arrays was carried out using the robust multi-array average (RMA) expression measure (346) with the resulting summarised intensities being expressed in log base 2. FIOS carried out group to group comparisons using linear modeling and subsequent empirical Bayesian approaches. This included vertical p value adjustment for multiple testing (within a given comparison) in order to control for false discovery rate (Benjamini Hochberg). The Bioconductor package limma was used (347). The statistical significance level chosen for carrying data on for functional analysis was adjusted $p < 0.01$. The functional analysis of significant genes found in these comparisons was carried out by examining KEGG pathway membership and Gene Ontology (GO). Assessment for up and down regulation was carried out for enrichment p values of < 0.05 for KEGG and of < 0.001 for GO.

Results

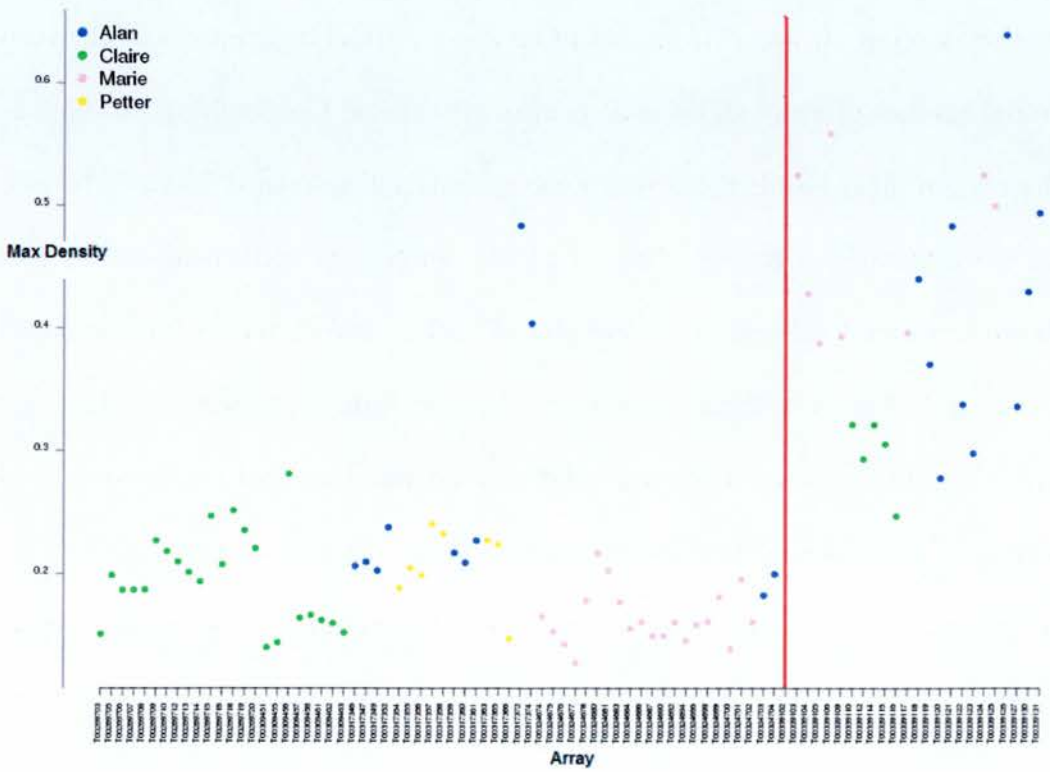
Quality Concerns

FIOS genomics expressed some concerns about the quality of microarray data received. Firstly there was some concern about background signals being high. Secondly, there was concern expressed over the degree of variability between samples and in particular a lack of mid-range expression genes in some samples.

Thirdly, there was concern that there seemed to be a distinct divide of the arrays into two groups (not related to the experimental groups of control, infected and possible), leading to the suggestion that two separate analyses may be appropriate. We, as a research group, discussed these concerns. The background issue was not new and arrays had been successfully analysed in spite of this before. I would postulate, at least in part, that this was due to the first batch of arrays having a problem with migration of glue (that was holding the removable cover in situ). It would appear that these concerns were not confined to the first batch however and this may be considered to be a drawback of the CodeLink™ set-up. The large number of differences between samples had been noted for our neonatal samples run on CodeLink™ previously, and confirmed on Affymetrix chips (data not shown). The concern over differences of mid-range expression genes was postulated to be due to a processing step, possibly due to the use of Cy-5 for labelling. The normalisation method used was chosen with the potential for biological variation in our sample population in mind. It was evident that there was a distinct difference seen between the initial batches of array slides used compared with the last batch of 24 that was purchased. In other words there was some manufacturing-related batch difference which was affecting our results. This difference can be seen in the peak density plot displayed in figure 5.1.1 and in the Boxplot of background corrected data displayed in figure 5.1.2. For each, the affected batch is on the right of the red line. This was clearly frustrating to us as it was not related to anything we had control over. I consider this to be an unfortunate consequence of the fact that we had scaled up the study gradually as the pilot studies were seen to be promising and more samples were obtained. To split the results into two groups for analysis would not only

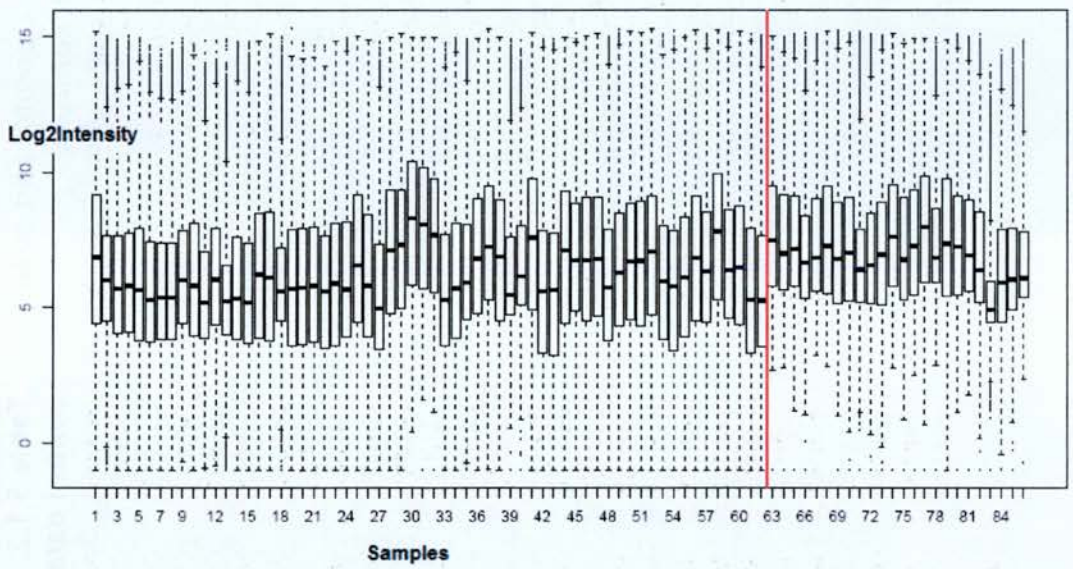
reduce the power of the study drastically but it would also skew the data. All of the possible infection samples and more of the controls had been run on earlier batches than on the 24 latest slides. Our initial intention had been to analyse this data with a view to publishing the results but with the concerns described above we felt that results would no longer be of high enough quality as they stood. We therefore decided to proceed with data analysis on a look and see basis, analysing all of the results (except the two arrays excluded as significant outliers) as a single batch. It was decided to re-run as many samples as possible as a single batch on another microarray platform (presented in chapter 5.2) and that any corroboration of results between platforms would be considered positive. Although several secondary comparisons were carried out by FIOS genomics, given the concerns regarding the quality of the data, I have not presented any of these results here.

Figure 5.1.1: ***Density Plot Peak Values***



Acknowledgement: Figures 5.1.1 and 5.1.2 were generated by Al Ivens of FIOS genomics and modified by me.

Figure 5.1.2: ***Boxplot of Background Corrected Raw Data***



Samples: Final data set

86 microarray scans were sent to FIOS genomics with the intention of analysis consisting of 27 controls, 27 infected and 27 "possible infection". The extra scans were sent to allow potential substitution for scans that failed on quality control. Two scans were failed on data quality control due to lack of correlation with the other samples. In the actual event, 27 infected samples, 27 controls and 30 possible infection samples were analysed – balance of operator was maintained.

Table 5.1.1: *Operator and quality control data: controls.*

Sample	Operator	RIN	A_{260}:A_{280}
csb001	Claire	10	1.95
csb006	Claire	10	1.86
csb014	Claire	9.9	1.76
csb017	Claire	8.5	1.73
csb028	Claire	9.4	2.06
csb042	Claire	8.5	1.80
csb049	Claire	9	1.65
csb051	Claire	9.2	2.00
csb058	Claire	9.3	2.00
csb062	Marie	9.4	1.73
csb063	Marie	9.1	1.77
csb065	Marie	9.3	1.80
csb067	Marie	9.4	1.61
csb069	Marie	9.6	1.55
csb071	Marie	7.2	1.71
csb072	Marie	9.6	1.63
csb073	Alan	9.6	1.94
csb079	Marie	9.3	1.87
csb081	Alan	8.8	1.81
csb086	Alan	8.8	1.29
csb087	Alan	10	1.83
csb096	Alan	8.6	2.04
csb165	Marie	9.2	2.29
csb179	Alan	9.2	2.31
csb181	Alan	9.4	2.16
csb200	Alan	9.3	2.19
csb202	Alan	8.8	2.17

Table 5.1.2: *Operator and quality control data: infected.*

Sample	Operator	RIN	A_{260}:A_{280}
csb004	Claire	9.1	1.84
csb012	Claire	9.2	1.80
csb032	Claire	6.8	2.50
csb047	Claire	8.9	1.68
csb075	Marie	8.7	1.91
csb082	Alan	na	1.13
csb083	Alan	na	2.45
csb084	Marie	na	2.14
csb089	Alan	9.6	1.90
csb091	Claire	9.4	2.00
csb102	Claire	9.1	2.05
csb107	Claire	5	1.92
csb112	Claire	7.1	2.03
csb114	Claire	7.9	2.02
csb118	Alan	na	2.14
csb125	Alan	na	2.12
csb132	Marie	9.5	1.94
csb137	Marie	9.3	2.00
csb138	Marie	9.3	1.95
csb152	Marie	8.7	2.34
csb155	Marie	9.2	2.21
csb159	Marie	9.4	2.19
csb161	Marie	9.1	2.40
csb162	Alan	8.5	2.75
csb164	Alan	na	3.19
csb191	Alan	6.5	2.42
csb198	Alan	9.6	2.91

Table 5.1.3: *Operator & quality control data: possible infection.*

Sample	Operator	RIN	A_{260}:A_{280}
csb002	Marie	9.2	1.90
csb005	Marie	10	1.80
csb008	Claire	9.5	2.19
csb010	Claire	9.8	2.03
csb011	Claire	6.6	1.76
csb013	Marie	10	1.72
csb019	Marie	9.5	1.98
csb020	Petter	8.6	1.70
csb024	Marie	9.2	1.98
csb025	Alan	9.6	2.04
csb026	Claire	8.7	2.05
csb027	Claire	9.1	2.25
csb033	Claire	9.9	2.47
csb034	Claire	10	2.27
csb035	Petter	10	2.15
csb036	Marie	10	2.31
csb037	Petter	9.1	1.40
csb038	Marie	9.7	1.98
csb039	Alan	9.8	1.94
csb041	Petter	8.9	1.58
csb044	Petter	9.9	1.84
csb053	Claire	9.4	2.01
csb054	Alan	8.2	2.00
csb055	Claire	9.4	2.00
csb057	Petter	7.4	2.00
csb059	Claire	9.5	2.08
csb061	Marie	9.3	1.47
csb064	Marie	9.2	1.96
csb068	Petter	9.4	1.83
csb074	Marie	8.3	1.99

Samples: RNA quality and operator

Tables 5.1.1, 5.1.2 and 5.1.3 display each of the groups of samples in the final dataset stating operator and RIN values for each sample. 6 samples did not generate a RIN. For these and all samples with RIN less than 8, the bioanalyzer graphs were studied qualitatively and were felt to be acceptable. Of the 78 samples with RIN values, the range of RIN was 5-10 with a mean of 9.0 and a median of 9.3.

Samples: Clinical Characteristics

Table 5.1.4 displays the gender, age and weight of the infants sampled in each group. The control group had a greater proportion of boys than the infected group and possible group with 63 % compared to 52 % and 53 % respectively. The infected group was notably more premature and therefore both of lower birthweight and older when sampled than the controls.

Table 5.1.4: *Demographics of Samples Used*

	Control (n=27)	Infected (n=27)	Possible Infection (n=30)
Male	17 (63 %)	14 (52 %)	16 (53 %)
Gestational Age (completed weeks)	26-42 Mean 37.2 Median 39 (1 sample gestational age not stated)	23-40 Mean 29 Median 29	25-41 Mean 34.5 Median 35
Age (day of postnatal life)	2-56 Mean 12.7 Median 7	1-75 Mean 18.3 Median 11	1-32 Mean 5.1 Median 1
Corrected gestational age (weeks + days)	31+1 to 44+1 Median 40+2	25+0 to 41+4 Median 30+6	27+3 to 41+6 Median 35+1
Birthweight (g)	650-4350 Mean 2902 Median 3260 (1 sample birthweight not stated)	430-3880 Mean 1254 Median 963	680-4840 Mean 2413 Median 2205

The possible infection group was more premature than the control group but less so than the infected group. They tended to be younger in terms of number of postnatal days than either the infected or control groups. The breakdown of reasons for blood sampling in the control group is presented in table 5.1.5.

Table 5.1.5: *Clinical Reasons for Blood Sampling in Control Group (n)*	
Maternal Thyroid Disease	9
Jaundice	4
Screening for Congenital Adrenal Hyperplasia	3
Coombs positive	1
Urea and electrolytes: repeat after previous abnormality (dehydration, hypo or hypernatraemia)	3
Guthrie screening	1
"Routine" neonatal screening (preterms)	5
Hypoxic ischaemic encephalopathy	1

Table 5.1.6 lists the micro-organisms isolated from the infected group. The majority of samples isolated were Coagulase negative staphylococci. One infant had a positive growth of yeast (*Candida albicans*) from blood culture and also had *Klebsiella* species isolated from peritoneal fluid. One infant had CMV isolated from both blood and urine. One infant had Enterovirus meningitis and another had *Listeria* meningitis. The remaining infants had positive blood cultures with other pathogenic bacteria.

Table 5.1.6: *Microorganisms Isolated from Infected Group*		
Organism	**n**	**Comments**
Coagulase negative staphylococci	12	
Candida albicans	1	*Klebsiella* was also isolated from peritoneal fluid
Enterobacter species	1	
Enterococci (*E.faecalis*)	3	1 sample also had coagulase negative staphylococcus in CSF
Group B streptococcus	2	1 sample also had coagulase negative staphylococcus in CSF
Pseudomonas aeruginosa	1	Coagulase negative staphylococcus in blood too
Cytomegalovirus (CMV)	1	Isolated in blood and urine
Klebsiella species	3	
Listeria monocytogenes	1	From CSF only
Enterovirus	1	From CSF only
Escherichia coli	1	

All of the infected samples included were considered to be convincing cases of infection. Where the organism was viral, was isolated from CSF rather than blood or where the isolated organism was coagulase negative staphylococcus, there had to be supporting clinical evidence of significant infection. Evidence for this is presented in table 5.1.7.

Table 5.1.7: ***Supporting Evidence for Infection in Viral, Meningitis and Coagulase Negative Staphylococcal Cases.***

Organism	**Supporting Evidence** (days of antibiotics)
Coagulase negative staphylococcus	• Re-intubated, neutrophilia (5) • Bradycardias, pyrexia, low platelets (15) • Apnoeas, bradycardia, temperature instability, lethargy (2) • Repeat isolate obtained the following day, hypotension, hyperglycaemia, thrombocytopenia, died (15) • Bradycardias, desaturations, long line in situ (11) • Apnoeas, bradycardias, leucopoenia, thrombocytopenia (10) • Previous isolate, bradycardias, desaturations, mottled (18) • Previous isolate, long line in situ: also isolated from line tip, bradycardias, desaturations, reduced perfusion (18) • Bradycardias, desaturations, reintubated, deranged sugars (10) • Previous isolate, hypotension, hypoglycaemia, oedema, died (16) • Isolated in 2 separate blood cultures taken at same time, bradycardia, desaturations, reintubated, temperature instability (5) • Apnoeas, pyrexia, reduced perfusion, elevated white cell count, infant had two further significant episodes of infection associated with coagulase negative staphylococcus (>10)
Enterovirus	Poor perfusion, elevated CSF white cell count, poor handling, poor colour, rash (10)
Listeria monocytogenes	Pyrexia, hypertonia, elevated CSF white cells, neutrophilia in blood (14)
Cytomegalovirus	Profound bradycardias and desaturations (nil)

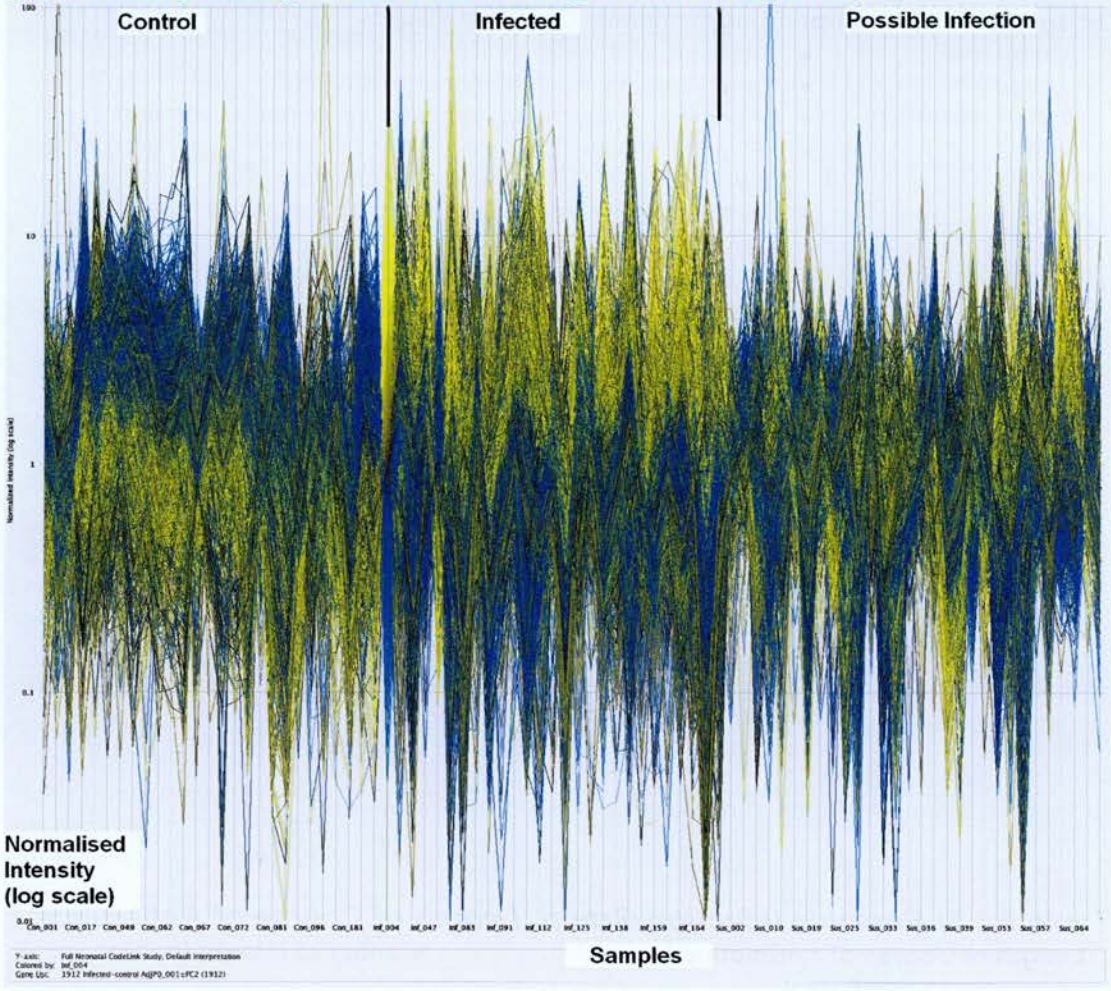
Table 5.1.8: ***Possible Infection Group: Length of Course of Antibiotics***

Length of Course of Antibiotics (days)	**Number of infants (n=30)**
None	5
Up to 2 days	0
2 or more days but less than 5 days	12
5 days	11
More than 5 days	2

Table 5.1.8 displays the length of course of antibiotics received for infants in the possible infection group. The range of length of course of antibiotics was zero to

fifteen days. This table is included to give an indication of the degree of concern over these infants. For the decision for no antibiotics to be given in 5 cases there must have been minimal clinical concern. Conversely, those infants with 5 or more days of antibiotics would be expected to have had significant clinical concern or significant risk factors for infection.

Figure 5.1.3: ***Genespring expression profile of significantly differentially expressed genes (p < 0.01, fold change > 2: 1912 genes).***



Acknowledgement: The above figure was generated from Genespring by Paul Dickinson.

The Genespring expression profile of significantly differentially expressed genes (between infected and control groups) pictured here as figure 5.1.3 shows each individual sample on the x-axis, with controls on the left, infected samples in the

middle and possible infection samples on the right. The y-axis is the normalised log intensity. Blue lines represent significantly down-regulated genes and yellow lines represent significantly up-regulated genes (1912 genes). There is a discernable difference in response seen between infected and control infants. For example, the control infants have a blue band of genes that are expressed at a higher intensity than a yellow band of genes, the converse being true for the infected infants. The pattern for the possible infection group appears to be somewhere in between that for the infected group and that for the control group with some blue peaks and some yellow peaks.

Infected Compared to Control

Significant Differential Expression

3660 features were significantly differentially expressed with an adjusted p value of less than 0.01: 967 up-regulated and 2693 down-regulated.

GO Term Enrichment

In the significantly up-regulated genes there were 52 enriched GO terms and in significantly down-regulated genes there were 94 enriched GO terms. These are displayed in tables 5.1.9 (a and b) and 5.1.10 (a to c). The ontology categories are BP (biological process), MF (molecular function) and CC (cellular component). These abbreviations are used throughout this section. The terms are listed in order of significance with the most significant first. I have highlighted in red the terms that are immediately recognisable as being involved in immune response.

Table 5.1.9a: *GO Terms for Up-regulated Genes. Infected compared to Control*

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0009611	response to wounding	4.67E-10
BP	GO:0002376	immune system process	1.11E-07
BP	GO:0009605	response to external stimulus	1.33E-07
BP	GO:0006955	immune response	1.61E-07
BP	GO:0006952	defence response	7.15E-07
BP	GO:0006954	inflammatory response	1.81E-06
BP	GO:0006950	response to stress	2.96E-06
BP	GO:0007599	haemostasis	4.94E-06
BP	GO:0007596	blood coagulation	1.09E-05
BP	GO:0048522	positive regulation of cellular process	1.16E-05
BP	GO:0050817	coagulation	1.24E-05
BP	GO:0050878	regulation of body fluid levels	1.33E-05
BP	GO:0048518	positive regulation of biological process	1.67E-05
BP	GO:0006935	chemotaxis	4.82E-05
BP	GO:0042330	taxis	4.82E-05
BP	GO:0007626	locomotory behaviour	5.59E-05
BP	GO:0007155	cell adhesion	0.000104544
BP	GO:0022610	biological adhesion	0.000104544
BP	GO:0065007	biological regulation	0.000117914
BP	GO:0007165	signal transduction	0.000194926
BP	GO:0007610	behaviour	0.000279811
BP	GO:0042060	wound healing	0.000285928
BP	GO:0051272	positive regulation of cell motion	0.000353055
BP	GO:0051270	regulation of cell motion	0.00036052
BP	GO:0031325	positive regulation of cellular metabolic process	0.000406484
BP	GO:0051707	response to other organism	0.000411483
BP	GO:0050896	response to stimulus	0.000430816
BP	GO:0006928	cell motion	0.000504015
BP	GO:0051674	localisation of cell	0.000504015
BP	GO:0012501	programmed cell death	0.000506126
BP	GO:0042221	response to chemical stimulus	0.000549588
BP	GO:0008284	positive regulation of cell proliferation	0.00060892
BP	GO:0031328	positive regulation of cellular biosynthetic process	0.000661952
BP	GO:0008219	cell death	0.000672349
BP	GO:0065008	regulation of biological quality	0.000678745
BP	GO:0016265	death	0.000747207
BP	GO:0006915	apoptosis	0.000760179
BP	GO:0009891	positive regulation of biosynthetic process	0.000827454
BP	GO:0009893	positive regulation of metabolic process	0.00095447
BP	GO:0009607	response to biotic stimulus	0.000977322

Table 5.1.9b: *GO Terms for Up-regulated Genes. Infected compared to Control*			
MF	GO:0032403	protein complex binding	0.000157948
MF	GO:0005509	calcium ion binding	0.00022189
MF	GO:0004871	signal transducer activity	0.000369437
MF	GO:0060089	molecular transducer activity	0.000369437
MF	GO:0019865	immunoglobulin binding	0.000810849
MF	GO:0005100	Rho GTPase activator activity	0.000969558
CC	GO:0042405	nuclear inclusion body	5.53E-05
CC	GO:0031091	platelet alpha granule	0.000190492
CC	GO:0005886	plasma membrane	0.000208567
CC	GO:0005615	extracellular space	0.000393516
CC	GO:0016020	membrane	0.000643486
CC	GO:0031225	anchored to membrane	0.00071491

Table 5.1.10a: *GO Terms for Down-Regulated Genes: Infected compared to Control*			
Ontology	**GO ID**	**GO Term**	**Enriched p Value**
BP	GO:0010467	gene expression	8.81E-24
BP	GO:0006414	translational elongation	3.37E-20
BP	GO:0034645	cellular macromolecule biosynthetic process	1.43E-18
BP	GO:0009059	macromolecule biosynthetic process	2.14E-18
BP	GO:0006412	Translation	2.41E-18
BP	GO:0043170	macromolecule metabolic process	3.03E-18
BP	GO:0034961	cellular biopolymer biosynthetic process	3.36E-18
BP	GO:0043284	biopolymer biosynthetic process	3.36E-18
BP	GO:0043283	biopolymer metabolic process	9.96E-18
BP	GO:0034960	cellular biopolymer metabolic process	1.37E-17
BP	GO:0044260	cellular macromolecule metabolic process	2.87E-17
BP	GO:0008152	metabolic process	2.02E-15
BP	GO:0044238	primary metabolic process	4.84E-15
BP	GO:0044237	cellular metabolic process	8.75E-15
BP	GO:0044249	cellular biosynthetic process	1.08E-13
BP	GO:0009058	biosynthetic process	1.46E-13
BP	GO:0016070	RNA metabolic process	4.32E-12
BP	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.25E-11
BP	GO:0022613	ribonucleoprotein complex biogenesis	3.82E-10
BP	GO:0034660	ncRNA metabolic process	4.55E-10
BP	GO:0034470	ncRNA processing	1.03E-09

Table 5.1.10b: ***GO Terms for Down-Regulated Genes: Infected compared to Control***

BP	GO:0042254	ribosome biogenesis	1.04E-08
BP	GO:0006396	RNA processing	1.78E-08
BP	GO:0006350	Transcription	6.85E-08
BP	GO:0006364	rRNA processing	1.66E-07
BP	GO:0016072	rRNA metabolic process	3.81E-07
BP	GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	4.71E-07
BP	GO:0006351	transcription, DNA-dependent	2.99E-06
BP	GO:0009987	cellular process	3.11E-06
BP	GO:0032774	RNA biosynthetic process	3.38E-06
BP	GO:0000387	spliceosomal snRNP biogenesis	4.17E-06
BP	GO:0060255	regulation of macromolecule metabolic process	6.90E-06
BP	GO:0010556	regulation of macromolecule biosynthetic process	7.10E-06
BP	GO:0051252	regulation of RNA metabolic process	1.33E-05
BP	GO:0006355	regulation of transcription, DNA-dependent	1.43E-05
BP	GO:0010468	regulation of gene expression	1.55E-05
BP	GO:0045449	regulation of transcription	1.97E-05
BP	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.86E-05
BP	GO:0031326	regulation of cellular biosynthetic process	5.45E-05
BP	GO:0009889	regulation of biosynthetic process	5.55E-05
BP	GO:0019222	regulation of metabolic process	8.73E-05
BP	GO:0031323	regulation of cellular metabolic process	0.000104809
BP	GO:0042110	T cell activation	0.000138833
BP	GO:0002696	positive regulation of leukocyte activation	0.000254711
BP	GO:0006399	tRNA metabolic process	0.000287827
BP	GO:0009451	RNA modification	0.000315192
BP	GO:0050867	positive regulation of cell activation	0.00037817
BP	GO:0001510	RNA methylation	0.000689262
BP	GO:0019538	protein metabolic process	0.000812769
BP	GO:0044267	cellular protein metabolic process	0.000875299
BP	GO:0046649	lymphocyte activation	0.000921125
MF	GO:0003735	structural constituent of ribosome	1.21E-17
MF	GO:0003676	nucleic acid binding	3.90E-14
MF	GO:0003723	RNA binding	1.17E-08
MF	GO:0032395	MHC class II receptor activity	1.43E-06
MF	GO:0008168	methyltransferase activity	1.45E-05

Table 5.1.10c: ***GO Terms for Down-Regulated Genes: Infected compared to Control***

MF	GO:0016741	transferase activity, transferring one-carbon groups	2.27E-05
MF	GO:0003677	DNA binding	2.59E-05
MF	GO:0008270	zinc ion binding	5.43E-05
MF	GO:0046914	transition metal ion binding	0.000284282
MF	GO:0045182	translation regulator activity	0.000942185
CC	GO:0005622	intracellular	3.31E-22
CC	GO:0044424	intracellular part	5.25E-19
CC	GO:0005840	ribosome	2.23E-18
CC	GO:0030529	ribonucleoprotein complex	3.01E-17
CC	GO:0043231	intracellular membrane-bounded organelle	1.16E-15
CC	GO:0043227	membrane-bounded organelle	1.34E-15
CC	GO:0043229	intracellular organelle	4.55E-15
CC	GO:0043226	organelle	5.00E-15
CC	GO:0005634	nucleus	1.30E-13
CC	GO:0033279	ribosomal subunit	7.20E-13
CC	GO:0044464	cell part	1.23E-12
CC	GO:0005623	cell	1.33E-12
CC	GO:0022626	cytosolic ribosome	1.47E-11
CC	GO:0005829	cytosol	3.16E-09
CC	GO:0005739	mitochondrion	4.45E-09
CC	GO:0042613	MHC class II protein complex	1.61E-08
CC	GO:0022625	cytosolic large ribosomal subunit	6.67E-08
CC	GO:0005737	cytoplasm	6.87E-08
CC	GO:0015934	large ribosomal subunit	8.60E-08
CC	GO:0044445	cytosolic part	2.46E-07
CC	GO:0032991	macromolecular complex	5.93E-07
CC	GO:0015935	small ribosomal subunit	1.86E-06
CC	GO:0044444	cytoplasmic part	4.82E-06
CC	GO:0044428	nuclear part	8.11E-06
CC	GO:0044429	mitochondrial part	1.46E-05
CC	GO:0070013	intracellular organelle lumen	2.46E-05
CC	GO:0042611	MHC protein complex	3.97E-05
CC	GO:0031974	membrane-enclosed lumen	5.78E-05
CC	GO:0043233	organelle lumen	6.83E-05
CC	GO:0042101	T cell receptor complex	0.000103417
CC	GO:0031981	nuclear lumen	0.000111046
CC	GO:0022627	cytosolic small ribosomal subunit	0.000117166
CC	GO:0005730	nucleolus	0.000366846

KEGG Pathway Enrichment

There were found to be 11 enriched KEGG pathways in significantly up-regulated genes (displayed in Table 5.1.11) and 19 enriched KEGG pathways in significantly down-regulated genes (displayed in Table 5.1.12).

Table 5.1.11: ***Kegg Pathways for Up-Regulated Genes. Infected compared to Control.***

Pathway ID	Pathway Description	p Value
4060	Cytokine-cytokine receptor interaction	0.002389079
4610	Complement and coagulation cascades	0.005371016
530	Aminosugars metabolism	0.006040276
4640	Hematopoietic cell lineage	0.007304653
4910	Insulin signaling pathway	0.010435325
4510	Focal adhesion	0.014617217
4620	Toll-like receptor signalling pathway	0.016858824
4512	ECM-receptor interaction	0.021752537
4210	Apoptosis	0.024699901
4810	Regulation of actin cytoskeleton	0.04115909
51	Fructose and mannose metabolism	0.049818179

Table 5.1.12: ***Kegg Pathways for Down-Regulated Genes. Infected compared to Control.***

Pathway ID	Pathway Description	p Value
3010	Ribosome	2.15E-15
5332	Graft-versus-host disease	4.68E-08
4660	T cell receptor signalling pathway	2.34E-06
5330	Allograft rejection	3.86E-06
5310	Asthma	7.46E-06
4640	Hematopoietic cell lineage	1.32E-05
4940	Type I diabetes mellitus	2.52E-05
5320	Autoimmune thyroid disease	0.000152039
970	Aminoacyl-tRNA biosynthesis	0.00065215
4612	Antigen processing and presentation	0.001981931
5340	Primary immunodeficiency	0.00621127
4650	Natural killer cell mediated cytotoxicity	0.012287985
4370	VEGF signaling pathway	0.020515548
240	Pyrimidine metabolism	0.022336428
290	Valine, leucine and isoleucine biosynthesis	0.023052917
4310	Wnt signaling pathway	0.025003562
4514	Cell adhesion molecules (CAMs)	0.028900544
3020	RNA polymerase	0.029943213
4120	Ubiquitin mediated proteolysis	0.033662806

Fold Change

Largest Fold Changes

Before statistical filtering there were 626 features exhibiting greater than 2-fold up-regulation and 1818 features exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 220 features with greater than 2-fold up-regulation and 986 with greater than 2-fold down regulation. The genes for the features with the twenty largest fold changes are described in tables 5.1.13(a and b) for up-regulated genes and 5.1.14 (a and b) for down-regulated genes (the full lists of statistically filtered genes with fold changes of greater than 2 can be found on the enclosed CD). Each line represents one feature on the array with the largest fold changes being listed first. Phrases readily recognisable as being immune function related are highlighted in red.

Table 5.1.13a:
Table of Fold Changes (> Log 2) for Up-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).

Feature ID	Description	Pathway Description	Fold Change
GE61410	CD177 molecule	-	18.10285
GE572841	ankyrin repeat domain 22	-	11.19002
GE484349	NA	-	9.810082
GE79252	G protein-coupled receptor 84	-	8.679239
GE79166	resistin	-	7.14978
GE57919	tumor necrosis factor, alpha-induced protein 6	-	6.769111
GE60514	Fc fragment of IgG, high affinity Ia, receptor (CD64)	Hematopoietic cell lineage, Systemic lupus erythematosus	6.387135
GE537401	ankyrin repeat domain 22	-	5.953429
GE545552	chromosome 19 open reading frame 59	-	5.879541
GE61045	DnaJ (Hsp40) homolog, subfamily C, member 5	-	5.783771
GE58106	carbonic anhydrase IV	Nitrogen metabolism	5.585605
GE57860	annexin A3	-	5.584148

Table 5.1.13b:
Table of Fold Changes (> Log 2) for Up-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).

GE78980	BMX non-receptor tyrosine kinase	-	5.419599
GE59735	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	-	5.214402
GE54140	maltase-glucoamylase (alpha-glucosidase)	Galactose metabolism, Starch and sucrose metabolism	5.173496
GE53097	dishevelled associated activator of morphogenesis 2	Wnt signaling pathway	4.979599
GE59851	interleukin 1 receptor, type II	MAPK signalling pathway, Cytokine-cytokine receptor interaction, Hematopoietic cell lineage	4.815815
GE57525	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	Leukocyte transendothelial migration, Bladder cancer	4.810045
GE60380	proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)	Asthma	4.769481
GE62094	SAM domain, SH3 domain and nuclear localization signals 1	-	4.754757

Table 5.1.14a: ***Table of Fold Changes (> Log 2) for Down-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

Feature ID	Description	Pathway Description	Fold Change
GE79288	lectin, galactoside-binding, soluble, 2	-	-11.4879
GE61347	ribosomal protein L23a	Ribosome	-7.32457
GE79244	ribosomal protein L7a	-	-6.82837
GE62103	tubulin polymerization-promoting protein family member 3	-	-6.80987
GE481282	hypothetical protein LOC129293	-	-6.57359
GE61920	melanoma antigen family H, 1	-	-6.182
GE612668	Fc receptor-like 6	-	-5.89332
GE555144	zinc finger protein 683	-	-5.71583
GE61029	sarcosine dehydrogenase	Glycine, serine and threonine metabolism	-5.70711

Table 5.1.14b: ***Table of Fold Changes (> Log 2) for Down-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***			
GE81458	CD8b molecule	Cell adhesion molecules (CAMs), Antigen processing and presentation, Hematopoietic cell lineage, T cell receptor signalling pathway, Primary immunodeficiency	-5.64136
GE80046	zeta-chain (TCR) associated protein kinase 70kDa	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway, Primary immunodeficiency	-5.48596
GE61919	transmembrane protein 204	-	-5.44077
GE85839	NA	-	-5.43182
GE539361	BTB and CNC homology 1, basic leucine zipper transcription factor 2	-	-5.30172
GE53988	leucine rich repeat neuronal 3	-	-5.23077
GE59630	interleukin 23, alpha subunit p19	Cytokine-cytokine receptor interaction, Jak-STAT signalling pathway	-5.17507
GE519859	OCIA domain containing 2	-	-4.96511
GE81550	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	MAPK signalling pathway, T cell receptor signalling pathway	-4.92475
GE62569	signal-regulatory protein gamma	-	-4.90356
GE60283	src kinase associated phosphoprotein 1	-	-4.8916

The whole list of genes with fold change > 2 and $p < 0.001$ was further examined by gene ontology using the online tool DAVID 2008 (348, 349). The result for the DAVID gene ontology lists for up-regulated genes is found in table 5.1.15 and for down-regulated genes in table 5.1.16 (a and b).

Table 5.1.15: ***DAVID GO Terms for Fold-Change > 2: Up-Regulated Genes. Infected compared to control.***

Ontology	GO Term	No. of Genes (from 220)	p value
BP	Response to wounding	18	4.9E-7
BP	Response to external stimulus	20	7.5E-6
BP	Inflammatory response	12	1.1E-4
BP	Response to stress	24	1.4E-4
BP	Regulation of body fluid levels	8	1.4E-4
BP	Blood coagulation	7	3.1E-4
BP	Coagulation	7	3.5E-4
BP	Hemostasis	7	4.1E-4
BP	Response to other organism	9	5.6E-4
BP	Defence response	15	9.2E-4
MF	Calcium ion binding	20	8.1E-1

Table 5.1.16a: ***DAVID GO Terms for Fold-Change > 2: Down-Regulated Genes. Infected compared to control.***

Ontology	GO Term	No. of Genes (from 986)	p value
BP	Gene expression	197	3.3E-12
BP	RNA metabolic process	152	2.5E-7
BP	Macromolecule metabolic process	305	2.7E-7
BP	Primary metabolic process	341	3.0E-7
BP	Cellular process	482	9.5E-7
BP	Regulation of cell activation	15	1.6E-6
BP	T cell activation	17	2.4E-6
BP	Translation	46	2.4E-6
BP	RNA processing	37	3.0E-6
BP	Metabolic process	365	3.3E-6
BP	Regulation of lymphocyte activation	14	4.5E-6
BP	Lymphocyte activation	21	5.6E-6
BP	Biopolymer metabolic process	232	8.5E-6
BP	Regulation of T cell activation	12	9.8E-6
BP	Cellular metabolic process	332	1.1E-5
BP	Leukocyte activation	22	1.1E-5
BP	Ribosome biogenesis and assembly	14	1.3E-5
BP	Regulation of biological process	208	2.1E-5
BP	Cell activation	23	2.4E-5
BP	Cellular biosynthetic process	67	2.6E-5
BP	Nucleobase, nucleoside, nucleotise and nucleic acid metabolic process	181	3.1E-5
BP	Regulation of cellular process	192	8.8E-5
BP	Macromolecule biosynthetic process	53	2.1E-4
BP	Biological regulation	219	2.1E-4

Table 5.1.16b: ***DAVID GO Terms for Fold-Change > 2: Down-Regulated Genes. Infected compared to control.***			
BP	Transcription	125	2.2E-4
BP	rRNA processing	10	3.1E-4
BP	rRNA metabolic process	10	4.4E-4
BP	Transcription, DNA-dependent	115	4.9E-4
BP	RNA biosynthetic process	115	5.1E-4
BP	T cell differentiation in the thymus	5	5.4E-4
BP	Positive regulation of lymphocyte activation	9	6.8E-4
BP	Regulation of transcription, DNA-dependent	112	7.0E-4
BP	Regulation of metabolic process	132	8.8E-4
BP	Biosynthetic process	76	9.2E-4
CC	Cytosolic ribosome (sensu Eukaryota)	20	1.7E-11
CC	Ribosomal subunit	23	6.9E-9
CC	Cytosolic large ribosomal subunit (sensu Eukaryota)	13	7.1E-9
CC	Intracellular	443	9.5E-9
CC	Cytosolic part	23	1.2E-8
CC	Ribonucleoprotein complex	50	3.6E-8
CC	Large ribosomal subunit	16	6.0E-8
CC	Ribosome	34	1.0E-7
CC	Immunological synapse	9	2.4E-7
CC	Intracellular part	413	8.1E-7
CC	Cell part	564	9.6E-6
CC	Cell	564	9.9E-6
CC	Nucleus	216	1.3E-5
CC	Intracellular organelle	350	1.4E-5
CC	Organelle	350	1.4E-5
	T cell receptor complex	6	3.9E-5
CC	Intracellular membrane-bound organelle	305	9.9E-5
CC	Membrane-bound-organelle	305	1.0E-4
CC	Cytoplasm	267	2.1E-4
CC	Cytosol	35	7.5E-4
MF	Structural constituent of ribosome	33	7.7E-9
MF	Binding	469	1.7E-5
MF	Nucleic acid binding	163	2.2E-5
MF	Zinc ion binding	114	3.4E-5
MF	RNA binding	46	7.9E-5
MF	Transition metal ion binding	131	8.7E-5
MF	MHC class II receptor acticity	5	2.1E-4

Infected compared to possible samples

Significant Differential Expression

1616 features were significantly differentially expressed with an adjusted p value of less than 0.01: 770 up-regulated and 846 down-regulated.

GO Term Enrichment

In the significantly up-regulated genes there were 48 enriched GO terms and in significantly down-regulated genes there was 1 enriched GO term. These are displayed in tables 5.1.17(a and b) and 5.1.18. The terms are listed in order of significance with the most significant first. I have highlighted in red the terms that are immediately recognisable as being involved in immune response.

Table 5.1.17a: ***GO Terms for Up-Regulated Genes. Infected compared to possible.***

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0002376	immune system process	8.70E-10
BP	GO:0006955	immune response	1.05E-06
BP	GO:0009611	response to wounding	2.28E-06
BP	GO:0002682	regulation of immune system process	8.62E-06
BP	GO:0001775	cell activation	1.19E-05
BP	GO:0045321	leukocyte activation	1.23E-05
BP	GO:0006954	inflammatory response	3.69E-05
BP	GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	4.79E-05
BP	GO:0002250	adaptive immune response	5.36E-05
BP	GO:0009605	response to external stimulus	5.96E-05
BP	GO:0006952	defence response	8.11E-05
BP	GO:0034097	response to cytokine stimulus	9.50E-05
BP	GO:0046649	lymphocyte activation	9.71E-05
BP	GO:0002684	positive regulation of immune system process	0.000131
BP	GO:0006935	chemotaxis	0.000141
BP	GO:0042330	taxis	0.000141
BP	GO:0031325	positive regulation of cellular metabolic process	0.000164
BP	GO:0001817	regulation of cytokine production	0.000194

Table 5.1.17b: *GO Terms for Up-Regulated Genes. Infected compared to possible.*			
BP	GO:0007155	cell adhesion	0.00021
BP	GO:0022610	biological adhesion	0.00021
BP	GO:0007596	blood coagulation	0.000292
BP	GO:0001816	cytokine production	0.00031
BP	GO:0002252	immune effector process	0.00031
BP	GO:0050817	coagulation	0.000321
BP	GO:0042088	T-helper 1 type immune response	0.000368
BP	GO:0031328	positive regulation of cellular biosynthetic process	0.000374
BP	GO:0009893	positive regulation of metabolic process	0.000391
BP	GO:0009891	positive regulation of biosynthetic process	0.000466
BP	GO:0007599	Hemostasis	0.000507
BP	GO:0002449	lymphocyte mediated immunity	0.000514
BP	GO:0030307	positive regulation of cell growth	0.000537
BP	GO:0010604	positive regulation of macromolecule metabolic process	0.000551
BP	GO:0032649	regulation of interferon-gamma production	0.000591
BP	GO:0042035	regulation of cytokine biosynthetic process	0.000666
BP	GO:0048518	positive regulation of biological process	0.000676
BP	GO:0006950	response to stress	0.000719
BP	GO:0048247	lymphocyte chemotaxis	0.000736
BP	GO:0032609	interferon-gamma production	0.000743
BP	GO:0019724	B cell mediated immunity	0.000749
BP	GO:0050878	regulation of body fluid levels	0.000762
BP	GO:0002520	immune system development	0.000843
BP	GO:0045087	innate immune response	0.000951
BP	GO:0010557	positive regulation of macromolecule biosynthetic process	0.000968
MF	GO:0005518	collagen binding	0.000205
MF	GO:0005548	phospholipid transporter activity	0.000257
MF	GO:0005515	protein binding	0.000274
MF	GO:0005319	lipid transporter activity	0.00063
CC	GO:0005615	extracellular space	0.000573

Table 5.1.18: *GO Terms for Down-Regulated Genes. Infected compared to possible.*			
Ontology	**GO ID**	**GO Term**	**Enriched p Value**
CC	GO:0045211	postsynaptic membrane	0.000428

KEGG Pathway Enrichment

There were found to be 5 enriched KEGG pathways in significant up-regulated genes (displayed in Table 5.1.19) and 6 enriched KEGG pathways in significantly down-regulated genes (displayed in Table 5.1.20).

Table 5.1.19: ***KEGG Pathways for Up-Regulated Genes. Infected compared to possible.***

Pathway ID	Pathway Description	p Value
4610	Complement and coagulation cascades	0.004513
4620	Toll-like receptor signaling pathway	0.014177
4520	Adherens junction	0.029785
5220	Chronic myeloid leukemia	0.029785
5120	Epithelial cell signaling in Helicobacter pylori infection	0.039417

Table 5.1.20: ***KEGG Pathways for Down-Regulated Genes. Infected compared to possible.***

Pathway ID	Pathway Description	p Value
480	Glutathione metabolism	0.024297
3010	Ribosome	0.024861
4920	Adipocytokine signaling pathway	0.030363
590	Arachidonic acid metabolism	0.035783
4120	Ubiquitin mediated proteolysis	0.040034
430	Taurine and hypotaurine metabolism	0.046659

Largest Fold Changes

Before statistical filtering there were 555 features exhibiting greater than 2-fold up-regulation and 941 features exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 140 features with greater than 2-fold up-regulation and 222 with greater than 2-fold down regulation. The twenty features with the largest fold changes are described in tables 5.1.21 for up-regulated genes and 5.1.22 for down-regulated genes (the full lists can be found on the enclosed CD).

Table 5.1.21: ***Table of Fold Changes > 2 for Up-Regulated Genes. Infected compared to possible. (Each line represents 1 feature on array)***

Feature ID	Description	Pathway Description	Fold Change
GE81123	heterogeneous nuclear ribonucleoprotein A2/B1	-	40.32271
GE57919	tumor necrosis factor, alpha-induced protein 6	-	6.427991
GE61107	interferon-induced protein with tetratricopeptide repeats 3	-	6.276232
GE61045	DnaJ (Hsp40) homolog, subfamily C, member 5	-	6.230537
GE83057	melanoma inhibitory activity 2	-	5.436336
GE59642	chemokine (C-X-C motif) ligand 10	Cytokine-cytokine receptor interaction, Toll-like receptor signalling pathway	5.410446
GE572841	ankyrin repeat domain 22	-	5.346185
GE60514	Fc fragment of IgG, high affinity Ia, receptor (CD64)	Hematopoietic cell lineage, Systemic lupus erythematosus	5.139712
GE53687	hect domain and RLD 5	-	5.006804
GE79166	Resistin	-	4.41192
GE79252	G protein-coupled receptor 84	-	4.306733
GE81553	Epstein-Barr virus induced 3	-	4.184662
GE492768	NA	-	4.078122
GE59320	interleukin 1 receptor antagonist	-	4.037408
GE484349	NA	-	3.729362
GE537401	ankyrin repeat domain 22	-	3.570259
GE59697	Pleckstrin	-	3.513284
GE521522	phospholipid scramblase 1	-	3.345852
GE56131	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	O-Glycan biosynthesis, Glycan structures - biosynthesis 1	3.295413
GE57501	chemokine (C-C motif) ligand 4	Cytokine-cytokine receptor interaction, Toll-like receptor signalling pathway	3.290582

Table 5.1.22: *Table of Fold Changes > 2 for Down-Regulated Genes. Infected compared to possible. (Each line represents 1 feature on array)*

Feature ID	Description	Pathway Description	Fold Change
GE79550	histone cluster 1, H2ad	Systemic lupus erythematosus	-15.1675
GE62790	similar to protein x 013	-	-9.41351
GE83044	tubulin, alpha 4b (pseudogene)	-	-6.21929
GE84069	ribosomal protein L10a	Ribosome	-6.05651
GE82889	mannosidase, endo-alpha	-	-5.96556
GE59782	glycophorin E	-	-5.85163
GE57167	aquaporin 2 (collecting duct)	-	-5.22818
GE79164	ubiquitin A-52 residue ribosomal protein fusion product 1	-	-5.13446
GE82187	intelectin 1 (galactofuranose binding)	-	-5.06674
GE82545	four and a half LIM domains 5	-	-4.98957
GE86577	olfactory receptor, family 1, subfamily N, member 2	Olfactory transduction	-4.97609
GE84956	NA	-	-4.71615
GE61347	ribosomal protein L23a	Ribosome	-4.68224
GE897195	NA	-	-4.60281
GE83109	ATPase, class V, type 10B	-	-4.58334
GE56530	microphthalmia-associated transcription factor	Melanogenesis, Melanoma	-4.56203
GE872322	lamin A/C	Cell junctions	-4.5045
GE81548	tetraspanin 5	-	-4.43985
GE79195	ermin, ERM-like protein	-	-4.2776
GE866295	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2	-	-4.24457

The significant GO terms as determined using DAVID 2008 (348, 349) from the full lists of up- and down-regulated genes with fold changes > 2 and p values of less than 0.001, are displayed in tables 5.1.23 and 5.1.24 respectively.

Table 5.1.23: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Up-Regulated Genes. Infected compared to possible***

Ontology	GO Term	No. of Genes Involved (from 140)	p value
BP	Response to stress	20	9.1E-6
BP	Response to wounding	12	3.4E-5
BP	Response to external stimulus	14	7.3E-5
BP	Immune system process	17	8.6E-4
MF	Protein binding	64	1.7E-5

Table 5.1.24: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Down-Regulated Genes. Infected compared to possible.***

Ontology	GO Term	No. of Genes (from 222)	p value
CC	Cytoplasm	69	2.9E-4

Possible compared to control samples

Significant Differential Expression

4517 features were significantly differentially expressed with an adjusted p value of less than 0.01: 2000 up-regulated and 2517 down-regulated.

GO Term Enrichment

In the significantly up-regulated genes there were 20 enriched GO terms and in significantly down-regulated genes there were 131 enriched GO terms. These are

displayed in tables 5.1.25 and 5.1.26 (a to d). The terms are listed in order of significance with the most significant first.

Table 5.1.25: *GO Terms for Up-regulated Genes. Possible compared to controls.*

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0007185	transmembrane receptor protein tyrosine phosphatase signaling pathway	9.20E-06
BP	GO:0030036	actin cytoskeleton organization	0.000156
BP	GO:0030029	actin filament-based process	0.000281
BP	GO:0015669	gas transport	0.000327
BP	GO:0051056	regulation of small GTPase mediated signal transduction	0.000415
BP	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.000833
BP	GO:0007010	cytoskeleton organization	0.0009
BP	GO:0055072	iron ion homeostasis	0.000944
MF	GO:0008092	cytoskeletal protein binding	4.27E-07
MF	GO:0003779	actin binding	3.92E-06
MF	GO:0030695	GTPase regulator activity	7.32E-05
MF	GO:0005096	GTPase activator activity	0.000142
MF	GO:0005085	guanyl-nucleotide exchange factor activity	0.000377
MF	GO:0005083	small GTPase regulator activity	0.000645
MF	GO:0008601	protein phosphatase type 2A regulator activity	0.000789
MF	GO:0005310	dicarboxylic acid transmembrane transporter activity	0.000926
CC	GO:0005886	plasma membrane	6.78E-06
CC	GO:0044459	plasma membrane part	0.000278
CC	GO:0005833	hemoglobin complex	0.000466
CC	GO:0016020	membrane	0.000926

Table 5.1.26a: *GO Terms for Down-regulated Genes. Possible compared to controls.*

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0010467	gene expression	5.97E-21
BP	GO:0043170	macromolecule metabolic process	4.03E-19
BP	GO:0043283	biopolymer metabolic process	7.44E-19
BP	GO:0044260	cellular macromolecule metabolic process	8.39E-19
BP	GO:0034960	cellular biopolymer metabolic process	1.01E-18
BP	GO:0044237	cellular metabolic process	1.33E-17
BP	GO:0008152	metabolic process	7.12E-17

Table 5.1.26b: ***GO Terms for Down-regulated Genes. Possible compared to controls.***			
BP	GO:0044238	primary metabolic process	3.36E-16
BP	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.97E-15
BP	GO:0016070	RNA metabolic process	3.96E-13
BP	GO:0009059	macromolecule biosynthetic process	8.04E-13
BP	GO:0034645	cellular macromolecule biosynthetic process	9.14E-13
BP	GO:0034961	cellular biopolymer biosynthetic process	1.28E-12
BP	GO:0043284	biopolymer biosynthetic process	1.28E-12
BP	GO:0006396	RNA processing	1.36E-11
BP	GO:0006412	translation	1.52E-11
BP	GO:0044249	cellular biosynthetic process	7.24E-10
BP	GO:0009058	biosynthetic process	9.33E-10
BP	GO:0022613	ribonucleoprotein complex biogenesis	1.02E-07
BP	GO:0006605	protein targeting	1.20E-07
BP	GO:0060255	regulation of macromolecule metabolic process	6.95E-07
BP	GO:0034660	ncRNA metabolic process	9.39E-07
BP	GO:0042110	T cell activation	1.21E-06
BP	GO:0006397	mRNA processing	1.80E-06
BP	GO:0010468	regulation of gene expression	1.88E-06
BP	GO:0008380	RNA splicing	2.27E-06
BP	GO:0010556	regulation of macromolecule biosynthetic process	4.21E-06
BP	GO:0009987	cellular process	5.27E-06
BP	GO:0019222	regulation of metabolic process	6.81E-06
BP	GO:0016071	mRNA metabolic process	8.62E-06
BP	GO:0034470	ncRNA processing	8.75E-06
BP	GO:0046649	lymphocyte activation	1.18E-05
BP	GO:0006399	tRNA metabolic process	1.65E-05
BP	GO:0006350	transcription	1.82E-05
BP	GO:0031323	regulation of cellular metabolic process	1.85E-05
BP	GO:0009889	regulation of biosynthetic process	2.67E-05
BP	GO:0031326	regulation of cellular biosynthetic process	3.49E-05
BP	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	5.62E-05
BP	GO:0006351	transcription, DNA-dependent	7.20E-05
BP	GO:0032774	RNA biosynthetic process	8.01E-05
BP	GO:0017038	protein import	8.10E-05
BP	GO:0034621	cellular macromolecular complex subunit organization	9.50E-05
BP	GO:0043933	macromolecular complex subunit organization	9.51E-05
BP	GO:0045321	leukocyte activation	0.00011
BP	GO:0065003	macromolecular complex assembly	0.00011
BP	GO:0051252	regulation of RNA metabolic process	0.000111
BP	GO:0045449	regulation of transcription	0.00012
BP	GO:0006996	organelle organization	0.000138
BP	GO:0043368	positive T cell selection	0.000156

Table 5.1.26c: *GO Terms for Down-regulated Genes. Possible compared to controls.*			
BP	GO:0006607	NLS-bearing substrate import into nucleus	0.000201
BP	GO:0006355	regulation of transcription, DNA-dependent	0.000208
BP	GO:0031341	regulation of cell killing	0.000231
BP	GO:0006886	intracellular protein transport	0.000291
BP	GO:0006414	translational elongation	0.000292
BP	GO:0022618	ribonucleoprotein complex assembly	0.000294
BP	GO:0006974	response to DNA damage stimulus	0.00031
BP	GO:0008033	tRNA processing	0.000331
BP	GO:0034622	cellular macromolecular complex assembly	0.000337
BP	GO:0034613	cellular protein localization	0.000348
BP	GO:0031343	positive regulation of cell killing	0.000364
BP	GO:0001775	cell activation	0.000408
BP	GO:0022607	cellular component assembly	0.000529
BP	GO:0006913	nucleocytoplasmic transport	0.000597
BP	GO:0010608	posttranscriptional regulation of gene expression	0.000638
BP	GO:0006606	protein import into nucleus	0.000638
BP	GO:0006413	translational initiation	0.000687
BP	GO:0051170	nuclear import	0.000738
BP	GO:0016043	cellular component organization	0.000744
BP	GO:0051169	nuclear transport	0.000748
BP	GO:0034984	cellular response to DNA damage stimulus	0.000805
BP	GO:0000018	regulation of DNA recombination	0.000878
MF	GO:0003676	nucleic acid binding	5.71E-11
MF	GO:0003723	RNA binding	1.57E-10
MF	GO:0003735	structural constituent of ribosome	3.96E-07
MF	GO:0008270	zinc ion binding	2.32E-06
MF	GO:0005488	binding	7.80E-06
MF	GO:0005515	protein binding	2.35E-05
MF	GO:0046914	transition metal ion binding	3.92E-05
MF	GO:0003743	translation initiation factor activity	0.000107
MF	GO:0045182	translation regulator activity	0.00017
MF	GO:0004549	tRNA-specific ribonuclease activity	0.000531
CC	GO:0005622	intracellular	1.65E-29
CC	GO:0044424	intracellular part	2.77E-29
CC	GO:0043231	intracellular membrane-bounded organelle	2.28E-25
CC	GO:0043227	membrane-bounded organelle	2.74E-25
CC	GO:0043229	intracellular organelle	2.86E-23
CC	GO:0043226	organelle	3.22E-23
CC	GO:0005634	nucleus	1.96E-21
CC	GO:0030529	ribonucleoprotein complex	4.04E-14
CC	GO:0005840	ribosome	1.84E-11
CC	GO:0044428	nuclear part	3.02E-11
CC	GO:0032991	macromolecular complex	3.25E-11
CC	GO:0005739	mitochondrion	5.76E-11
CC	GO:0044464	cell part	1.60E-10
CC	GO:0005623	cell	1.71E-10

Table 5.1.26d: *GO Terms for Down-regulated Genes. Possible compared to controls.*			
CC	GO:0044429	mitochondrial part	2.04E-08
CC	GO:0005737	cytoplasm	5.35E-08
CC	GO:0070013	intracellular organelle lumen	7.72E-08
CC	GO:0031975	envelope	1.38E-07
CC	GO:0031967	organelle envelope	2.40E-07
CC	GO:0044422	organelle part	2.79E-07
CC	GO:0043233	organelle lumen	2.88E-07
CC	GO:0044446	intracellular organelle part	3.81E-07
CC	GO:0031974	membrane-enclosed lumen	8.48E-07
CC	GO:0015935	small ribosomal subunit	1.43E-06
CC	GO:0031981	nuclear lumen	1.55E-06
CC	GO:0044444	cytoplasmic part	1.81E-06
CC	GO:0005829	cytosol	1.85E-06
CC	GO:0031966	mitochondrial membrane	7.04E-06
CC	GO:0033279	ribosomal subunit	8.23E-06
CC	GO:0000313	organellar ribosome	1.08E-05
CC	GO:0005761	mitochondrial ribosome	1.08E-05
CC	GO:0005654	nucleoplasm	2.90E-05
CC	GO:0005740	mitochondrial envelope	4.00E-05
CC	GO:0005759	mitochondrial matrix	4.78E-05
CC	GO:0031980	mitochondrial lumen	4.78E-05
CC	GO:0042101	T cell receptor complex	9.20E-05
CC	GO:0044455	mitochondrial membrane part	0.000194
CC	GO:0043234	protein complex	0.000211
CC	GO:0044451	nucleoplasm part	0.000213
CC	GO:0005743	mitochondrial inner membrane	0.000226
CC	GO:0044445	cytosolic part	0.000244
CC	GO:0019866	organelle inner membrane	0.000255
CC	GO:0005681	spliceosome	0.000269
CC	GO:0022627	cytosolic small ribosomal subunit	0.00051
CC	GO:0005643	nuclear pore	0.000602
CC	GO:0022626	cytosolic ribosome	0.000602
CC	GO:0005635	nuclear envelope	0.000631
CC	GO:0000314	organellar small ribosomal subunit	0.000893
CC	GO:0005763	mitochondrial small ribosomal subunit	0.000893
CC	GO:0042613	MHC class II protein complex	0.000899

KEGG Pathway Enrichment

There were found to be 7 enriched KEGG pathways in significant up-regulated genes (displayed in Table 5.1.27) and 24 enriched KEGG pathways in significant down-regulated genes (displayed in Table 5.1.28).

Table 5.1.27: ***KEGG Pathways for Up-Regulated Genes. Possible compared to controls.***

Pathway ID	Pathway Description.	p Value
860	Porphyrin and chlorophyll metabolism	0.005517
4520	Adherens junction	0.008852
4140	Regulation of autophagy	0.021283
4810	Regulation of actin cytoskeleton	0.02878
5213	Endometrial cancer	0.03712
2010	ABC transporters - General	0.04242
5223	Non-small cell lung cancer	0.04514

Table 5.1.28: ***KEGG Pathways for Down-Regulated Genes. Possible compared to controls.***

Pathway ID	Pathway Description	p Value
4660	T cell receptor signalling pathway	2.56E-05
5330	Allograft rejection	0.000117
5012	Parkinson's disease	0.000535
4514	Cell adhesion molecules (CAMs)	0.000811
790	Folate biosynthesis	0.001682
5332	Graft-versus-host disease	0.001682
5010	Alzheimer's disease	0.001832
4940	Type I diabetes mellitus	0.001908
190	Oxidative phosphorylation	0.005443
5340	Primary immunodeficiency	0.006211
5320	Autoimmune thyroid disease	0.006402
3010	Ribosome	0.006488
51	Fructose and mannose metabolism	0.00767
970	Aminoacyl-tRNA biosynthesis	0.007771
3050	Proteasome	0.009325
240	Pyrimidine metabolism	0.010528
310	Lysine degradation	0.010622
632	Benzoate degradation via CoA ligation	0.016508
5310	Asthma	0.020213
620	Pyruvate metabolism	0.022467
290	Valine, leucine and isoleucine biosynthesis	0.023053
4612	Antigen processing and presentation	0.028692
20	Citrate cycle (TCA cycle)	0.029485
3020	RNA polymerase	0.029943

Largest Fold Changes

Before statistical filtering there were 1431 features exhibiting greater than 2-fold up-regulation and 1242 features exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 392 features with greater than 2-fold up-

regulation and 547 with greater than 2-fold down regulation. The twenty features with the largest fold changes are described in tables 5.1.29 for up-regulated genes and 5.1.30 for down-regulated genes (the full lists can be found on the enclosed CD).

Table 5.1.29: ***Table of Fold Changes > 2 for Up-Regulated Genes. Possible compared to controls. (Each line represents 1 feature on array)***

Feature ID	Description	Pathway Description	Fold change
GE79550	histone cluster 1, H2ad	Systemic lupus erythematosus	6.341961
GE56370	transmembrane protein 158	-	5.710848
GE782559	insulin-like growth factor 2 (somatomedin A)	-	5.653595
GE59782	glycophorin E	-	5.12556
GE59884	Rh blood group, CcEe antigens	-	4.928375
GE81117	glycophorin B (MNS blood group)	-	4.921044
GE53214	transmembrane and coiled-coil domain family 2	-	4.880993
GE498697	tripartite motif-containing 10	-	4.754302
GE59837	myosin, light chain 4, alkali; atrial, embryonic	-	4.635194
GE56530	microphthalmia-associated transcription factor	Melanogenesis, Melanoma	4.565678
GE55112	NA	-	4.537659
GE56647	RUN domain containing 3A	-	4.473495
GE501834	erythrocyte membrane protein band 4.2	-	4.447314
GE57937	carbonic anhydrase I	Nitrogen metabolism	4.413507
GE846526	pre-B-cell leukemia homeobox 1	-	4.361147
GE80892	erythrocyte membrane protein band 4.2	-	4.335315
GE490042	zinc finger, CCHC domain containing 5	-	4.292945
GE905584	thioesterase superfamily member 5	-	4.280413
GE56496	tensin 1	-	4.269189
GE60301	chloride intracellular channel 2	-	4.249896

Table 5.1.30: ***Table of Fold Changes > 2 for Down-Regulated Genes. Possible compared to controls (Each line represents 1 feature on array)***

Feature ID	Description	Pathway Description	Fold Change
GE79288	lectin, galactoside-binding, soluble, 2	-	-5.20748
GE59630	interleukin 23, alpha subunit p19	Cytokine-cytokine receptor interaction, Jak-STAT signalling pathway	-4.49206
GE62103	tubulin polymerization-promoting protein family member 3	-	-4.02624
GE62569	signal-regulatory protein gamma	-	-3.97887
GE82786	poliovirus receptor related immunoglobulin domain containing	-	-3.84301
GE80046	zeta-chain (TCR) associated protein kinase 70kDa	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway, Primary immunodeficiency	-3.83167
GE481282	hypothetical protein LOC129293	-	-3.81244
GE53976	ABI family, member 3	-	-3.75101
GE61110	interleukin 32	-	-3.6856
GE85839	NA	-	-3.64741
GE60308	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2	-	-3.60975
GE59854	transcription factor 7 (T-cell specific, HMG-box)	Wnt signalling pathway, Adherens junction, Melanogenesis, Colorectal cancer, Endometrial cancer, Prostate cancer, Thyroid cancer, Basal cell carcinoma, Acute myeloid leukemia	-3.60748
GE81458	CD8b molecule	Cell adhesion molecules (CAMs), Antigen processing and presentation, Hematopoietic cell lineage, T cell receptor signalling pathway, Primary immunodeficiency	-3.50889
GE655722	GRB2-related adaptor protein	-	-3.50407
GE82784	protein kinase C, eta	Tight junction	-3.49746
GE59645	adenosine deaminase	Purine metabolism, Primary immunodeficiency	-3.49409
GE60283	src kinase associated phosphoprotein 1	-	-3.43815
GE80132	family with sequence similarity 113, member B	-	-3.42685
GE60502	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	-	-3.40997
GE527765	agmatine ureohydrolase (agmatinase)	Urea cycle and metabolism of amino groups	-3.39487

The full lists of genes with fold changes of greater than 2 and p value of less than 0.001 were examined further using DAVID 2008 (348, 349). The list of significant GO terms for up-regulated genes is displayed in table 5.1.31 and for down-regulated genes in table 5.1.32 (a and b).

Table 5.1.31: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Up-Regulated Genes. Possible compared to control***

Ontology	GO Term	No. of Genes (from 392)	p value
MF	Transporter activity	36	3.0E-4
MF	Active transmembrane transporter activity	15	6.8E-4
MF	Secondary active transmembrane transporter activity	11	9.2E-4

Table 5.1.32a: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Down-Regulated Genes. Possible compared to control***

Ontology	GO Term	No. of Genes (from 392)	p value
BP	Primary metabolic process	223	1.1E-7
BP	Lymphocyte activation	18	4.3E-7
BP	Macromolecule metabolic process	198	5.7E-7
BP	T cell activation	14	1.2E-6
BP	Biopolymer metabolic process	155	2.4E-6
BP	Leukocyte activation	18	2.5E-6
BP	Cellular metabolic process	217	2.8E-6
BP	Metabolic process	235	3.3E-6
BP	Regulation of lymphocyte activation	11	1.0E-5
BP	Cell activation	18	1.3E-5
BP	Regulation of cell activation	11	1.7E-5
BP	Protein metabolic process	112	2.4E-5
BP	Cellular process	301	2.8E-5
BP	Regulation of T cell activation	301	5.9E-5
BP	Gene expression	108	1.7E-4
BP	Biopolymer modification	63	2.4E-4
BP	Cellular protein metabolic process	102	2.9E-4

Table 5.1.32b: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Down-Regulated Genes. Possible compared to control***			
BP	Peptidyl-amino acid modification	11	3.1E-4
BP	Cellular macromolecule metabolic process	103	3.2E-4
BP	Positive regulation of biological process	42	3.4E-4
BP	Positive regulation of cellular process	38	6.4E-4
BP	Protein modification process	59	7.8E-4
CC	Immunological synapse	11	4.6E-12
CC	Intracellular part	269	3.3E-7
CC	Intracellular	281	5.9E-7
CC	T cell receptor complex	6	4.0E-6
CC	Cell part	355	2.5E-5
CC	Cell	355	2.6E-5
CC	Cytoplasm	179	2.9E-5
CC	Nuclear part	43	2.7E-4
CC	Membrane-bound organelle	194	6.4E-4
CC	Receptor complex	9	7.2E-4
CC	Ribonucleoprotein complex	27	8.0E-4
CC	Intracellular membrane-bound organelle	193	9.0E-4
MF	Protein binding	210	1.5E-9
MF	Binding	311	7.3E-6
MF	Identical protein binding	20	5.3E-4
MF	Protein dimerization activity	18	6.7E-4

Discussion

When the population characteristics displayed in table 5.1.4 are considered, overall the infected group can be seen to be more premature and therefore of lower birth weight and also older in terms of postnatal days when sampled than the controls. This was partly because in the early stages of the study all of the controls were term babies attending the neonatal unit for blood screening to exclude thyroid or adrenal problems in the early days of life. In the latter stages of the study care was taken to ensure that the control group also included preterm infants and infants more than a few days old. The possible infection group was also more premature than the control

group but less so than the infected group. The possible infection group tended to be younger in terms of number of postnatal days than either the infected or control groups. These findings may be a reflection of the low threshold for investigation for infection in premature infants during the early days of life and the relatively low yield of microbiological investigations in this group. It is not surprising that premature infants account for a significant proportion of those sampled due to suspicion of infection because these are the infants that are more at risk for infection and who make up a large proportion of the population of the neonatal unit. The inclusion of a wide range of gestations in the control group and the fact that the infected and possible infected samples are representative of the neonatal population most at risk of infection mean that the comparisons between groups presented in this chapter are valid and useful.

The particular organisms responsible for the infections in the infected group are listed in table 5.1.6. The particular infections included were always going to be dependent on the infections encountered during the study period. As would be expected, not all infants who had clinical samples taken for suspected infection would have had study samples taken due to absence of parental consent, insufficient blood for the study sample or oversight on the part of the clinician taking the sample. It is therefore reassuring to note that the range of organisms in our infected group is representative of that encountered in the neonatal unit. I was very careful to ensure that all of the infected samples included were clearly infected and supporting evidence for inclusion of some of the samples is presented in table 5.1.7. In addition, the decision for inclusion or exclusion of culture-positive samples in the infected

group was independently verified by a neonatal consultant (Ben Stenson). Any sample where there was doubt was not included in the infected group and not run on microarray.

The Genespring expression profile (figure 5.1.3) for significantly differentially expressed genes between the infected and control group (fold change greater than 2 and $p < 0.01$) is very encouraging. There is a discernable difference in response seen between infected and control infants. What is particularly interesting and exciting is that the possible infection groups seems to be composed of some samples that look more like the control samples and others that look more like the infected group. This looks promising in terms of future studies of gene expression profiling being able to distinguish between infected and non-infected infants.

Infected compared to control samples

From the examination of significantly differentially expressed genes ($p < 0.01$) for the infected compared to control group, a notable proportion of significantly up-regulated GO terms in the infected group are immune related (table 5.1.9). This is especially true for the terms that are most significantly enriched but there are immune related terms throughout the list. The most enriched terms included response to wounding, immune system process, response to external response, immune response, defence response, inflammatory response and response to stress. There is also enrichment of immune-related pathways within the up-regulated genes on KEGG analysis (table 5.1.11), in particular cytokine-cytokine receptor interaction,

complement and coagulation pathways and TLR signalling. This concurs with the findings of Wong and colleagues when comparing paediatric patients with septic shock with controls (152). On examination of the GO terms for the down-regulated genes (table 5.1.10), there are immune-related terms but not in such a sizeable proportion than was seen in the up-regulated terms. The down-regulated immune terms include antigen processing via MHC II, T cell/leukocyte/lymphocyte activation and MHC II receptor activity. KEGG analysis (table 5.1.12) mapped down-regulated genes to T cell receptor signalling, antigen processing/presentation and natural killer cell mediated cytotoxicity related pathways. This again concurs with findings in the literature with respect to down-regulated genes in paediatric septic shock (152, 153, 282). The enriched down-regulated KEGG pathways also included terms related to allergy, autoimmunity and graft versus host disease. Simplistically, it may be that such immune processes are damped down during infection to enable the immune system to concentrate on responding to the invading pathogen.

When fold change is taken into consideration, and genes with fold change of greater than 2 for significance level of $p < 0.001$ are examined, there are both up- and down-regulated genes related to immune functions. For up-regulated genes (table 5.1.13) these included genes related to cluster of differentiation molecules CD177 (human neutrophil alloantigen), CD64 (related to the Fc portion of IgG) and CD 55 (decay accelerating factor for complement). Up-regulated genes related to cytokines and cytokine receptors included those for TNFα induced protein 6, IL-1receptor antagonist, IL-1 receptor associated kinase 3 and IL-17 receptor A. In addition there was up-regulation of genes involved in the interferon response, of TLR 8, of

chemokine (C-C) ligands 3 and 4 and of proteoglycan 2 (involved in NK activation). Down-regulated genes included cluster of differentiation molecules involved in T cell function, NK cell mediated immunity and antigen presentation (CD1c, CD3(d,z) CD7, CD 8b, CD28, CD40 ligand, CD96, CD 226, CD 247 and CD 248). Down-regulated genes related to cytokines and cytokine receptors included IL-23 (involved in cytokine interactions and the Jak-Stat pathway), IL-7 receptor, IL-21 receptor and TNF receptor associated factor 1. Chemokine (C-C) ligands 1 and 9 were also down-regulated. When the whole list of genes with fold change > 2 and $p < 0.001$ is further examined by gene ontology using the online tool DAVID 2008 (348, 349) it is clear that there are a notable proportion of immune related processes related to these genes. The up-regulated genes ontology (table 5.1.15) showed enrichment of terms such as response to wounding, response to external stimulus, inflammatory response, response to stress, response to other organisms and defence response. The down-regulated genes (table 5.1.16) showed enrichment of terms related to T cell/lymphocyte activation and differentiation and MHC II receptor activity. This again concurs with findings in the literature (152, 153).

I have chosen a few of the genes listed above to consider in light of their function and in the context of existing literature. This is not meant to be an exhaustive examination of genes involved, but rather to help put the findings into context. In terms of the up-regulated genes, TNFα induced protein 6 was found to be part of the common response to infection by Huang and colleagues in their study of dendritic cells (145). TNF is pro-inflammatory and is itself elevated in infection (1, 45). TLR 8 has been previously been described as being up-regulated in surgical patients with

SIRS who go on to develop sepsis (156). Interleukin-1 receptor antagonist has previously been described as being elevated in paediatric sepsis (350) and prior to the diagnosis of infection in neonates (351). Chemokine (C-C) ligands 3 and 4 have been described as being up-regulated in the common host response to infection as described by both Hossain and Jenner (146, 159). These are chemokines that are pro-inflammatory and involved in chemotaxis. CD 55 antigen (also known as DAF) is involved in protecting the host from tissue damage due to complement. For the down-regulated genes, CD28 antigen has previously been noted to be down-regulated in surgical patients with SIRS which was subsequently found to be sepsis (156). Neonatal monocytes have been previously shown to have lower expression of MHC II in response to bacterial antigens compared to adult monocytes (78). Neonatal monocytes have also been shown to be selectively defective in MHC II antigen presentation not related to expression (352). In particular HLA-DR expression has been described as being reduced in infected neonates (353-355).

While there is a paucity of data in the literature describing the neonatal host response to infection, it is really exciting that many of the results presented in this chapter corroborate work found in the literature regarding other groups of patients with infection. Where I have not referenced other groups with respect to our findings it is because I found nothing in the literature regarding these genes in this context. These previously undecribed genes are likely to reflect results that will be corroborated as more studies emerge and some unique neonatal responses. It is gratifying that our results make physiological sense and affirm that neonates are capable of mounting a substantial immune response to infection. Our results show differential expression

not only of genes involved in a wide range of the innate immune responses (cytokines, chemokines, Toll like receptors, complement, natural killer cells) but also genes involved in T cell and antibody responses.

In terms of the findings above, it could be argued that the infected group were more premature than the control group and that some of the differences seen could be gestation-related rather than infection-related. Given that immunological immaturity is more marked in premature infants and given that immune-related gene expression changes are expected in infection, I would expect that the findings are genuinely infection related. Studies of the effect of gestational age on gene expression are however needed to be completely sure of this and would likely enhance understanding of neonatal immunology greatly.

Infected compared to possible infected group

From the examination of significantly differentially expressed genes ($p < 0.01$) for the infected compared to possible infection group, a notable proportion of significantly up-regulated GO terms in the infected group are immune related (table 5.1.17). GO terms including immune system process, immune response, response to wounding, regulation of immune response and leukocyte activation are most enriched but there are several other cytokine related, interferon related, B cell related and innate immune response related terms that are also significantly enriched. KEGG pathway analysis of the up-regulated genes (table 5.1.19) mapped them to complement and coagulation pathways and TLR related pathways. Significantly

down-regulated genes did not have obvious immune related functions when enriched GO terms (table 5.1.18) or KEGG analysis results (table 5.1.20) were examined.

When fold change is taken into consideration and genes with a fold change of greater than 2 at significance level $p < 0.001$ are examined, some up-regulated genes can be immediately recognised as being linked to immune function but this is not the case for down-regulated genes (tables 5.1.21 and 5.1.22). Examples of immune related genes in the up-regulated groups include TNFα induced protein 6, interferon induced protein, C-X-C ligand 10, Fc of IgG, IL-1 receptor antagonist, NFkB, and C-C ligand 4. When these genes are examined by gene ontology using DAVID it becomes even more apparent that the up-regulated genes (table 5.1.23) can be linked to immune function (response to stress, response to wounding, response to external stimulus, immune system process) but this is still not obviously the case for down-regulated genes (table 5.1.24).

The implication from this is that the infected group are showing more in the way of immune response than the possible infection group. As the possible group is probably a mix of cases of infection and cases of non-infection this would make sense. The lack of comparative down-regulation of immune related genes suggests that the possible infection group is different from the control group.

Possible infection compared to control group

From the examination of significantly differentially expressed genes ($p < 0.01$) for the possible infection compared to control group, a small proportion of significantly

down-regulated GO terms in the possible infection group are immune related (table 5.1.26) but this is not the case for significantly up-regulated genes (table 5.1.25). The down-regulated terms that are immune related include T cell activation, leukocyte activation and MHC class II. Overall, in comparison with the results already examined above there is a marked absence of immune related enriched ontology terms.

There are no clearly immune-related pathways enriched for the up-regulated genes on KEGG analysis (table 5.1.27). For the down-regulated genes (table 5.1.28) the immune related pathways include T cell receptor signalling, primary immune deficiency and antigen presentation and processing. There is also down-regulation of pathways related to autoimmune disease such as diabetes and thyroid disease and of graft versus host disease.

When fold change is taken into consideration and genes with a fold change of greater than 2 at significance level $p < 0.001$ are examined, none of the gene descriptions for the up-regulated genes (table 5.1.29) are immediately recognisable as immune related although on examination of the full list (on enclosed CD) IL-28a (interferon related) and TLR4 were up-regulated. Some down-regulated genes (table 5.1.30) are readily seen to be involved in immune function. These genes include IL-23 (involved in cytokine receptor activation and Jak Stat pathways), IL-32, CD 7 (T and B cell interaction), CD8b (involved in antigen processing and presentation, T cell signalling), CD 28 (T cell signalling) and CD 248. When DAVID was used to examine these genes for ontology, there were no immune related terms enriched in

up-regulated genes (table 5.1.31) but terms related to T cell activation, T cell regulation and T cell receptor were enriched in down-regulated genes (table 5.1.32).

These findings again make the possible infection group look like something in between the control and infected groups. This is as would be expected as the possible infected group is heterogeneous and likely to included both infected and non-infected infants. Ideally, if a set of candidate genes for use in a diagnostic tool could be identified, the possible infected group could be used as a training set to try and differentiate into infected and uninfected groups.

Summary

Overall, despite the concerns regarding the data quality, it can still be clearly seen that there is immune-related differential gene expression between infected and control infants. This is very encouraging. Up-regulated genes in the infected group include those involved in cytokine, complement, interferon and TLR related processes. Down-regulated genes in the infected group include genes involved in antigen processing, MHC II activity and T cell activation and signalling. It is exciting that the possible infection group appears to be made up of some samples which have gene expression profiles looking more like the infected group and others more like the control group (as seen in the Genespring diagram, figure 5.1.3). This bodes well for the use of these samples as a training set in the future and looks promising in terms of being able to distinguish infected from non-infected infants.

Conclusions

Infected infants show significant up-regulation and down-regulation of genes involved in immune processes when compared to controls. These results provide corroboration of some already published findings for adult and paediatric populations. The results from this chapter may provide useful corroboration of results obtained from other microarray platforms.

Section 5.2: Investigation of Neonatal Infection Using Illumina® Microarrays

Background

Following the problems experienced with the CodeLink™ platform it was decided, where possible, to hybridise the samples again using an Illumina® platform. This decision provided several advantages. Temporally the hybridisations would be performed together therefore minimising technical variation, the samples would be run on one batch of arrays minimising inter-array variation and it would be possible to compare results with those achieved using CodeLink™ for corroboration. In addition, the hybridisation and scanning of the arrays would be performed by a single operator, so any inter-operator variability would be eliminated.

The Division of Pathway Medicine is not currently running Illumina® arrays and therefore the hybridisation and scanning of the arrays was contracted out to a company: Gen-Probe, formerly Tepnel.

Ideally, I would have liked to have run the same samples on Illumina® as those that had been run on CodeLink™, i.e. 27 positive, 27 control and 30 possible infection samples. This would have allowed direct comparison between platforms and any corroboration of results would have been very encouraging. Unfortunately this was not possible for a number of reasons. Firstly, for cost and platform reasons we were able to commission the running of only 72 samples. This meant that we were not able to run the requisite target of at least 27 for each subgroup of infected, control and possible infection groups. It was therefore decided that none of the "possible

infection" samples would be run in order to concentrate on identifying differences between positives and controls. Secondly, six of the original samples (3 controls, 3 infected) had been used up in previous work and were therefore not available to run. However, the Illumina® platform required a smaller starting amount of RNA, which gave the potential to run some samples that had been too small to run on Codelink™.

Aim

To examine differences in blood RNA expression profiles between infected and control infants using Illumina® microarrays.

Acknowledgements

RNA extractions of the samples in the final dataset were performed by myself (32), Alan Ross (30) and by Paul Dickinson (1). Quality control was carried out initially by myself and Alan Ross. Paul Dickinson ran repeat quality control steps for many of the samples included in this analysis.

Marie Craigon and Paul Dickinson adjusted the volume/concentration of any samples that did not meet the starting requirements set out by Gen-Probe.

Hybridisation and scanning of the arrays was carried out by Gen-Probe.

Microarray data was sent to Al Ivens of FIOS Genomics for data quality control, data filtering and normalisation, statistical comparisons and functional analysis. The results presented here include data from FIOS Genomics along with analysis and interpretation of the results that I have carried out using the raw and normalised data.

Methods

RNA that had previously been collected and extracted for the Codelink™ study was used. Although quality control steps had been carried out previously any samples that had been run on the BioMate spectrophotometer were re-run using the Nanodrop spectrophotometer in order to maximise uniformity of data.

37 infected samples (using the definition on page 236) and 36 control samples were selected. They were transported on ice directly to Gen-Probe's laboratory where further quality control steps were carried out and the samples run on Illumina® Human Whole-Genome Expression BeadChips each comprising 48802 features. Raw data obtained from Gen-Probe was then passed to FIOS genomics for analysis. In addition, the raw data was used to perform unsupervised analysis and I have presented the unsupervised results below before presenting the filtered data.

The quality control analysis carried out by FIOS genomics used the array quality metrics package in Bioconductor (356). Arrays were scored using maplot, boxplot and heatmap. Although the arrays contained 48802 features, many of these features did not give a signal on any of the arrays and were therefore removed from analysis. This left 23342 features for which normalisation was carried out. The raw data was transformed using a variance stabilising transformation prior to being subjected to robust spline normalisation resulting in intensities in log base 2. In addition, a gender check of each sample was performed using Y-chromosome specific loci to identify male samples – all of the samples were correctly assigned.

FIOS Genomics carried out a series of comparisons using linear modelling and subsequent empirical Bayesian approaches. This included vertical p value adjustment for multiple testing (within a given comparison) in order to control for false discovery rate (Benjamini Hochberg). The Bioconductor package limma was used (347). The statistical significance level chosen for carrying data on for functional analysis was adjusted $p < 0.001$. The functional analysis of significant genes found in these comparisons was carried out by examining KEGG pathway membership and Gene Ontology (GO). Assessment for up and down regulation was carried out for enrichment p values of < 0.05 for KEGG and of < 0.001 for GO.

Results

Samples: Quality control and exclusion

73 samples were sent to Gen-Probe – 36 controls and 37 positives (two positives were very small and only sufficient for one reaction, the extra positive was sent as back up). A minimum starting amount of 50 ng in 11 µl was required; we aimed for 100 ng in 11 µl and sent 250 ng of each sample where possible. Despite several attempts, two samples did not generate a RIN – the bioanalyzer traces for each case were felt to be acceptable however. The RIN values for each sample are summarised in table 5.2.1. The range of RIN values for the samples selected was 5 to 10 (those with values of less than 8 had the bioanalyzer graphs reviewed and were felt to be acceptable). The mean RIN of the 71 samples for which an RIN was generated was 8.7 and the median 9.

Table 5.2.1: ***RIN values of samples sent for hybridisation***

Key:
Final dataset i.e. run on array and data analysed
Sample sent to Gen-probe but not run on array
Sample run on array but excluded from data analysis

Sample	Category	RIN	Sample	Category	RIN
csb001	Control	10	**csb009**	infected	9.1
csb017	Control	8.5	**csb016**	infected	9.5
csb021	Control	9.5	**csb047**	infected	8.9
csb022	Control	8.8	**csb075**	infected	8.7
csb028	Control	9.4	**csb082**	infected	8.6
csb029	Control	8.2	**csb083**	infected	8.5
csb040	Control	7.2	**csb084**	infected	Null
csb042	Control	8.5	**csb089**	infected	9.6
csb043	Control	9.4	**csb091**	infected	9.4
csb046	Control	8.8	**csb098**	infected	7.8
csb048	Control	8.3	**csb102**	infected	9.1
csb049	Control	9	**csb107**	infected	5
csb051	Control	9.2	**csb108**	infected	5.9
csb058	Control	9.3	**csb111**	infected	8.4
csb062	Control	9.4	**csb112**	infected	7.1
csb063	Control	9.1	**csb114**	infected	7.9
csb065	Control	9.3	**csb116**	infected	9.7
csb067	Control	9.4	**csb118**	infected	9
csb069	Control	9.6	**csb119**	infected	7.4
csb071	Control	7.2	**csb125**	infected	5
csb072	Control	9.6	**csb132**	infected	9.5
csb073	Control	9.6	**csb133**	infected	8.8
csb077	Control	8	**csb137**	infected	9.3
csb079	Control	9.3	**csb138**	infected	9.3
csb081	Control	8.8	**csb145**	infected	9.7
csb086	Control	8.8	**csb149**	infected	9.6
csb087	Control	10	**csb152**	infected	8.7
csb096	Control	8.6	**csb155**	infected	9.2
csb097	Control	7.9	**csb157**	infected	9.4
csb165	Control	9.2	**csb159**	infected	9.4
csb179	Control	9.2	**csb161**	infected	9.1
csb181	Control	9.4	**csb162**	infected	8.5
csb200	Control	9.3	**csb164**	infected	7.3
csb202	Control	8.8	**csb191**	infected	6.5
csb206	Control	8	**csb198**	infected	9.6
csb207	Control	9.1	**csb201**	infected	Null
			csb203	infected	7.2

84 microarray scan results were generated by Gen-Probe from 70 samples (9 samples were run on either 2 or 3 scans as technical replicates) and had data quality control carried out. Those samples for which we did not receive array results presumably failed Gen-Probe's quality control checks or failed to hybridise.

10 scan results failed on two or three of the quality control metrics – maplot, boxplot and heatmap – and were therefore excluded from analysis. 4 scan results were removed from analysis as they exhibited a mean correlation, compared with all of the other samples, of less than 0.65. 2 further scan results were removed on the basis of poor sample relation QC plots. Sample csb201 was excluded from analysis at my request. Although it had been chosen to fill a space on the arrays for interest it did not fully fit the criteria because the only positive culture was a clean catch urine culture. As culture of clean catch urine in neonates can easily be affected by contaminants it could not be considered a robust infected sample but neither was it a control. (Interestingly, it clustered with the controls on analysis of raw data –data not shown). Where more than one scan was received for a sample, the scan with the highest mean correlation to all of the other samples was retained and the others were discarded prior to normalisation. In other words, no technical replicates were used for analysis. This gave a final dataset of 63 samples – 28 infected samples and 35 control samples. The 28 infected samples were taken from 23 infants. 5 infants had two samples included in the analysis – in each case these were samples taken during separate episodes of infection.

Table 5.2.2: *Demographics of Samples Used*		
	Controls (n=35)	**Infected (n=28)**
Male	22 (63 %)	25 (54 %)
Gestational Age (completed weeks)	26-42 (mean 37.9)	24-38 (mean 28.5)
Age (day of postnatal life)	2-56 (mean 11.9)	1-75 (mean 18.2)
Corrected gestational age (weeks + days)	31+1 to 44+1 (median 40+4)	26+1 to 39+2 (median 30+4)
Birthweight (g)	650-4570 (mean 3080, median 3420)	430-3880 (mean 1126, median 963)

Table 5.2.2 displays the gender, age and weight of the infants sampled in each group. The control group had a greater proportion of boys than the infected group with 63 % compared to 54 %. The breakdown of reasons for blood sampling in the control group is presented in table 5.2.3. Of the control group, 2 had previously had a documented episode of infection. 9 had previously had antibiotics but none were on antibiotics at the time of sampling. Within the dataset presented here, none of the control samples were duplicated from the same baby.

Table 5.2.3: *Clinical Reasons for Blood Sampling in Control Group (n)*	
Maternal Thyroid Disease	17
Jaundice	5
Screening for Congenital Adrenal Hyperplasia	2
Coombs positive	1
Urea and electrolytes: repeat after previous abnormality (hypo or hypernatraemia)	3
Guthrie screening	1
"Routine" neonatal screening (preterms)	5
Hypoxic ischaemic encephalopathy	1

The organisms isolated from the infected infants are listed in table 5.2.4. Coagulase negative staphylococci were the organisms most frequently isolated. 26 of the 28 infected samples were bacterial with the exceptions being one case of candida

albicans, although this patient also had Klebsiella isolated from their peritoneal fluid, and one case of cytomegalovirus. 25 of the samples had positive blood cultures, the case of CMV had virus isolated from blood and urine, one case of Coagulase Negative Staphylococcus was an isolate from an abscess (this patient had also had recent Coagulase Negative Staphylococcus in blood culture) and one patient had negative blood cultures but had *Klebsiella* isolated from cerebrospinal fluid.

Table 5.2.4: ***Microorganisms Isolated from Infected Group***

Organism	n	Comments
Coagulase negative staphylococci	15	One of these was an isolate from an abscess
Candida albicans	1	Klebsiella was also isolated from peritoneal fluid
Enterobacter	1	
Enterococci (*E faecalis*)	4	
Group B streptococcus	2	
Pseudomonas aeruginosa	1	Coagulase negative staphylococcus in blood too
Cytomegalovirus (CMV)	1	Isolated in blood and urine
Klebsiella	2	One of these was isolated from CSF only
Escherichia coli	1	

Of the infected group, 14 had previously had a documented episode of infection. Only 2 had not previously had antibiotics. 17 were on antibiotics at the time of sampling. 5 babies each had two samples included in the dataset presented here. These samples were taken on separate occasions of clinical concern of infection. Three of these infants grew Coagulase Negative Staphylococci on both occasions, one grew *Enterococcus faecalis* on both occasions, and the fifth grew Coagulase Negative Staphylococcus on one occasion and *Klebsiella* species on the other. One of the infected infants did not have any antibiotics (the infant with CMV); another had only two days of antibiotics, the remaining 26 infants were treated with 5 to more than 25 days of antibiotics. With regard to the infants with Coagulase Negative

Staphylococcal infection and the infant with CMV, I have presented supporting evidence of infection in table 5.2.5.

Table 5.2.5: ***Supporting Evidence for Infection in Viral and Coagulase Negative Staphylococcal Cases.***

Organism	**Supporting Evidence** (days of antibiotics)
Coagulase Negative Staphylococcus	• Also on long line tip. Pyrexia, tachycardia, respiratory distress, low platelets (>21) • Bradycardias, pyrexia, low platelets (15) • Apnoeas, bradycardia, temperature instability, lethargy (2) • Isolated in 2 blood cultures. Apnoeas, bradycardias, lethargy, poor perfusion – 2 fluid boluses given (9) • Bradycardias, desaturations, long line in situ (11) • Red hot fluctuant abscess (14) • Apnoeas, bradycardias, low temperature, lethargy, low platelets, ventilated, poor colour (18) • Previous isolate, bradycardias, desaturations, mottled (18) • Apnoeas, bradycardias, jaundice, hyperglycaemia (10) • Bradycardias, desaturations, reintubated, deranged sugars, long line, poor handling (10) • Previous isolate, hypotension, hypoglycaemia, oedema, died (16) • Lethargy, wound abscess, central line, recent surgery for NEC, low platelets (16) • Isolated in 2 separate blood cultures taken at same time, bradycardia, desaturations, reintubated, temperature instability (5) • Apnoeas, pyrexia, reduced perfusion, elevated white cell count, infant had two further significant episodes of infection associated with coagulase negative staphylococcus (>10) • Apnoeas, pyrexia, tachypneoa, bradycardia, poor perfusion, poor colour, lethargy, low platelets (6)
Cytomegalovirus	Profound bradycardias and desaturations (nil)

More detailed information for each sample regarding gender, gestation at birth (completed weeks), corrected gestation at time of sampling (completed weeks) and infecting organism (if relevant) is displayed in table 5.2.6 (a and b). The sample numbers that are coloured link the samples that were taken from the same baby.

Table 5.2.6a: ***Gestation, age and condition of each sample.***

Sample	Organsim/Control	Gender	Gestation at birth (weeks)	Corrected Gestation (weeks)
csb001	control	Boy	31	35
csb009	Coagulase negative staphylococcus	Girl	28	33
csb017	control	Boy	40	41
csb021	control	Girl	40	41
csb022	control	Girl	41	41
csb028	control	Girl	36	37
csb029	control	Boy	39	39
csb040	control	Boy	39	40
csb042	control	Boy	42	43
csb043	control	Boy	41	42
csb046	control	Girl	39	40
csb047	Candida	Boy	29	30
csb048	control	Boy	38	39
csb049	control	Girl	41	41
csb051	control	Boy	39	40
csb058	control	Boy	39	40
csb062	control	Girl	41	42
csb063	control	Boy	40	44
csb065	control	Girl	41	42
csb067	control	Boy	39	40
csb069	control	Boy	40	40
csb071	control	Boy	42	43
csb072	control	Boy	40	41
csb073	control	Boy	36	37
csb075	CMV	Girl	26	37
csb077	control	Girl	41	42
csb079	control	Girl	40	40
csb081	control	Girl	41	42
csb082	*Enterobacter cloacae*	Boy	29	30
csb083	*E faecium*	Girl	29	35
csb084	Coagulase negative staphylococcus	Boy	28	30
csb086	control	Boy	39	43
csb087	control	Girl	34	34
csb089	Coagulase negative staphylococcus	Boy	27	28
csb091	Group B Streptococcus	Girl	38	39
csb096	control	Boy	37	38
csb097	control	Boy	42	42
csb102	*Pseudomonas aeuruginosa*	Girl	27	32
csb111	Coagulase negative staphylococcus	Boy	28	30
csb112	Group B Streptococcus	Girl	24	26
csb114	Coagulase negative staphylococcus	Girl	26	27

Table 5.2.6b: *Gestation, age and condition of each sample.*				
Sample	**Organsim/Control**	**Gender**	**Gestation at birth (weeks)**	**Corrected Gestation (weeks)**
csb116	Coagulase negative staphylococcus	Boy	28	30
csb119	Coagulase negative staphylococcus	Girl	28	30
csb125	Coagulase negative staphylococcus	Girl	28	29
csb132	*E faecalis*	Boy	27	30
csb133	Coagulase negative staphylococcus	Boy	27	28
csb138	Coagulase negative staphylococcus	Girl	25	27
csb145	*Enterococcus faecalis*	Boy	27	29
csb149	*Klebsiella* species	Girl	27	29
csb152	Coagulase negative staphylococcus	Boy	28	31
csb157	Coagulase negative staphylococcus	Girl	28	31
csb159	Enterococcus	Boy	30	31
csb162	Coagulase negative staphylococcus	Boy	27	27
csb164	*Klebsiella oxytoca*	Girl	28	32
csb165	control	Girl	28	31
csb181	control	Girl	26	34
csb191	*E coli*	Boy	37	37
csb198	Coagulase negative staphylococcus	Boy	32	33
csb200	control	Boy	32	35
csb202	control	Boy	31	35
csb203	Coagulase negative staphylococcus	Boy	32	39
csb206	control	Boy	29	33
csb207	control	Boy	41	42

Figure 5.2.1 is a boxplot of samples before normalisation and figure 5.2.2 a boxplot of samples after normalisation. This shows good data consistency following normalisation.

Acknowledgement: Figures 5.2.1 and 5.2.2 were generated by Paul Dickinson.

Figure 5.2.1: ***Boxplot of Raw Data***

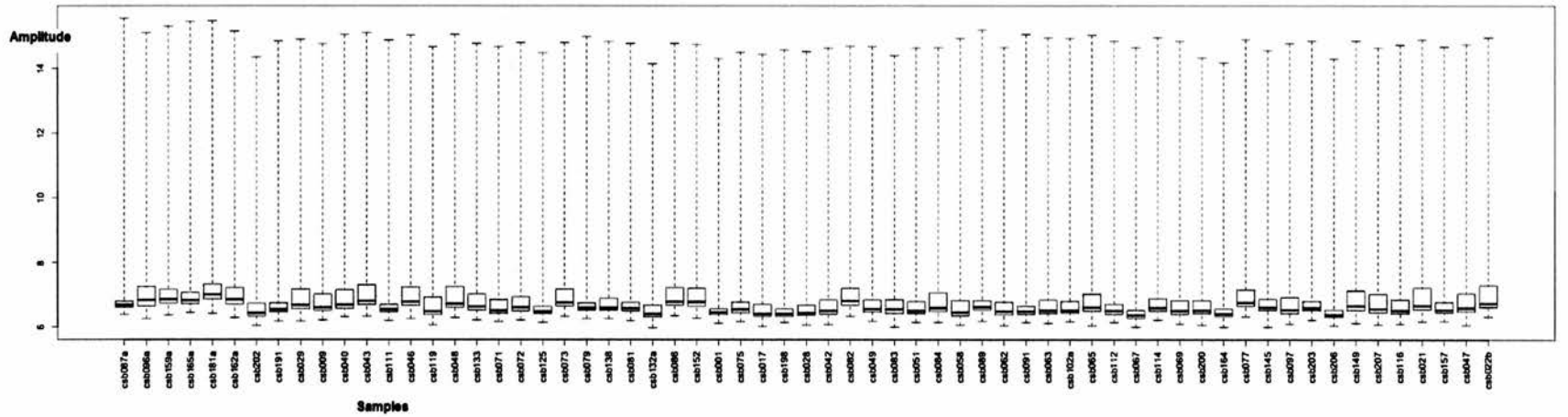
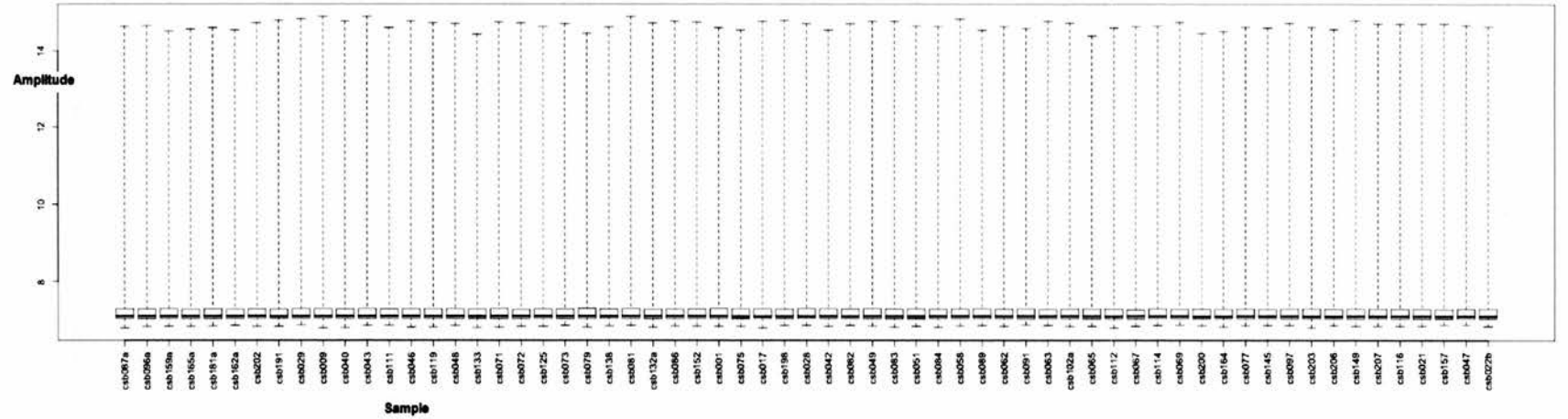


Figure 5.2.2: ***Boxplot of Quantile Normalised Data***



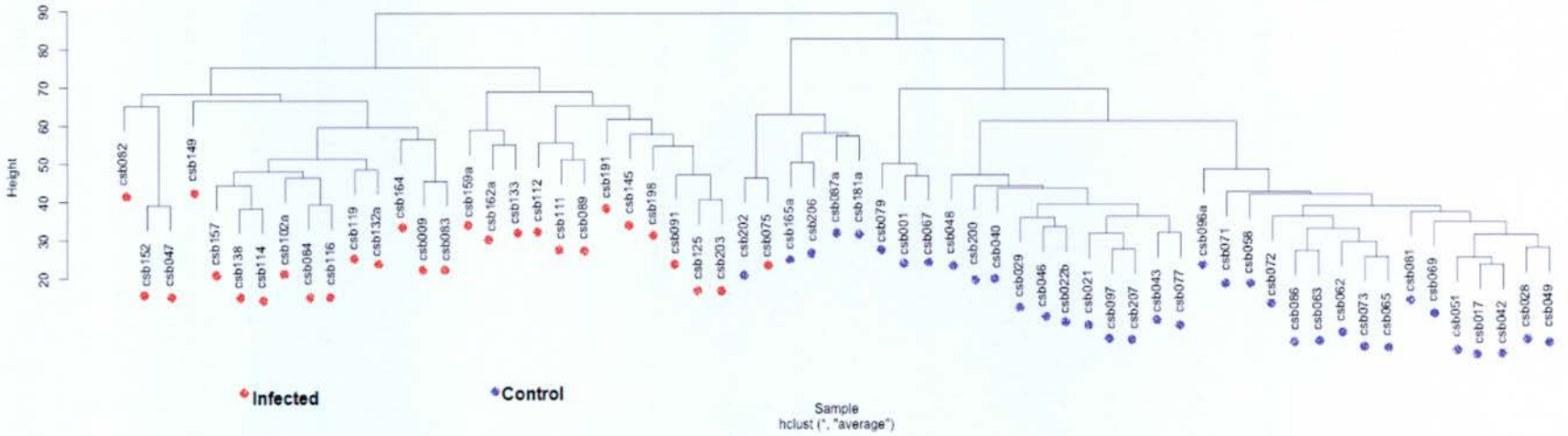
Infected compared to control samples

Figure 5.2.3 is a sample relation graph, grouping samples with respect to similarity of gene variability. This was before statistical filtering of data. It was very encouraging to see that even on this unfiltered data, there was a split between control and infected groups. The only sample that did not group neatly was that of csb75. This sample is from the infant with CMV infection and it is seen to group with the controls.

Figure 5.2.4 is a heat map. This is after filtering of genes and is displaying intensity values of genes significantly differentially expressed (fold change of > 2 and $p < 0.01$) between control and infected samples. The intensity values are in log 2 scale of expression, per gene normalised across the dataset. Genes that are up-regulated in the infected group compared to the control group are yellow and those that are comparatively down-regulated are blue. The more yellow or the bluer, the more up or down regulated respectively. Again, there is a separation of patterns between control and infected samples with the CMV infected sample being the exception. There seem to be 3 distinct groups of genes (marked at left of figure). The infected samples show down-regulation of group 1 and up-regulation of groups 2 and 3 whereas most of the controls show the converse. There is group of nine samples (8 controls and the CMV infected sample) however that shows up-regulation of groups 1 and 3 and down-regulation of group 2.

Acknowledgement: Figures 5.2.3 and 5.2.4 were generated by Paul Dickinson and modified by me.

Figure 5.2.3: ***Sample Relations Based on 10206 Genes with SD/mean > 0.1***



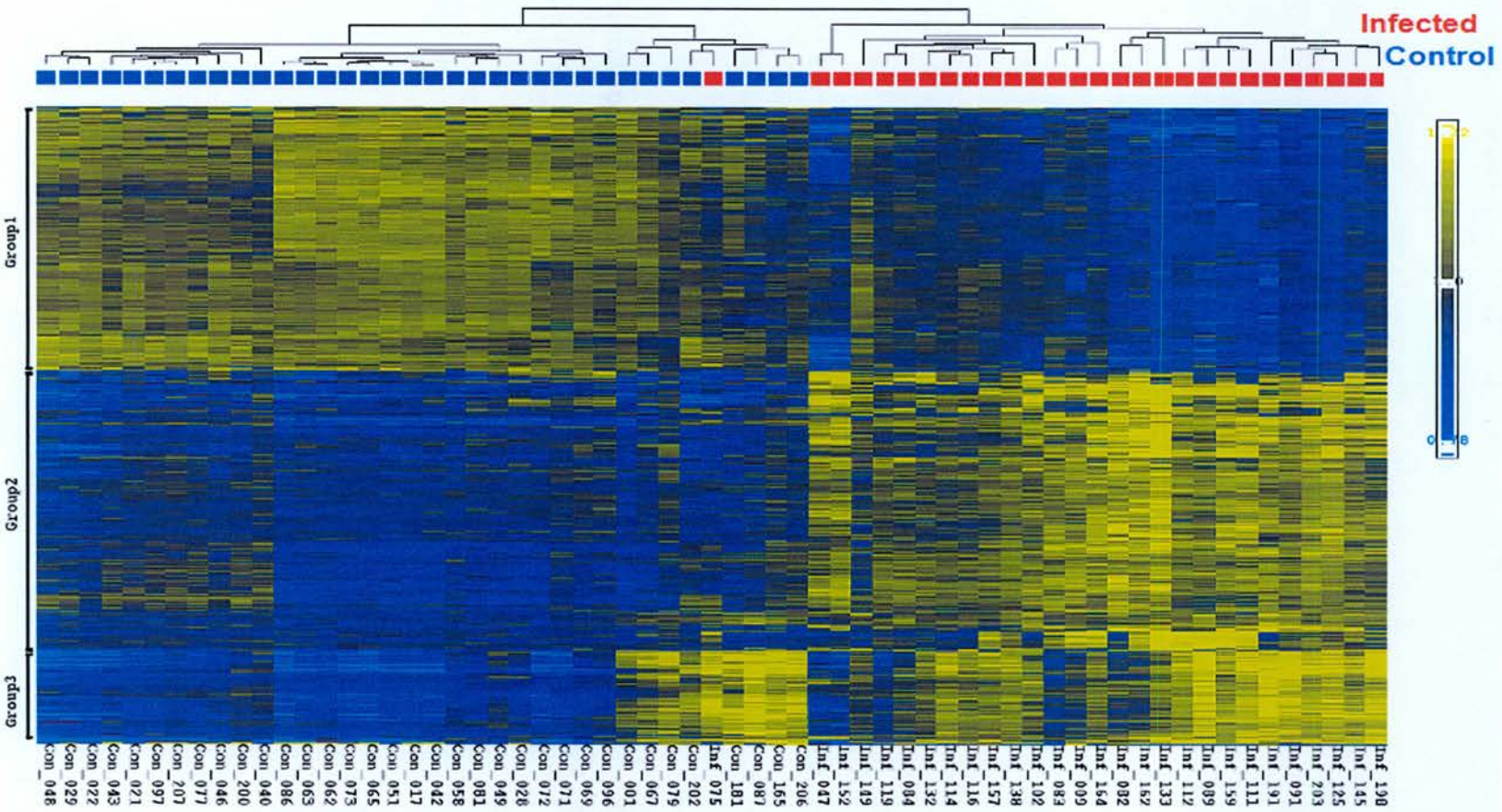


Figure 5.2.4: ***Heatmap of Significantly Differentially Expressed Genes (p < 0.01)***

Significant Differential Expression

With an adjusted p value of < 0.001, there were 6,221 features significantly differentially expressed: 2,550 up-regulated and 3,671 down-regulated.

GO Term Enrichment

In the significantly up-regulated genes there were 232 enriched GO terms and in the significantly down-regulated genes there were 152. The 232 GO terms for up-regulated genes consisted of 140 Biological Process ontology terms, 36 Molecular Function terms and 56 Component terms. The Molecular Function and Cellular Component terms are too vague to be able to be attributed to immune-related functions or not, e.g. "kinase activity", "cell part". I have therefore not displayed them here. The Biological Processes Ontology terms are displayed in Table 5.2.7 (a to d). The terms are listed in order of significance with the most significant first. I have highlighted in red the terms that are immediately recognisable as being involved in immune response.

Table 5.2.7a: ***GO Terms(Biological Processes) for Up-regulated Genes Infected compared to control.***

GO ID	GO Term	Enriched p Value
GO:0002376	immune system process	3.67E-17
GO:0006952	defence response	2.15E-16
GO:0009605	response to external stimulus	3.77E-16
GO:0007243	protein kinase cascade	2.06E-15
GO:0007242	intracellular signalling cascade	3.79E-15
GO:0009611	response to wounding	5.04E-15
GO:0006950	response to stress	7.46E-13
GO:0006954	inflammatory response	2.04E-12
GO:0006955	immune response	4.01E-12
GO:0012501	programmed cell death	1.56E-10
GO:0007249	I-kappaB kinase/NF-kappaB cascade	1.68E-10
GO:0015031	protein transport	1.87E-10
GO:0045184	establishment of protein localization	2.07E-10
GO:0006915	Apoptosis	2.69E-10
GO:0016192	vesicle-mediated transport	4.10E-10

Table 5.2.7b: ***GO Terms (Biological Processes) for Up-regulated Genes. Infected compared to control.***		
GO:0010627	regulation of protein kinase cascade	7.29E-10
GO:0008104	protein localization	1.10E-09
GO:0043122	regulation of I-kappaB kinase/NF-kappaB cascade	1.87E-09
GO:0033036	macromolecule localization	4.69E-09
GO:0008219	cell death	5.11E-09
GO:0016265	death	5.11E-09
GO:0016044	membrane organization	6.66E-09
GO:0009967	positive regulation of signal transduction	7.04E-09
GO:0043123	Positive regulation of I-kappaB kinase/NF-kappaB cascade	2.24E-08
GO:0050793	regulation of developmental process	3.25E-08
GO:0051707	response to other organism	8.66E-08
GO:0051179	localization	9.63E-08
GO:0042981	regulation of apoptosis	9.97E-08
GO:0043067	regulation of programmed cell death	1.02E-07
GO:0065008	regulation of biological quality	1.03E-07
GO:0009607	response to biotic stimulus	1.13E-07
GO:0006897	endocytosis	1.27E-07
GO:0010324	membrane invagination	1.27E-07
GO:0051704	multi-organism process	1.33E-07
GO:0006793	phosphorus metabolic process	4.49E-07
GO:0006796	phosphate metabolic process	4.49E-07
GO:0030099	myeloid cell differentiation	5.72E-07
GO:0006464	protein modification process	6.33E-07
GO:0016310	Phosphorylation	1.15E-06
GO:0043687	post-translational protein modification	1.67E-06
GO:0006629	lipid metabolic process	1.79E-06
GO:0007264	small GTPase mediated signal transduction	2.22E-06
GO:0002274	myeloid leukocyte activation	2.33E-06
GO:0043412	biopolymer modification	2.39E-06
GO:0044255	cellular lipid metabolic process	3.06E-06
GO:0051234	establishment of localization	3.11E-06
GO:0006468	protein amino acid phosphorylation	3.71E-06
GO:0006810	Transport	4.16E-06
GO:0009966	regulation of signal transduction	5.87E-06
GO:0050896	response to stimulus	6.41E-06
GO:0048518	positive regulation of biological process	7.22E-06
GO:0009615	response to virus	8.44E-06
GO:0043066	negative regulation of apoptosis	8.86E-06
GO:0030097	hemopoiesis	9.29E-06
GO:0042221	response to chemical stimulus	1.22E-05
GO:0043069	negative regulation of programmed cell death	1.29E-05
GO:0045087	innate immune response	1.36E-05
GO:0048522	positive regulation of cellular process	1.41E-05
GO:0051641	cellular localization	1.63E-05
GO:0044419	interspecies interaction between organisms	1.73E-05
GO:0001816	cytokine production	1.81E-05
GO:0006928	cell motion	1.85E-05
GO:0051674	localization of cell	1.85E-05

Table 5.2.7c: *GO Terms(Biological Processes) for Up-regulated Genes Infected compared to control.*		
GO:0006643	membrane lipid metabolic process	2.20E-05
GO:0008654	phospholipid biosynthetic process	2.46E-05
GO:0048534	hemopoietic or lymphoid organ development	2.49E-05
GO:0030036	actin cytoskeleton organization	2.84E-05
GO:0065009	regulation of molecular function	2.86E-05
GO:0007599	Hemostasis	3.11E-05
GO:0050878	regulation of body fluid levels	3.19E-05
GO:0051649	establishment of localization in cell	3.98E-05
GO:0046467	membrane lipid biosynthetic process	4.08E-05
GO:0016050	vesicle organization	4.18E-05
GO:0007596	blood coagulation	4.21E-05
GO:0002520	immune system development	4.65E-05
GO:0048872	homeostasis of number of cells	5.02E-05
GO:0050817	Coagulation	5.12E-05
GO:0002237	response to molecule of bacterial origin	5.50E-05
GO:0042089	cytokine biosynthetic process	5.95E-05
GO:0000165	MAPKKK cascade	6.36E-05
GO:0006935	Chemotaxis	6.57E-05
GO:0042330	Taxis	6.57E-05
GO:0042107	cytokine metabolic process	7.45E-05
GO:0007010	cytoskeleton organization	7.83E-05
GO:0008610	Lipid biosynthetic process	8.04E-05
GO:0006886	intracellular protein transport	8.44E-05
GO:0006914	Autophagy	9.65E-05
GO:0034101	erythrocyte homeostasis	9.97E-05
GO:0009617	response to bacterium	0.000101952
GO:0019221	cytokine-mediated signaling pathway	0.000109354
GO:0032640	tumor necrosis factor production	0.00011891
GO:0008286	insulin receptor signaling pathway	0.000120966
GO:0030029	actin filament-based process	0.000121824
GO:0006066	cellular alcohol metabolic process	0.000136509
GO:0019318	hexose metabolic process	0.000138255
GO:0042060	wound healing	0.000141475
GO:0006826	Iron ion transport	0.000142536
GO:0001775	cell activation	0.000170662
GO:0007265	Ras protein signal transduction	0.000178475
GO:0043068	positive regulation of programmed cell death	0.000186198
GO:0006916	anti-apoptosis	0.000190149
GO:0044275	cellular carbohydrate catabolic process	0.000200538
GO:0051094	positive regulation of developmental process	0.000207026
GO:0007165	signal transduction	0.000219967
GO:0050790	regulation of catalytic activity	0.000248993
GO:0016052	carbohydrate catabolic process	0.000249272
GO:0042108	positive regulation of cytokine biosynthetic process	0.000252165
GO:0000041	transition metal ion transport	0.000258967
GO:0046631	alpha-beta T cell activation	0.000266475
GO:0043065	positive regulation of apoptosis	0.000267844
GO:0005996	monosaccharide metabolic process	0.0002738

Table 5.2.7d: ***GO Terms(Biological Processes) for Up-regulated Genes Infected compared to control.***		
GO:0051093	negative regulation of developmental process	0.000284405
GO:0044262	cellular carbohydrate metabolic process	0.000290782
GO:0030100	regulation of endocytosis	0.000291985
GO:0005975	carbohydrate metabolic process	0.00031263
GO:0030218	erythrocyte differentiation	0.000322403
GO:0009991	response to extracellular stimulus	0.000326692
GO:0045080	positive regulation of chemokine biosynthetic process	0.000335141
GO:0002252	immune effector process	0.00034676
GO:0042035	regulation of cytokine biosynthetic process	0.000350674
GO:0007202	activation of phospholipase C activity	0.000418559
GO:0000187	activation of MAPK activity	0.00047484
GO:0051050	positive regulation of transport	0.000514586
GO:0043406	positive regulation of MAP kinase activity	0.000574075
GO:0006644	phospholipid metabolic process	0.000625131
GO:0051336	regulation of hydrolase activity	0.000628101
GO:0046907	intracellular transport	0.000667671
GO:0006901	vesicle coating	0.000707542
GO:0050900	leukocyte migration	0.000755122
GO:0044260	cellular macromolecule metabolic process	0.000772095
GO:0002467	germinal centre formation	0.00077553
GO:0043300	regulation of leukocyte degranulation	0.00077553
GO:0031667	response to nutrient levels	0.000860368
GO:0051345	positive regulation of hydrolase activity	0.000878847
GO:0031347	regulation of defence response	0.000880902
GO:0042592	homeostatic process	0.000911753
GO:0019320	hexose catabolic process	0.000944357
GO:0007154	cell communication	0.00095694
GO:0003013	circulatory system process	0.000985052
GO:0008015	blood circulation	0.000985052

The 152 GO terms for down-regulated genes consisted of 75 Biological Process terms, 15 Molecular Function terms and 62 Cellular Component terms. Again, I have not displayed the Molecular Function or Cellular Component terms here. The Biological Processes Ontology terms for significantly down-regulated genes are displayed in Table 5.2.8 (a and b) with the most significant listed first.

Table 5.2.8a: *GO Terms (Biological Processes): Down-Regulated Genes. Infected compared to control.*

GO ID	GO Term	Enriched p Value
GO:0010467	gene expression	4.26E-65
GO:0006412	Translation	3.23E-57
GO:0043170	macromolecule metabolic process	1.60E-46
GO:0044237	cellular metabolic process	9.51E-42
GO:0008152	metabolic process	1.71E-39
GO:0006414	translational elongation	1.29E-36
GO:0044238	primary metabolic process	1.54E-35
GO:0006396	RNA processing	2.58E-35
GO:0009059	macromolecule biosynthetic process	8.21E-34
GO:0043283	biopolymer metabolic process	1.13E-33
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	7.52E-32
GO:0044249	cellular biosynthetic process	1.56E-28
GO:0022613	ribonucleoprotein complex biogenesis	4.91E-28
GO:0016070	RNA metabolic process	8.30E-27
GO:0009058	biosynthetic process	3.87E-25
GO:0008380	RNA splicing	8.76E-21
GO:0042254	ribosome biogenesis	1.49E-20
GO:0006397	mRNA processing	1.86E-19
GO:0016071	mRNA metabolic process	4.95E-19
GO:0034470	ncRNA processing	1.81E-18
GO:0000375	RNA splicing, via transesterification reactions	3.60E-16
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	3.60E-16
GO:0000398	nuclear mRNA splicing, via spliceosome	3.60E-16
GO:0043284	biopolymer biosynthetic process	3.73E-16
GO:0016072	rRNA metabolic process	9.98E-15
GO:0006364	rRNA processing	1.11E-14
GO:0044267	cellular protein metabolic process	3.52E-12
GO:0019538	protein metabolic process	1.18E-11
GO:0044260	cellular macromolecule metabolic process	2.16E-11
GO:0022618	ribonucleoprotein complex assembly	4.07E-09
GO:0006399	tRNA metabolic process	2.69E-08
GO:0009987	cellular process	3.17E-08
GO:0010468	regulation of gene expression	4.56E-08
GO:0006350	Transcription	1.08E-07
GO:0010556	regulation of macromolecule biosynthetic process	1.41E-07
GO:0009889	regulation of biosynthetic process	5.33E-07
GO:0060255	regulation of macromolecule metabolic process	8.89E-07
GO:0006119	oxidative phosphorylation	1.42E-06
GO:0008033	tRNA processing	2.00E-06
GO:0032774	RNA biosynthetic process	2.66E-06
GO:0006351	transcription, DNA-dependent	4.02E-06
GO:0045449	regulation of transcription	4.18E-06
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	9.17E-06
GO:0006413	translational initiation	1.03E-05

Table 5.2.8b: ***GO Terms (Biological Processes): Down-Regulated Genes. Infected compared to control.***		
GO:0031323	regulation of cellular metabolic process	1.31E-05
GO:0019222	regulation of metabolic process	1.45E-05
GO:0006355	regulation of transcription, DNA-dependent	1.53E-05
GO:0051252	regulation of RNA metabolic process	1.68E-05
GO:0000387	spliceosomal snRNP biogenesis	1.78E-05
GO:0006457	protein folding	3.88E-05
GO:0009451	RNA modification	5.62E-05
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	5.65E-05
GO:0009260	ribonucleotide biosynthetic process	6.79E-05
GO:0032259	Methylation	8.84E-05
GO:0034621	cellular macromolecular complex subunit organization	0.000110568
GO:0010608	posttranscriptional regulation of gene expression	0.000174097
GO:0034622	cellular macromolecular complex assembly	0.000194224
GO:0006417	regulation of translation	0.000203268
GO:0015985	energy coupled proton transport, down electrochemical gradient	0.000208861
GO:0015986	ATP synthesis coupled proton transport	0.000208861
GO:0006446	regulation of translational initiation	0.000343353
GO:0009152	purine ribonucleotide biosynthetic process	0.000348601
GO:0050657	nucleic acid transport	0.000372816
GO:0050658	RNA transport	0.000372816
GO:0051236	establishment of RNA localization	0.000372816
GO:0051028	mRNA transport	0.000411664
GO:0006403	RNA localization	0.000454121
GO:0009060	aerobic respiration	0.00051651
GO:0006921	cell structure disassembly during apoptosis	0.000604717
GO:0022411	cellular component disassembly	0.000604717
GO:0034220	transmembrane ion transport	0.000604717
GO:0006886	intracellular protein transport	0.000662901
GO:0015931	nucleobase, nucleoside, nucleotide and nucleic acid transport	0.000682094
GO:0006164	purine nucleotide biosynthetic process	0.000728556
GO:0009259	ribonucleotide metabolic process	0.000889407

KEGG Pathway Enrichment

There were found to be 40 enriched KEGG pathways for significantly up-regulated genes (displayed in Table 5.2.9) and 31 enriched KEGG pathways for significantly down-regulated genes (displayed in Table 5.2.10). Immune related terms are highlighted in red.

Table 5.2.9 ***Kegg Pathways: Up-Regulated Genes Infected compared to control.***

Pathway ID	Pathway Description	p Value
4620	Toll-like receptor signalling pathway	4.82E-07
4920	Adipocytokine signalling pathway	5.32E-05
5221	Acute myeloid leukaemia	5.64E-05
5215	Prostate cancer	8.40E-05
4210	Apoptosis	8.54E-05
5220	Chronic myeloid leukaemia	0.000133
4540	Gap junction	0.000317
4630	Jak-STAT signalling pathway	0.000631
5120	Epithelial cell signalling in Helicobacter pylori infection	0.00074
4520	Adherens junction	0.000862
4810	Regulation of actin cytoskeleton	0.001137
4010	MAPK signalling pathway	0.001282
5213	Endometrial cancer	0.001851
5211	Renal cell carcinoma	0.001995
600	Sphingolipid metabolism	0.002874
3320	PPAR signalling pathway	0.003028
4910	Insulin signalling pathway	0.004092
4662	B cell receptor signalling pathway	0.004584
5040	Huntington's disease	0.005612
51	Fructose and mannose metabolism	0.005703
5216	Thyroid cancer	0.006916
5212	Pancreatic cancer	0.009193
4610	Complement and coagulation cascades	0.010138
770	Pantothenate and CoA biosynthesis	0.011861
4012	ErbB signalling pathway	0.014274
4650	Natural killer cell mediated cytotoxicity	0.016634
4510	Focal adhesion	0.017056
900	Terpenoid biosynthesis	0.017814
5223	Non-small cell lung cancer	0.018249
561	Glycerolipid metabolism	0.020073
5130	Pathogenic Escherichia coli infection - EHEC	0.023406
5131	Pathogenic Escherichia coli infection - EPEC	0.023406
564	Glycerophospholipid metabolism	0.025439
4370	VEGF signalling pathway	0.025439
1032	Glycan structures – degradation	0.02743
4670	Leukocyte transendothelial migration	0.028729
4664	Fc epsilon RI signalling pathway	0.028771
521	Streptomycin biosynthesis	0.031077
5214	Glioma	0.036488
5219	Bladder cancer	0.038773

Table 5.2.10 *Kegg Pathways: Down-Regulated Genes Infected compared to control.*		
Pathway ID	**Pathway Description**	**p Value**
3010	Ribosome	7.39E-34
190	Oxidative phosphorylation	1.09E-07
970	Aminoacyl-tRNA biosynthesis	0.000133
4660	T cell receptor signalling pathway	0.000258
5310	Asthma	0.000526
5332	Graft-versus-host disease	0.001934
280	Valine, leucine and isoleucine degradation	0.001976
271	Methionine metabolism	0.002979
5330	Allograft rejection	0.00309
5340	Primary immunodeficiency	0.00309
240	Pyrimidine metabolism	0.003286
230	Purine metabolism	0.003905
340	Histidine metabolism	0.003934
4940	Type I diabetes mellitus	0.003934
4612	Antigen processing and presentation	0.006232
4120	Ubiquitin mediated proteolysis	0.00882
3020	RNA polymerase	0.009959
620	Pyruvate metabolism	0.01146
130	Ubiquinone biosynthesis	0.012051
5060	Prion disease	0.01222
450	Selenoamino acid metabolism	0.012671
440	Aminophosphonate metabolism	0.01508
785	Lipoic acid metabolism	0.016278
5320	Autoimmune thyroid disease	0.017677
640	Propanoate metabolism	0.019212
20	Citrate cycle (TCA cycle)	0.020264
4640	Hematopoietic cell lineage	0.030615
520	Nucleotide sugars metabolism	0.030752
5322	Systemic lupus erythematosis	0.041064
720	Reductive carboxylate cycle (CO2 fixation)	0.04154
71	Fatty acid metabolism	0.04668

Largest Fold Changes

Before statistical filtering there were 490 features exhibiting greater than 2-fold up-regulation and 344 features exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 448 features with greater than 2-fold up-regulation and 341 with greater than 2-fold down regulation. The genes for the forty features with the largest fold-changes are described in tables 5.2.11 (a to c) for up-

regulated genes and 5.2.12 (a to c) for down-regulated genes (the full lists of genes are found on the enclosed CD). Each line represents one feature on the array with the largest fold changes being listed first. Phrases readily recognisable as being immune function related are highlighted in red.

Table 5.2.11a: ***Fold Changes > 2 for Up-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

Feature ID	Description	Pathway Description	Fold Change
fn3Hpm4vVouL4FK6FA	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	Leukocyte transendothelial migration, Bladder cancer	13.55488
EqT7npVH9Hpffr.qqA	interleukin 1 receptor, type II	MAPK signalling pathway, Cytokine-cytokine receptor interaction, Hematopoietic cell lineage	9.604756
KqU013001301 55716U	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	-	9.389889
HN97InJZXh5fSF6V7I	spermine oxidase	-	9.338627
WUTbfV7VDYUuzYaeLk	lipocalin 2	-	8.308514
ijfUpc6MjSik0eKUok	interferon induced transmembrane protein 3 (1-8U)	-	7.971189
upH6Ifs4COnuXeXt1U	interleukin 1 receptor antagonist	-	7.741772
3mxHd1KQUncUXTUg_k	peptidoglycan recognition protein 1	Peptidoglycan biosynthesis	7.730001
Qz8qQwkk_l6LIIm0XU	resistin	-	7.621582
HdbE7q7ZTu5XtHN7p4	chromosome 19 open reading frame 59	-	7.083687
ok0Fe53OSU4FID1ICQ	G protein-coupled receptor 84	-	6.81959
HIf_t7qU.BejIFP86o	ankyrin repeat domain 22	-	6.723585
6pVadNuztOpruIO0iM	syntaxin binding protein 2	-	6.63982
0255m6FSUH9Z5w1Lv0	hemoglobin, mu	-	6.364718
Z0We_XII_ZL4du3igE	hexokinase 3 (white cell)	Glycolysis / Gluconeogenesis, Fructose and mannose metabolism, Galactose metabolism, Starch and sucrose metabolism, Streptomycin biosynthesis, Aminosugars metabolism	6.065425

Table 5.2.11b: ***Fold Changes > 2 for Up-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

T.OKlGre72HS XSoYd4	basigin (Ok blood group)	-	6.060933
QVuVUflxXe15f vPsLk	spermine oxidase	-	6.022997
TY47U6f7qCMe 6l4n1l	tumor necrosis factor, alpha-induced protein 6	-	5.973717
fr.cFeEaHuTH9 XTU54	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	-	5.93226
KXq9Q16d7p_c nO4X4k	myeloperoxidase	-	5.886758
cW9Lul4ARTu5 QvUvRk	erythrocyte membrane protein band 4.2	-	5.746367
EX243qjuUVl.1e H30Y	proteinase 3	-	5.737136
T5n6RXVVFEldI 46CP8	aminolevulinate, delta-, synthase 2	Glycine, serine and threonine metabolism, Porphyrin and chlorophyll metabolism	5.66826
Q0R16neBERlJ 13eCYc	alkaline phosphatase, liver/bone/kidney	gamma-Hexachlorocyclohexane degradation, Folate biosynthesis	5.585284
xKXCKf3Uoo8n 6L3cx0	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	-	5.543602
Qkl1OrSbvUloB bBfkc	NA	-	5.536282
oXed_glbkhVE MCDCTo	ankyrin repeat domain 22	-	5.501625
fG6ilCJP2dH54 10qEU	orosomucoid 1	-	5.469193
xeSJevQDS0dX 0gkhOk	glycogenin 1	-	5.343584
oSEiVHyftROjd DlDXU	brain abundant, membrane attached signal protein 1	-	5.17689
l1Cl6OcNCOBK 2H9CD0	S100 calcium binding protein A12	-	5.087631
QlF3hK7l5VTd9 4_e8E	chromosome 5 open reading frame 32	-	5.04204
x5.XReVeVxfEV Un4fc	erythrocyte membrane protein band 4.9 (dematin)	-	5.026276
QXFcOUvUn03 Ffi0z4o	formyl peptide receptor 2	Neuroactive ligand-receptor interaction	4.960542
KpCYUm43RIk QFdLdQU	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	-	4.956158

Table 5.2.11c: ***Fold Changes > 2 for Up-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

iXftIjXcUSE8MniSiU	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5	O-Glycan biosynthesis, Glycan structures - biosynthesis 1	4.915085
cuUXXtx0__TiV7eUlQ	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)	-	4.831054
BnWcV6qY_yd4Tej_B0	selenium binding protein 1	-	4.778226
oTl3uX2dOe.fWHoOX0	paroxysmal nonkinesigenic dyskinesia	-	4.659756
iZauVnkv36hMomQIJw	CCAAT/enhancer binding protein (C/EBP), delta	-	4.657713

Table 5.2.12a: ***Fold Changes > 2 for Down-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

Feature ID	Description	Pathway Description	Fold Change
3Hnm08SpEpTo6AGd3k	CD247 molecule	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway	-5.62584
rAVtDscox55tPEqRKU	CD247 molecule	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway	-5.33643
lLAVMzUgV4V5bsUnil	NA	-	-4.98617
9GoohS1VFSWxSAwKfc	CD7 molecule	Hematopoietic cell lineage	-4.66544
cn_ot95.noTiHoKe3o	CD3d molecule, delta (CD3-TCR complex)	Hematopoietic cell lineage, T cell receptor signalling pathway, Primary immunodeficiency	-4.52427
Znl3dSA4HuOuhEnO4o	leucine rich repeat neuronal 3	-	-4.48762
uqdSIKVm3AwJlJko4A	Lck interacting transmembrane adaptor 1	-	-4.34556
NnpUIV8gV4uekDdHnU	GRB2-related adaptor protein	-	-4.34222
EWX22SIXfnq.QQG5T4	integral membrane protein 2A	-	-4.30725
6UX6KAep2gQuBeleuk	CD3d molecule, delta (CD3-TCR complex)	Hematopoietic cell lineage, T cell receptor signalling pathway, Primary immunodeficiency	-4.23239
EQl6SwNYM1lFeFeW7E	NA	-	-4.08084
EAVtFSNtJlil6rlkrg	NA	-	-4.04609
6dFQSN.UitTrolYwV4	mal, T-cell differentiation protein	-	-4.04272

Table 5.2.12b: ***Fold Changes > 2 for Down-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

lXb5HniJCOr RSSex6l	ribosomal protein S29	Ribosome	-4.03884
cdeqd0u5TloS QeuPv0	major histocompatibility complex, class II, DM beta	Cell adhesion molecules (CAMs), Antigen processing and presentation, Type I diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease	-4.03427
ErilgoBfqKKE 8evvek	src kinase associated phosphoprotein 1	-	-3.95345
rl1Ll9VboMBr u14gFE	CD79b molecule, immunoglobulin-associated beta	B cell receptor signalling pathway	-3.76408
Boq4S6WVCl Rifvf738	chemokine (C-C motif) receptor 7	Cytokine-cytokine receptor interaction	-3.62719
QBhFJQFJJV 0k1JBMJo	CD52 molecule	-	-3.61014
ES3J4_9RQI NSXnnz2I	killer cell lectin-like receptor subfamily B, member 1	-	-3.58958
B_5R5EdQ3q 4CFiHWXQ	CD96 molecule	-	-3.56901
QVuhFUheYK ONlYwlaM	megakaryocyte-associated tyrosine kinase	-	-3.51789
cUncUlXQo2II 4UsSUU	CD3g molecule, gamma (CD3-TCR complex)	Hematopoietic cell lineage, T cell receptor signalling pathway	-3.51363
Bon3BW069D ET33.lJk	NA	-	-3.4625
rMFE5ACCNZ O0VF9KaU	hypothetical protein LOC731985	-	-3.45002
uVoZUoVTpL qADsoh78	lymphotoxin beta (TNF superfamily, member 3)	Cytokine-cytokine receptor interaction	-3.39272
ZvGex87PV7 F4Efvnl0	family with sequence similarity 117, member B	-	-3.32231
QjupcAJNtCi5 qBZ57k	GTPase, IMAP family member 7	-	-3.31405
Ql1H6oV5du2 383lzx0	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	-	-3.30565

Table 5.2.12c: ***Fold Changes > 2 for Down-Regulated Genes. Infected compared to control. (Each line represents 1 feature on array).***

ZfEgo719KKC SRwCR3g	NA	-	-3.30072
6tRdEndeoxA npOFRNc	NA	-	-3.29883
Qun3e4Pl7BE y6fdZX4	lymphoid enhancer-binding factor 1	Wnt signalling pathway, Adherens junction, Melanogenesis, Colorectal cancer, Endometrial cancer, Prostate cancer, Thyroid cancer, Basal cell carcinoma, Acute myeloid leukemia	-3.29077
9nj9cUl.l319D uv1Oo	NA	-	-3.26661
NE_8qgdKcu0 m36.qiA	family with sequence similarity 102, member A	-	-3.24589
lERLzbg6VSK BXmiofQ	similar to rat ribosomal protein L9 homologue	-	-3.23499
0eE_BUvnSrS VPfH1eo	interleukin 7 receptor	Cytokine-cytokine receptor interaction, Jak-STAT signalling pathway, Hematopoietic cell lineage, Primary immunodeficiency	-3.20688
0..gjSdbqnr.rU STBI	OCIA domain containing 2	-	-3.15584
3njnpE0.56lv. zMeyU	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	Renal cell carcinoma	-3.12153
cCUgLove1n mJeiOqBI	PAS domain containing serine/threonine kinase	-	-3.10824
ToESkBW4.6 51_wdJXk	B-cell CLL/lymphoma 11B (zinc finger protein)	-	-3.10712

The Gene Ontology for these genes was examined using the online tool DAVID 2008 (348, 349). The results are displayed in tables 5.2.13 (a and b) for up-regulated genes and 5.2.14 (a and b) for down regulated genes.

Table 5.2.13a: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Up-Regulated Genes. Infected compared to control.***

Ontology	GO Term	No. of Genes (from 448)	p value
BP	Defence response	56	1.1E-23
BP	Inflammatory response	38	8.5E-20
BP	Response to wounding	39	2.4E-15
BP	Response to external stimulus	46	2.9E-14
BP	Immune system process	62	2.3E-13
BP	Response to stress	53	4.1E-10
BP	Response to stimulus	102	2.6E-9
BP	Immune response	47	2.7E-9
BP	Taxis	16	1.6E-7
BP	Chemotaxis	16	1.6E-7
BP	Response to other organism	18	3.0E-7
BP	Multi-organism process	22	8.7E-7
BP	Response to biotic stimulus	21	2.4E-6
BP	Response to chemical stimulus	30	4.9E-6
BP	Locomotory behaviour	16	5.9E-6
BP	Apoptosis	35	6.2E-6
BP	Programmed cell death	35	7.4E-6
BP	Regulation of biological quality	37	1.2E-5
BP	Localisation of cell	23	1.9E-5
BP	Cell motility	23	1.9E-5
BP	Death	35	2.2E-5
BP	Cell death	35	2.2E-5
BP	Response to bacterium	11	2.4E-5
BP	Negative regulation of apoptosis	16	3.5E-5
BP	Negative regulation of programmed cell death	16	4.1E-5
BP	Cell proliferation	33	5.2E-5
BP	Myeloid cell differentiation	10	5.3E-5
BP	Acute inflammatory response	9	8.7E-5
BP	Intracellular signalling cascade	49	2.1E-4
BP	Cellular chemical homeostasis	14	2.3E-4
BP	Cellular ion homeostasis	14	2.3E-4
BP	Chemical homeostasis	16	2.6E-4
BP	I-kappaB kinase/NF-kappaB cascade	11	2.9E-4
BP	Elevation of cytosolic calcium ion concentration	8	3.0E-4
BP	Cytosolic calcium ion homeostasis	8	3.0E-4
BP	Immune effector process	10	3.0E-4
BP	Acute phase response	6	3.2E-4
BP	Defence response to bacterium	9	3.4E-4
BP	Signal transduction	98	4.2E-4
BP	Immunoglobulin mediated immune response	7	5.4E-4
BP	Behaviour	17	5.6E-4
BP	Response to virus	9	5.9E-4
BP	Ion homeostasis	14	6.0E-4
BP	Regulation of apoptosis	23	6.4E-4
BP	B cell mediated immunity	7	6.6E-4
BP	Erythrocyte differentiation	6	7.2E-4
BP	Regulation of programmed cell death	23	7.4E-4
BP	Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	8	8.1E-4

Table 5.2.13b: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001: Up-Regulated Genes. Infected compared to control.***			
BP	Leukocyte mediated immunity	8	8.1E-4
BP	Adaptive immune response	8	8.1E-4
BP	Cellular di-, tri-valent inorganic cation homeostasis	11	9.8E-4
CC	Cytoplasm	155	5.7E-6
CC	Vacuole	16	2.0E-5
CC	Integral to plasma membrane	45	3.3E-5
CC	Intrinsic to plasma membrane	45	4.4E-5
CC	Membrane	168	1.7E-4
MF	Binding	271	6.9E-8
MF	Protein binding	173	4.4E-7
MF	Interleukin receptor activity	7	2.3E-5
MF	Interleukin binding	7	6.2E-5
MF	Cytokine binding	9	2.3E-4

Table 5.2.14a: ***DAVID GO Terms for Fold-Change > 2 and p < 0.001 Down-Regulated Genes. Infected compared to control.***			
Ontology	**GO Term**	**No. of Genes (from 341)**	**p Values**
BP	Translation	45	1.2E-22
BP	Macromolecule biosynthetic process	47	6.1E-18
BP	Cellular biosynthetic process	50	5.4E-16
BP	Biosynthetic process	51	6.5E-12
BP	Gene expression	77	2.6E-8
BP	Protein metabolic process	76	6.7E-8
BP	Cellular protein metabolic process	71	3.9E-7
BP	Cellular macromolecule metabolic process	71	7.1E-7
BP	Immune system process	31	2.8E-5
BP	Immune reponse	27	3.0E-5
BP	Cell activation	12	1.1E-4
BP	Ribonucleoprotein complex biogenesis and assembly	11	1.3E-4
BP	Hemopoietic or lymphoid organ development	11	1.5E-4
BP	Macromolecule metabolic process	109	1.5E-4
BP	Immune system development	11	2.3E-4
BP	Cellular defence response	7	2.3E-4
BP	RNA processing	16	2.8E-4
BP	Lymphocyte activation	10	3.1E-4
BP	Lymphocyte differentiation	7	3.3E-4
BP	T cell activation	16	4.2E-4
BP	Leukocyte activation	10	7.7E-4
CC	Cytosolic ribosome (sensu Eukaryota)	22	4.0E-23

Table 5.2.14b: *DAVID GO Terms for Fold-Change > 2 and p < 0.001 Down-Regulated Genes. Infected compared to control.*			
CC	Ribosome	35	4.8E-22
CC	Ribnucleoprotein complex	43	6.1E-21
CC	Ribosomal subunit	24	1.3E-19
CC	Cytosolic part	23	4.3E-18
CC	Immunological synapse	10	8.0E-13
CC	Macromolecular complex	74	1.1E-12
CC	Cytosolic large ribosomal subunit (sensu Eukaryota)	11	2.9E-11
CC	Cytosolic small ribosomal subunit (sensu Eukaryota)	10	2.9E-10
CC	Small ribosomal subunit	12	6.1E-10
CC	Large ribosomal subunit	12	1.8E-9
CC	Cytosol	25	3.0E-8
CC	T cell receptor complex	6	2.0E-7
CC	Intracellular non-membrane bound organelle	49	1.6E-6
CC	Non-membrane bound organelle	49	1.6E-6
CC	Cytoplasm	107	1.2E-5
CC	Cytoplasmic part	72	2.9E-5
CC	Receptor complex	8	9.0E-5
CC	Organelle part	65	3.5E-4
CC	Intracellular organelle part	64	5.7E-4
MF	Structural constituent of ribosome	35	3.3E-24
MF	RNA binding	36	6.9E-13
MF	Structural molecule activity	37	2.0E-11
MF	MHC class II receptor activity	4	1.6E-4

Secondary Analyses

FIOS Genomics carried out statistical comparisons of groups other than infected and control using clinical data that was supplied for the samples. Bearing in mind the risks of over-interpreting differences in groups other than those for primary analysis, I have selected only 2 comparisons to present here. I present the comparison made between whether or not samples were frozen. In addition I have presented the data on male compared to female. I have not presented any comparisons where there were obvious confounders nor have I presented comparisons of clinical features, e.g.

whether had antenatal steroids or not, whether ventilated or not. I felt that the comparisons of clinical features were not of any practical use and in some cases (despite my best efforts) the datasets risked being incomplete. An example of potential incomplete data is that administration of antenatal steroids to the mother was not always clearly documented in the infant's notes and therefore there could be overlap between groups. Also, some differences between groups could be explained at least in part by gestational age and unfortunately gestational age was not examined statistically during analysis by FIOS genomics.

Comparison of Samples that were Frozen Compared to those Not Frozen

Significant Differential Expression

With an adjusted p value of < 0.001, there were 145 features significant differentially expressed: 22 up-regulated and 123 down-regulated.

GO Term Enrichment

In the significantly up-regulated genes there were 7 enriched GO terms and in significantly down-regulated genes there were 65. The ontology terms are displayed in Table 5.2.15. for up-regulated and Table 5.2.16(a and b) for down-regulated genes. The terms are listed in order of significance with the most significant first.

Table 5.2.15.: *GO Terms for Up-regulated Genes. Frozen compared to not frozen groups.*

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0006898	receptor-mediated endocytosis	1.12E-05
BP	GO:0006897	endocytosis	0.000921
BP	GO:0010324	membrane invagination	0.000921
MF	GO:0004402	histone acetyltransferase activity	0.000204
MF	GO:0004468	lysine N-acetyltransferase activity	0.000204
MF	GO:0043426	MRF binding	0.000899
MF	GO:0051577	MyoD binding	0.000899

Table 5.2.16a: *GO Terms for Down-regulated Genes. Frozen compared to not frozen groups.*

Ontology	GO ID	GO Term	Enriched p Value
BP	GO:0006412	translation	3.52E-13
BP	GO:0010467	gene expression	3.75E-08
BP	GO:0006396	RNA processing	2.04E-05
BP	GO:0006414	translational elongation	2.07E-05
BP	GO:0044237	cellular metabolic process	4.19E-05
BP	GO:0006120	mitochondrial electron transport, NADH to ubiquinone	5.24E-05
BP	GO:0042773	ATP synthesis coupled electron transport	0.000143
BP	GO:0042775	mitochondrial ATP synthesis coupled electron transport	0.000143
BP	GO:0008380	RNA splicing	0.000176
BP	GO:0043170	macromolecule metabolic process	0.000179
BP	GO:0022904	respiratory electron transport chain	0.000184
BP	GO:0022900	electron transport chain	0.000199
BP	GO:0006119	oxidative phosphorylation	0.000592
BP	GO:0044249	cellular biosynthetic process	0.000737
BP	GO:0009059	macromolecule biosynthetic process	0.000756
BP	GO:0008152	metabolic process	0.000783
MF	GO:0003735	structural constituent of ribosome	2.05E-16
MF	GO:0005198	structural molecule activity	7.77E-08
MF	GO:0003723	RNA binding	2.45E-05
MF	GO:0003954	NADH dehydrogenase activity	6.35E-05
MF	GO:0008137	NADH dehydrogenase (ubiquinone) activity	6.35E-05
MF	GO:0050136	NADH dehydrogenase (quinone) activity	6.35E-05
MF	GO:0016655	oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor	9.56E-05
MF	GO:0016651	oxidoreductase activity, acting on NADH or NADPH	0.00051
CC	GO:0030529	ribonucleoprotein complex	1.45E-21

Table 5.2.16b: *GO Terms for Down-regulated Genes. Frozen compared to not frozen groups.*

CC	GO:0005840	ribosome	3.08E-17
CC	GO:0032991	macromolecular complex	2.10E-13
CC	GO:0033279	ribosomal subunit	3.54E-12
CC	GO:0005739	mitochondrion	7.53E-12
CC	GO:0000313	organellar ribosome	3.36E-10
CC	GO:0005761	mitochondrial ribosome	3.36E-10
CC	GO:0044429	mitochondrial part	2.05E-09
CC	GO:0044444	cytoplasmic part	3.12E-09
CC	GO:0015934	large ribosomal subunit	4.95E-08
CC	GO:0043229	intracellular organelle	2.29E-07
CC	GO:0043226	organelle	2.34E-07
CC	GO:0005759	mitochondrial matrix	2.51E-07
CC	GO:0031980	mitochondrial lumen	2.51E-07
CC	GO:0015935	small ribosomal subunit	1.33E-06
CC	GO:0044446	intracellular organelle part	3.08E-06
CC	GO:0044422	organelle part	3.32E-06
CC	GO:0000314	organellar small ribosomal subunit	3.91E-06
CC	GO:0005763	mitochondrial small ribosomal subunit	3.91E-06
CC	GO:0044424	intracellular part	4.14E-06
CC	GO:0005737	cytoplasm	8.48E-06
CC	GO:0043231	intracellular membrane-bounded organelle	9.70E-06
CC	GO:0043227	membrane-bounded organelle	9.81E-06
CC	GO:0005622	intracellular	1.50E-05
CC	GO:0043228	non-membrane-bounded organelle	3.03E-05
CC	GO:0043232	intracellular non-membrane-bounded organelle	3.03E-05
CC	GO:0005746	mitochondrial respiratory chain	3.66E-05
CC	GO:0031974	membrane-enclosed lumen	3.67E-05
CC	GO:0043233	organelle lumen	3.67E-05
CC	GO:0044455	mitochondrial membrane part	4.41E-05
CC	GO:0022626	cytosolic ribosome	4.98E-05
CC	GO:0005829	cytosol	6.02E-05
CC	GO:0000315	organellar large ribosomal subunit	6.24E-05
CC	GO:0005762	mitochondrial large ribosomal subunit	6.24E-05
CC	GO:0005747	mitochondrial respiratory chain complex I	6.91E-05
CC	GO:0030964	NADH dehydrogenase complex	6.91E-05
CC	GO:0045271	respiratory chain complex I	6.91E-05
CC	GO:0044445	cytosolic part	0.000123
CC	GO:0005681	spliceosome	0.000152
CC	GO:0030532	small nuclear ribonucleoprotein complex	0.000421
CC	GO:0005743	mitochondrial inner membrane	0.000923

KEGG Pathway Enrichment

There were found to be 5 enriched KEGG pathways in significantly up-regulated genes (displayed in Table 5.2.17) and 7 enriched KEGG pathways in significantly down-regulated genes (displayed in Table 5.2.18).

Table 5.2.17: ***Kegg Pathways: Up-Regulated Genes. Frozen compared to not frozen.***

Pathway ID	Pathway Description.	p Value
4520	Adherens junction	0.002616
4630	Jak-STAT signalling pathway	0.010456
5040	Huntington's disease	0.035285
1510	Neurodegenerative diseases	0.041789
4330	Notch signalling pathway	0.048258

Table 5.2.18: ***Kegg Pathways: Down-Regulated Genes. Frozen compared to not frozen.***

Pathway ID	Pathway Description	p Value
3010	Ribosome	1.40E-07
190	Oxidative phosphorylation	3.01E-05
3050	Proteasome	0.005904
3020	RNA polymerase	0.00832
3440	Homologous recombination	0.009669
785	Lipoic acid metabolism	0.011151
730	Thiamine metabolism	0.043891

Largest Fold Changes

Before statistical filtering there were 233 features exhibiting greater than 2-fold up-regulation and 39 features exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 8 features with greater than 2-fold up-regulation and 8 with greater than 2-fold down regulation. These are displayed in tables 5.2.19 and 5.2.20 respectively.

Table 5.2.19: ***Up-regulated genes with fold change > 2, p < 0.001 Frozen compared to not frozen.***

Feature ID	Description	Pathway Description	Fold Change
fmpZa4DFHHeNv.0eFY	NA	-	3.879695
3fSd576nru6EgoOgJQ	prokineticin 2	-	3.289418
Bh8SXmPMQo.ETn13rk	tubulin, beta 1	Gap junction, Pathogenic Escherichia coli infection - EHEC, Pathogenic Escherichia coli infection - EPEC	2.819807
lXyXe.lHerNt1zpAlM	NA	-	2.691123
Nee665U3z14x54l6q4	sortilin-related receptor, L(DLR class) A repeats-containing	-	2.686725
0HMVdH0F1dFPpJckfk	NA	-	2.376661
HXVeluKF01hTsPFPZ8	chromosome 7 open reading frame 41	-	2.248996
upZTAWUjQXoncZdcb8	NA	-	2.037282

Table 5.2.20: ***Down-regulated genes with fold change > 2, p < 0.001 Frozen compared to not frozen.***

Feature ID	Description	Pathway Description	Fold Change
lXb5HniJCOrRSSex6l	ribosomal protein S29	Ribosome	-3.71109
rMFE5ACCNZO0VF9KaU	hypothetical protein LOC731985	-	-2.84094
u_VFUJ_oJu.uXs_UKo	hypothetical LOC283412	Ribosome	-2.29035
lCiieaASejheCuie1Q	similar to ribosomal protein L35	Ribosome	-2.24241
BD74JXg6Qorpp0Lgp0	Parkinson disease (autosomal recessive, early onset) 7	Neurodegenerative diseases, NA	-2.12063
lnnAvUkACNFlmLACnU	ribosomal protein L14	Ribosome	-2.1088
9Yllno5wtY7RQUn7l8	COMM domain containing 6	-	-2.0862
ZqKJpxpLGalIKUlJeQ	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	-	-2.02171

When the genes with significant fold changes were examined using DAVID 2008 (348, 349) there were no GO terms that were significantly enriched ($p < 0.001$).

Comparison of Male and Female Sample Groups

Significant Differential Expression

With an adjusted p value of < 0.001, there were 14 features significant differentially expressed: 11 up-regulated and 3 down-regulated.

GO Term Enrichment

There were no enriched GO terms in either significantly up-regulated or in significantly down-regulated genes.

KEGG Pathway Enrichment

There were found to be 10 enriched KEGG pathways in significantly up-regulated genes (displayed in Table 5.2.21) and no enriched KEGG pathways in significantly down-regulated genes.

Table 5.2.21. : *Kegg Pathways: Up-Regulated Genes. Male compared to female..*

Pathway ID	Pathway Description.	p Value
4742	Taste transduction	0.011404
4340	Hedgehog signalling pathway	0.012522
5110	Vibrio cholerae infection	0.013193
4720	Long-term potentiation	0.015206
4540	Gap junction	0.021243
4912	GnRH signalling pathway	0.021691
4916	Melanogenesis	0.022138
4910	Insulin signalling pathway	0.030411
4310	Wnt signalling pathway	0.031977
4020	Calcium signalling pathway	0.038462

Largest Fold Changes

Before statistical filtering there were 4 features exhibiting greater than 2-fold up-regulation and 1 feature exhibiting greater than 2-fold down-regulation. After statistical filtering ($p < 0.001$) there were 4 features with greater than 2-fold up-

regulation and one with greater than 2-fold down regulation. These up-regulated genes are displayed in table 5.2.22 and the down-regulated gene in table 5.2.23. When these features were examined further using DAVID 2008 (348, 349) there were no GO terms that were significantly enriched ($p < 0.001$).

Table 5.2.22.: ***Fold Changes > 2 for Up-Regulated Genes. Male compared to female. (Each line represents 1 feature on array).***

Feature ID	Description	Chromosome	Fold Change
QEpD_7OuProlBdrbvo	NA	-	26.43253
WeKOvcS_o.6U00FCCI	eukaryotic translation initiation factor 1A, Y-linked	Y	15.82515
BqhAqD1EXrh0452BT0	ribosomal protein S4, Y-linked 2	Y	2.216063
reXjDTnq9wwlcfUXDc	eukaryotic translation initiation factor 1A, Y-linked	Y	2.149208

Table 5.2.23.: ***Fold Changes > 2 for Down-Regulated Genes. Male compared to female. (Each line represents 1 feature on array).***

Feature ID	Description	Chromosome	Fold Change
T7rNJ6HSwRV.fSeqOo	X (inactive)-specific transcript (non-protein coding)	X	3.186006

Chapter 5.2: Discussion

Samples and population

The sample population in this chapter consisted entirely of infected and control infants. The samples used for these groups were predominantly but not entirely the same samples used in the previous section. The demographics of the study population overall therefore are similar and are presented in table 5.2.2. In other words, the infected group was notably more premature and therefore of lower birth weight and was also older in terms of postnatal age when sampled than the controls. As explained in chapter 5.1, in the early stages of the study all of the controls were term babies attending the neonatal unit for blood screening to exclude thyroid or adrenal problems in the early days of life. In the latter stages of the study care was taken to ensure that the control group also included preterm infants and infants more than a few days old. The population studied is representative of the neonatal unit population whilst ensuring that the control group contained infants from a wide range of gestations.

In terms of the organisms isolated in the infected group, coagulase negative staphylococci were the organisms most frequently isolated, reflecting the prematurity and vulnerability of the neonatal unit population. The full list of organisms isolated is displayed in table 5.2.4. The range of organisms isolated from this group is representative of the organisms encountered most commonly in our neonatal unit. Samples with positive cultures were only included in the infected group if the diagnosis of infection was felt to be robust. I was very careful to select only cases where there was clear clinical signs of infection and looked for repeat culture results,

isolates from other sites and haematological or biochemical supporting evidence wherever possible. I have presented supporting evidence for the use of cases of coagulase negative staphylococci and CMV in table 5.2.5. To be clear for the infant with CMV; CMV can be asymptomatic in neonates but this infant was clearly symptomatic and underwent repeated investigations. In all cases, the decision for inclusion or exclusion of culture-positive samples in the infected group was independently verified by a neonatal consultant (Ben Stenson). Any sample where there was doubt was not included in the infected group and not run on microarray.

Infected compared to control samples

It was very encouraging to see that the sample relation graph shown as figure 5.2.3 showed a split between control and infected groups. The only sample that did not group neatly was that of csb75. This sample was from the infant with CMV infection and was seen to group with the controls. This was an interesting finding in itself given that all of the other infected infants had bacterial infections. These observations are even more pleasing when you consider that this figure was generated before statistical filtering.

After statistical filtering of genes ($p < 0.01$) it can be seen from the heat map in figure 5.2.4 that there is again a separation of patterns between control and infected samples with the CMV infected sample being the exception. There seem to be 3 distinct groups of genes (marked at left of figure). The infected samples show down-regulation of group 1 and up-regulation of groups 2 and 3 whereas most of the controls show the converse. There is group of nine samples however that shows up-

regulation of groups 1 and 3 and down-regulation of group 2. This is very interesting as this group is made up of the majority of the preterm controls, one term baby, one growth restricted baby with neonatal encephalopathy and the infant with CMV. It is particularly interesting that the preterm controls have grouped together and are distinct from the infected preterms. This is very exciting and confirms that future studies on effects of gestation are warranted.

From examining the gene ontology terms for significantly up-regulated genes in the infected group compared to controls (table 5.2.7), it can be seen that a significant proportion are immune related. The most significantly enriched terms include immune system process, defence response, response to external stimulus, response to wounding, response to stress and inflammatory response. There are immune related terms throughout the list however and some that may be worth noting in particular are I-kappa B kinase/NK kappa B cascade, cytokine production, TNF production and αβ T cell activation. When KEGG analysis of up-regulated genes is examined (table 5.2.9) pathways related to TLR signalling, apoptosis, Jak-Stat and MAPK pathways are among the most enriched pathways with B cell receptor signalling, complement/coagulation and NK cell mediated cytotoxicity pathways also featuring. Many of these findings concur with other groups who have investigated sepsis in children and presepsis in adults and I list these here. NFkB and NFkB signalling have been noted to be up regulated in both adult presepsis and paediatric septic shock (152, 156) and up regulated as part of the common host response as described by Hossain and by Jenner (146, 159). TLRs 1, 4, 5 and 8 have been found to be up-regulated in adult presepsis (155, 156) and TLR signalling to be up-regulated in

children with septic shock (152). Up-regulation of MAPK14 was found to be up regulated in adults with sepsis (357) and other MAPK genes described as being part of the common host response (146, 159). TNF and cytokines such as IL1β, IL-6, and IL-8 have been described as being up regulated in the context of the common host response to infection (146, 159). In addition, it is known that blood levels of cytokine such as IL-1β, IL-6, IL-8 and TNFα are increased in sepsis (358). It is very encouraging that our findings concur with existing knowledge.

The gene ontology for most significantly down regulated genes is displayed in table 5.2.8 with "antigen processing and presentation of antigen via MHC II" being the only term that is clearly immune related. When KEGG analysis is performed the significantly down-regulated genes can be seen to map to T cell receptor signalling and antigen processing and presentation pathways (table 5.2.10). Down-regulation of T cell signalling, antigen processing and NK cell signalling have been described previously in children with septic shock (152, 153) and T cell response as being reduced in adult sepsis (282). As for with the CodeLink™ data, the down-regulated genes also map to pathways related to graft versus host disease, asthma and autoimmune diseases such as diabetes and thyroid disease (highlighted in green). It could be suggested that suppression of such processes may expected in order to allow response to infection.

When fold change is taken into account (fold change of greater than 2 and p value of less than 0.001) the genes related to immune functions that are up-regulated (table 5.2.11) include IL-1 receptor II which is involved in cytokine signalling, interferon

induced transmembrane protein, interleukin 1 receptor antagonist and TNFα induced protein 6. In addition to those with the largest fold changes listed above, other genes that mapped to these terms included (not exhaustively) Toll Like receptors 5 and 8, IL-1β, IL-8 receptors α and β, IL-18 receptor 1, IL-4 receptor, NFkB , GCSF 3, chemokine (C-C) receptors 1 and 3, interferon regulatory factor 7, metallopeptidase and complement components C5aR1 and7 C3aR1. In addition the following cluster of differentiation molecules were also up-regulated: CD55 (decay accelerating factor for complement), CD14 (involved in innate immunity), CD24 (cell adhesion molecule found on B cells), CD 82, CD97 (found on leukocytes) and CD163 (soluble form has anti-inflammatory role). It is reassuring to note that many of these were also up-regulated in the CodeLink™ study. In addition, CD14 has been previously identified as being involved in the common host response (146) and Jiang and colleagues observed elevated levels of IL-1β, IL-4 receptor and GCSF on stimulation of cord blood monocytes with LPS (78). IL-4 receptor and IL-18 receptor 1 have previously been found to be up-regulated in adult presepsis (155). In addition, both TLR 5 and TLR 8 have been previously described as up-regulated in adult presepsis (155, 156). TLR 8 agonists have been shown to activate neonatal antigen presenting cells (94) and to increase the magnitude and reduce duration of early systemic inflammatory response in neonatal mice (359). Interleukin 1 receptor antagonist levels have been noted to be elevated in children with sepsis (350) and to be elevated up to 2 days before the diagnosis of infection in neonates (351). When the up-regulated genes are examined by ontology (table 5.2.13) there are many enriched terms seen – the most significantly enriched being defence response, inflammatory response, response to wounding, response to external stimulus, immune system

process and response to stress. Other notable up-regulated terms include NFkB cascade, acute phase response, immunoglobulin mediated immune response, B cell mediated immunity, leucocyte mediated immunity, interleukin receptor activity, interleukin binding and cytokine binding.

The genes with fold change of greater than 2 and $p < 0.001$ that are down regulated and are related to immune functions (table 5.2.12) include MHC II, IL-7 receptor, IL11 receptor alpha, IL-2 inducible T cell kinase, TNF receptor factor 5, Fc receptors and chemokine (C-C) receptor 7 and ligand 5. In addition, there is down regulation of many T cell related cluster of differentiation molecules (CD2, CD3d, CD6, CD7, CD8a, CD52, CD96 and CD 247). Of these CD 2 and CD96 are also involved in NK cell function. CD79a and CD79b which are B cell receptor related were also down-regulated, as was CD27 which is a TNF receptor. These findings again concur with previous findings of reduced MHC II expression (78, 153, 353-355) and of down-regulated T cell functioning (152, 153, 282) in adult and paediatric sepsis. In addition, chemokine (C-C) ligand 5 has been found to be down-regulated in adult presepsis (155). Interestingly however, IL-7 receptor in the work of Jiang and colleagues on stimulating cord blood monocytes with LPS showed up-regulation (78). Jiang's work reflects in vitro stimulation. Our population were exposed to a range of pathogens in vivo and there could be pathogen related reasons or host physiological reasons why our samples show the opposite. The immune-related GO terms that are enriched for down-regulated genes (table 5.2.14) include lymphocyte/T cell/leukocyte activation, T cell receptor complex and MHC II receptor activity.

There is in summary, up-regulation of components of the immune system involved in both innate and adaptive immunity. There is up-regulation of cytokine and chemokine components, kinases and NF-kB signalling representing both pro- and anti- inflammatory processes. Complement related genes, Toll like receptors (specifically 5 and 8) and interferon related genes are also up-regulated. There is down-regulation of adaptive immune function particularly in relation to T cell function and regulation, as well as down regulation of antigen presentation and of MHC II activity.

Overall these results are very exciting. It looks likely that, with larger studies and with examination of training sets of data, immune gene expression signatures for neonatal infection will be able to be defined. The work presented here is novel in that it is describing host gene expression profiling of infection in a neonatal population. A proportion of the neonates being studied will be encountering infection for the first time. In the context of a developing immune system and possible gestational effects on immune response, it is very encouraging to see differential gene expression that has previously been described in older subjects. It is likely however that with further work expression patterns that are seen only in neonatal infection and not in older populations will be found. Some of the findings described in this chapter are not yet found in the literature and this probably represents a combination of this being a new field of research and expression patterns unique to neonates. Carrying out work such as presented in this chapter is an important step forward in moving towards using microarrays to diagnose neonatal

infection using host response patterns. The results presented here are part of ongoing work and I am pleased to have been able to do so in more in depth than originally intended in the context of this thesis. The findings so far are really encouraging and I am optimistic that our ongoing work will continue to lead to progress in this area.

Comparison of samples that were frozen with those that were not frozen.

For significantly differentially expressed genes, there were none of the gene ontology terms (tables 5.2.15 and 5.2.16) that were specifically immune related. Table 5.2.17 shows KEGG analysis of significantly up-regulated genes and of these only the Jak-Stat pathway was enriched (enriched p value of 0.0104). None of the enriched pathways for significantly down-regulated genes were immune-related (table 5.2.18). When taking fold change into account (fold change of greater than 2, $p < 0.001$) only one gene term in the up-regulated group was identifiable as immune related (Pathogenic E coli related) and none in the down-regulated group (tables 5.2.19 and 5.2.20). There were no GO terms identified for genes differentially expressed with fold change greater than 2. The lack of immune related differential gene expression between samples that were frozen prior to RNA extraction in relation to those not frozen is encouraging. This supports use of either fresh or frozen samples for studies of neonatal blood. Laboratory processing of samples can therefore be batched efficiently. It would also facilitate expedient extraction of local samples while allowing samples transported from elsewhere to be frozen if a multi-centre study were to follow. These findings agree with the findings of Vartanian and colleagues who found little or no difference in expression profiling from PAXgene™ blood samples whether they were fresh or frozen when RNA was extracted (360).

Comparison of samples from males compared to females.

It was both expected and encouraging that there was little difference in gene expression between the male and female groups. It is reassuring that the genes with the largest fold changes of differential expression, displayed in tables 5.2.22 and 5.2.23, were either found on the Y chromosome (relatively up-regulated in boys) or the X chromosome (relatively down-regulated in boys). This provides confidence that the arrays were correctly differentiating between samples. There were no GO terms enriched for differentially expressed genes either using significance alone or when considering fold change. The KEGG pathways enriched for up-regulated genes are displayed in table 5.2.21 and of these, only "Vibrio cholerae infection" relates to immune function (p value 0.013). The actual significance of this is not clear.

Note on Gestational Age

Although gestational age was not examined statistically for this data, I believe that the effect of gestational age on gene expression in response to infection is an area that needs to be studied further. The heat-map at the beginning of this section showing clustering of premature infants confirms this. This indicated that preterm control infants showed different gene expression profiles from infected preterm infants and from term controls. I do not think that the differences seen between the control and infected groups in this chapter can be explained by gestational differences as it would be expected that premature infants would be less able to show significant up-regulation of immune genes than less premature infants if there indeed

were to be any difference at all. This data does however show that neonates are capable of producing significant immune responses to infection.

Conclusions

Differences in gene expression can be detected between infected and control neonates. Many of these differences are immune-function related. The ability to detect such differences could have diagnostic potential but this would clearly require further investigation. Such data also provides evidence that neonates are capable of mounting a substantial immune response.

Freezing samples prior to RNA extraction does not lead to significant differences in immune related gene expression compared to fresh samples. This has positive implications for expanding future studies to include other centres.

Gestational age may affect the patterns of RNA expression profile seen and this needs further investigation.

Section 5.3: Comparison of CodeLink™ and Illumina® Findings

Background

Having run infected and control samples on two different microarray platforms, I present here a comparison of the ontology of significantly up- and down-regulated genes for each platform. The infected and control groups for each platform consist of mostly but not entirely the same samples. In other words, this is not a direct comparison and therefore some differences would be expected. Also, as would be expected, the two platforms are not looking at exactly the same gene fragments. This makes it very difficult to compare the genes themselves and is why I have opted to look at the ontology and pathway comparisons. It should also be noted that the cut-offs for significance in order to carry the genes of interest on for ontology and pathway examinations differed between groups: $p < 0.001$ for the Illumina® samples and $p < 0.01$ for CodeLink™ samples.

Aim

To provide a brief comparison of the results obtained for the CodeLink™ and Illumina® studies (chapters 5.1 and 5.2) with the focus being on shared findings.

Comparison of Gene Ontology

The top ten most significantly enriched GO terms for significantly up-regulated genes for each platform are displayed in table 5.3.1 and those for significantly down-regulated genes in table 5.3.2. Seven terms are shared between platforms in the top

ten up-regulated group and six terms in the ten down-regulated terms. Shared terms are highlighted in pink.

Table 5.3.1: ***Top 10 most significantly enriched GO terms for each platform: Up-regulated genes.***

CodeLink™		Illumina®	
GO ID	**GO Term**	**GO ID**	**GO Term**
GO:0009611	response to wounding	GO:0002376	immune system process
GO:0002376	immune system process	GO:0006952	defence response
GO:0009605	response to external stimulus	GO:0009605	response to external stimulus
GO:0006955	immune response	GO:0007243	protein kinase cascade
GO:0006952	defence response	GO:0007242	intracellular signalling cascade
GO:0006954	inflammatory response	GO:0009611	response to wounding
GO:0006950	response to stress	GO:0006950	response to stress
GO:0007599	haemostasis	GO:0006954	inflammatory response
GO:0007596	blood coagulation	GO:0006955	immune response
GO:0048522	positive regulation of cellular process	GO:0012501	programmed cell death

Table 5.3.2: ***Top 10 most significantly enriched GO terms for each platform: Down-regulated genes.***

CodeLink™		Illumina®	
GO ID	**GO Term**	**GO ID**	**GO Term**
GO:0010467	gene expression	GO:0010467	gene expression
GO:0006414	translational elongation	GO:0006412	translation
GO:0034645	cellular macromolecule biosynthetic process	GO:0043170	macromolecule metabolic process
GO:0009059	macromolecule biosynthetic process	GO:0044237	cellular metabolic process
GO:0006412	Translation	GO:0008152	metabolic process
GO:0043170	macromolecule metabolic process	GO:0006414	translational elongation
GO:0034961	cellular biopolymer biosynthetic process	GO:0044238	primary metabolic process
GO:0043284	biopolymer biosynthetic process	GO:0006396	RNA processing
GO:0043283	biopolymer metabolic process	GO:0009059	macromolecule biosynthetic process
GO:0034960	cellular biopolymer metabolic process	GO:0043283	biopolymer metabolic process

Comparison of KEGG Pathway Enrichment

The top ten most significantly enriched KEGG pathways for significantly up-regulated genes for each platform are displayed in table 5.3.3 and those for significantly down-regulated genes in table 5.3.4. Two pathways are shared between

platforms in the top ten up-regulated group and six of the top ten down-regulated pathways.

Table 5.3.3: ***Top 10 most significantly enriched KEGG Pathways for each platform: Up-regulated genes.***

CodeLink™		Illumina®	
Pathway ID	**Pathway Description**	**Pathway ID**	**Pathway Description**
4060	Cytokine-cytokine receptor interaction	4620	Toll-like receptor signalling pathway
4610	Complement and coagulation cascades	4920	Adipocytokine signalling pathway
530	Aminosugars metabolism	5221	Acute myeloid leukaemia
4640	Hematopoietic cell lineage	5215	Prostate cancer
4910	Insulin signalling pathway	4210	Apoptosis
4510	Focal adhesion	5220	Chronic myeloid leukaemia
4620	Toll-like receptor signalling pathway	4540	Gap junction
4512	ECM-receptor interaction	4630	Jak-STAT signalling pathway
4210	Apoptosis	5120	Epithelial cell signalling in Helicobacter pylori infection
4810	Regulation of actin cytoskeleton	4520	Adherens junction

Table 5.3.4: ***Top 10 most significantly enriched KEGG Pathways for each platform: Down-regulated genes.***

CodeLink™		Illumina®	
Pathway ID	**Pathway Description**	**Pathway ID**	**Pathway Description**
3010	Ribosome	3010	Ribosome
5332	Graft-versus-host disease	190	Oxidative phosphorylation
4660	T cell receptor signalling pathway	970	Aminoacyl-tRNA biosynthesis
5330	Allograft rejection	4660	T cell receptor signalling pathway
5310	Asthma	5310	Asthma
4640	Hematopoietic cell lineage	5332	Graft-versus-host disease
4940	Type I diabetes mellitus	280	Valine, leucine and isoleucine degradation
5320	Autoimmune thyroid disease	271	Methionine metabolism
970	Aminoacyl-tRNA biosynthesis	5330	Allograft rejection
4612	Antigen processing and presentation	5340	Primary immunodeficiency

Comparison by Largest Fold-Change

For the largest fold-change, the statistical significance cut-offs were the same for both groups: fold change of greater than log 2 and $p < 0.001$. Due to the difference in probes used for each platform it may be difficult to compare the gene-sets directly. On direct comparison of genes with significant fold changes between platforms there were 61 shared up-regulated genes and 87 shared down-regulated genes – found by matching Entrez IDs. These shared genes are displayed in tables 5.3.5 (a to c) and 5.3.6(a to d).

Table 5.3.5a: ***Shared Genes Between Microarray Platforms with Significant Fold Changes.***
Up-regulated Genes

Entrez ID	Description	Pathway Description
249	alkaline phosphatase, liver/bone/kidney	gamma-Hexachlorocyclohexane degradation, Folate biosynthesis
306	annexin A3	-
366	aquaporin 9	-
383	arginase, liver	Urea cycle and metabolism of amino groups, Arginine and proline metabolism
602	B-cell CLL/lymphoma 3	-
634	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	-
762	carbonic anhydrase IV	Nitrogen metabolism
1604	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	Complement and coagulation cascades, Hematopoietic cell lineage
1647	growth arrest and DNA-damage-inducible, alpha	MAPK signalling pathway, Cell cycle, p53 signaling pathway
2180	acyl-CoA synthetase long-chain family member 1	Fatty acid metabolism, PPAR signalling pathway, Adipocytokine signalling pathway
2358	formyl peptide receptor 2	Neuroactive ligand-receptor interaction
2710	glycerol kinase	Glycerolipid metabolism, PPAR signalling pathway
3101	hexokinase 3 (white cell)	Glycolysis / Gluconeogenesis, Fructose and mannose metabolism, Galactose metabolism, Starch and sucrose metabolism, Streptomycin biosynthesis, Aminosugars metabolism

Table 5.3.5b: ***Shared Genes Between Microarray Platforms with Significant Fold Changes.***
Up-regulated Genes

3142	H2.0-like homeobox	-
3422	isopentenyl-diphosphate delta isomerase 1	Biosynthesis of steroids, Terpenoid biosynthesis
3557	interleukin 1 receptor antagonist	-
3687	integrin, alpha X (complement component 3 receptor 4 subunit)	Regulation of actin cytoskeleton
3726	jun B proto-oncogene	-
4318	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	Leukocyte transendothelial migration, Bladder cancer
4792	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	Apoptosis, Toll-like receptor signalling pathway, T cell receptor signalling pathway, B cell receptor signalling pathway, Adipocytokine signalling pathway, Epithelial cell signalling in Helicobacter pylori infection, Prostate cancer, Chronic myeloid leukemia,
5208	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	Fructose and mannose metabolism
5341	pleckstrin	-
5553	proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)	Asthma
6005	Rh-associated glycoprotein	-
6272	sortilin 1	-
6283	S100 calcium binding protein A12	-
6478	seven in absentia homolog 2 (Drosophila)	-
7057	thrombospondin 1	Cell junctions, p53 signalling pathway, TGF-beta signalling pathway, Focal adhesion, ECM-receptor interaction, Bladder cancer
7130	tumor necrosis factor, alpha-induced protein 6	-
7357	UDP-glucose ceramide glucosyltransferase	Sphingolipid metabolism, Glycan structures - biosynthesis 2
7850	interleukin 1 receptor, type II	MAPK signalling pathway, Cytokine-cytokine receptor interaction, Hematopoietic cell lineage
8291	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	-
8530	cystatin F (leukocystatin)	-
8778	sialic acid binding Ig-like lectin 5	-
9050	proline-serine-threonine phosphatase interacting protein 2	-
9334	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	O-Glycan biosynthesis, Glycan structures - biosynthesis 1

Table 5.3.5c: ***Shared Genes Between Microarray Platforms with Significant Fold Changes.***
Up-regulated Genes

9586	cAMP responsive element binding protein 5	Prostate cancer
10079	ATPase, class II, type 9A	-
10226	mannose-6-phosphate receptor binding protein 1	-
10410	interferon induced transmembrane protein 3 (1-8U)	-
10900	RUN domain containing 3A	-
23569	peptidyl arginine deiminase, type IV	-
23765	interleukin 17 receptor A	Cytokine-cytokine receptor interaction
51311	toll-like receptor 8	Toll-like receptor signalling pathway
51312	solute carrier family 25, member 37	-
51327	erythroid associated factor	-
53831	G protein-coupled receptor 84	-
54498	spermine oxidase	-
55331	phytoceramidase, alkaline	Sphingolipid metabolism
55432	YOD1 OTU deubiquinating enzyme 1 homolog (S. cerevisiae)	Benzoate degradation via CoA ligation, Limonene and pinene degradation, Biosynthesis of unsaturated fatty acids
57568	signal-induced proliferation-associated 1 like 2	-
60675	prokineticin 2	-
64092	SAM domain, SH3 domain and nuclear localization signals 1	-
80216	alpha-kinase 1	-
84418	chromosome 5 open reading frame 32	-
91543	radical S-adenosyl methionine domain containing 2	-
116496	family with sequence similarity 129, member A	-
116844	leucine-rich alpha-2-glycoprotein 1	-
118932	ankyrin repeat domain 22	-
120892	leucine-rich repeat kinase 2	Neurodegenerative diseases
199675	chromosome 19 open reading frame 59	-

Table 5.3.6a: ***Shared Genes Between Microarray Platforms with Significant Fold Changes.***
Down-regulated Genes

Entrez ID	Description	Pathway Description
274	bridging integrator 1	-
914	CD2 molecule	Cell adhesion molecules (CAMs), Hematopoietic cell lineage
915	CD3d molecule, delta (CD3-TCR complex)	Hematopoietic cell lineage, T cell receptor signalling pathway, Primary immunodeficiency
917	CD3g molecule, gamma (CD3-TCR complex)	Hematopoietic cell lineage, T cell receptor signaling pathway
919	CD247 molecule	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway
924	CD7 molecule	Hematopoietic cell lineage
1039	cerebellar degeneration-related protein 2, 62kDa	-
1043	CD52 molecule	-
1178	Charcot-Leyden crystal protein	-
1937	eukaryotic translation elongation factor 1 gamma	-
1938	eukaryotic translation elongation factor 2	-
2113	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	-
3003	granzyme K (granzyme 3; tryptase II)	-
3094	histidine triad nucleotide binding protein 1	Neuroactive ligand-receptor interaction
3109	major histocompatibility complex, class II, DM beta	-
3113	major histocompatibility complex, class II, DP alpha 1	Cell adhesion molecules (CAMs), Antigen processing and presentation, Type I diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease
3399	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	Cell adhesion molecules (CAMs), Antigen processing and presentation, Hematopoietic cell lineage, Type I diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease
3575	interleukin 7 receptor	Cell adhesion molecules (CAMs), Antigen processing and presentation, Hematopoietic cell lineage, Type I diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease

Table 5.3.6b: ***Shared Genes Between Microarray Platforms with Significant Fold Changes.***
Down-regulated Genes

3702	IL2-inducible T-cell kinase	Cytokine-cytokine receptor interaction, Jak-STAT signalling pathway, Hematopoietic cell lineage
3820	killer cell lectin-like receptor subfamily B, member 1	T cell receptor signalling pathway, Leukocyte transendothelial migration
3945	lactate dehydrogenase B	Calcium signalling pathway, Phosphatidylinositol signalling system, Gap junction, Long-term potentiation, Long-term depression, Taste transduction, GnRH signalling pathway
3983	actin binding LIM protein 1	Glycolysis / Gluconeogenesis, Cysteine metabolism, Pyruvate metabolism, Propanoate metabolism
4068	SH2 domain protein 1A	Cytokine-cytokine receptor interaction, Antigen processing and presentation, Type I diabetes mellitus
4118	mal, T-cell differentiation protein	Cytokine-cytokine receptor interaction
4145	megakaryocyte-associated tyrosine kinase	Natural killer cell mediated cytotoxicity
5583	protein kinase C, eta	Focal adhesion, Long-term potentiation, Regulation of actin cytoskeleton, Insulin signalling pathway
5859	glutaminyl-tRNA synthetase	-
6015	ring finger protein 1	-
6124	ribosomal protein L4	-
6137	ribosomal protein L13	Ribosome
6138	ribosomal protein L15	Ribosome
6146	ribosomal protein L22	Ribosome
6152	ribosomal protein L24	-
6160	ribosomal protein L31	Ribosome
6208	ribosomal protein S14	Ribosome
7188	TNF receptor-associated factor 5	-
7423	vascular endothelial growth factor B	-
8445	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2	-
8631	src kinase associated phosphoprotein 1	-
8725	chromosome 19 open reading frame 2	-
9214	Fas apoptotic inhibitory molecule 3	Ribosome
9452	integral membrane protein 2A	-
9806	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	-
9987	heterogeneous nuclear ribonucleoprotein D-like	-

Table 5.3.6c: ***Shared Genes Between Microarray Platforms with Significant Fold Changes. Down-regulated Genes***		
10125	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	-
10225	CD96 molecule	-
10412	TGF beta-inducible nuclear protein 1	-
10480	eukaryotic translation initiation factor 3, subunit M	-
10801	septin 9	-
23180	raftlin, lipid raft linker 1	-
23521	ribosomal protein L13a	-
25777	unc-84 homolog B (C. elegans)	-
26053	autism susceptibility candidate 2	-
27040	linker for activation of T cells	-
28955	dexamethasone-induced transcript	-
29121	C-type lectin domain family 2, member D	Natural killer cell mediated cytotoxicity, T cell receptor signalling pathway, Fc epsilon RI signalling pathway
51020	HD domain containing 2	-
51176	lymphoid enhancer-binding factor 1	-
51275	chromosome 12 open reading frame 47	-
51466	Enah/Vasp-like	-
54463	family with sequence similarity 134, member B	-
54674	leucine rich repeat neuronal 3	-
55013	coiled-coil domain containing 109B	-
55340	GTPase, IMAP family member 5	-
59338	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1	-
79058	alveolar soft part sarcoma chromosome region, candidate 1	-
79961	DENN/MADD domain containing 2D	-
80212	coiled-coil domain containing 92	-
80342	TRAF3 interacting protein 3	-
81606	limb bud and heart development homolog (mouse)	-
84678	F-box and leucine-rich repeat protein 10	-
84881	RNA pseudouridylate synthase domain containing 4	-
84969	TOX high mobility group box family member 2	-
85315	progestin and adipoQ receptor family member VIII	-
91523	family with sequence similarity 113, member B	-
114932	Morf4 family associated protein 1-like 1	-
132299	OCIA domain containing 2	-

Table 5.3.6d: *Shared Genes Between Microarray Platforms with Significant Fold Changes.* *Down-regulated Genes*		
154075	sterile alpha motif domain containing 3	-
155066	ATPase, H+ transporting V0 subunit e2	-
168537	GTPase, IMAP family member 7	Oxidative phosphorylation, Vibrio cholerae infection, Epithelial cell signalling in Helicobacter pylori infection
170622	COMM domain containing 6	-
220433	similar to 40S ribosomal protein S4, X isoform	-
387751	GTPase, very large interferon inducible 1	-
389289	chromosome 5 open reading frame 39	-
399665	family with sequence similarity 102, member A	-
439949	hypothetical gene supported by AY007155	-

Of these shared genes with significant fold changes those highlighted in red can be easily identified as having immune related-functions. Tables 5.3.7 (up-regulated genes) and table 5.3.8 a and b (down-regulated genes) display the top DAVID derived GO terms (biological process terms) from all of the genes with significant fold changes from each platform. In the top up-regulated terms the platforms shared four terms and in the top down-regulated terms they shared seven. It is also clear that many of the other terms listed are similar but not exactly the same.

Table 5.3.7: ***DAVID GO terms(Biological process) for each platform: Significantly up-regulated genes with fold change > 2***

Illumina®			CodeLink™		
GO Term	**No of genes**	**P value**	**GO Term**	**No of genes**	**P value**
Defence response	56	1.1E-23	Response to wounding	18	4.9E-7
Inflammatory response	38	8.5E-20	Response to external stimulus	20	7.5E-6
Response to wounding	39	2.4E-15	Inflammatory response	12	1.1E-4
Response to external stimulus	46	2.9E-14	Response to stress	24	1.4E-4
Immune system process	62	2.3E-13	Regulation of body fluid levels	8	1.4E-4
Response to stress	53	4.1E-10	Blood coagulation	7	3.1E-4
Response to stimulus	102	2.6E-9	Coagulation	7	3.5E-4
Immune response	47	2.7E-9	Hemostasis	7	4.1E-4
Taxis	16	1.6E-7	Response to other organism	9	5.6E-4
Chemotaxis	16	1.6E-7			

Table 5.3.8a: ***DAVID GO terms(Biological process) for each platform: Significantly down-regulated genes with fold change > 2***

Illumina®			CodeLink™		
GO Term	**No of genes**	**P value**	**GO Term**	**No of genes**	**P value**
Translation	45	1.2E-22	Gene expression	197	3.3E-12
Macromolecule biosynthetic process	47	6.1E-18	RNA metabolic process	152	2.5E-7
Cellular biosynthetic process	50	5.4E-16	Macromolecule metabolic process	305	2.7E-7
Biosynthetic process	51	6.5E-12	Primary metabolic process	341	3.0E-7
Gene expression	77	2.6E-8	Cellular process	482	9.5E-7
Protein metabolic process	76	6.7E-8	Regulation of cell activation	15	1.6E-6
Cellular protein metabolic process	71	3.9E-7	T cell activation	17	2.4E-6

Table 5.3.8b: ***DAVID GO terms(Biological process) for each platform: Significantly down-regulated genes with fold change > 2***

Illumina®			CodeLink™		
GO Term	**No of genes**	**P value**	**GO Term**	**No of genes**	**P value**
Cellular macromolecule metabolic process	71	7.1E-7	Translation	46	2.4E-6
Immune system process	31	2.8E-5	RNA processing	37	3.0E-6
Immune reponse	27	3.0E-5	Metabolic process	365	3.3E-6
Cell activation	12	1.1E-4	Regulation of lymphocyte activation	14	4.5E-6
Ribonucleoprotein complex biogenesis and assembly	11	1.3E-4	Lymphocyte activation	21	5.6E-6
Hemopoietic or lymphoid organ development	11	1.5E-4	Biopolymer metabolic process	232	8.5E-6
Macromolecule metabolic process	109	1.5E-4	Regulation of T cell activation	12	9.8E-6
Immune system development	11	2.3E-4	Cellular metabolic process	332	1.1E-5
Cellular defence response	7	2.3E-4	Leukocyte activation	22	1.1E-5
RNA processing	16	2.8E-4	Ribosome biogenesis and assembly	14	1.3E-5
Lymphocyte activation	10	3.1E-4	Regulation of biological process	208	2.1E-5
Lymphocyte differentiation	7	3.3E-4	Cell activation	23	2.4E-5
T cell activation	16	4.2E-4	Cellular biosynthetic process	67	2.6E-5
Leukocyte activation	10	7.7E-4	Nucleobase, nucleoside, nucleotise and nucleic acid metabolic process	181	3.1E-5

Chapter 5.3: Discussion

When comparing the results of the work presented in 5.1 with that in 5.2, some differences in expression profiles dependent on the microarray platform used are to be expected. In addition, the study cohorts run on each platform, although made up of predominantly the same samples were not identical. It is encouraging that so many of the most significantly enriched functional terms for both up-regulated and down-regulated genes are shared by both platforms.

The up-regulated terms consistently include immune response, inflammatory response, response to stimulus, response to stress and response to wounding (tables 5.3.1 and 5.3.7). The common up-regulated pathways included apoptosis and TLR signalling (table 5.3.3). Shared up-regulated genes with fold change greater than 2 include CD55 (involved in complement and coagulation functions), MAPK, IL-1 receptor antagonist, matrix metallopeptidase 9, integrin, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha, TNF alpha induced protein 6, IL-1 receptor type 2, interferon induced transmembrane protein 3, IL-17 receptor antagonist and TLR 8. Many of these genes have been discussed already in chapters 5.1 and 5.2. Up-regulation of such genes makes physiological sense in a host fighting infection, with several of the genes having been observed to be up-regulated in infection in adults and children with sepsis (146, 152, 156, 357).

KEGG pathways that were enriched for significantly down-regulated genes in both platforms included T cell receptor signalling pathway, graft versus host disease,

allograft rejection and asthma (table 5.3.6). The shared down-regulated ontology terms in table 5.3.2 are not clearly specific to immune function but when fold change is taken into account the shared down-regulated ontology terms are seen to include T cell activation, lymphocyte activation and leukocyte activation (table 5.3.8). Down-regulated genes with fold changes of over two on both platforms included cluster of differentiation molecules CD2, CD3, CD247, CD7, and CD5. MHC II, interleukin 7 receptor, interleukin-2 inducible T cell kinase, Killer cell lectin-like receptor B, mal T cell differentiation protein and TNF receptor associated factor 5 were also down-regulated. The down-regulation of T cell activity and MHC II function has been previously been described in paediatric septic shock (152, 153).

Results shared between platforms are likely to be of particular interest when looking at potential signature sets of genes. Ongoing genomic and pathway analysis within the department is likely to include further examination of these terms.

Conclusion

This comparative data acts as corroboration that neonatal samples can be run on different microarray platforms and that neonates with infection show differential gene expression in immune related genes compared to control neonates.

Chapter 6

Overall Conclusions

And

Future Applications

My central hypothesis was that distinct patterns of host transcriptional response detected from neonatal whole blood may enable diagnosis of neonatal infection. When I embarked on this project there was no literature regarding the use of neonatal whole blood for gene expression profiling. In addition, potential clinical applications of gene expression profiling were only just starting to appear, making this a very exciting area of research to be involved in. In order to address this hypothesis the questions of whether indeed it would be possible to use neonatal whole blood for expression profiling in the first place, and, whether differences in gene expression profiles between infected and control infants could be detected using microarray technology needed to be resolved. It is these questions that the work in this thesis addresses.

It is very satisfying on reaching the end of this thesis to have achieved more than I originally set out to do. From the outset, for financial reasons, this was planned to be a small-scale study using in the region of 6-12 microarrays as a feasibility study leading to something potentially larger. I have been immensely privileged to have had the opportunity, on achieving my initial aims, to be able to carry out further microarray studies examining differences in RNA expression profiles between infected and control neonates.

Due to the necessity of blood samples from neonates being small, there was always the possibility that such samples could not yield sufficient RNA for microarray work or that poor RNA quality may hamper progress. Fortunately, this was not the case and I have demonstrated that high quality RNA can consistently be obtained from

small neonatal samples. Furthermore, I was able to determine that the PAXgene™ Blood RNA system yielded the highest quality RNA of the tubes and methods investigated. In addition, I was able to set parameters of incubation time and sample storage for use in future studies. In generating gene expression profiles from the neonatal samples, this study has yielded lots of information, examining multiple markers simultaneously from low volumes of blood. To gain this volume of information using other techniques would require prohibitive amounts of blood, time and labour. In addition, I have shown it is possible to obtain samples to achieve such data without disrupting clinical care or altering the practice of clinical sampling. In refining the techniques used, this work, in addition to the potential diagnostic benefits, will undoubtedly be useful for future clinical research.

As with any investigative tool, there are also disadvantages to the approach we have adopted. The expression profiles obtained represent a snapshot in time and changes may occur depending on the exact time point of sampling within the course of the illness. The samples described in this thesis were taken at the point of clinical suspicion of infection. As this is also the point where clinical decision making becomes necessary, our samples represent an important clinical point in time. Future studies into the progression of neonatal expression profiles within the time course of an infection would be informative and indeed, other groups have begun doing this with adult and paediatric populations (153-155, 285). In addition, host transcriptional profiling does not reflect every aspect of biology but it does present a large amount of increased information. One aspect of biology that is not examined is

that of protein function. It may be that in future combined approaches will be adopted both in terms of research and in clinical application.

Due to the paucity of previous work in this area it was never certain that our microarray data would be informative. On progressing to gene expression profiling it could have been the case that the results obtained would be too variable to allow any response to infection to be elucidated. Also, given the known susceptibility of neonates to infection, it could have been the case that very little in the way of immune response be seen at all. In fact, this was not true. What I have presented in this thesis is evidence of a very clear, substantial immune response being detected. It is exciting that many of the differentially expressed genes identified in this work have immune and inflammatory functions and that the profiles being detected are therefore clearly disease related. It is also encouraging that many of our findings concur with those found in studies of adult and paediatric sepsis (146, 152, 153, 155, 156, 159, 282). Our results also show differential expression of genes that have not previously been described in studies of host transcriptional response to infection. There are likely to be differences in the profiles found in a neonatal population compared to older populations, reflecting the developing immune system and in some cases, the first encounter with infection. Deeper analyses which were beyond the scope of this study and further neonatal studies will help elucidate this further. In addition, there are still relatively few publications in this area. It would be expected that as more publications emerge, even greater understanding of the host transcriptional response to infection will be achieved. Definition of a signature set of genes for infection is now closer to reality.

The work presented in this thesis is clearly supportive of the hypothesis that distinct patterns of host transcriptional response detected from neonatal whole blood can enable diagnosis of neonatal infection. More work however needs to be done to prove this. Validation of microarray findings is necessary – for example confirming the presence of genes of interest with RT-PCR would provide useful corroboration. The next steps would be to define a set of signature genes and to test this with expression profiles from future samples. In addition, as collaboration, potential signature sets of genes could be tested against sets of infected and control profiles from other groups. Closer examination of the group of samples with suspected infection (the possible infection group from chapter 5.1) would also be valuable. It would be in this group especially, for whom more accurate diagnosis could well change clinical management, that clear and accurate division into infected and not infected would be greatly beneficial. Unfortunately such analyses are beyond the time and budget set out for this project but there are plans for future work in this area. Also, greater understanding of the effects of factors such as gestation, postnatal age, drugs including antibiotics and steroids, and co-existent medical conditions is needed. This will require large-scale studies.

Furthermore, translation of the use of diagnostic host expression signatures from research tools to clinical tools will require further work. Currently, the costs involved in the use of microarray technology are prohibitive and the detection of host expression patterns takes several days. Once a robust set of genes is validated as being able to diagnose infection then production of a custom microarray would be possible, therefore reducing the cost of further studies. Results of such studies could

then be used to enable development of a more closed system that would be rapid and streamlined, potentially giving results within a few hours. The final diagnostic tests need not be microarray based. Markers identified and validated from microarray work could be detected using a PCR platform for example. In addition, if the signatures pointed towards protein markers of infection then use of more traditional techniques of protein detection may be appropriate. Ultimately, I would hope that rapid bedside testing for infection would be possible. From a clinical viewpoint, the aim would be bedside devices that could determine not only the presence or absence of infection but if the infection is caused by bacteria, virus, fungus and the specific microorganism responsible. It may be the case that host gene expression profiles can be used to determine not only the presence of infection but also the group of organisms that the pathogen responsible belongs to or even to identify the individual organism. However, this may not be necessary if host gene expression profiling is used in conjunction with organism detection. Methods of rapid pathogen detection using gene expression profiling are emerging and look very promising in terms of diagnostics. I would envisage that bedside diagnostic tests would incorporate both pathogen detection and host response profiling. This would enable specific diagnosis of presence of a pathogenic microorganism, potentially give specific information on sensitivity or resistance to antimicrobials and also determine whether the organism was causing infection as opposed to being a contaminant or being carried without symptoms. It may be that host expression profiling will also yield valuable information regarding prognosis. Any diagnostic test would need to have very high sensitivity and specificity as the consequences of missing neonatal infection can be catastrophic. Neonatologists are therefore likely to need complete confidence in any

test in order for it to be accepted into clinical practise and used as a gold standard. Successful development of such tools that could diagnose neonatal infection more accurately and rapidly than is currently possible would be greatly beneficial. Although such tests are likely to be financially expensive to begin with, the costs associated with production would be expected to come down with time and should be off-set with savings made due to expected reduction in morbidity and mortality, reduced hospital stay, more tailored antibiotic therapy and reduced iatrogenic complications. The benefits of reduced parental anxiety and the neonatal and long-term benefits to the babies' emotional as well as physical wellbeing are immeasurable.

In summary, the work presented in this thesis has shown that it is possible to consistently obtain high quality RNA from neonatal blood in sufficient quantity to use for microarray work. It has also shown that it is possible to detect differences in host gene expression profile between infected and non-infected neonates. Many of these differences are seen to be related to immune functions. The results are highly supportive of the hypothesis that distinct patterns of host transcriptional response detected from neonatal whole blood may enable diagnosis of neonatal infection. This is an exciting time in the world of infection diagnostics and I believe that the work presented in this thesis represents some significant strides towards microarray diagnosis of infection in the newborn.

References

1. Greenough A. Bacterial sepsis and meningitis. Seminars in Neonatology. 1996;1:147-159.
2. Stoll BJ, Hansen NI, Adams-Chapman I, Fanaroff AA, Hintz SR, Vohr B, Higgins R. Neurodevelopmental and Growth Impairment Among Extremely Low-Birth-Weight Infants with Neonatal Infection. Journal of the American Medical Association. 2004;292:2357-2365.
3. ISD Scotland. Scottish Perinatal Morbidity and Mortality Report 2004. ISD Scotland. 2004.
4. Stoll BJ, Hansen NI, Fanaroff AA, Wright LL, Carlo WA, Ehrenkranz RA, Lemons JA, Donovan EF, Stark AR, Tyson JE, Oh W, Bauer CR, Korones SB, Shankaran S, Laptook AR, Stevenson DK, Papile LA, Poole WK. Late-onset sepsis in very low birth weight neonates: the experience of the NICHD neonatal research network. Pediatrics. 2002;110(2):285-291.
5. Spitzer AR, Kirkby S, Kornhauser M. Practice Variation in Suspected Neonatal Sepsis: A Costly Problem in Neonatal Intensive Care. Journal of Perinatology. 2005;25(4):265-269.
6. Aggarwal R, Sarkar N, Deorari AK, Paul VK. Sepsis in the newborn. Indian Journal of Pediatrics. 2001;68(12):1143-1147.
7. Zaidi AKM, Awasthi S, deSilva HJ. Burden of infectious diseases in South Asia. British Medical Journal. 2004;328:811-815.
8. Lever A, Mackenzie I. Sepsis: definition, epidemiology and diagnosis. British Medical Journal. 2007;335:879-883.
9. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G, For the International Sepsis Definitions Conference. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Critical Care Medicine. 2003;31(4):1250-1256.
10. Goldstein B, Giroir B, Randolph A, and the Members of the International Consensus Conference on Pediatric Sepsis. International pediatric consensus conference: Definitions for sepsis and organ dysfunction in pediatrics. Pediatric Critical Care Medicine. 2005;6(1):2-8.
11. Medicine ACoCPaSoCC. American College of Chest Physicians and Society of Critical Care Medicine Concensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Critical Care Medicine. 1992;20:864-874.
12. Brook I. Unexplained fever in young children: how to manage severe bacterial infection. British Medical Journal. 2003;327:1094-1097.
13. The Merck Manual. Neonatal Infections. In section 19: Paediatrics, Chapter 260 Disturbances in Newborns and Infants. The Merck Manual of Diagnosis and Therapy; 2006.
14. Osrin D, Vergnano S, Costello A. Serious bacterial infections in newborn infants in developing countries. Current Opinion in Infectious Diseases. 2004;17(3):217-224.
15. Burgner D, Isaacs D. Nosocomial neonatal infections. Seminars in Neonatology. 1996;1:169-175.

16. Stoll BJ, Gordon T, Korones SB, Shankaran S, Tyson JE, Bauer CR, Fanaroff AA, Lemons JA, Donovan EF, Oh W, Stevenson DK, Ehrenkranz RA, Papile LA, Verter J, Wright LL. Early-onset sepsis in very low birth weight neonates: A report from the National Institute of Child Health and Human Development Neonatal Research Network. The Journal of Pediatrics. 1996;129(1):72-80.
17. Turow J, Spitzer AR. Group B Streptococcal Infection Early Onset Disease Contoversies in Prevention Guidelines, and Management Strategies for the Neonate. Clinical Pediatrics. 2000;39:317-326.
18. Rittichier KR, Bryan PA, Bassett KE, Taggart EW, Enriquez FR, Hillyard DR, Byington CL. Diagnosis and outcomes of enterovirus infection in young infants. The Paediatric Infectious Disease Journal. 2005;24(6):546-550.
19. Gerdes JS. Diagnosis and management of bacterial infections in the neonate. Pediatric Clinics of North America. 2004;51:939-959.
20. Gerdes JS. Clinicopathologic Approach to the Diagnosis of Neonatal Sepsis. Clinics in Perinatology. 1991;18(2):361-381.
21. Kumar Y, Qunibi M, Neal TJ, Yoxall CW. Time to positivity of neonatal blood cultures. Archives of Disease in Childhood Fetal and Neonatal Edition. 2001;85(3):F182-F186.
22. BacT/ALERT(R) PF. Information Sheet. BioMerieux 2000. March 2002;43-02443.
23. Buttery JP. Blood cultures in newborns and children: optimising an everyday test. Archives of Disease in Childhood Fetal and Neonatal Edition. 2002;87:F25-F28.
24. Sarkar S, Bhagat I, DeCristofaro JD, Wiswell TE, Spitzer AR. A study of the role of multiple site blood cultures in the evaluation of neonatal sepsis. Journal of Perinatology. 2006;26(1):18-22.
25. Hansen A, Forbes P, Buck R. Potential substitution of cord blood for infant blood in the neonatal sepsis evaluation. Biology of the Neonate. 2005;88(1):12-18.
26. Herson VC, Block C, McLaughlin JC, Tetreault J, Eisenfeld LI, Krause PJ. Placental blood sampling: an aid to diagnosis of neonatal sepsis. Journal of Perinatology. 1998;18(2):135-137.
27. Fredricks DN, Relman DA. Application of Polymerase Chain Reaction to the Diagnosis of Infectious Disease. Clinical Infectious Diseases. 1999;29:475-488.
28. Krediet T, Gerards L, Fleer A, van Stekelenburg G. The predictive value of CRP and I/T ratio in neonatal infection. Journal of Perinatal Medicine. 1992;20(6):479-485.
29. Howard MR, Smith RA. Early diagnosis of septicaemia in preterm infants from examination of peripheral blood films. Clinical and Laboratory Haematology. 1999;21(5):365-368.
30. Garland SM, Bowman ED. Reappraisal of C-reactive protein as a screening tool for neonatal sepsis. Pathology. 2003;35:240-243.
31. Squire EN, Reich HM, Merenstein GB, Favara BE, Todd JK. Criteria for the discontinuation of antibiotic therapy during presumptive treatment of

suspected neonatal infection. Pediatric Infectious Diseases Journal. 1982;1(2):85-90.
32. Turner MA, Power S, Emmerson AJB. Gestational age and C reactive protein response. Archives of Disease in Childhood Fetal and Neonatal Edition. 2004;89(3):F272-F273.
33. Mezzasoma L, Bacarese-Hamilton T, Di Cristina M, Rossi R, Bistoni F, Crisanti A. Antigen Microarrays for Serodiagnosis of Infectious Diseases. Clinical Chemistry. 2002;48(1):121-130.
34. Bianchi D, Weinschenk N. Diagnostic blood test to identify infants at risk for sepsis. http://wwwfreepatentsonlinecom/5830679html. 1997:United States Patent 5830679.
35. Gouriet F, Samson L, Delaage M, Mainardi J-L, Meconi S, Drancourt M, Raoult D. Multiplexed whole bacterial antigen microarray, a new format for the automation of serodiagnosis: The culture-negative endocarditis paradigm. Clinical Microbiology and Infection. 2008;14(12):1112-1118.
36. Ng PC, Lam HS. Diagnostic markers for neonatal sepsis. Current Opinion in Pediatrics. 2006;18(2):125-131.
37. Skogstrand K, Thorsen P, Norgaard-Pederson B, Schendel DE, Sorensen LC, Hougaard DM. Simultaneous measurement of 25 inflammatory markers and neurotrophins in neonatal dried blood spots by immunoassay with xMAP technology. Clinical Chemistry. 2005;51(10):1854-1866.
38. Kite P, Millar MR, Gorham P, Congdon P. Comparison of five tests used in diagnosis of neonatal bacteraemia. Archives of Disease in Childhood. 1988;63(6):639-643.
39. Ng PC. Diagnostic markers of infection in neonates. Archives of Disease in Childhood Fetal and Neonatal Edition. 2004;89:F229-F235.
40. Ng PC, Cheng SH, Chui KM, Fok TF, Wong MY, Wong W, Wong RP, Cheung KL. Diagnosis of late onset sepsis with cytokines, adhesion molecule, and C-reactive protein in preterm very low birthweight infants. Archives of Disease in Childhood Fetal and Neonatal Edition. 1997;77(3):F221-F227.
41. Gonzales BE, Mercado CK, Johnson L, Brodsky NL, Bhandari V. Early markers of late-onset sepsis in premature neonates: clinical, hematological and cytokine profile. Journal of Perinatal Medicine. 2003;31(1):60-68.
42. Kafetzis DA, Tigani GS, Costalos C. Immunological markers in the neonatal period: Diagnostic value and accuracy in infection. Expert Review of Molecular Diagnostics. 2005;5(2):231-239.
43. Weirich E, Rabin RL, Maldonado Y, Benitz W, Modler S, Herzenberg LA, Herzenberg LA. Neutrophil CD11b expression as a diagnostic marker for early-onset neonatal infection. Journal of Pediatrics. 1998;132(3 Pt 1):445-451.
44. Ng PC, Li G, Chui KM, Chu WCW, Li K, Wong RPO, Chik KW, Wong E, Fok TF. Neutrophil CD64 Is a Sensitive Diagnostic Marker for Early-Onset Neonatal Infection. Pediatric Research. 2004;56(5):796-803.
45. Kowalik K, Czeszynska MB, Celewicz Z. Evaluation of diagnostic usefulness of the cord blood TNF-alpha levels as a marker of early onset neonatal infection (abstract only: full article in Polish). Ginekologia Polska. 2003;74(6):439-445.

46. Atici A, Satar M, Alparslan N. Serum interleukin-1 beta in neonatal sepsis. Acta Paediatrica. 1996;85(3):371-374.
47. Reyes CS, Garcia-Munoz F, Reyes D, Gonzalez G, Dominguez C, Domenach E. Role of cytokines (interleukin-1beta, 6, 8, tumour necrosis factor-alpha, and soluble receptor of interleukin-2) and C-reactive protein in the diagnosis of neonatal sepsis. Acta Paediatrica. 2007;92:221-227.
48. Laborada G, Rego M, Jain A, Guliano M, Stavola J, Ballabh P, Krauss AN, Auld PA, Nesin N. Diagnostic Value of Cytokines and C-reactive Protein in the First 24 Hours of Neonatal Sepsis. American Journal of Perinatology. 2003;20(8):491-501.
49. Franz AR, Bauer K, Schalk A, Garland SM, Bowman ED, Rex K, Nyholm C, Norman M, Bougatef A, Kron M, Mihatsch WA, Pohlandht F for the International IL-8 Study Group. Measurement of Interleukin 8 in Combination with C-Reactive Protein Reduced Unnecessary Antibiotic Therapy in Newborn Infants: A Multicenter, Randomized, Controlled Trial. Pediatrics. 2004;114(1):1-8.
50. Resch B, Gusenleitner W, Muller WD. Procalcitonin and Interleukin-6 in the diagnosis of early-onset sepsis of the neonate. Acta Paediatrica. 2003;92:243-245.
51. Distefano G, Curreri R, Betta P, Romeo MG, Amato M. Procalcitonin serum levels in perinatal bacterial and fungal infection of preterm infants. Acta Paediatrica. 2004;93:216-219.
52. Vazzalwar R, Pina-Rodrigues E, Puppala BL, Angst DB, Schweig L. Procalcitonin as a Screening Test for Late-Onset Sepsis in Preterm Very Low Birth Weight Infants. Journal of Perinatology. 2005;25:397-402.
53. Casado-Flores J, Blanco-Quiros A, Asensio J, Arranz E, Garrote JA, Nieto M. Serum procalcitonin in children with suspected sepsis: A comparison with C-reactive protein and neutrophil count. Pediatric Critical Care Medicine. 2003;4(2):190-195.
54. Joram N, Boscher C, Denizot S, Loubersac V, Winer N, Roze JC, Gras-Le Guen C. Umbilical cord blood procalcitonin and C reactive protein concentrations as markers for early diagnosis of very early onset neonatal infection. Archives of Disease in Childhood Fetal and Neonatal Edition. 2006;91:F65-F66.
55. Verboon-Maciolek MA, Thijsen SF, Hemels MA, Menses M, van Loon AM, Krediet TG, Gerards LJ, Fleer A, Vorrbij HA, Rijkers GT. Inflammatory mediators for the diagnosis and treament of sepsis in earl infancy. Pediatric Research. 2006;59(3):457-461.
56. Lopez Sastre JB, Perez Solis D, Roques Serradilla V, Fernandez Colomer B, Coto Cotallo GD, Krauel Vidal X, Narbona Lopez E, Garcia del Rio M, Sanchez Luna M, Belaustgui Cueto A, Moro Serrano M, Urbon Artero A, Alvero Iglesias E, Cotero Lavin A et al. Procalcitonin is not sufficiently reliable to be the sole marker of neonatal sepsis of nosocomial origin. BMC Pediatrics. 2006;18(6):16.
57. Carrol ED, Thomson APJ, Hart CA. Procalcitonin as a marker of sepsis. International Journal of Antimicrobial Agents. 2002;20:1-9.

58. Zilow EP, Huack W, Linderkamp O, Zilow G. Alternative pathway activation of the complement system in preterm infants with early onset infection. Pediatric Research. 1997;41(3):334-339.
59. Peakman M, Senaldi G, Liossis G, Gamsu HR, Vergani D. Complement activation in neonatal infection. Archives of Disease in Childhood 1992;67(7):802-807.
60. Kennon C, Overturf G, Bessman S, Sierra E, Smith KJ, Brann B. Granulocyte colony-stimulating factor as a marker for bacterial infection in neonates. Journal of Pediatrics. 1996;128(6):765-769.
61. Edgar JD, Wilson DC, McMillan SA, Crockard AD, Halliday MI, Gardiner KR, Rowlands BJ, Halliday HL, McNeill TA. Predictive value of soluble immunological mediators in neonatal infection. Clinical Science (London). 1994;87(2):165-171.
62. Distefano G, Betta P, Rodono A, Tina LG, Sciotto A, Sciacca A, Romeo MG. Concentration of blood ICAM-1s in the newborn at risk of infection. A prospective study in the first 2 weeks of life. (Abstract only: full article in Italian) La Pediatria Medica et Chirurgica. 1997;19(1):27-30.
63. Davis BH. Improved diagnostic approaches to infection/sepsis detection. Expert Reviews of Molecular Diagnostics. 2005;5(2):193-207.
64. Baek YW, Brokat S, Padbury JF, Pinar H, Hixon DC, Lim Y-P. Inter-[alpha] inhibitor proteins in infants and decreased levels in neonatal sepsis. Journal of Pediatrics. 2003;143(1):11-15.
65. Griffin MP, Lake DE, Moorman JR. Heart Rate Characteristics and Laboratory Tests in Neonatal Sepsis. Pediatrics. 2005;115(4):937-941.
66. Mautone A, Montagna O, Giusti A, De Simone B, Balducci G, Santoro A. Proposed scoring system for early diagnosis of infection in neonates. (Abstract: full article in Italian). Minerva Pediatrica. 1989;41(3):143-145.
67. Bang AT, Bang RA, Stoll BJ, Baitule SB, Reddy HM, Deshmukh MD. Is Home-Based Diagnosis and Treatment of Neonatal Sepsis Feasible and Effective? Seven Years of Intervention in the Gadchiroli Field Trial (1996 to 2003). Journal of Perinatology. 2005;25(Supplement 1):S62-S71.
68. Chiu CH, Lin TY, Bullard MJ. Identification of febrile neonates unlikely to have bacterial infections. Pediatric Infectious Diseases Journal. 1997;16(1):59-63.
69. Malik A, Hui CP, Pennie RA, Kirpalani H. Beyond the Complete Blood Cell Count and C-Reactive Protein: a systematc review of modern diagnostic tests for neonatal sepsis. Archives of Pediatric and Adolescent Medicine. 2003;157(6):511-516.
70. Hassan J, Reen DJ. Reduced primary antigen specific T cell precursor frequencies in neonates is associated with deficient IL-2 production. Immunology. 1996;87:604-608.
71. Levy O. Innate immunity of the newborn: basic mechanisms and clinical correlates. Nature Reviews Immunology. 2007;7:379-390.
72. Walker VP, Akinbi HT, Meinzen-Derr J, Narendran V, Visscher M, Hoath SB Host defense proteins on the surface of neonatal skin: implications for innate immunity. Journal of Pediatrics. 2008;152(6):777-781.
73. Christensen RD, Rothstein G. Pre- and postnatal development of granulocytic stem cells in the rat. Pediatric Research. 1984;18:599-602.

74. Yost CC, Cody MJ, Harris ES, Thornton NL, McInturff AM, Martinez ML, Chandler NB, Rodesch CK, Albertine KH, Petti CA, Weyrich AS, Zimmerman GA. Impaired neutrophil extracellular trap (NET) formation: a novel innate immune deficiency of human neonates. Blood. 2009;113(25):6419-6427.
75. Wu Y-C, Huang Y-F, Lin C-H, Shieh C-C. Detection of defective granulocyte function with flow cytometry in newborn infants. Journal of Microbiology, Immunology and Infection. 2005;38(1):17-24.
76. Hanna N, Vasquez P, Pham P, Heck DE, Laskin JD, Laskin DL, Weinberger B. Mechanisms underlying reduced apoptosis in neonatal neutrophils. Pediatric Research. 2005;57(1):56-62.
77. Satwani P, Morris E, van de Ven C, Cairo MS. Dysregulation of Expression of Immunoregulatory and Cytokine Genes and Its Association with the Immaturity in Neonatal Phagocytic and Cellular Immunity. Biology of the Neonate. 2005;88:214-227.
78. Jiang H, Van de Ven C, Satwani P, Baxi LV, Cairo MS. Differential Gene Expression Patterns by Oligonucleotide Microarray of Basal versus Lipopolysaccharide-Activated Monocytes from Cord Blood versus Adult Peripheral Blood. The Journal of Immunology. 2004;172:5870-5879.
79. Levy O. Innate immunity of the human newborn: distinct cytokine responses to LPS and other Toll-like receptor agonists. Journal of Endotoxin Research. 2005;11(2):113-116.
80. Marodi L. Innate cellular immune responses in newborns. Clinical Immunology. 2006;118(2-3):137-144.
81. Koch L, Linderkamp O, Ittrich C, Benner A, Poeschl J. Gene expression profiling of adult peripheral and cord blood mononuclear cells altered by lipopolysaccharide. . Neonatology. 2008;93(2):87-100.
82. Tatad AMF, Nesin M, Peoples J, Cheung S, Lin H, Sison C, Perlman J, Cunningham-Rundles S. Cytokine expression in response to bacterial antigens in preterm and term infant cord blood monocytes. Neonatology. 2008;94(1):8-15.
83. Aksoy E, Albarani V, Nguyen M, Laes J-F, Ruelle J-L, De Wit D, Willems F, Goldman M, Goriely S. Interferon regulatory factor 3-dependent responses to lipopolysaccharide are selectively blunted in cord blood cells. Blood. 2007;109(7):2887-2893.
84. Lawrence S, Tang Y, Frank Mb, Dozmorov I, Jiang K, Chen Y, Cadwell C, Turner S, Centola M, Jarvis JN. A dynamic model of gene expression in monocytes reveals differences in immediate/early response genes between adult and neonatal cells. Journal of Inflammation. 2007;4:Article number 4.
85. Jiang H, van de Ven C, Baxi L, Satwani P, Cairo MS. Differential gene expression signatures of adult peripheral blood vs cord blood monocyte-derived immature and mature dendritic cells. Experimental Hematology. 2009;37(10):1201-1215.
86. Zaghouani H, Hoeman CM, Adkins B. Neonatal immunity: faulty T-helpers and the shortcomings of dendritic cells. Trends in Immunology. 2009;30(12):585-591.

87. Koga Y, Matsuzaki A, Suminoe A, Hattori H, Hara T. Expression of cytokine-associated genes in dendritic cells (DCs): Comparison between adult peripheral blood- and umbilical cord blood-derived DCs by cDNA microarray. Immunology Letters. 2008;116(1):55-63.
88. Krumbiegel D RJ, Schmidtke P, Knuf M, Zepp F, Meyer CU. Efficient maturation and cytokine production of neonatal DCs requires combined proinflammatory signals. Clinical and Developmental Immunology. 2005;12(2):99-105.
89. Velilla PA, Rugeles MT, Chougnet CA. Defective antigen-presenting cell function in human neonates. Clincal Immunology. 2006;121(3):251-259.
90. Perez A, Gurbindo MD, Resino S, Aguaron A, Munoz-Fernandez MA. NK cell increase in neonates from the preterm to the full-term period of gestation. Neonatology. 2007;92(3):158-163.
91. Fleer A, Krediet TG. Innate immunity: toll-like receptors and some more. A brief history, basic organization and relevance for the human newborn. Neonatology. 2007;92(3):145-157.
92. Sadeghi K, Berger A, Langgartner M, Prusa AR, Hayde M, Herkner K, Pollak A, Spittler A, Forster-Waldl E. Immaturity of infection control in preterm and term newborns is associated with impaired toll-like receptor signalling. Journal of Infectious Diseases. 2007;195(2):296-302.
93. Angelone DF, Wessels MR, Coughlin M, Suter EE, Valentini P, Kalish LA, Levy O. Innate immunity of the human newborn is polarized towards a high ratio of IL-6/TNF-alpha production in vitro and in vivo. Pediatric Research. 2006;60(2):205-209.
94. Levy O, Suter EE, Miller RL, Wessels MR. Unique efficacy of Toll-like receptor 8 agonists in activating human neonatal antigen-presenting cells. Blood. 2006;108(4):1284-1290.
95. Belderbos ME, van Bleek GM, Levy O, Blanken ML, Houben ML, Schuijff L, Kimpen JL, Bont L. Skewed pattern of Toll-like receptor 4-mediated cytokine production in human neonatal blood: low LPS-induced IL-12p70 and high IL-10 persist through the first month of life. Clinical Immunology. 2009;133(2):228-237.
96. Siegrist CA. Neonatal and early life vaccinology. Vaccine. 2001;19:3331-3346.
97. Wilson CB, Kollmann TR. Induction of antigen-specific immunity in human neonates and infants. Nestle Nutritional Workshop Series Pediatric Program. 2008;61:183-195.
98. Hassan J, Dooley S, Hall W. Immunological response to cytomegalotvirus in congenitally infected neonates. Clinical and Experimental Immunology. 2007;147:465-471.
99. Omar SA, Wooliever D, Herrington CJ, Lorens JM, Rawas F, Huard T. Expression of naive and memory T-cells in newborn infants with early-onset sepsis. Journal of Pediatric Infectious Diseases. 2009;4(3):233-239.
100. Aygun AD, Citak Kurt AN, Godekmerden A, Kurt A, Akarsu S, Gurgoze MK, Yilmaz E. Neonates with culture proven sepsis have lower amounts and percentage of CD45RA+ T cells. Inflammation. 2008;31(4):222-226.

101. Schaub B, Schleich I, Hoppler S, Sattler C, von Mutius E. Impairment of T helper and T regulatory cell response at birth. Allergy. 2008;63(11):1438-1447.
102. Zhao J, Kim KD, Yang X, Auh S, Fu YX, Tang H. Hyper innate responses in neonates lead to increased morbidity and mortality after infection. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(21):7528-7533.
103. Gibbons DL, Haque SF, Silberzahn T, Hamilton K, Langford C, Ellis P, Carr R, Hayday AC. Neonates harbour highly active gammadelta T cells with selective impairments in preterm infants. European Journal of Immunology. 2009;39(7):1794-1806.
104. Hedges JF, Lubick KJ, Jutila MA. gammadelta T cells respond directly to pathogen-associated molecular patterns. Journal of Immunology. 2005;174(10):6045-6053.
105. Stoll BJ, Lee FK, Hale E, Schwartz D, Holmes R, Ashby R, Czerkinsky C, Nahmias AJ. Immunoglobulin secretion by the normal and the infected newborn infant. Journal of Pediatrics. 1993;122(5 Pt 1):780-786.
106. Newburg DS, Walker WA. Protection of the neonate by the innate immune system of developing gut and of human milk. Pediatric Research. 2007;61(1):2-8.
107. Le Bouder E, Rey-Nores JE, Raby AC, Affolter M, Vidal K, Thornton CA, Labeta MO. Modulation of neonatal microbial recognition: TLR-mediated innate immune responses are specifically and differentially modulated by human milk. Journal of Immunology. 2006;176(6):3742-3752.
108. Krumbiegel D, Anthogalidis- Voss C, Markus H, Zepp F, Meyer CU. Enhanced expression of IL-27 mRNA in human newborns. Pediatric Allergy and Immunology. 2008;19(6):513-516.
109. Schultz C, Strunk K, Temming P, Matzke N, Hartel C. Reduced IL-10 production and -recepor expression in neonatal T lymphocytes. Acta Paediatrica. 2007;96(8):1122-1125.
110. Skogstrand K, Hougaard DM, Schendel DE, Brent NP, Svaerke C, Thorsen P. Association of preterm birth with sustained postnatal inflammatory response. Obstetrics and Gynecology. 2008;111(5):1118-1128.
111. Rizos D, Protonotariou E, Malamitsi-Puchner A, Sarandakou A, Trakakis E, Salamalekis E. Cytokine concentrations during the first days of life. European Journal of Obstetrics and Gynecology and Reproductive Biology. 2007;131(1):32-35.
112. Mestan K, Yu Y, Thorsen P, Skogstrand K, Matoba N, Liu X, Kumar R, Hougaard DM, Gupta M, Pearson C, Ortiz K, Bauchner H, Wang X. Cord blood biomarkers of the fetal inflammatory response. Journal of Maternal-Fetal and Neonatal Medicine. 2009;22(5):379-387.
113. Gotsch F, Romero R, Kusanovic JP, Mazaki-Tovi S, Pineles BL, Erez O, Espinoza J, Hassan SS. The fetal inflammatory response syndrome. Clinical Obstetrics and Gynecology. 2007;50(3):652-683.
114. Malamitsi-Puchner A, Protonotariou E, Boutsikou T, Makrakis E, Sarandakou A, Creatsas G. The influence of the mode of delivery on circulating cytokine concentrations in the perinatal period. Early Human Development. 2005;81(4):387-392.

115. Ly NP, Ruiz-Perez B, Onderdonk AB, Tzianabos AO, Litonjua AA, Liang C, Laskey D, Delaney ML, DuBois AM, Levy H, Gold DR, Ryan LM, Weiss ST, Celedon JC. Mode of delivery and cord blood cytokines: A birth cohort study. Clinical and Molecular Allergy. 2006;4:Article 13.
116. Schaub B, Liu J, Hopper S, Schleich I, Huehn J, Olek S, Wieczorek G, Illi S, von Mutius E. Maternal farm exposure modulates neonatal immune mechanisms through regulatory T cells. Journal of Allergy and Clinical Immunology. 2009;123(4):774-782.
117. Reber AJ, Donovan DC, Gabbard J, Galland K, Aceves-Avila M, Holbert KA, Marshall L, Hurley DJ. Transfer of maternal colostral leukocytes promotes development of the neonatal immune system. Part II. Effects on neonatal lymphocytes. Veterinary Immunology and Immunopathology. 2008;123(3-4):305-313.
118. Merkerova M, Vasikova A, Bruchova H, Libalova H, Topinka J, Balascak I, Sram RJ, Brdicka R. Differential gene expression in umbilical cord blood and maternal peripheral blood. European Journal of Haematology. 2009;83(3):183-190.
119. Strunk T, Richmond P, Simmer K, Currie A, Levy O. Burgner D. Neonatal immune responses to coagulase-negative staphylococci. Current Opinion in Infectious Diseases. 2007;20(4):370-375.
120. Hartel C, Osthues I, Rupp J, Haase B, Roder K, Gopel W, Herting E, Schultz C. Characterisation of the host inflammatory response to Staphylococcus epidermidis in neonatal whole blood. Archives of Disease in Childhood Fetal and Neonatal Edition. 2008;93(2):F140-F145.
121. Piersanti S, Martina Y, Cherubini G, Avitabile D, Saggio I. Use of DNA microarrays to monitor host response to virus and virus-derived gene therapy vectors. American Journal of Pharmacogenomics. 2004;4(6):345-356.
122. Hinton JC, Hautefort I, Eriksson S, Thompson A, Rhen M. Benefits and pifalls of using microarrays to monitor bacterial gene expression during infection. Current Opinion in Microbiology. 2004;7(3):277-282.
123. Bryant PA, Venter D, Robins- Browne R, Curtis N. Chips with everything: DNA microarrays in infectious diseases. Lancet Infectious Diseases. 2004;4(2):100-111.
124. Bodrossy L, Sessitsch A. Oligonucleotide microarrays in microbial diagnostics. Current Opinion in Microbiology. 2004;7(3):245-254.
125. Bochud PY, Calandra T. Pathogenesis of sepsis: new concepts and implications for future treatment. British Medical Journal. 2003;326:262-266.
126. Talwar S, Munson PJ, Barb J, Fiuza C, Cintron AP, Logan C, Tropea M, Khan S, Reda D, Shelhamer JH, Danner RL, Suffredini AF. Gene expression profiles of peripheral blood leukocytes after endotoxin challenge in humans. Physiological Genomics. 2006;25(2):203-215.
127. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression expression monitoring of 1000 genes. Proceedings of the National Academy of Science of the United States of America. 1996;93:10614-10619.
128. Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T. Cellular gene expression altered by human cytomegalovirus: global monitoring with

oligonucleotide arrays. Proceedings of the National Academy of Science of the United States of America. 1998;95(14470-14475).
129. Hussein S, Michael P, Brabant D, Omri A, Narain R, Passi K, Ramana CV, Parrillo JE, Kumar A, Parissenti A, Kumar A. Characterization of human septic sera induced gene expression modulation in human monocytes. International Journal of Clinical and Experimental Medicine. 2009;2(2):131-148.
130. Alizadeh A, Eisen M, Botstein D, Brown PO, Staudt LM. Probing lymphocyte biology by genomic-scale gene expression analysis. Journal of Clinical Immunology. 1998;18:373-379.
131. Der SD, Zhou A, Williams BR, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta or gamma using olionucleotide arrays. Proceedings of the National Academy of Science of the United States of America. 1998;95:15623-15628.
132. Brown PO, L Hartwell. Genomics and human disease - variations on variation. Nature Genetics. 1998;18 91-93.
133. Bolt G, Berg K, Blixenkrone-Moller M. Measles virus-induced modulation of host-cell gene expression. Journal of General Virology. 2002;83:1157-1165.
134. Feezor RJ, Oberholzer C, Baker HV, Novick D, Rubenstein M, Moldawer LL, Pribble J, Souza S, Dinarello C, Ertel W, Oberholzer A. Molecular Characterization of the Acute Inflammatory Response to Infections with Gram-negative versus Gram-positive Bacteria. Infection and Immunity. 2003;71(10):5803-5813.
135. Israel DA, Salama N, Arnold CN, Moss SF, Ando T, Wirth HP, Tham KT, Camorlinga M, Blaser MJ, Falkow S, Peek RM Jr. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. The Journal of Clinical Investigation 2001;107(5):611-620.
136. Blumenthal A, Lauber J, Hoffmann R, Ernst M, Keller C, Buer J, Ehlers S, Reiling N. Common and Unique Gene Expression Signatures of Human Macrophages in Response to Four Strains of Mycobacterium avium That Differ in Their Growth and Persistence Characteristics. Infection and Immunity. 2005;73(6):3330-3341.
137. Mans JJ, Lamont RJ, Handfield M. Microarray analysis of human epithelial cell responses to bacterial interaction. Infectious Disorders - Drug Targets. 2006;6(3):299-309.
138. Tsai K-N, Tsai T-Y, Chen C-M. A new approach to deciphering pathways of enterovirus 71-infected cells: An integration of microarray data, gene ontology, and pathway database. Biomedical Engineering Applications Basis Communications. 2006;18(6):337-342.
139. Munir S, Kapur V. Regulation of host cell transcriptional physiology by the avian pneumovirus provides key insights into host-pathogen interactions. Journal of Virology. 2003;77(8):4899-4910.
140. Coombes BK, Mahony JB. cDNA Array Analysis of Altered Gene Expression in Human Endothelial Cells in Response to *Chlamydia pneumoniae* Infection. Infection and Immunity. 2001;69(3):1420-1427.

141. Kim HS, Choi EH,, Khan J, Roilides E, Francesconi A, Kasai M, Sein T, Schaufele RL, Sakurai K, Son CG, Greer BT, Chanock S, Lyman CA, Walsh TJ. Expression of Genes Encoding Innate Host Defense Molecules in Normal Human Monocytes in Response to *Candida Albicans*. Infection and Immunity. 2005;73(6):3714-3724.
142. Fredericksen BL, Smith M, Katze MG, Shi P-Y, Gale M Jr. The Host Response to West Nile Virus Infection Limits Viral Spread through the Activation of the Interferon Regulatory Factor 3 Pathway. Journal of Virology. 2004;78(14):7737-7747.
143. Korth MJ, Katze MG. Unlocking the Mysteries of Virus-Host Interactions. Does Functional Genomics Hold the Key? Annals of the New York Academy of Science. 2002;975:160-168.
144. Schnappinger D, Schoolnik GK, Ehrt S. Expression profiling of host pathogen interactions: how *Mycobacterium tuberculosis* and the macrophage adapt to one another. Microbes and Infection. 2006;8(4):1132-1140.
145. Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, Young RA, Lander ES, Hacohen N. The plasticity of dendritic cell responses to pathogens and their components. Science. 2001;294(5543):870-875.
146. Hossain H, Tchatalbachev S, Chakraborty T. Host gene expression profiling in pathogen-host interactions. Current Opinion in Immunology. 2006;18(4):422-429.
147. Ren Q, Robertson SJ, Howe D, Barrows LF, Heinzen RA. Comparitive DNA Microarray Analysis of Host Cell Transcriptional Responses to Infection by *Coxiella burnetii* or *Chlamydia trachomatis*. Annals of the New York Academy of Sciences. 2006;990:701-713.
148. Das R, Hammamieh R, Neill R, Ludwig GV, Eker S, Lincoln P, Ramamoorthy P, Dhokalia A, Mani S, Mendis C, Cummings C, Kearney B, Royaee A, Huang X-Z, Paranavitana C, Smith L, Peel S, Kanesa-Thasan N, Hoover D, Lindler LE, Yang D, Henchall E, Jett M. Early indicators of exposure to biological threat agents using host gene profiles in peripheral blood mononuclear cells. BMC Infectious Diseases. 2008;8:104.
149. Chinnaiyan AM, Huber-Lang M, Kumar-Sinha C, Barrette TR, Shankar-Sinha S, Sarma VJ, Padgaonkar VA, Ward PA. Molelcular signatures of sepsis: multiorgan gene expression profiles of systemic inflammation. Americal Journal of Pathology. 2001;159(4):1199-1209.
150. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF & the Inflammation and Host Response to Injury Collaberation. A network-based analysis of systemic inflammation in humans. Nature. 2005;437:1032-1037.
151. Reghunathan R, Jayapal M, Hsu L-Y, Chng H-H, Tai D, Leung BP, Melendez AJ. Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome. BMC Immunology. 2005;6:2.
152. Wong HR, Shanley TP, Sakthivel B, Cyjanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M ,Tofil NM, Penfil S. Monaco M, Tagavilla MA, Odoms K, Dunsmore K, Barnes M, Aronow BJ, Genomics of Pediatric SIRS/Septic Shock Investigators. Genome-level expression profiles

in pediatric septic shock indicates a role for altered zinc homeostasis in poor outcome. Physiological Genomics. 2007;30(2):146-155.
153. Shanley TP, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, Odoms K, Barnes M, Sakthivel B, Aronow BJ, Wong HR. Genome-level longitudinal expression of signalling pathways and gene networks in pediatric septic shock. Molecular Medicine. 2007;13(9-10):495-508.
154. Wong HR, Cvijanovich N, Allen GL, Lin R, Anas N, Meyer K, Freishtat RJ, Monaco M, Odoms K, Sakthivel B, Shanley TP, Genomics of Pediatric SIRS/Septic Shock Investigators. Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis and septic shock spectrum. Critical Care Medicine. 2009;37(5):1558-1566.
155. Johnson SB, Lissauer M, Bochicchio GV, Moore R, Cross AS, Scalea TM. Gene expression profiles differentiate between sterile SIRS and early sepsis. Annals of Surgery. 2007;245(4):611-621.
156. Lissauer ME, Johnson SB, Bochicchio GV, Feild CJ, Cross AS, Hasday JD, Whiteford CC, Nussbaumer WA, Towns M, Scalea TM. Differential expression of toll-like receptor genes: Sepsis compared with sterile inflammation 1 day before a septic diagnosis. Shock. 2009;31(3):238-244.
157. Long HT, Hibberd ML, Hein TT, Dung NM, Van Ngoc T, Farrar J, Wills B, Simmons CP. Patterns of gene transcript abundance in the blood of children with severe or uncomplicated dengue highlight the differences in disease evolution and host responses to dengue viral infection. Journal of Infectious Diseases. 2009;199(4):537-546.
158. Howrylak JA, Dolinay T, Lucht L, Wang Z, Christiani DC, Sethi JM, Xing EP, Donahoe MP, Choi AMK. Discovery of the gene signature for acute lung injury in patients with sepsis. Physiological Genomics. 2009;37(2):133-139.
159. Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. Nature Reviews in Microbiology. 2005;3:281-294.
160. Manger ID, Relman DA. How the host 'sees' pathogens: global gene expression responses to infection. Current Opinion in Immunology. 2000;12(2):215-218.
161. Takeuchi O, Akira S. Toll-like receptors; their physiological role and signal transduction system. International Immunopharmacology. 2001;1:625-635.
162. Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H, Ogawa T, Takeda K, Akira S. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity. 1999;11:443-451.
163. Zeytun A, van Velkinburgh JC, Pardington PE, Cary RR, Gupta G. Pathogen-specific innate immune response. Advances in Experimental Medicine and Biology. 2007;598:342-457.
164. Hagberg H, Mallard C. Effect of inflammation on central nervous system develpoment and vulnerability. Current Opinion in Neurology. 2005;18:117-123.

165. Klein MB, Silver G, Gamelli RL, Gibran NS, Herndon DN, Hunt JL, Tompkins RG and the Inflammation and the Host Response Investigators. Inflammation and the Host Response to Injury: An Overview of the Multicenter Study of the Genomic and Proteomic Response to Burn Injury. Journal of Burn Care Research. 2006;27:448-451.
166. Oberholzer A, Steckholzer U, Kurimoto M, Trentz O, Ertel W. Interleukin- 18 plasma levels are increased in patients with sepsis compared to severely injured patients. Shock. 2001;16:411-414.
167. Chung TP, Laramie JM, Meyer DJ, Downey T, Tam LH, Ding H, Buchman TG, Karl I, Stormo GD, Hotchkiss RS, Cobb JP. Molecular diagnostics in sepsis: from bedside to bench. Journal of the American College of Surgery. 2006;203(5):585-598.
168. Petrakou E, Mouchtouri A, Levi E, Lipsou N, Xanthou M, Fotopoulos S. Interleukin-8 and monocyte chemotactic protein-1 mRNA expression in perinatally infected and asphyxiated preterm neonates. Neonatology. 2007;91(2):107-113.
169. Burgner D, Levin M. Genetic susceptibility to infectious diseases. The Pediatric Infectious Disease Journal. 2003;22(1):1-6.
170. Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, Keers S, Turnbull DM, Howell N, Chinnery PF. Mitochondrial DNA and survival after sepsis: a prospective study. The Lancet. 2005;366:2118-2121.
171. Baier RJ, Loggins J, Yanamandra K. IL-10, IL-6 and CD14 polymorphisms and sepsis outcome in ventilated very low birth weight infants. BMC Medicine. 2006;4:10.
172. Kerr JR, Kaushik N, Fear D, Baldwin DA, Nawaysir EF, Adcock IM. Single-nucleotide polymorphisms associated with symptomatic infection and differential human gene expression in healthy seropositive persons each implicate the cytoskeleton, intergrin signalling and oncosuppression in the pathogenesis of human parvovirus B19 infection. The Journal of Infectious Diseases. 2005;192(2):276-287.
173. Wunderink RG, Waterer GW. Genetic Effects on Sepsis and Pneumonia. Clinical Pulmonary Medicine. 2004;11(3):143-153.
174. Kwiatkowski D. Susceptibility to infection. British Medical Journal. 2000;321:1061-1064.
175. Papathanassoglou ED, Giannakopoulou MD, Bozas E. Genomic variations and susceptibility to sepsis. AACN Advanced Critical Care. 2006;17(4):394-422.
176. Clementi M, Di Gianatonio E. Genetic susceptibility to infectious diseases. Reproductive Toxicology. 2006;21(4):345-349.
177. Balding J, Healy CM, Livingstone WJ, White B, Mynett-Johnson L, Caffersky M, Smith OP. Genomic polymorphic profiles in an Irish population with meningococcaemia: is it possible to predict severity and outcome of disease? Genes and Immunity. 2003;4:533-540.
178. Roy S, Hill AVS, Knox K, Griffiths D, Crook D. Association of common genetic variant with susceptibility to invasive pneumococcal disesase. British Medical Journal. 2002;324:1369.
179. Frakking FNJ, Brouwer N, Van Eijkelenburg NK, Merkus MP, Kuijpers TW, Offringa M, Dolman KM. Low mannose-binding lectin (MBL) levels in

neonates with pneumonia and sepsis. Clinical and Experimental Immunology. 2007;150(2):255-262.
180. De Benedetti F, Auriti C, D'Urbano LE, Ronchetti MP, Rava L, Tozzi A, Ugazio AG, Orzalesi MM. Low serum levels of mannose binding lectin are a risk factor for neonatal sepsis. Pediatric Research. 2007;61(3):325-328.
181. Cotten CM, Ginsburg GS, Goldberg RN, Speer MC. Genomic analyses: A neonatology perspective. The Journal of Pediatrics. 2006;148(6):720-726.
182. Diehn M, Relman DA. Comparing functional genomic datasets: lessons from DNA microarray analyses of host-pathogen interactions. Current Opinion in Microbiology. 2001;4(1):95-101.
183. Forster T, Roy D, Ghazal P. Experiments using microarray technology: limitations and standard operating procedures. Journal of Endocrinology. 2003;178(2):195-204.
184. McLaren PJ, Mayne M, Rosser S, Moffatt T, Becker KG, Plummer FA, Fowke KR. Antigen-Specific Gene expression Profiles of Peripeheral Blood Mononuclear Cells Do not Reflect THose of T-Lymphocyte Subsets Clinical and Diagnostic Laboratory Immunology. 2004;11(5):977-982.
185. Copland JA, Davies PJ, Shipley GL, Wood CG, Luxon BA, Urban RJ. The Use of DNA Microarrays to Assess Clinical Samples: The Transition from Bedside to Bench to Bedside. Recent Progress in Hormone Research. 2003;58(25-53).
186. Petricoin III EF, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, Woodcock J,Feigal DW Jr, Zoon KC, Sistare FD. Medical applications of microarray technologies: a regulatory science perspective. Nature Genetics Supplement. 2002;32:474-479.
187. Reinhart K, Uhlig T. Differential gene expression: a (new) topic in the intensive care unit. Current Opinion in Anaesthesiology. 2004;17(2):93-95.
188. Bates MD. The potential of DNA microarrays for the care of children. The Journal of Pediatrics. 2003;142:235-239.
189. Hall AG, Coulthard SA, Irving J. Molecular diagnostics: A healthcare perspective. Expert Review of Molecular Diagnostics. 2003;3(1):13-16.
190. Toder R. DNA arrays as diagnostic tools in human healthcare. Expert Review of Molecular Diagnostics. 2002;2(5):422-428.
191. Simon R. Using DNA microarrays for diagnostic and prognostic prediction. Expert Review of Molecular Diagnostics. 2003;3(5):587-595.
192. Ardekani AM, Petricoin III EF, Hackett JL. Molecular diagnostics: An FDA perspective. Expert Review of Molecular Diagnostics. 2003;3(2):129-140.
193. Anthony RM, Brown TJ, French. DNA array technology and diagnostic microbiology. Expert Review of Molecular Diagnostics. 2001;1(1):30-38.
194. Pahl A. Gene expression profiling using RNA extracted from whole blood: technologies and clinical applications. Expert Review of Molecular Diagnostics. 2005;5(1):43-52.
195. Aitman TJ. DNA microarrays in medical practice. British Medical Journal. 2001;323:611-615.
196. Petrik J. Microarray technology: the future of blood testing? Vox Sanguis. 2001;80:1-11.

197. Jarvis JN, Centola M. Gene-expression profiling: time for clinical application? The Lancet. 2005;365:199-200.
198. Guttmacher AE, Collins FS. Welcome to the Genomics Era. New England Journal of Medicine. 2003;349(10):996-998.
199. Petrik J. Diagnostic applications of microarrays. Transfusion Medicine. 2006;16(4):233-247.
200. Boright AP, Kere JA, Scherer SW. Genomics and Pediatric Research. Pediatric Research. 2002;53(1):4-9.
201. Truog WE. Genomics and Complex Neonatal Disorders: Maybe We're Getting Somewhere. Journal of Perinatology. 2003;23:91-93.
202. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J,Jatkoe T, Berns EMJJ, Atkins D, Foekens JA. Gene-expression profiles to predict distinct metastasis of lymph-node-negative primary breast cancer. Lancet. 2005;365:671-679.
203. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loroid B, Bachelat L, Montfort J, Victorero G, Viret F, Ollendorf V, Fert V, Giovaninni M, Delpero J-R, Nguyen C, Viens P, Monges G, Birnbaum B, Houlgatte R. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene. 2004;23:1377-1391.
204. Ryder MI, Hyun W, Loomer P, Haqq C. Alteration of gene expression profile of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. Oral Microbiology and Immunology. 2004;19:39-49.
205. Shioda T. Application of DNA Microarray to Toxicological Research. Journal of Environmental Pathology, Toxicology and Oncology. 2004;23(1):13-31.
206. Wang Z, Neuburg D, Li C, Su L, Kim JY, Chen JC, Christiani DC. Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environmental Health Perspectives. 2005;113(2):233-241.
207. He B, Munson AE, Meade BJ. Analysis of gene expression induced by irritant and sensitizing chemical using oligonucleotide arrays. International Immunopharmacology. 2001;1:867-879.
208. Hershey AD, Tang Y, Powers SW, Kabbouche MA, Gilbert DL, Glauser TA, Sharp FR Genomic Abnormalities in Patients with Migraine and Chronic Migraine: Preliminary Blood Gene Expression Suggests Platelet Abnormalities. Headache. 2004;44(10):994-1004.
209. Tang Y, Glauser TA, Gilbert DL, Hershey AD, Privitera MD, Ficker DM, Szaflarski JP, Sharp FR. Valproic acid blood genomic expression patterns in children with epilepsy - a pilot study. Acta Neurologica Scandinavica. 2004;109(3):159-168.
210. Tang Y, Gilbert DL, Glauser TA, Hershey AD, Sharp FR. Blood Gene Expression Profiling of Neurologic Diseases: A Pilot Microarray Study. Archives of Neurology. 2005;62:210-215.
211. Schaffer LG, Kashork CD, Saleki R, Rorem E, Sundin K, Ballif BC, Bejjani BA. Targeted genomic microarray analysis for identification of

chromosome abnormalities in 1500 consecutive clinical cases. The Journal of Pediatrics. 2006;149(1):98-102.
212. Cheung VG, Spielman RS. The genetics of variation in gene expression. Nature Genetics Supplement. 2002;32:522-525.
213. Bittel DC, Kibiryeva N, O'Brien JE, Lofland GK, Butler MG. Gene expression in pediatric heart disease with emphasis on cotruncal defects. Progress in Pediatric Cardiology. 2005;20:127-141.
214. Watchko JF, Daood MJ, Biniwale M. Understanding neonatal hyperbilirubinaemia in the era of genomics. Seminars in Neonatology. 2002;7:143-152.
215. Kunz M, Ibrahim SM, Koczan D, Scheid S, Theisen HJ, Gross G. DNA microarray technology and its applications in dermatology. Experimental Dermatology. 2004;13(10):593-606.
216. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proceedings of the National Academy of Science of the United States of America. 1997;94:2150-2155.
217. Gladkevich A, Nelemans SA, Kauffman HF, Korf J. Microarray Profiling of Lymphocytes in Internal Diseases With an Altered Immune Response: Potential and Methodology. Mediators of Inflammation. 2005;6:317-330.
218. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, Via CS. Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. Clinical Immunology. 2002;102(3):283-290.
219. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas HD, Hersch SM, Hogarth P, Bouzou B, Jenson RV,Krainc D. Genome-wide expression profiling of human blood reveals Huntington's disease. Proceedings of the National Academy of Science of the United States of America. 2005;102(31):11023-11028.
220. Wagenaar GT, Ter Horst SAJ, Van Gastelen MA, Leijser LM, Mauad T, Van Der Velden PA, De Heer E, Hiemstra PS, Poorthuis BJHM, Walther FJ. Gene expression profile and histopathology of experimental bronchopulmonary dysplasia induced by prolonged oxadative stress. Free Radical Biology & Medicine. 2004;36(6):782-801.
221. Kaizer EC, Glaser CL, Chaussabel D, Banchererau J, Pascual V, White PC. Gene expression in peripheral blood mononuclear cells from children with diabetes. Journal of Clinical Endocrinology and Metabolism. 2007;92(9):3705-3711.
222. Grant SFA, Hakonarson H. Microarray technology and applications in the arena of genome-wide association. Clinical Chemistry. 2008;54(7):1116-1124.
223. Maynard JA, Myhre R, Roy B. Microarrays in infection and immunity. Current Opinion in Chemical Biology. 2007;11(3):306-315.
224. Hyatt G, Melamed R, Park R, Seguritan R, Laplace C, Poirot L, Zucchelli S, Obst R, Matos M, Venazi E, Goldrath A, Nguyen L, Luckey J, Yamagata T,, Herman A, Jacobs J, Mathis D, Benoist C. Gene expression

microarrays: glimpses of the immunological genome. Nature Immunology. 2006;7(7):686-691.
225. Tang Y, Lu A, Aronow BJ, Sharp FR. Blood Genomic Responses Differ After Stroke, Seizures, Hypoglycaemia, and Hypoxia: Blood Genomic Fingerprints of Disease. Annals of Neurology. 2001;50:699-707.
226. Dobrowolski SF, Banas RA, Naylor EW, Powdrill T, Thakkar D. DNA microarray technology for neonatal screening. Acta Paediatrica. 1999;432 (supplement):61-64.
227. Green NS, Pass KA. Neonatal screening by DNA microarray: Spots and chips. Nature Reviews Genetics. 2005;6(2):147-151.
228. Liu YT. A technological update of molecular diagnostics for infectious diseases. Infectious Disorders and Drug Targets. 2008;8(3):183-188.
229. Muldrew KL. Molecular diagnostics in infectious disease. Current Opinion in Pediatrics. 2009;21(1):102-111.
230. Weile J, Knabbe C. Current applications and future trends of molecular diagnostics in clinical bacteriology. Analytical and Bioanalytical Chemistry. 2009;394(3):731-742.
231. Klouche M, Schroder U. Rapid methods for diagnosis of bloodstream infections. Clinical Chemistry and Laboratory Medicine. 2008;46(7):888-908.
232. Mikhailovich V, Gryadunov D, Kolchinsky A, Makarov AA, Zasedatelev A. DNA microarrays in the clinic: Infectious diseases. BioEssays. 2008;30(7):673-682.
233. Baron EJ. Implications of new technology for infectious diseases practice. Clinical Infectious Diseases. 2006;43(10):1318-1323.
234. Call DR. Challenges and opportunities for pathogen detection using DNA microarrays. Critical Reviews in Microbiology. 2005;31(2):91-99.
235. Kumar S, Wang L, Fan J, Kraft A, Bose ME, Tiwari S, Van Dyke M, Haigis R, Luo T, Ghosh M, Tang H, Haghnia M, Mather EL, Weisberg WG, Henrickson KJ. Detection of 11 common viral and bacterial pathogens causing community-acquired pneumonia or sepsis in asymptomatic patients by using a multiplex reverse transcription-PCR assay with manual (enzyme hybridization) or automated (electronic microarray) detection. Journal of Clinical Microbiology. 2008;46(9):3063-3072.
236. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D et al. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proceedings of the National Academy of Science of the United States of America. 2002;99(19):12391-12396.
237. Kotra LP, Vakulenko S, Mobashery S. From genes to sequences to antibiotics: prospects for furture developments from microbial genomics. Microbes and Infection. 2000;2(6):651-658.
238. Weinstock GM. Genomics and Bacterial Pathogenesis. Emerging Infectious Diseases. 2000;6(5):496-504.
239. Jenks PJ. Microbial genome sequencing- beyond the double helix. British Medical Journal. 1998;317:1568-1571.

240. Bonhomme CJ, Renesto P, Raoult D. Bacterial protein microarrays for diagnosis of infectious diseases. Current Immunology Reviews. 2008;4(1):28-36.
241. Paolucci M, Capretti MG, Dal Monte P, Corvaglia L, Landini MP, Varani S, Pession A, Faldella G, Sambri V. Laboratory diagnosis of late-onset sepsis in newborns by multiplex real-time PCR. Journal of Medical Microbiology. 2009;58(4):533-534.
242. Palka-Santini M, Cleven BE, Eichinger L, Kronke M, Krut O. Large scale multiplex PCR improves pathogen detection by DNA microarrays. BMC Microbiology. 2009;9:1.
243. Allain JP. Will genome detection replace serology in blood screening for microbial agents? Balliere's Best Practice and Research Clinical Haematology. 2000;13(4):615-629.
244. Mothershed EA, Whitney AM. Nucleic acid-based methods for the detection of bacterial pathogens: Present and future considerations for the clinical laboratory. Clinical Chimica Acta. 2006;363(1-2):206-220.
245. Relman DA. Genome-wide responses of a pathogenic bacterium to its host. Commentary. Journal of Clinical Investigation. 2002;110:1071-1073.
246. Ke D, Bergeron MG. Molecular methods for rapid detection of group B streptococci. Expert Reviews of Molecular Diagnostics. 2001;1(2):175-181.
247. Tanel RE, Kao S-Y, Niemiec TM, Loeffelholz MJ, Holland DT, Shoaf LA, Stucky ER, Burns JC. Prospective Comparisons of Culture vs Genome Detection for Diagnosis of Enteroviral Meningitis in Childhood. Archives of Pediatrics and Adolescent Medicine. 1996;150(9):919-924.
248. Kim SR, Ki C-S, Lee NY. Rapid detection and identification of 12 respiratory viruses using a dual priming oligonucleotide system-based multiplex PCR assay. Journal of Virological Methods. 2009;156(1-2):111-116.
249. Chiu CY, Urisman A, Greenhow TL, Rouskin S, Yagi S, Schnurr D, Wright C, Drew WL, Wang D, Weintrub PS, DeRisi JL, Ganem D. Utility of DNA Microarrays for Detection of Viruses in Acute Respiratory Tract Infections in Children. Journal of Pediatrics. 2008;153(1):76-83.
250. Stramer SL. Nucleic acid testing for transfusion-transmissible agents. Current Opinion in Hematology. 2000;7(6):387-391.
251. Hanson EH, Niemeyer DM, Folio L, Agan BK, Rowley RK. Potential Use of Microarray Technology for Rapid Identification of Central Nervous System Pathogens. Military Medicine 2004;169(8):594-599.
252. Shang S, Chen G, Wa Y, Du L, Zhao Z. Rapid Diagnosis of Bacterial Sepsis with PCR Amplification and Microarray Hybridization in 16S rRNA Gene. Pediatric Research. 205;58(1):143-148.
253. Shang S, Chen Z, Yu X. Detection of bacterial DNA by PCR and reverse hybridization in the 16S rRNA gene with particular reference to neonatal septicaemia. Acta Paediatrica. 2001;90:179-183.
254. Jordan JA, Durso MB, Butchko AR, Jones JG, Brozanski BS. Evaluating the near-term infant for early onset sepsis: progress and challenges to consider with 16S rDNA polymerase chain reaction testing. Journal of Molecular Diagnostics. 2006;8(3):357-363.

255. Yadav AK, Wilson CG, Prasad PL, Menon PK. Polymerase chain reaction in rapid diagnosis of neonatal sepsis. Indian Pediatrics. 2005;42(7):681-685.
256. Edman CF, Mehta P, Press R, Spargo CA, Walker GT, Nerenberg M. Pathogen Analysis and Genetic Predisposition Testing Using Microelectronic Arrays and Isothermal Amplification. Journal of Investigative Medicine. 2000;48(2):93-101.
257. Francois P, Charbonnier Y, Jaquet J, Utinger D, Bento M, Lew D, Kresbach GM, Ehrat M, Schlegel W, Schrenzel J. Rapid bacterial identification using evanescent-waveguide oligonucleotide microarray classification. Journal of Microbiological Methods. 2006;65(3):390-403.
258. Keum KC, Yoo SM, Lee SY, Chang KH, Yoo NC, Yoo WM, Kim JM, Choi JY, Kim JS, Lee G. DNA microarray-based detection of nosocomial pathogenic *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. Molecular and Cellular Probes. 2006;20:42-50.
259. Wang RF, Beggs ML, Robertson LH, Cerniglia CE. Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. FEMS Microbiology Letters. 2002;213:175-182.
260. Liu Y, Han J-X, Huang H-Y, Zhu B. Development and evaluation of 16S rDNA microarray for detecting bacterial pathogens in cerebrospinal fluid. Experimental Biology and Medicine. 2005;230(8):587-591.
261. Relman DA. The Identification of Uncultured Microbial Pathogens. The Journal of Infectious Diseases. 1993;168:1-8.
262. Yoo SM, Lee SY, Chang KH, Yoo SY, Yoo NC, Keum KC, Yoo WM, Kim JM, Choi JY. High-throughput identification of clinically important bacterial pathogens using DNA microarray. Molecular and Cellular Probes. 2009;23(3-4):171-177.
263. Roth SB, Jalava J, Ruuskanen O, Ruohola A, Nikkari S. Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory tract infections. Journal of Clinical Microbiology. 2004;42(9):4268-4274.
264. Jarvinen AK, Laakso S, Piiparinen P, Aittakorpi A, Lindfors M, Huopaniemi L, Piiparinen H, Maki M. Rapid identification of bacterial pathogens using PCR- and microarray-based assay. BMC Microbiology. 2009;9:161.
265. Fukushima M, Kakinuma K, Hayashi H, Nagai H, Ito K, Kawaguchi R. Detection and Identification of *Mycobacterium* Species Isolates by DNA Microarray. Journal of Clinical Microbiology. 2003;41(6):2605-2615.
266. Cleven BE, Palka-Santini M, Gjelen J, Meembor S, Kronke M, Krut O. Identification and characterization of bacterial pathogens causing bloodstream infections by DNA microarray. Journal of Clinical Microbiology. 2006;44(7):2389-2397.
267. Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Molecular and Cellular Probes. 2002;16:119-127.
268. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL. Microarray-based detection and genotyping of viral pathogens.

Proceedings of the National Academy of Science of the United States of America. 2002;99(24):15687-15692.
269. Wiesinger-Mayr H, Vierlinger K, Pichler R, Kriegner A, Hirschi AM, Presterl E, Bodrossy L, Noehammer C. Identification of human pathogens isolated from blood using microarray hybridisation and signal pattern recognition. BMC Microbiology. 2007;7:78.
270. Lin B, Wang Z, Vora GJ, Thornton JA, Schnur JM, Thach DC, Blaney KM, Ligler AG, Malanoski AP, Santiago J, Walter EA, Agan BK, Metzgar D, Seto D, Daum LT, Kruzelock R, Rowley RK, Hanson EH, Tibbetts C, Stenger DA. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Research. 2006;16(4):527-535.
271. Kawaguchi K, Kaneko S, Honda M, Kawai HF, Shirota Y, Kobayashi K. Detection of Hepatitis B Virus DNA in Sera from Patients with Chronic Hepatitis B Virus Infection by DNA Microarray Method. Journal of Clincal Microbiology. 2003;41(4):1701-1704.
272. Livingston AD, Campbell CJ, Wagner EK, Ghazal P. Biochip sensors for the rapid and sensitive detection of viral disease. Genome Biology. 2005;6(6):112.
273. Palacios G, Quan PL, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, Zhai J, Renwick N, Hui J, Hegyi H, Grolla A, Strong JE, Towner JS, Geisbert TW, Jahrling PB, Buchen-Osmond C, Ellerbrok H, Sanchez-Seco MP, Lussier Y, Formenty P, Nichol ST, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerging Infectious Diseases. 2007;13(1):73-81.
274. Loy A, Bodrossy L. Highly parallel microbial diagnostics using oligonucleotide microarrays. Clinica Chimica Acta. 2006;363(1-2):106-119.
275. Yoo SM, Lee SY. Diagnosis of pathogens using DNA microarray. Recent Patents on Biotechnology. 2008;2(2):124-129.
276. Holland CA, Kiechle FA. Point-of-care molecular diagnostic systems - past, present and future. Current Opinion in Microbiology. 2005;8(5):504-509.
277. Kato-Maeda M, Gao Q, Small PM. Microarray analysis of pathogens and their interaction with hosts. Cellular Microbiology. 2001;3(11):713-719.
278. Fjaerli H-O, Buckholm G, Krog A, Skjaeret C, Holden M, Nakstad B. Whole blood gene expression in infants with respiratory syncytial virus bronchiolitis. BMC Infectious Diseases. 2006;6:175.
279. Kawada J, Kimura H, Kamachi Y, Nishikawa K, Taniguchi M, Nagaoka K, Kurahashi H, Kojima S, Morishima T. Analysis of gene-expression profiles by oligonucleotide microarray in children with influenza. Journal of General Virology. 2006;87(6):1677-1683.
280. Thach DC, Agan BK, Olsen C, Diao J, Lin B, Gomez J, Jesse M, Jenkins M, Rowley R, Hanson E, Tibbetts C, Stenger DA, Walter E. Surveillance of transcriptomes in basic military trainees with normal, febrile respiratory illness, and convalescent phenotypes. Genes and Immunity. 2005;6(7):588-595.
281. Ardura MI, Banchereau P, Mejias S, Di Pucchio T, Glaser C, Allantaz F, Pascual V, Banchereau J, Chaussabel D, Ramilo O. Enhanced monocyte response and decreased central memory T cells in children with invasive

Staphylococcus aureus infections. PLoS ONE (Electronic Resource). 2009;4(5):e5446.
282. Tang BM, McLean AS, Dawes IW, Huang SJ, Lin RC. Gene-expression profiling of peripheral blood mononuclear cells in sepsis. Critical Care Medicine. 2009;37(3):882-888.
283. Tang BM, McLean AS, Dawes IW, Huang SJ, Cowley MJ, Lin RC. Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients. Critical Care Medicine. 2008;36(4):1125-1128.
284. Tang BM, McLean AS, Dawes IW, Huang SJ, Lin RC. The use of gene-expression profiling to identify candidate genes in human sepsis. Americal Journal of Respiratory Critical Care Medicine. 2007;176(7):676-684.
285. Payen D, Lukaszewicz AC, Belikova I, Faivre V, Gelin C, Russwurm S, Launay JM, Sevenet N. Gene profiling in human blood leucocytes during recovery from septic shock. Intensive Care Medicine. 2008;34(8):1371-1376.
286. Pachot A, Lepape A, Vey S, Bienvenu J, Mougin B, Monneret G. Systemic transcriptional analysis in survivor and non-survivor septic shock patients: a preliminary study. Immunology Letters. 2006;106(1):63-71.
287. Fradin C, Kretschmar M, Nichterlein T, Gaillardin C, d'Enfert C, Hube B. Stage-specific gene expression of *Candida albicans* in human blood. Molecular Microbiology. 2003;47(6):1523-1543.
288. Aujame L, Burdin N, Vicari M. How Microarrays Can Improve Our Understanding of Immune Responses and Vaccine Development. Annals of the New York Academy of Sciences. 2002;975:1-23.
289. Waddell SJ, Butcher PD, Stoker NG. RNA profiling in host-pathogen interactions. Current Opinion in Microbiology. 2007;10(3):297-302.
290. Liu M, Popper SJ, Rubins KH, Relman DA. Early days: genomics and human responses to infection. Current Opinion in Microbiology. 2006;9(3):312-319.
291. Andeweg AC, Haagman BL, Osterhaus AD. Virogenomics: the virus-host interaction revisited. Current Opinion in Microbiology. 2008;11(5):461-466.
292. Schoolnik GK. Microarray Analysis of Bacterial Pathogenicity. Advances in Microbial Physiology. 2002;46:1-45.
293. Relman DA, Falkow S. The meaning and impact of the human genome sequence for microbiology. Trends in Microbiology. 2001;9(5):206-208.
294. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B, Banchereau J, Palucka AK, Chaussabel D. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood. 2007;109:2066-2077.
295. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. The Journal of Laboratory and Clinical Medicine. 2006;147:126-132.
296. Campbell C, Vernon SD, Karem KL, Nisenbaum R, Unger ER. Assessment of normal variability in peripheral blood gene expression. Disease Markers. 2002;18(4):201-206.

297. Whitney AR, Diehn M, Popper S, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(4):1896-1901.
298. Rainen L, Oelmueller U, Jurgensen S, Wyrich R, Ballas C, Schram J, Herdman C, Bankaitis-Davis D, Nicholls N, Trollinger D, Tryon V. Stabilization of mRNA expression in whole blood samples. Clinical Chemistry. 2002;48(11):1883-1890.
299. Spijker S, Van De Leemput JC, Hoeskstra C, Boomsma D, Smit AB. Profiling gene expression in whole blood samples following an in-vitro challenge. Twin Research. 2004;7(6):564-570.
300. Kagedal B, Lindqvist M, Farneback M, Lenner L, Peterson C. Failure of the PAXgene™ Blood RNA System to maintain mRNA stability in whole blood. Clinical and Chemical Laboratory Methods. 2005;43(11):1190-1192.
301. Members, of the Toxicogenomics Research Consortium, BK Weis. Standardizing global gene expression analysis between laboratories and across platforms. Nature Methods. 2005;2:351-356.
302. Surbek DV, Holzgreve W, Jansen W, Heim D, Garritsen H, Nissen C, Wodnar-Filipowicz A. Quantitative immunophenotypic characterization, cryopreservation, and enrichment of second- and third- trimester human fetal cord blood hematopoietic stem cells (progenitor cells). American Journal of Obstetrics and Gynecology. 1998;179(5):1228-1233.
303. Walker J, Rigley K. Gene expression profiling in human peripheral blood mononuclear cells using high-density filter-based cDNA microarrays. Journal of Immunological Methods. 2000;239:167-179.
304. Yuen PK, Kricka LJ, Fortina P, Panaro NJ, Sakazume T, Wilding P. Microchip Module for Blood Sample Preparation and Nucleic Acid Amplification Reactions. Genome Research. 2001;11(3):405-412.
305. Vachier I, Roux S, Chanez P, Loubatiere J, Terouanne B, Nicolas JC, Godard P. Glucocorticoids induced down-regulation of glucocorticoid receptor mRNA expression in asthma. Clinical and Experimental Immunology. 1996;103(2):311-315.
306. Laaksonen K, Waris M, Makela MJ, Terho EO, Savolainen J. In Vitro kinetics of allergen- and microbe-induced IL-4 and IFN-gamma mRNA expression in PBMC of pollen-allergic patients. Allergy. 2003;58(1):62-66.
307. Johansson M, Pisa EK, Tormanen V, Arstrand K, Kagedal B. Quantitative Analysis of Tyrosinase Transcripts in Blood. Clinical Chemistry. 2000;46(7):921-927.
308. Hrboticky N, Draude G, Hapfelmeier G, Lorenz R, Weber PC. Lovastatin Decreases the Receptor-Mediated Degradation of Acetylated and Oxidized LDLs in Human-Blood Monocytes During the Early Stage of Differentiation into Macrophages. Arteriosclerosis, Thrombosis, and Vascular Biology. 1999;19(5):1267-1275.
309. Rahman MT, De Ley M. Metallothionein isogene transcription in red blood cell precursors from human cord blood. European Journal of Biochemistry. 2001;268(3):849-856.

310. Karlsson MG, Ludvigsson J. Determination of mRNA expression for IFN-γ and IL-4 in lymphocytes from children with IDDM by RT-PCR technique. Diabetes Research and Clinical Practice. 1998;40:21-30.
311. Rozenfeld-Granot G, Toren A, Amariglio N, Nagler A, Rosenthal E, Biniaminov M, Brok-Simon F, Rechavi G. MAP kinase activation by mu opioid receptor in cord blood CD34+CD38- cells. Experimental Hematology. 2002;30:473-480.
312. Zhang HQ, Lu H, Enosawa S, Takahara S, Sakamoto K, Nakajima T, Saito H, Suzuki S. Microarray Analysis of Gene Expression in Peripheral Blood Mononuclear Cells Derived from Long-Surviving Renal Recipients. Transplantation Proceedings. 2002;34:1757-1759.
313. Gandini O, Celi FS, Magnati M, Gazzaniga P, Silvestri I, Conti B, Giuliani L, Mentuccia D, Agliano AM. A Rapid, Simple, and Inexpensive Step Facilitates RNA Extraction from Whole Blood Cells. Laboratory Investigation. 1999;79(12):1731-1732.
314. Chomczynski P, Sacchi N. Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. Analytical Biochemistry. 1987;162:156-159.
315. Ylikoski A, Sjoroos M, Lundwall A, Karp M, Lovgren T, Lija H, Iitia A. Quantitative Reverse Transcription-PCR Assay with an Internal Standard for the Detection of Prostate-specific Antigen mRNA. Clinical Chemistry. 1999;45(9):1397-1407.
316. Malarstig A, Tenno T, Jossan S, Aberg M, Siegbahn A. A quantitative real-time PCR method for tissue factor mRNA. Thrombosis Research. 2003;112:175-183.
317. Kruse N, Pette M, Toyka K, Rieckmann P. Quantification of cytokine mRNA expression by RT PCR in samples of previously frozen blood. Journal of Immunological Methods. 1997;210:195-203.
318. Tang Y, Lu A, Ran R, Aronow BJ, Schorry EK, Hopkin RJ, Gilbert DL, Glauser TA, Hershey AD, Richtand NW, Privitera M, Dalvi A, Sahay A, Szaflarski JP, Ficker DM, Ratner N, Sharp FR. Human blood genomics: distinct profiles for gender, age and neurofibromatosis type 1. Molecular Brain Research. 2004;132(2):155-167.
319. Sanchez-Ramos JR, Song S, Kamath SG, Zigova T, Willing A, Cardoza-Pelaez F, Stedeford T, Chopp M, Sanberg PR. Expression of Neural Markers in Human Umbilical Cord Blood. Experimental Neurology. 2001;171:109-115.
320. Chadderton T, Wilson C, Bewick M, Gluck S. Evaluation of three rapid RNA extraction reagents: relevance for use in RT-PCR's and measurement of low level gene expression in clinical samples. Cellular and Molecular Biology. 1997;43(8):1227-1234.
321. Maeno Y, Nakazawa S, Nagashima S, Sasaki J,Higo KM, Taniguchi K. Utility of the dried blood on filter paper as a source of cytokine mRNA for the analysis of immunoreactions in Plasmodium yoelii infection. Acta Tropica. 2003;87(2):295-300.
322. Bauer M, Patzelt D. A method for simultaneous RNA and DNA isolation from dried blood and semen stains. Forensic Science International. 2003;136(1-3):76-78.

323. Juusola J, Ballantyne J Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. Forensic Science International. 2003;135:85-96.
324. Suttorp M, Ritgen M, von Neuhoff N, Schoch R, Schmitz N. Blood on Filter Paper as a Readily Available Source of bcr-abl Rearranged mRNA. Blood. 1997;90(4):1713-1715.
325. Haak PT, Busik JV, Kort EJ, Tikhonenko M, Paneth N, Resau JH. Archived unfrozen neonatal blood spots are amenable to quantitative gene expression analysis. Neonatology. 2009;95(3):210-216.
326. Hardin J, Finnell RH, Wong D, Hogan ME, Horowitz J, Shu J, Shaw GM. Whole genome microarray analysis, from neonatal blood cards. BMC Genetics. 2009;10:38.
327. Thach DC, Lin B, Walter E, Kruzelock R, Rowley RK, Tibbetts C, Stenger DA. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. Journal of Immunological Methods. 2003;283:269-279.
328. Chai V, Vassilakos A, Lee Y, Wright JA, Young AH. Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. Journal of Clinical Laboratory Analysis. 2005;19(5):182-188.
329. Pahl A, Brune K. Stabilization of Gene Expression Profiles in Blood After Phlebotomy. Clinical Chemistry. 2002;48:2251-2253.
330. Muller MC, Merx K, Weisser A, Kreil S, Lahaye T, Hehlmann R, Hochhaus A. Improvement of molecular monitoring of residual disease in leukaemias by bedside RNA stabilization. Leukemia. 2002;16(12):2395-2399.
331. Debey-Pascher S, Eggle D, Schultze JL. RNA stabilization of peripheral blood and profiling by bead chip analysis. Methods in Molecular Biology. 2009;496:175-210.
332. Li L, Ying L, Naesens M, Xiao W, Sidgel T, Hsieh S, Martin J, Chen R, Liu K, Mindrinos M, Davis R, Sarwal M. Interference of globin genes with biomarker discovery for allograft rejection in peripheral blood samples. Physiological Genomics. 2008;32(2):190-197.
333. Technical note. Expression Analysis Technical Note. wwwexpressionanalysiscom. June 2005;919-405-2248.
334. GLOBINclear-Human Kit Manual. Version 0503.
335. Technical note. Globin-Reduction Protocol: A Method for Processing Whole Blood RNA Samples for Improved Array Results. wwwaffymetrixcom.
336. Hartel C, Bein G, Muller-Steinhardt M, Kluter H. Ex vivo induction of cytokine mRNA expression in human blood samples. Journal of Immunological Methods. 2001;249(1-2):63-71.
337. Affymetrix Technical Note. An Analysis of Blood Processing Methods to Prepare Samples for GeneChip Expression Profiling. Gene Expression Monitoring. 2003;www.affymetrix.com.
338. Tanner MA, Berk LS, Felten DL, Blidy AD, Bit SL, Ruff DW. Substantial Changes in Gene Expression Level Due To The Storage Temperature and Storage Duration of Human Whole Blood. Clinical and Laboratory Haematology. 2002;24(6):337-342.

339. Feezor RJ, Baker HV, Mindrinos M, Hayden D, Tannahill CL, Brownstein BH, Fay A, MacMillan S, Laramie J, Xiao W, Moldawer LL, Cobb, JP, Laudanski K et al. and the Inflammation and Host Response to Injury Large Scale Collaborative Research Program. Whole blood and leukocyte RNA isolation for gene expression analyses. Physiological Genomics. 2004;19(3):247-254.
340. Asare AL, Kolchinsky SA, Gao Z, Wang R, Raddassi K, Bourcier K, Seyfert-Margolis V. Differential gene expression profiles are dependent upon method of peripheral blood colection and RNA isolation. BMC Genomics. 2008;9:474.
341. Carrol ED, Salway F, Pepper SD, Saunders E, Mankhambo LA, Ollier WE, Hart CA, Day P. Successful downstream application of the Paxgene Blood RNA system from small blood samples in paediatric patients for quantitative PCR analysis. BMC Immunology. 2007;8:20.
342. Wang J, Robinson JF, Khan HMR, Carter DE, McKinney J, Miskie BA, Hegele RA. Optimizing RNA extraction yield from whole blood for microarray gene expression analysis. Clinical Biochemistry. 2004;37:741-744.
343. Manroe BL, Weinberg AG, Rosenfeld CR, Browne R. The neonatal blood count in health and disease. I. Reference values for neutrophilic cells. Journal of Pediatrics. 1979;95(1):89-98.
344. Mouzinho A, Rosenfeld CR, Sanchez PJ, Risser R. Revised reference ranges for circulating neutrophils in very-low-birth-weight neonates. Pediatrics. 1994;94(1):76-82.
345. Smith CL, Dickinson P, Forster T, Khondoker M, Craigon M, Ross A, Storm P, Burgess S, Lacaze P, Stenson BJ, Ghazal P. Quantitative assessment of human whole blood RNA as a potential biomarker for infectious disease. Analyst. 2007;132(1200-1209).
346. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-264.
347. Smyth GK. Chapter 23: limma:Linear Models for Microarray Data in Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Editors: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S.
: Springer; 2005.
348. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protocols. 2009;4(1):44-57.
349. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology. 2003;4(5):P3.
350. Fischer JE, Benn A, Harbarth S, Nadal S, Fanconi S. Diagnostic accuracy of G-CSF, IL-8, and IL-1Ra in critically ill children with suspected infection. Intensive Care Medicine. 2002;28(9):1324-1331.
351. Kuster H, Weiss M, Willeitner AE, Detlefsen S, Jeremias I, Zbojan J, Geiger R, Lopowsky G, Simbruner G. Interleukin-1 receptor antagonist and

interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation. Lancet. 1998;352(9136):1271-1277.
352. Canaday DH, Chakravarti S, Srivastava T, Tisch DJ, Cheruvu VK, Harding CV, Ramachandra L. Class II MHC antigen presentation defect in neonatal monocytes is not correlated with decreased MHC-II expression. Cell Immunology. 2006;243(2):96-106.
353. Birle A, Nebe CT, Gessler P. Age-related low expression of HLA-DR molecules on monocytes of term and preterm newborns with and without signs of infection. Journal of Perinatology. 2003;23(4):294-299.
354. Kanakoudi-Tsakalidou F, Debonera F, Drossou-Agakidou V, Sarafidis K, Tzimouli V, Taparkou A, Kremenopoulos G. Flow cytometric measurement of HLA-DR expression on circulating monocytes in healthy and sick neonates using monocyte negative selection. Clinical Experimental Immunology. 2001;123(3):402-407.
355. Hallwirth U, Pomberger G, Zaknun D, Szepfalusi Z, Horcher E, Pollak A, Roth E, Spittler A. Monocyte phagocytosis as a reliable parameter for predicting early-onset sepsis in very low birthweight infants. Early Human Development. 2002;67(1-2):1-9.
356. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25(3):415-416.
357. Prucha M, Ruryk A, Boriss H, Moller E, Zazula R, Herold I, Claus RA, Reinhart KA, Deigner P, Russwurm S. Expression profiling: toward an application in sepsis diagnostics. Shock. 2004;22(1):29-33.
358. Kurt AN, Aygun AD, Godekmerdan A, Kurt A, Dogan Y, Yilmaz E. Serum IL-1beta, IL-6, IL-8, and TNF-alpha levels in early diagnosis and management of neonatal sepsis. Mediators of Inflammation. 2007;2007:31397.
359. Wynn JL, Scumpia PO, Winfield RD, Delano MJ, Kelly-Scumpia K, Barker T, Ungaro R, Levy O, Moldawer LL. Defective innate immunity predisposes murine neonates to poor sepsis outcome but is reversed by TLR agonists. Blood. 2008;112(5):1750-1758.
360. Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA. Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. BMC Genomics. 2009;10:2.
361. ncbi. The published globin mRNA sequences wwwncbinlmnihgov/entrez/queryfcgi?db=Nucleotide(alpha)itool=toolby NM000558, NM 000518, NM000517, NM000559, NM000184.

Appendix 1

Short Notes on Investigation of Globin Reduction

Background

As gene expression profiling of whole blood has become more commonplace, methods to try and overcome the challenges presented by the heterogeneous nature of whole blood have emerged. In particular, methods for globin reduction have come onto the market. Globin mRNA transcripts have been estimated to constitute a significant proportion of whole blood mRNA. It is thought that these transcripts decrease the sensitivity of detecting the other mRNAs of interest when using microarray technology (333, 334). Indeed, some literature suggests that there is reduction in percent present cells and increased variability when looking at microarray data from whole blood mRNA compared to that extracted directly from white blood cells or compared to where globin reduction has been incorporated (333-335). Furthermore, globin reduction techniques have been put forward as ways to reduce this variability and to increase sensitivity of detecting less abundant mRNAs (333-335).

With such data emerging, I felt that it was important to consider using a globin reduction technique on my extracted RNA. However I was wary that adding an additional step would increase the time of sample processing and would introduce another possible source of variation between samples.

There was a significant problem when it came to considering the use of globin reduction techniques for my samples however. The globin reduction methods on the market were all targeted towards reduction of adult globin mRNAs, i.e. alpha and beta globin polypeptides whereas the haemoglobin in neonatal blood is predominantly fetal haemoglobin. Adult haemoglobin is a tetramer composed of two alpha and two beta subunits, fetal haemoglobin is a tetramer composed of two alpha and two gamma subunits. It is probably worth commenting on the fact that any blood transfusions given to newborns would be from adult donors. The donated blood is white cell depleted so should not affect the RNA expression profiles of interest when considering response to infection. The globin composition will differ from that in non-transfused infants however due to the higher Haemoglobin A in the transfused blood.

Through studying product literature (GeneChip® Globin-Reduction Kit, Troubleshooting Guide) and a personal communication with Ambion (GLOBINclear™-Human Kit) I confirmed that the available methods targeted alpha and beta globin mRNA and would not be expected to reduce gamma globin polypeptides. I further confirmed this by comparing the published target areas in company literature with the sequences published for each homo sapiens haemoglobin (haemoglobins Alpha 1, Alpha 2, Beta, Gamma, Gamma 2) on Entrez Nucleotide using NCBI Sequence Viewer v 2.0 (361). I also sought advice from GE who marketed the CodeLink™ arrays for any experience they had on the use of globin reduction.

Despite confirmation that the use of globin reduction techniques would only target a small proportion of neonatal haemoglobin, I decided to try out a globin reduction method. This was to try and get an indication as to whether or not globin reduction would be useful in my neonatal samples. Of the several globin reduction methods available I chose the GLOBINclear™-Human kit (Ambion®) because this method avoided use of RNase H treatments which we were concerned could degrade or even alter the expression profiles of our small precious samples. The other main contender was the GeneChip® Globin-Reduction method (Affymetrix®/PreAnalytiX) which utilises peptide nuceleic acid (PNA) oligonucleotides and was designed for use with the PAXgene™ system. There were two problems with the GeneChip® method with respect to this study. Firstly, it was meant for use with the target preparation protocol for GeneChip® arrays. Secondly, the yield of cDNA achieved with the GeneChip® was reduced and therefore the amount of input RNA needed for this protocol was higher at 8 µg. Pooling of PAXgene™ Blood RNA Tubes was recommended. This meant that the GeneChip® method was clearly not feasible for use with the neonatal samples. The GLOBINclear™ protocol also required an increased starting amount of RNA: 3 µg of RNA inputted into the globin reduction protocol would be required to provide 1 µg for the CodeLink™ labelling protocol. This level of reduction was however at least feasible to investigate further.

Due to financial constraints, at the time of carrying out this exploratory work we had only 6 Codelink™ Whole Human Genome Arrays available. It was therefore only possible to spare one array to investigate globin reduction. I therefore used a

GLOBINclear™-Human kit and protocol from Ambion® on sample csb1 which had already been run on microarray without globin reduction.

Acknowledgements

Running of the microarrays was carried out by me and supervised by Marie Craigon. The analysis and tabulation of comparisons of the arrays was carried out by Thorsten Forster. All of the other work in this section was carried out by myself.

Aim

To get an indication of potential benefit of using a globin reduction step with neonatal samples.

Method

Two microarrays of a single neonatal sample (csb1) were compared: one array without globin reduction, the other with globin reduction using the GLOBINclear protocol as found in the GLOBINclear™Kit Instruction Manual from Ambion®, Catalog #1980. Manual Version 0506.

Csb1 was the first of our RNA samples to be run on CodeLink™. It was initially labelled, hybridised and run on array (method 1) as described on page 101. On a separate occasion, 14 microlitres (1.82 micrograms) of the RNA sample was used for

globin reduction. No precipitation step was used. The same method of labelling, hybridisation and microarray analysis was then carried out.

Results

Table A.1.1 shows the spot quality data for each of the two arrays. From this it can be seen that there were more good quality spots, less contamination and less low signal spots with the sample that had not been subjected to globin reduction.

Table A.1.1: ***Comparison of Array Quality for a Neonatal Sample With and Without Globin Reduction***

Spot quality	T00269703 No globin reduction	T00269704 With globin reduction
Good	37473	35112
Contamination	77	196
MSR file but no intensity data	122	110
Low - below background intensity	16125	18972
Irregular shape	605	166
Spot manually excluded	0	error 1

Separately to the above experiment, when examining Bioanalyzer traces of cRNA from neonatal blood, there was an interesting observation made. In the literature (technical notes) (333, 335), the electrophoresis traces of PAXgene™ cRNA were described as having a significant peak due to globin transcripts, with this peak being reduced with globin reduction. In figure A.1.1 I have drawn an outline to show the type of bioanalyzer traces these notes describe. Figure A.1.2 shows Bioanalyzer graphs of cRNA for a selection of my neonatal blood samples. It was interesting to note that many of the neonatal samples had cRNA profiles which were closer to those having been subjected to globin reduction than those not. This may mean that the globin peak is not

such a prominent feature in neonatal blood, giving further doubt to the potential utility of globin reduction in neonatal samples.

Figure A.1.1: ***Expected shape of cRNA traces as described in the literature for PAXgene™ and Globin reduced samples.***

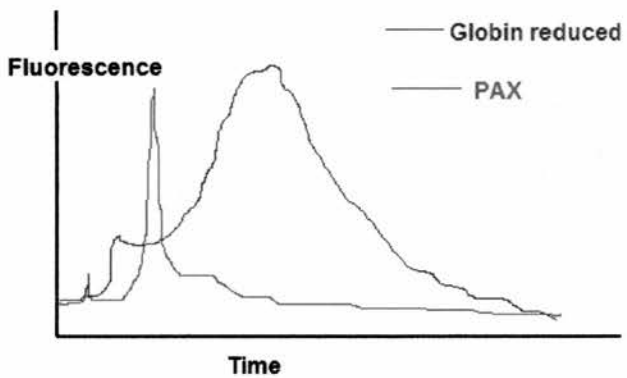
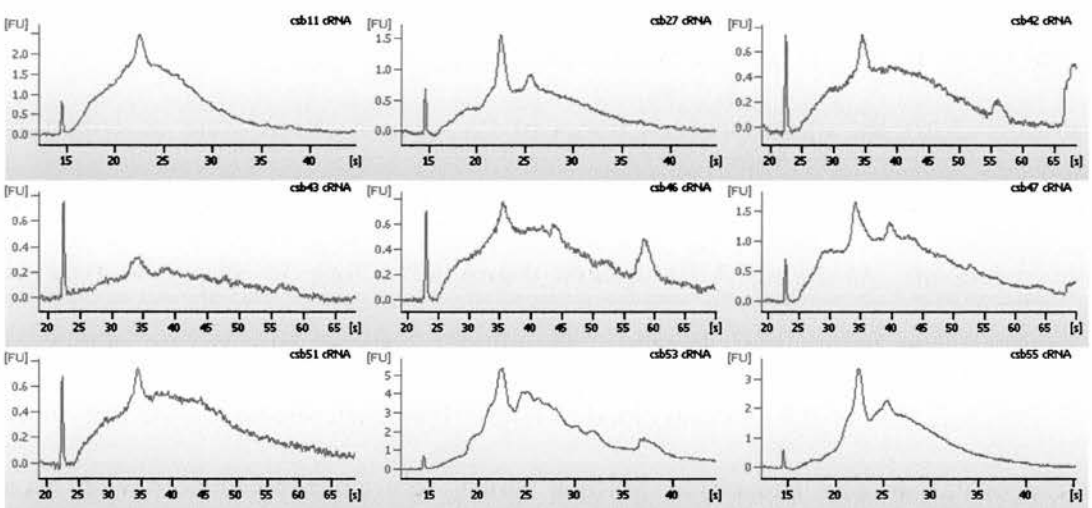


Figure A.1.2: ***Bioanalyzer traces for neonatal cRNA***



Discussion

This comparison was carried out on a look and see basis to give an indication of the usefulness of globin reduction. A comparison of two samples cannot be considered to give statistically significant results and caution must therefore be exercised in extrapolating too much from this data. It is possible that the differences seen could be due to normal technical variation. In addition, as both arrays were from the same sample it was not possible to assess inter-individual variability. While bearing these caveats in mind, the quality data presented here seems to indicate that the sample that had not had globin reduction carried out fared better in terms of the number of probes that gave good quality data. Why this should be the case is not clear and it may be a reflection that globin reduction based on adult sequences does not give full benefit to neonatal samples and may even have detrimental effects. The observation that the bioanalyzer traces obtained from neonatal blood appear closer in shape to globin reduced adult samples may suggest that the globin predominance is not such an issue in neonatal samples. Further work would clearly be required to elucidate this further.

Conclusions

This experiment did not provide evidence that globin reduction would be useful in neonatal samples. There was a suggestion that omitting the globin reduction step may actually give superior results. It was therefore decided that globin reduction would not be incorporated into the protocol for this study.

Appendix 2

The tables in this section contains raw data, Bioanalyzer traces and RIN values for the work carried out on umbilical cord blood RNA from chapter 3.

Notes on table 3.1.1

There were 5 individual tubes for which there are no data available. Sample cs1 Lithium Heparin tube had the wrong cap placed on it in error at the time of sampling. This meant that there would have been contamination from other chemicals in this tube and therefore RNA was not extracted. Sample cs3 PAXgene™ was eluted in the wrong buffer and therefore no valid data for comparison was available on this sample. In addition sample cs4 clotted in all three of the clinical blood tubes on route to the lab and these were therefore unable to be used. As the volume of blood required from each cord was fairly large, it proved difficult to achieve sufficient for all the tests on all occasions – this was especially true for the first few umbilical cords sampled and I got more proficient at obtaining more blood with experience. Where insufficient blood was obtained, all of the clinical and TRIzol®LS tubes were filled first and the PAXgene™ tube was injected with the remaining volume of blood (0.8 – 2 ml). 2.5 ml of blood was available for the PAXgene™ tube for each of the last 5 cords.

Table 3.1.1: *Raw data for samples according to collection tube.*							
	Blood Tube	**260/280 Ratio (in Tris)**	**A260**	**Total RNA (μg)**	**RNA/0.5 ml (μg)**	**RIN**	**Comments**
cs1	**EDTA**	1.01	0.07	11.70	11.70	1	
cs1	**Li Heparin**	-	.	-	-	-	Wrong cap on
cs1	**Na Citrate**	1.88	0.079	13.21	13.21	8.1	
cs1	**TRIzol®LS**	1.89	0.066	11.04	11.04	9	
cs1	**PAXgene™**	2.15	0.101	13.33	3.33	9.4	PAX: 2ml blood
cs2	**EDTA**	1.38	0.022	3.68	3.68	5.8	
cs2	**Li Heparin**	1.79	0.034	5.68	5.68	8.9	
cs2	**Na Citrate**	1.62	0.034	5.68	5.68	8.3	
cs2	**TRIzol®LS**	1.63	0.031	5.18	5.18	8.9	
cs2	**PAXgene™**	1.93	0.106	13.99	3.50	9.8	PAX: 2ml
cs3	**EDTA**	1.76	0.037	6.19	6.19	7.6	
cs3	**Li Heparin**	2.09	0.113	18.89	18.89	8.4	
cs3	**Na Citrate**	1.87	0.073	12.21	12.21	7.4	
cs3	**TRIzol®LS**	1.73	0.064	10.70	10.70	8.7	
cs3	**PAXgene™**	-	-	-	-	-	Elution error
cs4	**EDTA**	-	-	-	-	-	Blood clotted
cs4	**Li Heparin**	-	-	-	-	-	Blood clotted
cs4	**Na Citrate**	-	-	-	-	-	Blood clotted
cs4	**TRIzol®LS**	1.87	0.103	17.22	17.22	7.1	
cs4	**PAXgene™**	2.20	0.088	11.62	7.26	10	PAX: 0.8ml
cs5	**EDTA**	1.77	0.076	12.71	12.71	4.1	
cs5	**Li Heparin**	1.90	0.146	24.41	24.41	7.4	
cs5	**Na Citrate**	1.75	0.077	12.87	12.87	2.2	
cs5	**TRIzol®LS**	1.72	0.239	39.96	39.96	9.2	
cs5	**PAXgene™**	1.95	0.409	53.99	10.80	9.5	PAX: 2.5ml
cs6	**EDTA**	2.10	0.021	3.51	3.51	6.1	
cs6	**Li Heparin**	2.00	0.026	4.35	4.35	7.7	
cs6	**Na Citrate**	2.00	0.014	2.34	2.34	6.6	
cs6	**TRIzol®LS**	1.83	0.033	5.52	5.52	6.8	
cs6	**PAXgene™**	1.96	0.499	65.87	13.17	9.5	PAX:2.5ml
cs7	**EDTA**	1.93	0.058	9.70	9.70	7.3	
cs7	**Li Heparin**	1.97	0.059	9.86	9.86	7.7	
cs7	**Na Citrate**	2.00	0.076	12.71	12.71	2.3	
cs7	**TRIzol®LS**	2.02	0.127	21.23	21.23	8.1	
cs7	**PAXgene™**	2.01	0.358	47.26	9.45	9.7	PAX:2.5ml
cs8	**EDTA**	1.89	0.036	6.02	6.02	1.1	
cs8	**Li Heparin**	2.17	0.039	6.52	6.52	7.1	
cs8	**Na Citrate**	1.75	0.014	2.34	2.34	1.1	
cs8	**TRIzol®LS**	2.04	0.096	16.05	16.05	8.3	
cs8	**PAXgene™**	2.02	0.24	31.68	6.34	8.3	PAX:2.5ml
cs9	**EDTA**	2.01	0.149	24.91	24.91	7.3	
cs9	**Li Heparin**	2.01	0.183	30.60	30.60	7.9	
cs9	**Na Citrate**	1.95	0.121	20.23	20.23	8	
cs9	**TRIzol®LS**	2.01	0.221	36.95	36.95	7.4	
cs9	**PAXgene™**	1.99	0.432	57.02	11.40	9.6	PAX:2.5ml

Table 3.1.4a: *Bioanalyzer graphs and RIN values for RNA extracted after collection in EDTA, Lithium Heparin, Sodium Citrate, TRIzol®LS or PAXgene™ tubes.*

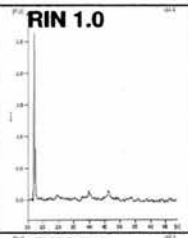
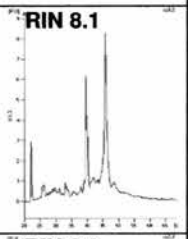
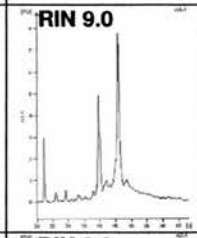
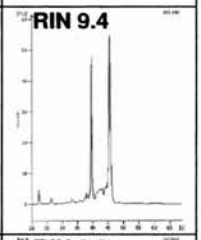
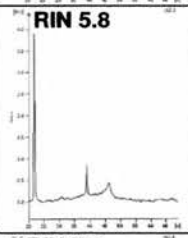
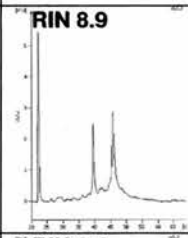
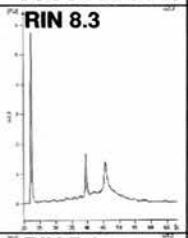
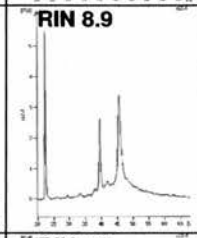
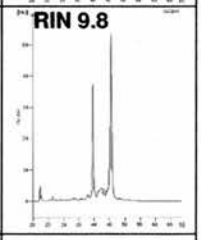
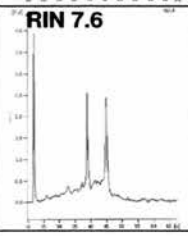
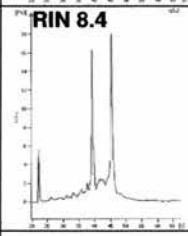
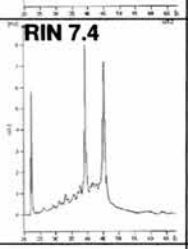
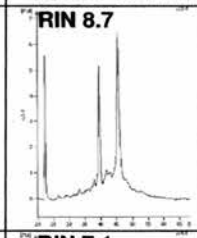
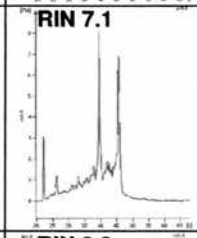
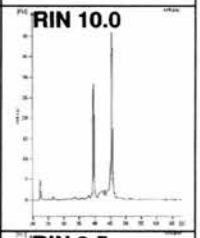
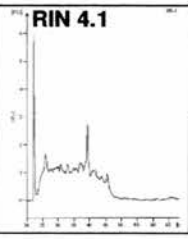
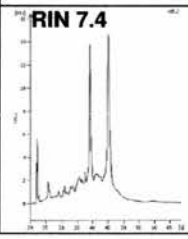
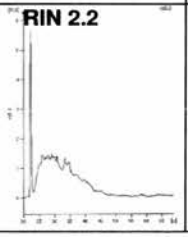
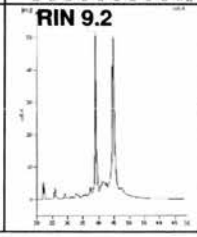
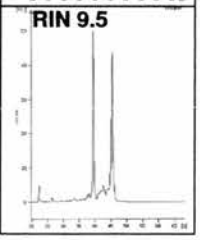
Sample	EDTA	Li Heparin	Na Citrate	TRIzol®LS	PAXgene™
cs1	 RIN 1.0	-	 RIN 8.1	 RIN 9.0	 RIN 9.4
cs2	 RIN 5.8	 RIN 8.9	 RIN 8.3	 RIN 8.9	 RIN 9.8
cs3	 RIN 7.6	 RIN 8.4	 RIN 7.4	 RIN 8.7	-
cs4	-	-	-	 RIN 7.1	 RIN 10.0
cs5	 RIN 4.1	 RIN 7.4	 RIN 2.2	 RIN 9.2	 RIN 9.5

Table 3.1.4b: *Bioanalyzer graphs and RIN values for RNA extracted after collection in EDTA, Lithium Heparin, Sodium Citrate, TRIzol®LS or PAXgene™ tubes.*

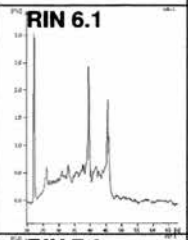
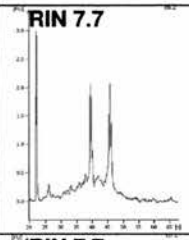
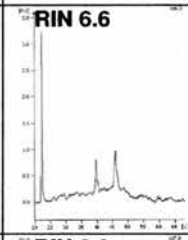
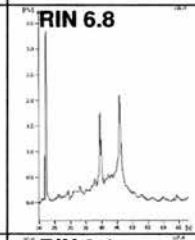
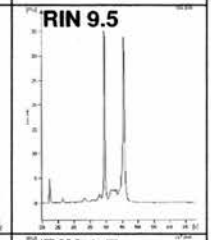
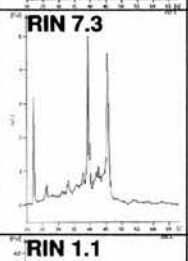
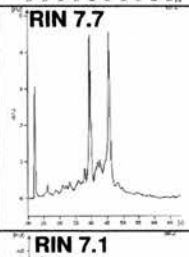
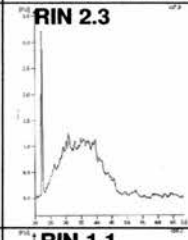
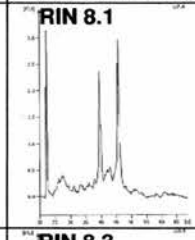
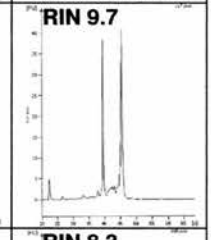
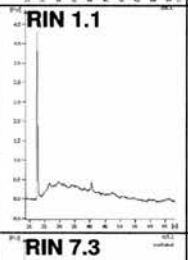
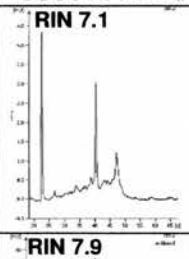
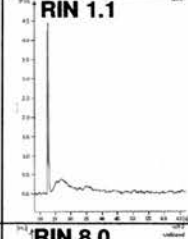
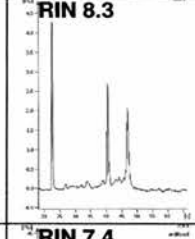
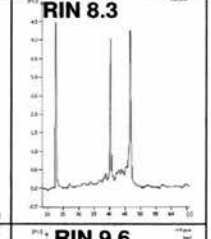
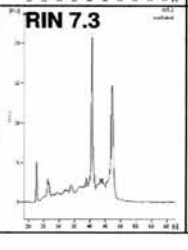
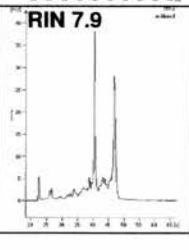
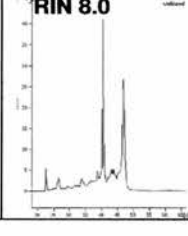
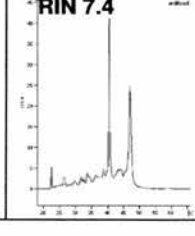
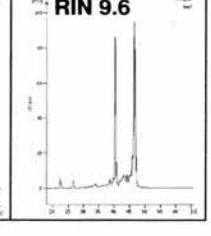
	EDTA	Li Heparin	Na Citrate	TRIzol®LS	PAXgene™
cs6	 RIN 6.1	 RIN 7.7	 RIN 6.6	 RIN 6.8	 RIN 9.5
cs7	 RIN 7.3	 RIN 7.7	 RIN 2.3	 RIN 8.1	 RIN 9.7
cs8	 RIN 1.1	 RIN 7.1	 RIN 1.1	 RIN 8.3	 RIN 8.3
cs9	 RIN 7.3	 RIN 7.9	 RIN 8.0	 RIN 7.4	 RIN 9.6

Table 3.2.1: *Raw data for each sample according to method of RNA extraction*

	Method of extraction	260/280	Repeat 260/280	A260	Total RNA (μg)	RNA/ 0.5 ml (μg)	RIN
cs10	**QIAamp®**	0 (E)	2.0	0.008	1.34	1.34	1
cs10	**Trizol®LS**	1.92	-	0.169	28.25	28.25	1
cs10	**Trizol®LS + QIAamp®**	2.02	-	0.105	17.56	17.56	7.1
cs10	**PAXgene™ (2.5 ml)**	2	-	0.43	56.76	11.35	7.1
cs11	**QIAamp®**	1.88	1.86	0.013	2.18	2.18	5.7
cs11	**Trizol®LS**	1.82	-	0.266	44.48	44.48	n/a
cs11	**Trizol®LS + QIAamp®**	1.95	-	0.152	25.41	25.41	7
cs11	**PAXgene™ (2.5 ml)**	1.98	-	0.631	83.29	16.66	5.5
cs12	**QIAamp®**	1.46	1.83	0.022	3.68	3.68	1.1
cs12	**Trizol®LS**	1.65	-	0.162	27.08	27.08	1.2
cs12	**Trizol®LS + QIAamp®**	1.97	-	0.057	9.53	9.53	6.4
cs12	**PAXgene™ (2.5 ml)**	1.96	-	0.167	22.04	4.41	8.2
cs13	**QIAamp®**	1.91	1.9	0.019	3.17	3.17	8.2
cs13	**Trizol®LS**	1.91	-	0.181	30.27	30.27	3.3
cs13	**Trizol®LS + QIAamp®**	1.99	-	0.135	22.57	22.57	7.4
cs13	**PAXgene™ (2.5 ml)**	1.99	-	0.358	47.26	9.45	7.5
cs14	**QIAamp®**	4.67	1.82	0.02	3.34	3.34	1.6
cs14	**Trizol® LS**	1.76	-	0.183	30.60	30.60	6.4
cs14	**Trizol®LS + QIAamp®**	1.92	-	0.096	16.06	16.06	7.8
cs14	**PAXgene™ (2.5 ml)**	1.95	-	0.296	39.08	7.82	8.9
cs15	**QIAamp®**	2.5	1.92	0.025	4.18	4.18	7.9
cs15	**Trizol®LS**	1.69	-	0.071	11.88	11.88	1
cs15	**Trizol®LS + QIAamp®**	1.5	-	0.054	9.03	9.03	n/a
cs15	**PAXgene™ (2.5 ml)**	2.01	-	0.335	44.22	8.84	7.3
cs16	**QIAamp®**	7	1.88	0.015	2.51	2.51	7.4
cs16	**Trizol®LS**	1.82	-	0.071	11.88	11.88	2.3
cs16	**Trizol®LS + QIAamp®**	2.02	-	0.083	13.88	13.88	6.6
cs16	**PAXgene™ (2.5 ml)**	2	-	0.236	31.15	6.23	8

Notes on table 3.2.1

The A_{260}:A_{280} ratios were noted to vary widely for the samples extracted by QIAamp® and on reflection I realised that these samples may have sat for longest after preparation for spectrophotometry prior to being analysed on the spectrophotometer. The QIAamp® samples were therefore re-run on the spectrophotometer after making up the dilutions freshly and ensuring that they were run as soon as possible after mixing by vortexing. Both sets of results for A_{260}:A_{280} ratios are displayed in table 3.2.1 for completeness. The A_{260} values displayed are those used for calculation of RNA yield. As always, the A_{260} values were considered valid if the spectrophotometry A_{260} values for each of the three cycles agreed within 0.002. In the unusual event of having repeated a sample and having two valid results, the highest A_{260} value would have been selected.

Table 3.2.4: ***Bioanalyzer Graphs and RIN values for RNA extracted using QIAamp®, TRIzol®LS, TRIzol®LS followed by QIAamp® clean-up and PAXgene™ methods.***

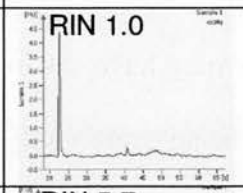
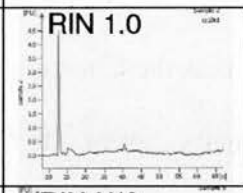
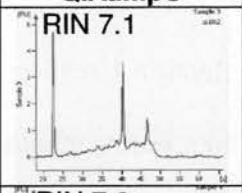
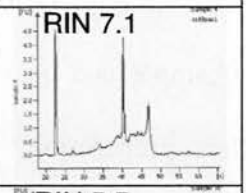
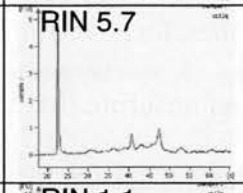
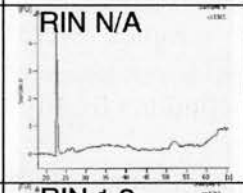
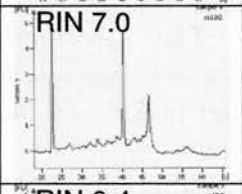
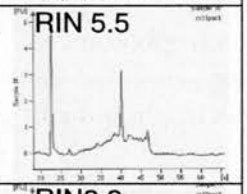
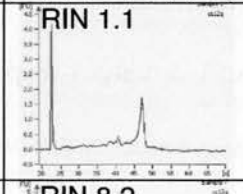
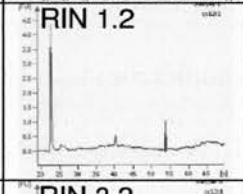
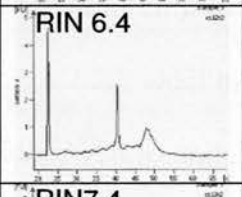
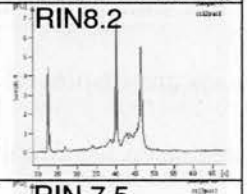
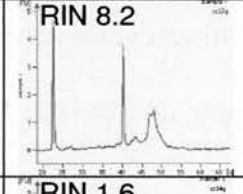
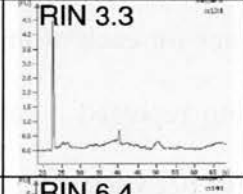
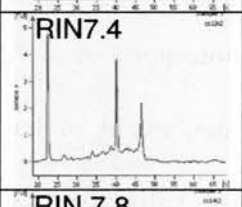
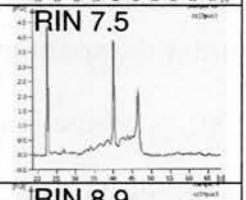
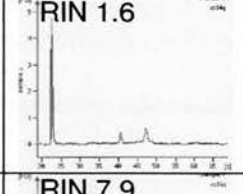
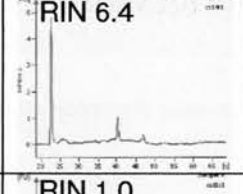
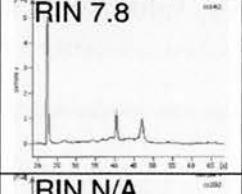
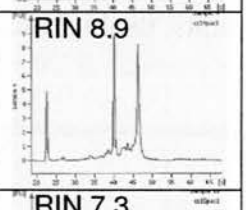
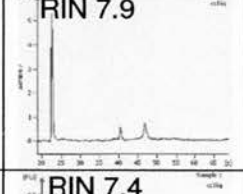
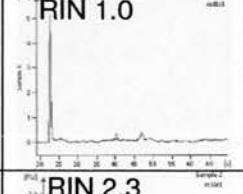
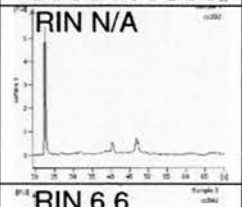
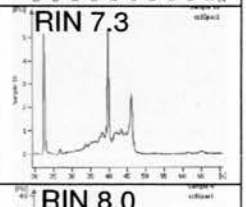
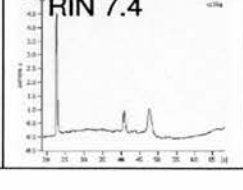
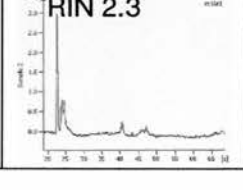
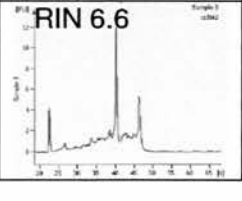
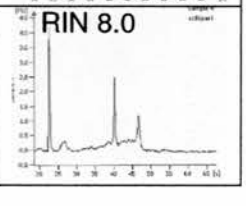
	QIAamp®	TRIzol®LS	TRIzol®LS & QIAamp®	PAXgene™
cs10	 RIN 1.0	 RIN 1.0	 RIN 7.1	 RIN 7.1
cs11	 RIN 5.7	 RIN N/A	 RIN 7.0	 RIN 5.5
cs12	 RIN 1.1	 RIN 1.2	 RIN 6.4	 RIN8.2
cs13	 RIN 8.2	 RIN 3.3	 RIN7.4	 RIN 7.5
cs14	 RIN 1.6	 RIN 6.4	 RIN 7.8	 RIN 8.9
cs15	 RIN 7.9	 RIN 1.0	 RIN N/A	 RIN 7.3
cs16	 RIN 7.4	 RIN 2.3	 RIN 6.6	 RIN 8.0

Table 3.3.4: *Bioanalyzer graphs and RIN values for RNA extracted using PAXgene™ and MagaZorb®.*

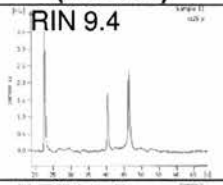
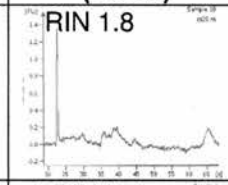
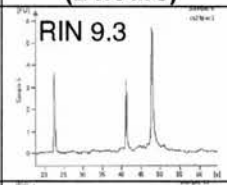
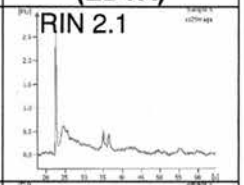
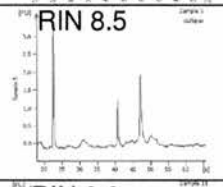
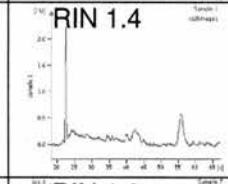
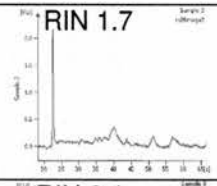
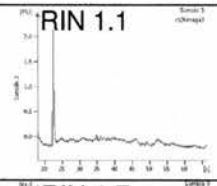
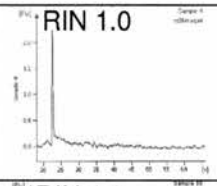
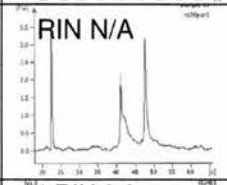
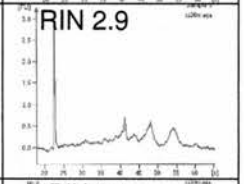
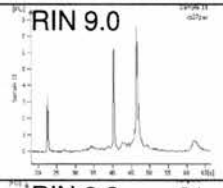
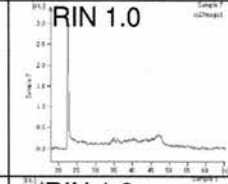
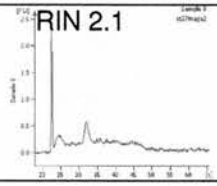
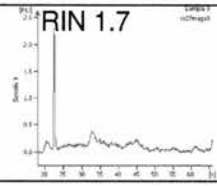
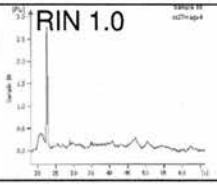
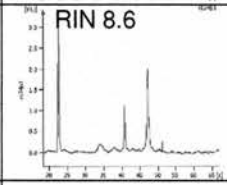
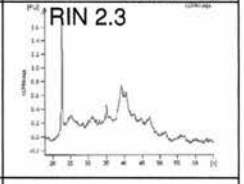
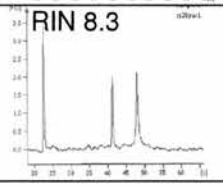
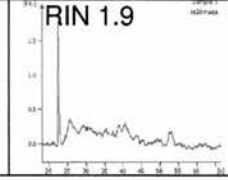
	PAXgene™ (2 hours)	MagaZorb® (EDTA)	MagaZorb® (Li Hep)	MagaZorb® (Na Citr)	MagaZorb® (Fl Oxal)		PAXgene™ (2 hours)	MagaZorb® (EDTA)
cs25	 RIN 9.4	 RIN 1.8	-	-	-	**cs29**	 RIN 9.3	 RIN 2.1
cs26	 RIN 8.5	 RIN 1.4	 RIN 1.7	 RIN 1.1	 RIN 1.0	**cs30**	 RIN N/A	 RIN 2.9
cs27	 RIN 9.0	 RIN 1.0	 RIN 2.1	 RIN 1.7	 RIN 1.0	**cs34**	 RIN 8.6	 RIN 2.3
cs28	 RIN 8.3	 RIN 1.9	-	-	-		-	-

Table 3.4.4: ***Bioanalyzer graphs and RIN values of RNA extracted using 2.5 ml of blood in PAXgene™ tube, 0.5 ml of blood in PAXgene™ tube or 0.5 ml of blood in an aliquot of PAXgene™ stabilization agent.***

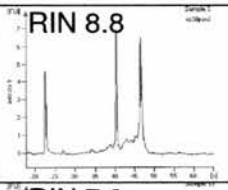
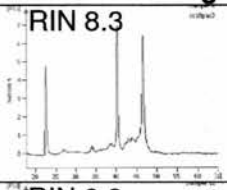
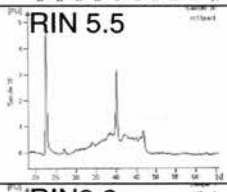
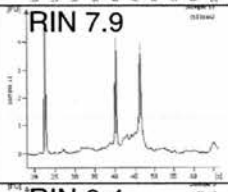
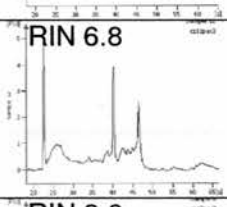
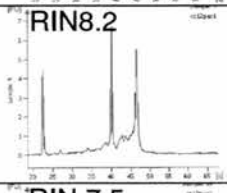
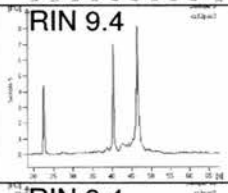
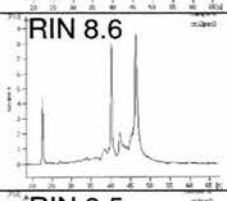
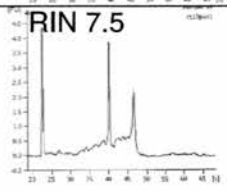
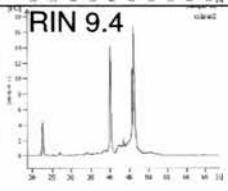
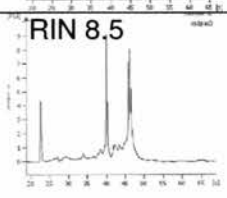
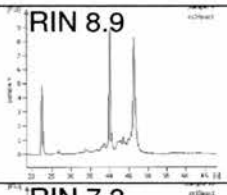
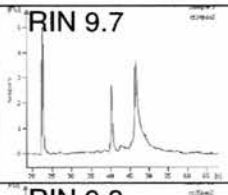
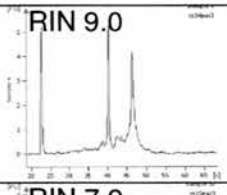
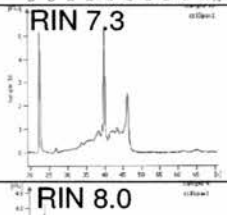
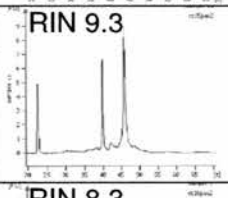
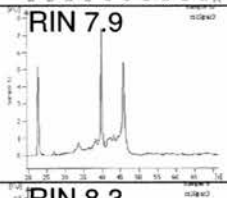
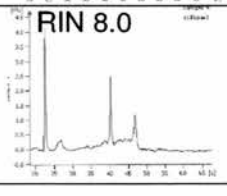
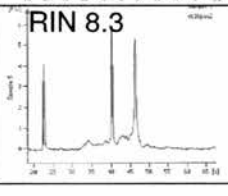
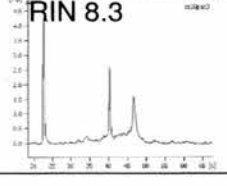
sample	**2.5 ml in PAXgene™ tube**	**0.5 ml in PAXgene™ tube**	**0.5 ml in aliquot of PAXgene™ stabilization agent**
cs10	 RIN 7.1	 RIN 8.8	 RIN 8.3
cs11	 RIN 5.5	 RIN 7.9	 RIN 6.8
cs12	 RIN8.2	 RIN 9.4	 RIN 8.6
cs13	 RIN 7.5	 RIN 9.4	 RIN 8.5
cs14	 RIN 8.9	 RIN 9.7	 RIN 9.0
cs15	 RIN 7.3	 RIN 9.3	 RIN 7.9
cs16	 RIN 8.0	 RIN 8.3	 RIN 8.3

Table 3.5.1: *Raw values for samples according to incubation time before RNA extraction: 2,24 or 72 hours.*

Sample	Hours Incubation	260/280 ratio	A260	Total RNA (μg)	RIN
cs19	2	1.3	0.026	3.43	9.5
cs19	24	1.66	0.068	8.98	8.6
cs19	72	1.47	0.044	5.81	7.9
cs20	2	1.41	0.055	7.26	8.8
cs20	24	1.47	0.047	6.20	7.7
cs20	72	1.35	0.042	5.54	7.3
cs21	2	1.13	0.026	3.43	7.6
cs21	24	1.4	0.049	6.47	6.8
cs21	72	1.28	0.037	4.88	n/a
cs22	2	2.16	0.067	8.84	8.6
cs22	24	1.19	0.259	34.19	7.2
cs22	72	2.04	0.051	6.73	8.1
cs23	2	1.62	0.021	2.77	9
cs23	24	2	0.038	5.02	8.7
cs23	72	1.87	0.043	5.68	7.6
cs24	2	1.34	0.047	6.20	9.9
cs24	24	1.71	0.036	4.75	9.1
cs24	72	1.85	0.037	4.88	8.6
cs28	2	2	0.016	2.11	8.3
cs28	24	2	0.026	3.43	8.5
cs28	72	1.46	0.019	2.51	7.6
cs29	2	1.92	0.023	3.04	9.3
cs29	24	2.43	0.017	2.24	7.3
cs29	72	2.16	0.041	5.41	7.8
cs30	2	2.29	0.016	2.11	n/a
cs30	24	2.36	0.026	3.43	9.4
cs30	72	2.5	0.015	1.98	6.8
cs31	2	2.67	0.008	1.06	9.4
cs31	24	4	0.008	1.06	8.4
cs31	72	1.67	0.01	1.32	7.8
cs32	2	2	0.014	1.85	8.6
cs32	24	1.87	0.028	3.70	8.5
cs32	72	1.26	0.024	3.17	8.2
cs33	2	av.1.26	av.0.020	2.64	9.4
cs33	24	av.1.49	av.0.017	2.24	8
cs33	72	av.1.32	av.0.034	4.49	7.8

Table 3.5.4a: ***Bioanalyzer graphs and RIN values for RNA extracted at 2, 24 and 72 hours.***

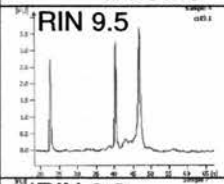
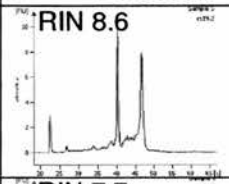
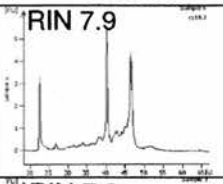
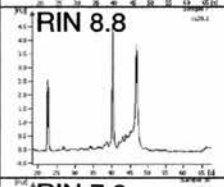
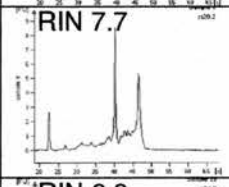
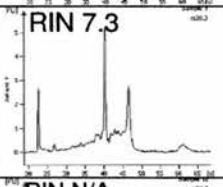
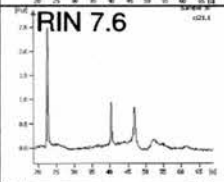
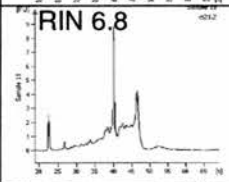
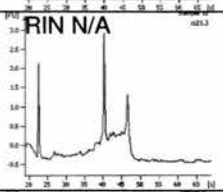
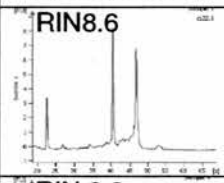
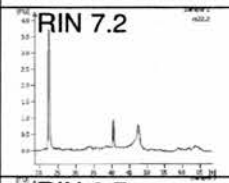
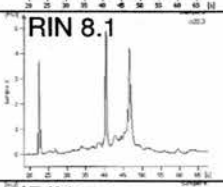
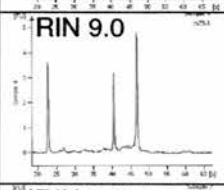
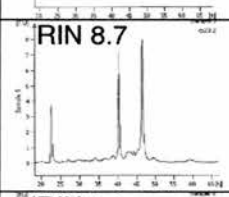
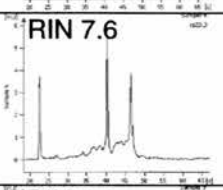
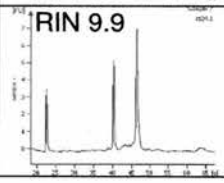
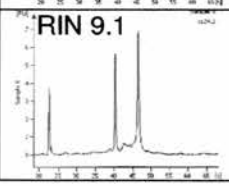
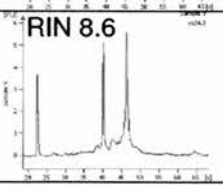
Sample	2 hours	24 hours	72 hours
cs19	 RIN 9.5	 RIN 8.6	 RIN 7.9
cs20	 RIN 8.8	 RIN 7.7	 RIN 7.3
cs21	 RIN 7.6	 RIN 6.8	 RIN N/A
cs22	 RIN8.6	 RIN 7.2	 RIN 8.1
cs23	 RIN 9.0	 RIN 8.7	 RIN 7.6
cs24	 RIN 9.9	 RIN 9.1	 RIN 8.6

Table 3.5.4b: ***Bioanalyzer graphs and RIN values for RNA extracted at 2, 24 and 72 hours.***

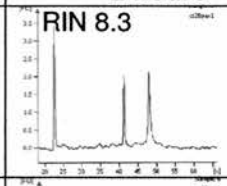
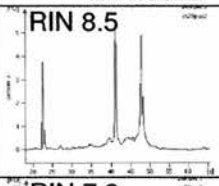
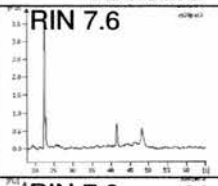
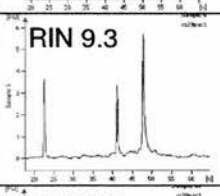
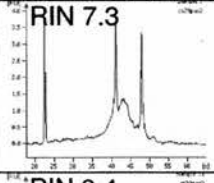
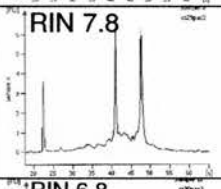
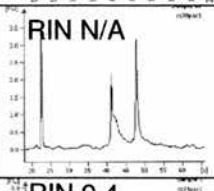
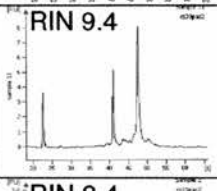
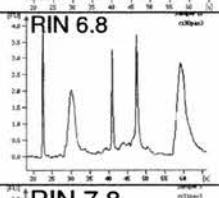
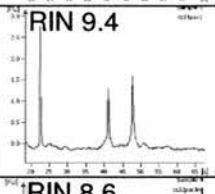
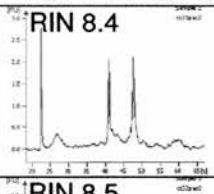
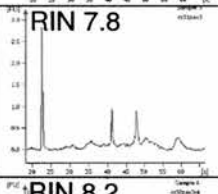
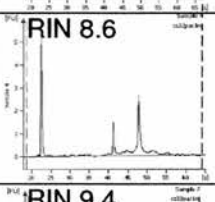
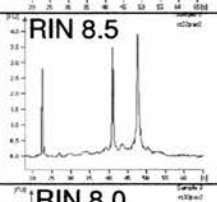
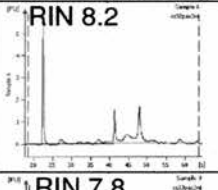
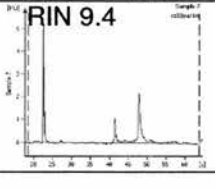
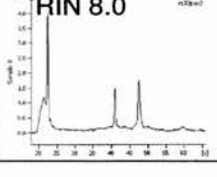
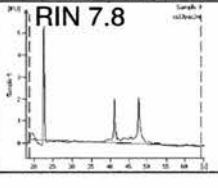
sample	2 hours	24 hours	72 hours
cs28	 RIN 8.3	 RIN 8.5	 RIN 7.6
cs29	 RIN 9.3	 RIN 7.3	 RIN 7.8
cs30	 RIN N/A	 RIN 9.4	 RIN 6.8
cs31	 RIN 9.4	 RIN 8.4	 RIN 7.8
cs32	 RIN 8.6	 RIN 8.5	 RIN 8.2
cs33	 RIN 9.4	 RIN 8.0	 RIN 7.8

Table 3.6.1a: *Raw data for each sample according to incubation period prior to RNA extraction: 2,4,6,8,10 or 12 hours.*

Sample	Incubation time (hours)	260/280 Ratio	A260	Total RNA (μg)	RIN
cs34	2	1.31	0.017	2.24	8.6
cs34	4	2.33	0.021	2.77	9
cs34	6	2.18	0.024	3.17	8.6
cs34	8	2.08	0.025	3.3	9.1
cs34	10	1.91	0.021	2.77	9.4
cs34	12	1.69	0.022	2.90	9.7
cs35	2	1.44	0.013	1.72	9
cs35	4	1.8	0.018	2.38	9.5
cs35	6	2	0.018	2.38	9.8
cs35	8	1.79	0.025	3.3	9.3
cs35	10	1.83	0.022	2.90	8.9
cs35	12	2.67	0.016	2.11	8.6
cs36	2	2	0.016	2.11	9.4
cs36	4	1.7	0.017	2.24	8.8
cs36	6	2.5	0.02	2.64	9.6
cs36	8	1.81	0.029	3.83	9.3
cs36	10	2.17	0.013	1.72	8.2
cs36	12	1.79	0.025	3.3	7.8
cs37	2	2	0.01	1.32	8.1
cs37	4	2.2	0.011	1.45	8.8
cs37	6	2	0.012	1.58	8.1
cs37	8	1.83	0.011	1.45	8.4
cs37	10	1.92	0.025	3.3	9.2
cs37	12	1.7	0.017	2.24	8
cs38	2	2	0.026	3.43	9.4
cs38	4	1.24	0.079	10.43	9.4
cs38	6	1.85	0.024	3.17	8.7
cs38	8	2	0.026	3.43	8.3
cs38	10	1.6	0.04	5.28	8.9
cs38	12	1.14	0.124	16.37	9.1

Table 3.6.1b: *Raw data for each sample according to incubation period prior to RNA extraction: 2,4,6,8,10 or 12 hours.*

Sample	Incubation time (hours)	260/280 Ratio	A260	Total RNA (μg)	RIN
cs39	2	1.78	0.016	2.11	8.7
cs39	4	1.5	0.012	1.58	n/a
cs39	6	1.67	0.015	1.98	9.3
cs39	8	3	0.015	1.98	9.2
cs39	10	1.8	0.018	2.38	8.1
cs39	12	2.33	0.014	1.85	10
cs40	2	4	0.008	1.06	9
cs40	4	3	0.012	1.58	8.7
cs40	6	1.93	0.029	3.83	9.4
cs40	8	2	0.018	2.38	8.7
cs40	10	2.1	0.021	2.77	8.7
cs40	12	1.87	0.028	3.70	8.6
cs41	2	2.07	0.029	3.83	9.4
cs41	4	1.8	0.027	3.56	10
cs41	6	1.73	0.038	5.02	9.7
cs41	8	1.93	0.029	3.83	9.8
cs41	10	2.14	0.03	3.96	8.4
cs41	12	1.56	0.028	3.70	9.6
cs42	2	1.78	0.016	2.11	9.9
cs42	4	1.68	0.037	4.88	9.9
cs42	6	2.07	0.029	3.83	9.7
cs42	8	1.67	0.025	3.3	9.3
cs42	10	1.78	0.016	2.11	n/a
cs42	12	2.29	0.032	4.22	9.6
cs43	2	1.91	0.18	23.76	n/a
cs43	4	1.9	0.156	20.59	6.9
cs43	6	1.9	0.129	17.03	7.1
cs43	8	1.9	0.131	17.29	6.5
cs43	10	1.86	0.19	25.08	n/a
cs43	12	1.84	0.171	22.57	4.9

Table 3.6.4a: *Bioanalyzer graphs and RIN values for RNA extracted at 2,4,6,8,10 &12 hours.*

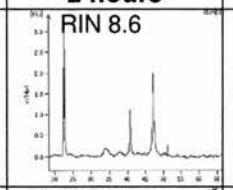
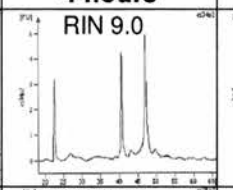
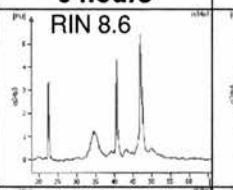
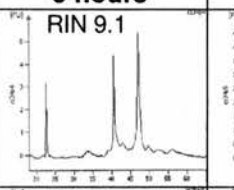
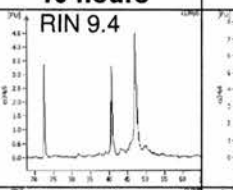
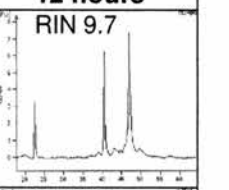
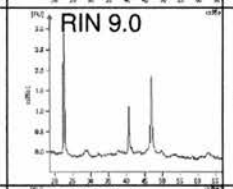
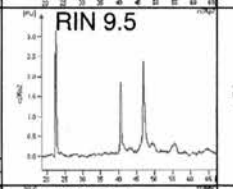
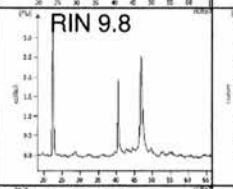
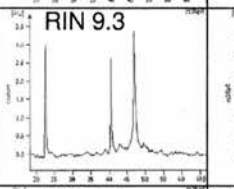
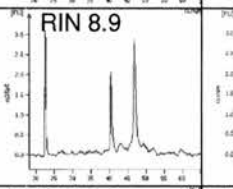
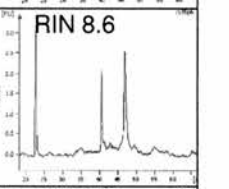
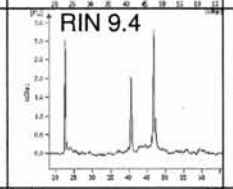
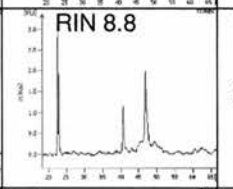
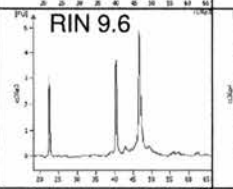
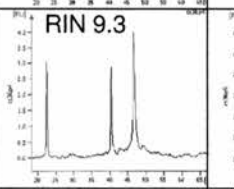
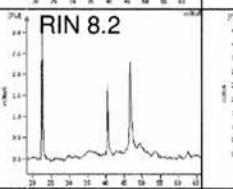
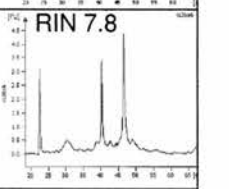
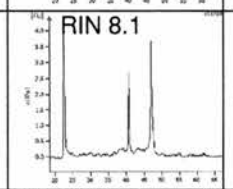
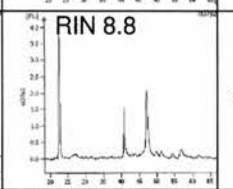
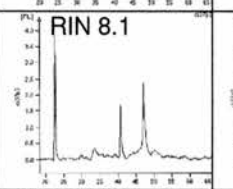
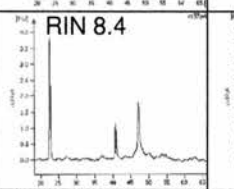
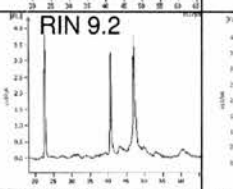
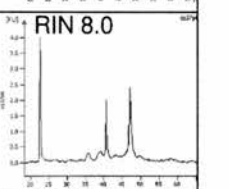
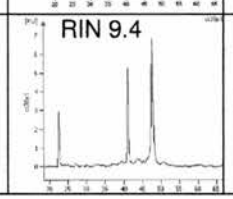
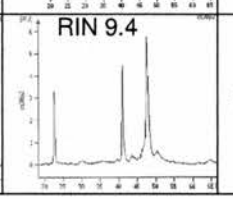
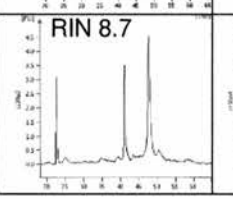
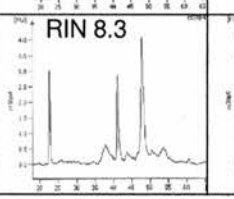
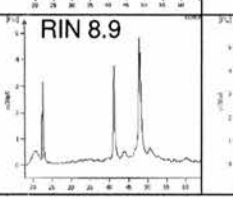
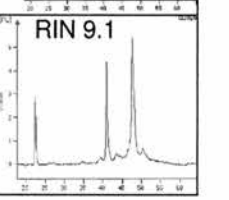
	2 hours	4 hours	6 hours	8 hours	10 hours	12 hours
cs34	 RIN 8.6	 RIN 9.0	 RIN 8.6	 RIN 9.1	 RIN 9.4	 RIN 9.7
cs35	 RIN 9.0	 RIN 9.5	 RIN 9.8	 RIN 9.3	 RIN 8.9	 RIN 8.6
cs36	 RIN 9.4	 RIN 8.8	 RIN 9.6	 RIN 9.3	 RIN 8.2	 RIN 7.8
cs37	 RIN 8.1	 RIN 8.8	 RIN 8.1	 RIN 8.4	 RIN 9.2	 RIN 8.0
38	 RIN 9.4	 RIN 9.4	 RIN 8.7	 RIN 8.3	 RIN 8.9	 RIN 9.1

Table 3.6.4b: ***Bioanalyzer graphs and RIN values for RNA extracted at 2,4,6,8,10 &12 hours.***

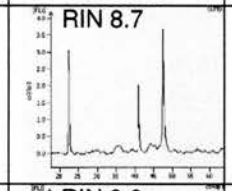
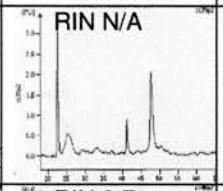
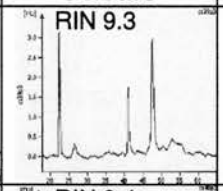
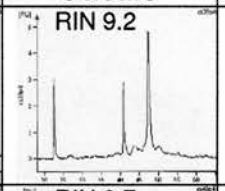
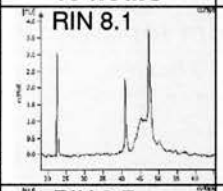
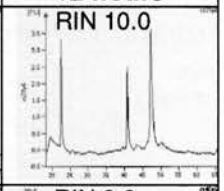
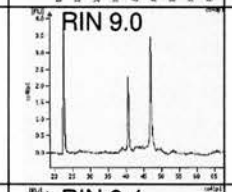
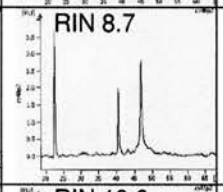
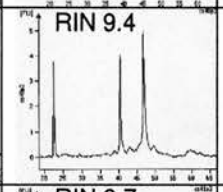
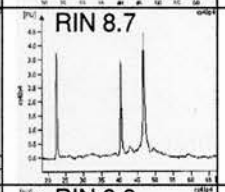
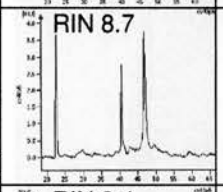
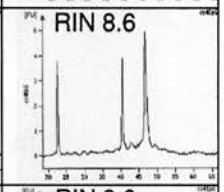
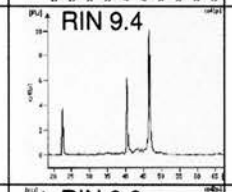
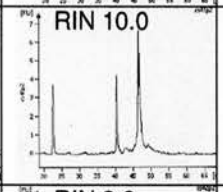
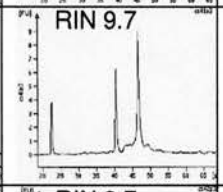
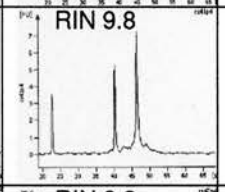
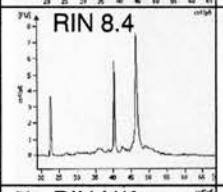
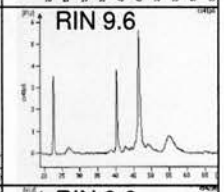
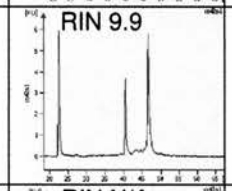
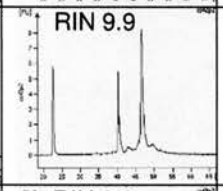
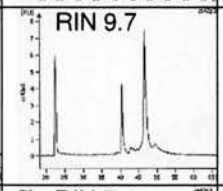
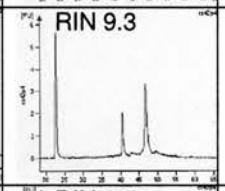
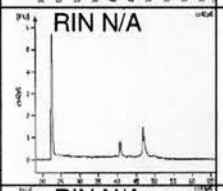
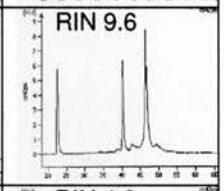
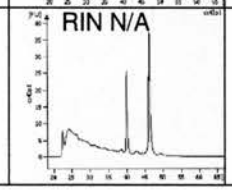
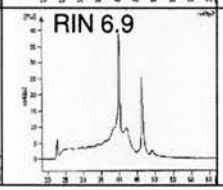
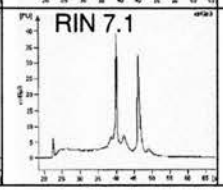
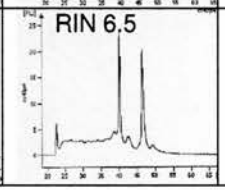
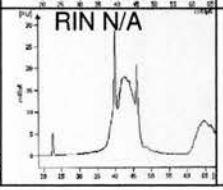
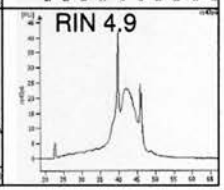
	2 hours	4 hours	6 hours	8 hours	10 hours	12 hours
cs39	 RIN 8.7	 RIN N/A	 RIN 9.3	 RIN 9.2	 RIN 8.1	 RIN 10.0
cs40	 RIN 9.0	 RIN 8.7	 RIN 9.4	 RIN 8.7	 RIN 8.7	 RIN 8.6
cs41	 RIN 9.4	 RIN 10.0	 RIN 9.7	 RIN 9.8	 RIN 8.4	 RIN 9.6
cs42	 RIN 9.9	 RIN 9.9	 RIN 9.7	 RIN 9.3	 RIN N/A	 RIN 9.6
cs43	 RIN N/A	 RIN 6.9	 RIN 7.1	 RIN 6.5	 RIN N/A	 RIN 4.9

Table 3.7.4a: ***Bioanalyzer graphs and RIN values of RNA extracted after 2 hours incubation at room temperature and after 15 hours, one day, three days, four days and seven days with incubation at -20 °C including 4½ hours defrosting/incubation at room temperature.***

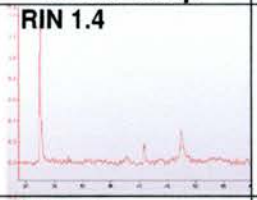
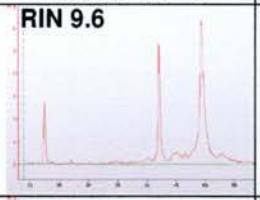
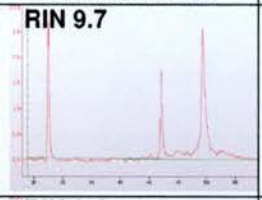
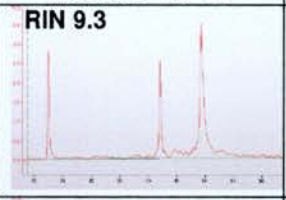
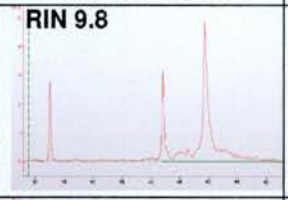
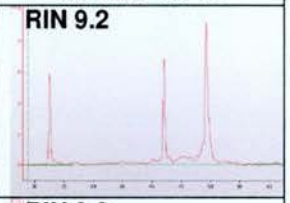
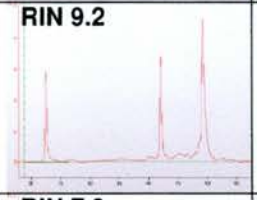
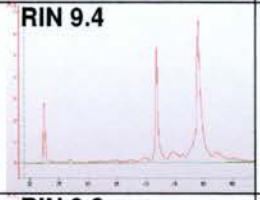
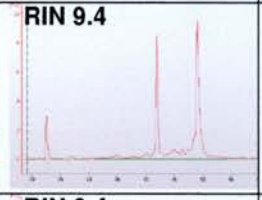
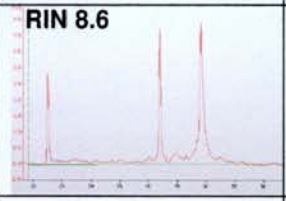
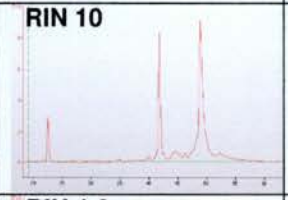
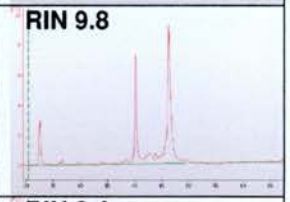
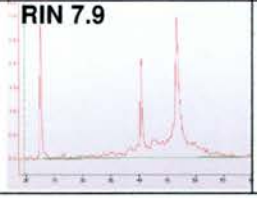
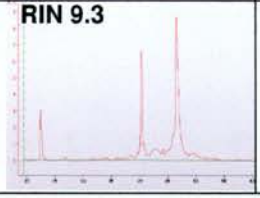
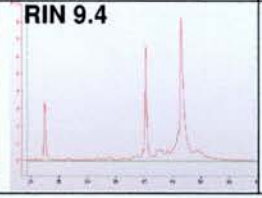
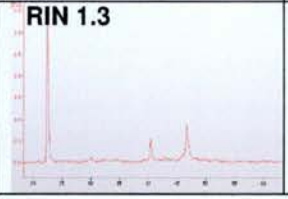
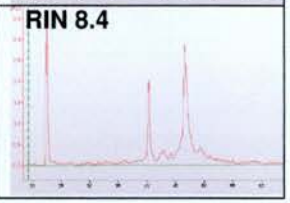
	2 hours Room Temp	15 hours Frozen -20 °C	1 day Frozen -20 °C	3 days Frozen -20 °C	4 days Frozen -20 °C	7 days Frozen -20 °C
cs47	RIN 1.4 	RIN 9.6 	RIN 9.7 	RIN 9.3 	RIN 9.8 	RIN 9.2 
cs48	RIN 9.2 	RIN 9.4 	RIN 9.4 	RIN 8.6 	RIN 10 	RIN 9.8 
cs49	RIN 7.9 	RIN 9.3 	RIN 9.4 	-	RIN 1.3 	RIN 8.4 

Table 3.7.4b: ***Bioanalyzer graphs and RIN values of RNA extracted after 2 hours incubation at room temperature and after 15 hours, one day, three days, four days and seven days with incubation at -20 °C including 4½ hours defrosting/incubation at room temperature.***

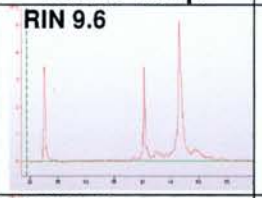
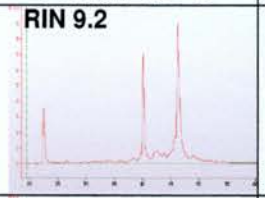
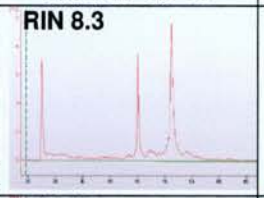
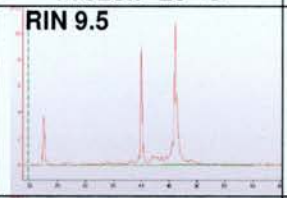
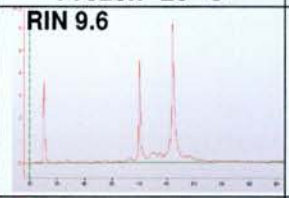
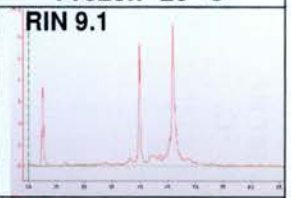
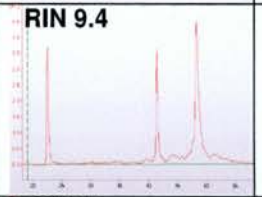
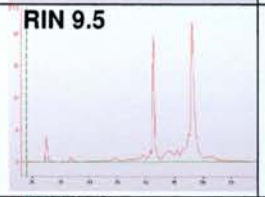
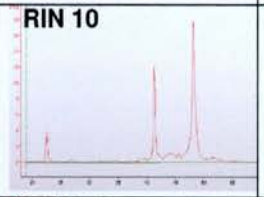
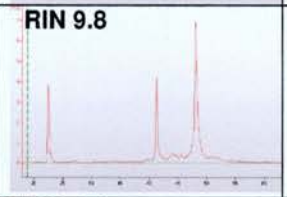
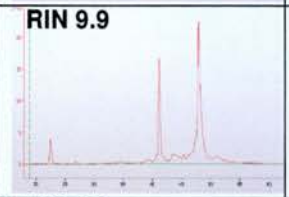
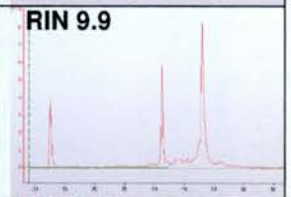
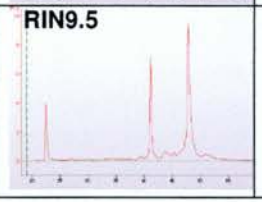
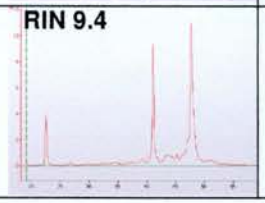
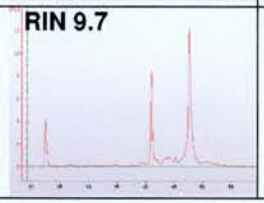
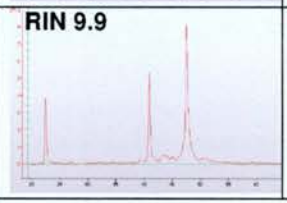
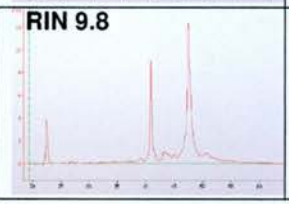
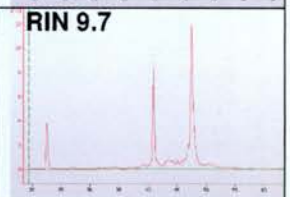
	2 hours Room Temp	15 hours Frozen -20 °C	1 day Frozen -20 °C	3 days Frozen -20 °C	4 days Frozen -20 °C	7 days Frozen -20 °C
cs50	RIN 9.6 	RIN 9.2 	RIN 8.3 	RIN 9.5 	RIN 9.6 	RIN 9.1 
cs51	RIN 9.4 	RIN 9.5 	RIN 10 	RIN 9.8 	RIN 9.9 	RIN 9.9 
cs52	RIN9.5 	RIN 9.4 	RIN 9.7 	RIN 9.9 	RIN 9.8 	RIN 9.7 

Table 3.8.4a: ***Bioanalyzer graphs and RIN values for RNA extracted after storage at 2 hours at room temperature, 1 week at -80 °C, 1 year at -80 °C, 1 week at -20 °C or 2 years at -20 °C.***

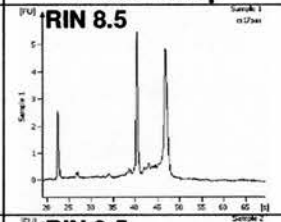
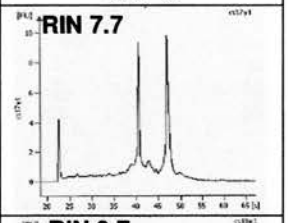
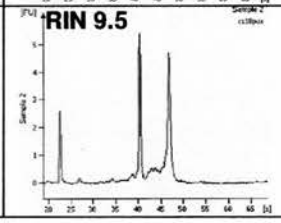
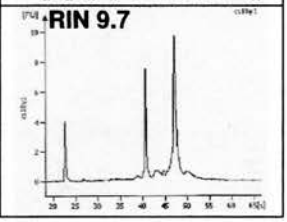
	2 hours Room Temp	1 year -80 °C
cs17		
cs18		

Table 3.8.4b can be found on the next page. Please note that the column headings for table 3.8.4b differ from those in table 3.8.4a.

Table 3.8.4b: ***Bioanalyzer graphs and RIN values for RNA extracted after storage at 2 hours at room temperature, 1 week at -80 °C, 1 year at -80 °C, 1 week at -20 °C or 2 years at -20 °C.***

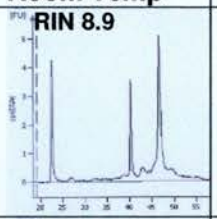
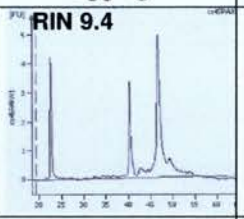
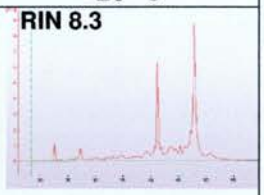
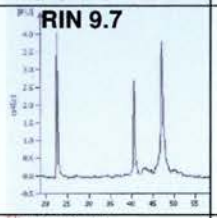
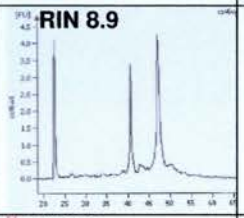
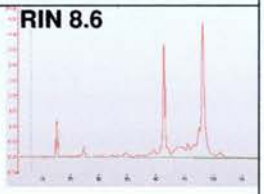
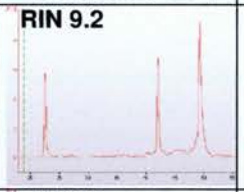
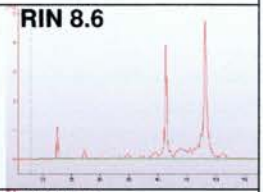
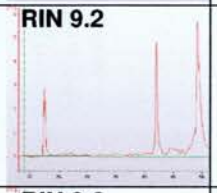
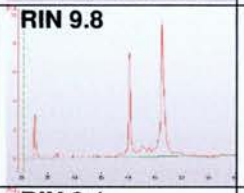
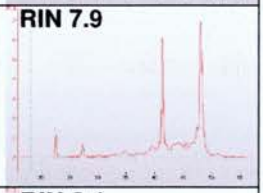
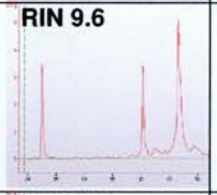
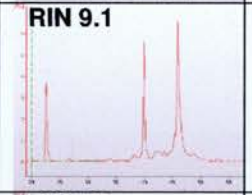
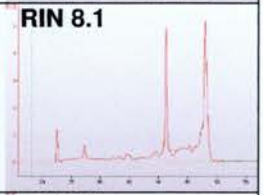
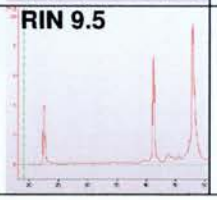
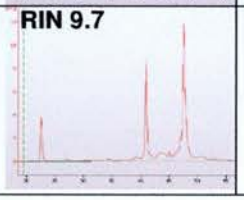
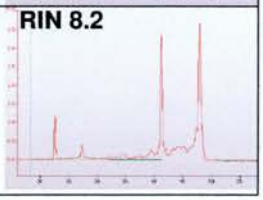
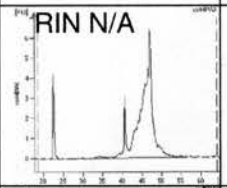
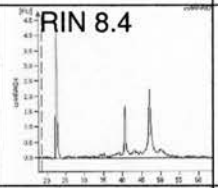
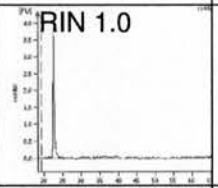
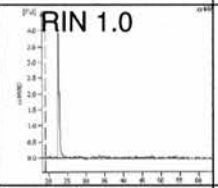
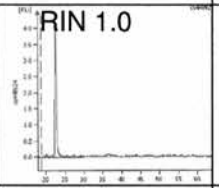
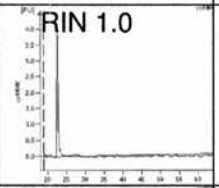
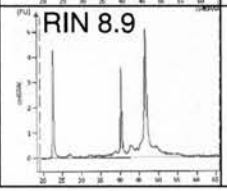
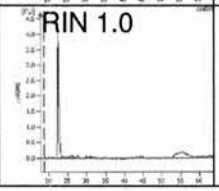
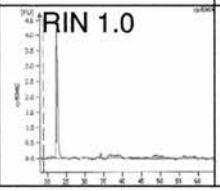
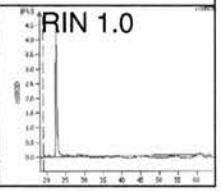
	2 hours Room Temp	1 week -80 °C	1 week -20 °C	2 years -20 °C
cs45	 RIN 8.9	 RIN 9.4	-	 RIN 8.3
cs46	 RIN 9.7	-	 RIN 8.9	 RIN 8.6
cs47	 RIN 1.4	-	 RIN 9.2	 RIN 8.6
cs48	 RIN 9.2	-	 RIN 9.8	 RIN 7.9
cs50	 RIN 9.6	-	 RIN 9.1	 RIN 8.1
cs52	 RIN 9.5	-	 RIN 9.7	 RIN 8.2

Table 3.9.4: *Bioanalyzer graphs and RIN values for RNA extracted using Whatman® technique compared with PAXgene™*

	PAXgene™ 2 hr	PAXgene™ 24 hr	Whatman® whole disc	Whatman® whole disc DNase	Whatman® whole disc 24 hr	Whatman® 1 punch (black)	Whatman® 2 punches (black)	Whatman® 2 punches (black) DNase
cs44	RIN N/A 	RIN 8.4 	RIN 1.0 	RIN 1.0 	RIN 1.0 	RIN 1.0 	-	-
cs45	RIN 8.9 	-	-	-	-	RIN 1.0 	RIN 1.0 	RIN 1.0 

Appendix 3

This section contains tables of Bioanalyzer traces and RIN values for the RNA of the neonatal blood samples described in chapter 4.

Table 4.1.3a: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

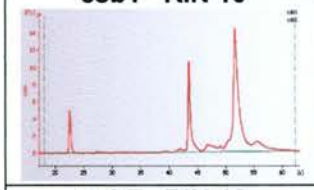
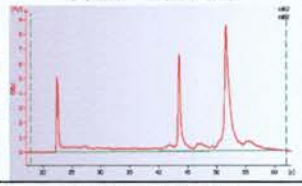
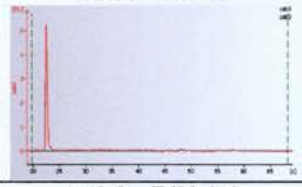
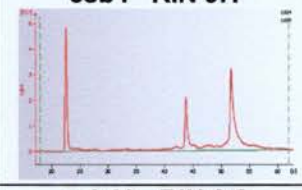
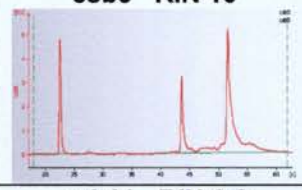
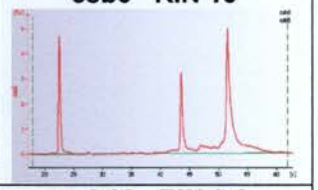
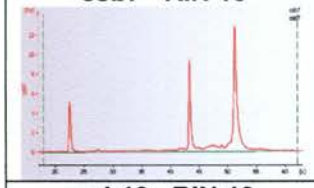
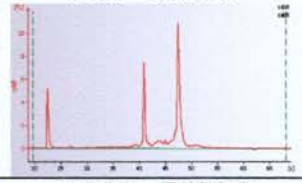
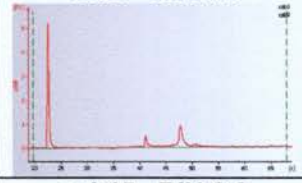
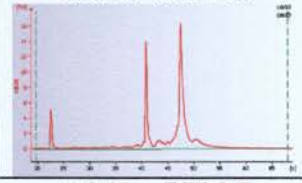
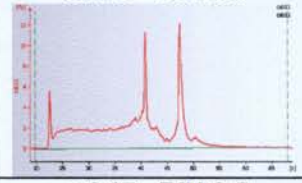
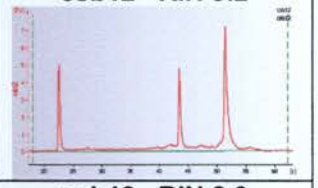
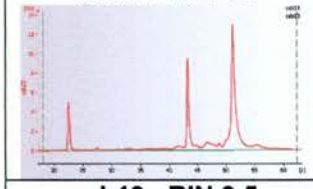
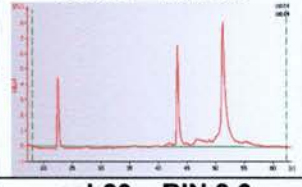
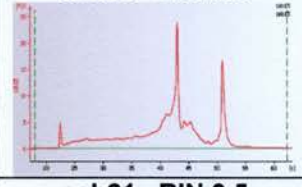
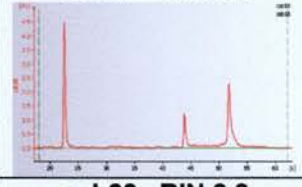
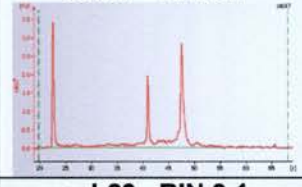
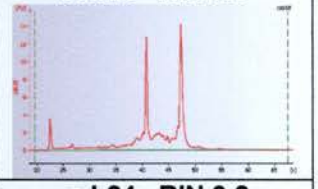
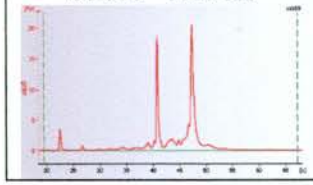
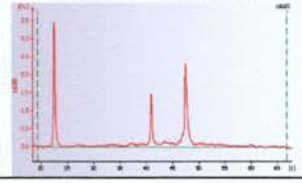
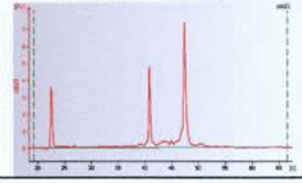
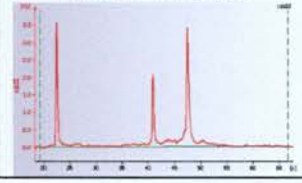
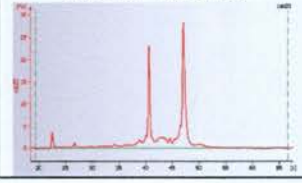
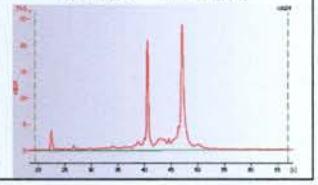
csb1 RIN 10 	csb2 RIN 9.2 	csb3 RIN 1 	csb4 RIN 9.1 	csb5 RIN 10 	csb6 RIN 10 
csb7 RIN 10 	csb8 RIN 9.5 	csb9 RIN 1.1 	csb10 RIN 9.8 	csb11 RIN 6.6 	csb12 RIN 9.2 
csb13 RIN 10 	csb14 RIN 9.9 	csb15 RIN 6.6 	csb16 RIN 9.5 	csb17 RIN 8.5 	csb18 RIN 8.3 
csb19 RIN 9.5 	csb20 RIN 8.6 	csb21 RIN 9.5 	csb22 RIN 8.8 	csb23 RIN 9.1 	csb24 RIN 9.2 

Table 4.1.3b: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

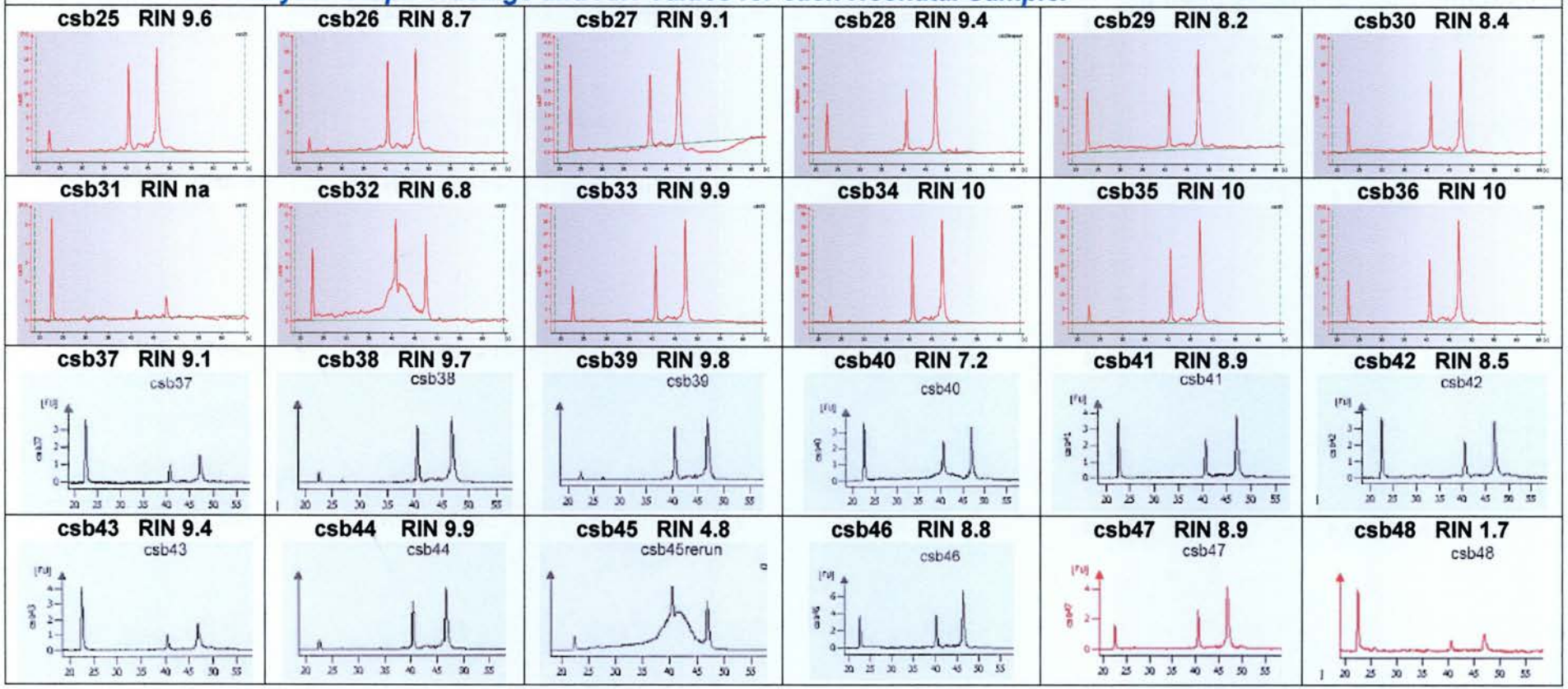
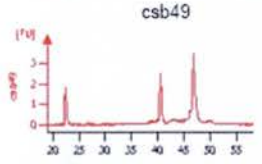


Table 4.1.3c: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

csb49 RIN 9	csb50 RIN 8.4	csb51 RIN 9.2	csb52 RIN na	csb53 RIN 9.4	csb54 RIN 8.2
csb55 RIN 9.4	csb56 RIN na	csb57 RIN 7.4	csb58 RIN 9.3	csb59 RIN 9.5	csb60 RIN 9.4
csb61 RIN 9.3	csb62 RIN 9.4	csb63 RIN 9.1	csb64 RIN 9.2	csb65 RIN 9.3	csb66 RIN 4.2
csb67 RIN 9.4	csb68 RIN 9.4	csb69 RIN 9.6	csb70 RIN 8.8	csb71 RIN 7.2	csb72 RIN 9.6



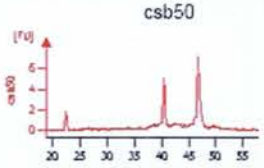


Table 4.1.3d: ***Bioanalyzer Graph Tracings and RIN Value for each Neonatal Sample***

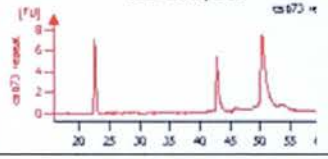
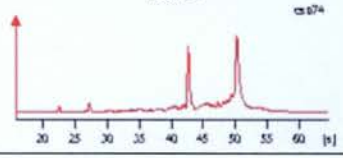
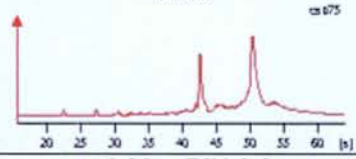
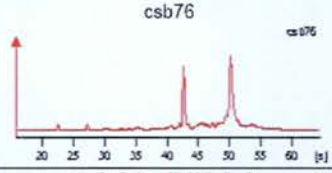
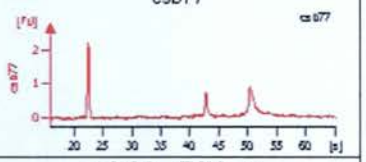
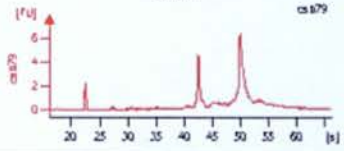
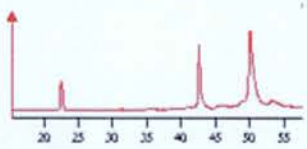
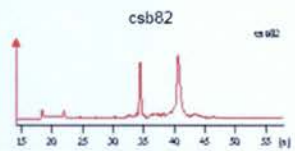
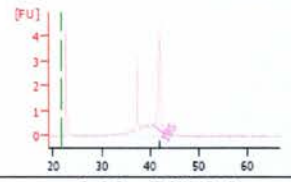
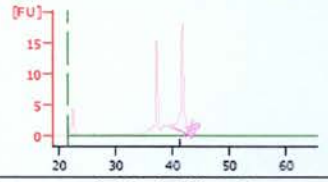
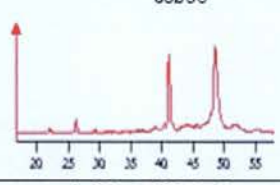
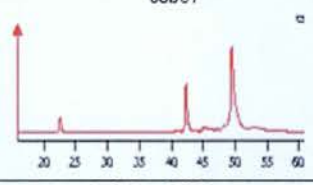
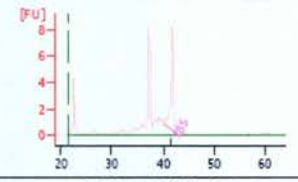
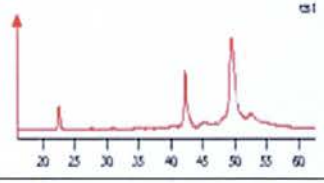
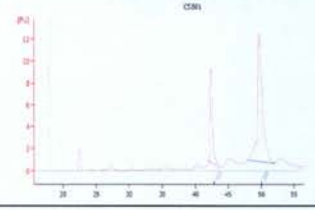
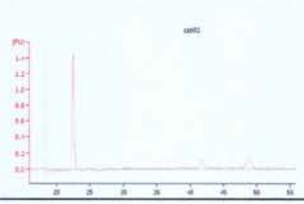
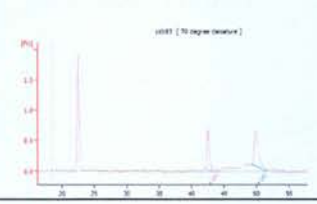
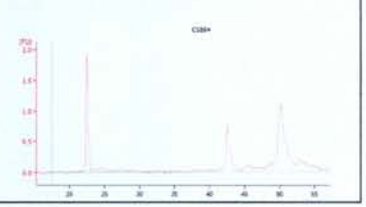
csb73 RIN 9.6 csb73 repeat 	**csb74 RIN 8.3** csb74 	**csb75 RIN 8.7** csb75 	**csb76 RIN 9** csb76 	**csb77 RIN 2.4** csb77 
csb78 RIN 1 csb78 	**csb79 RIN 9.3** csb79 	**csb80 RIN 9.8** csb80 	**csb81 RIN 8.8** csb81 	**csb82 RIN na** csb82 
csb83 RIN na 	**csb84 RIN na** 	**csb86 RIN 8.8** csb86 	**csb87 RIN 10** csb87 	**csb88 RIN na** 
csb89 RIN 9.6 csb89 	**csb91 RIN 9.4** 	**csb92 RIN 1.1** 	**csb93 RIN na** 	**csb94 RIN 9.3** 

Table 4.1.3e: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

csb95 RIN 8.6	csb96 RIN 8.6	csb97 RIN 7.9	csb98 RIN 2.2	csb99 RIN na
csb100 RIN 8.7	csb101 RIN 9.1	csb102 RIN 9.1	csb103 RIN 9.1	csb104 RIN 2.1
csb105 RIN na	csb106 RIN na	csb107 RIN 5	csb108 RIN 5.9	csb109 RIN 8.9
csb110 RIN 8.4	csb111 RIN 2.5	csb112 RIN 7.1	csb113 RIN 9.1	csb114 RIN 7.9

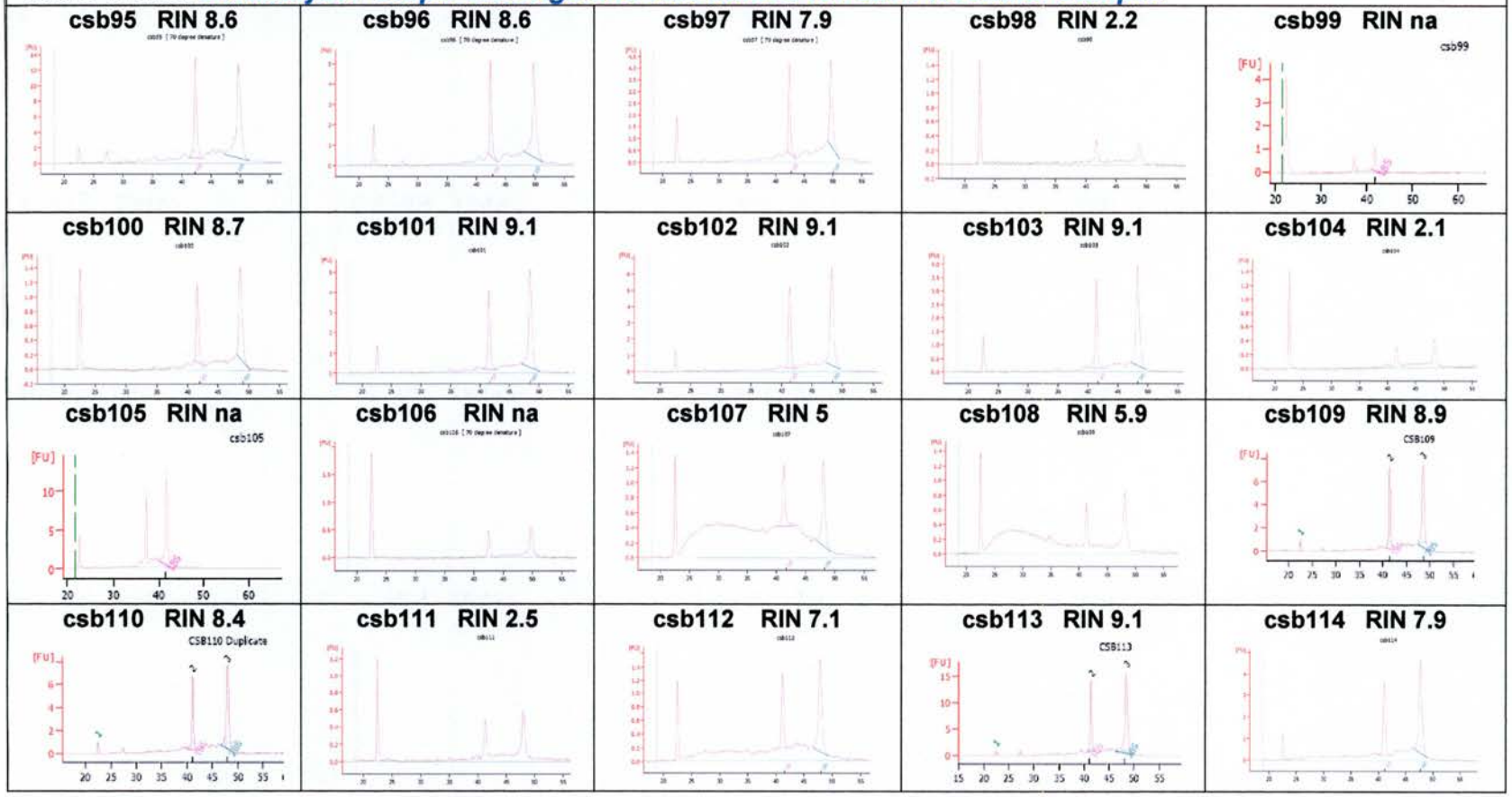


Table 4.1.3f: ***Bioanalyser Graph Tracings and RIN Values for each Neonatal Sample.***

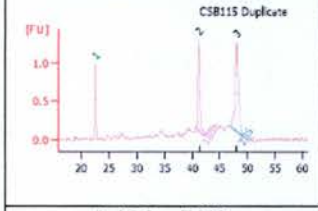
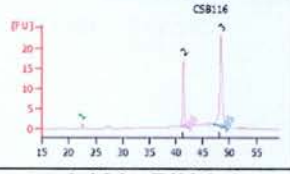
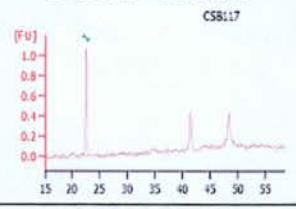
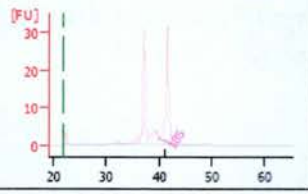
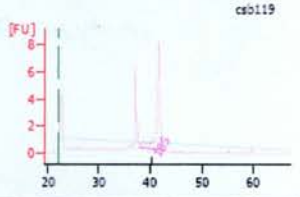
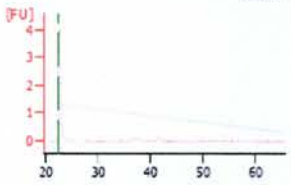
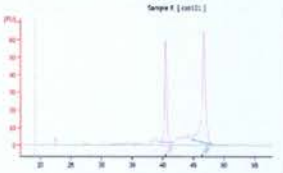
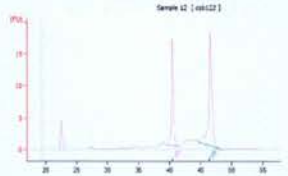
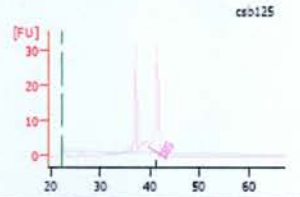
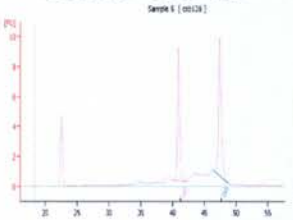
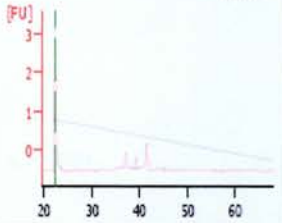
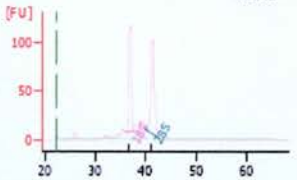
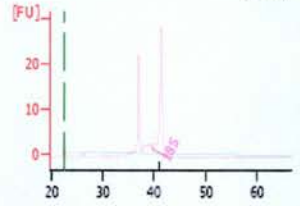
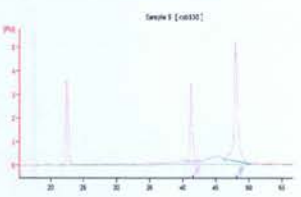
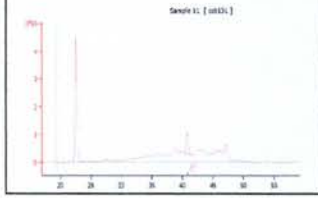
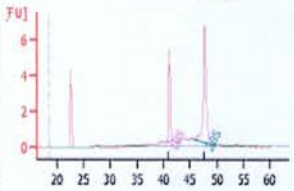
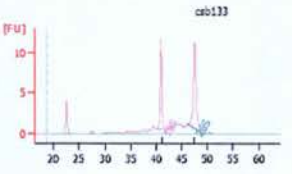
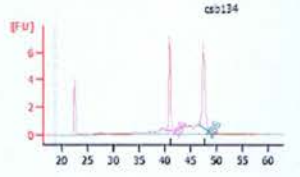
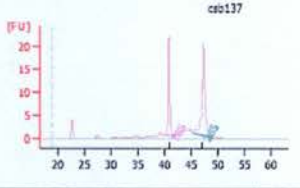
csb115 RIN 7.8	**csb116 RIN 9.7**	**csb117 RIN na**	**csb118 RIN na**	**csb119 RIN na**
				
csb120 RIN na	**csb121 RIN 9.4**	**csb122 RIN 9.1**	**csb124 RIN 8.9**	**csb125 RIN na**
				
csb126 RIN 8.8	**csb127 RIN na**	**csb128 RIN 8.8**	**csb129 RIN na**	**csb130 RIN 9.6**
				
csb131 RIN 4.7	**csb132 RIN 9.5**	**csb133 RIN 8.8**	**csb134 RIN 8.9**	**csb137 RIN 9.3**
				

Table 4.1.3g: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

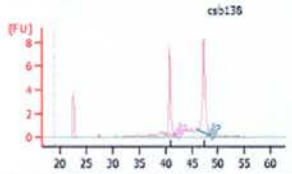
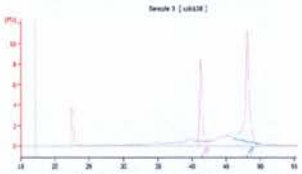
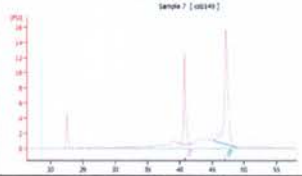
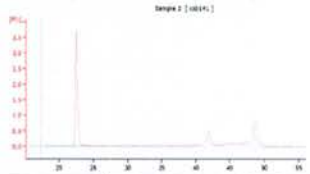
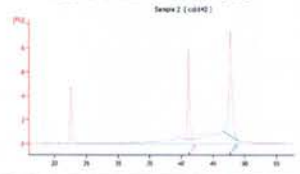
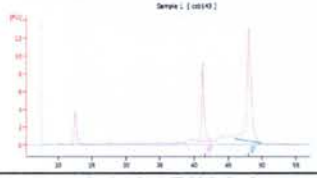
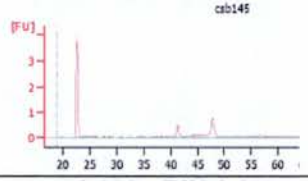
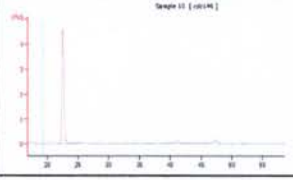
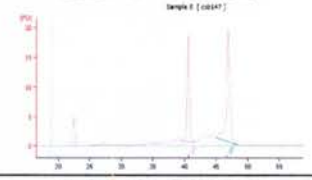
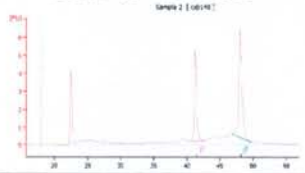
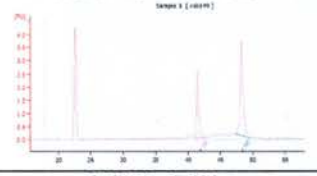
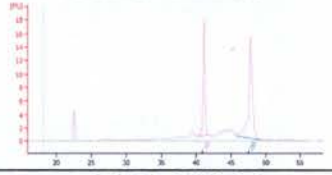
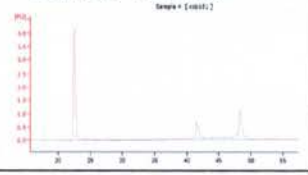
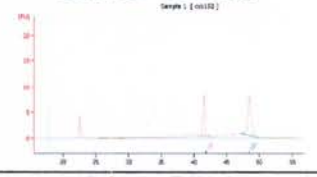
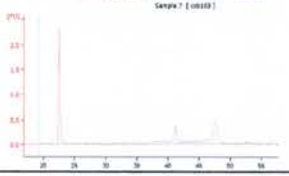
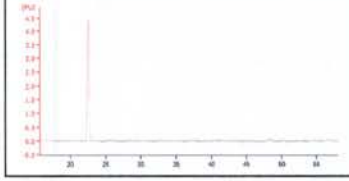
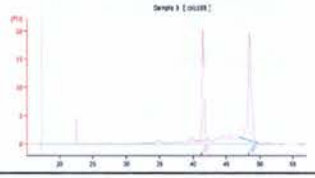
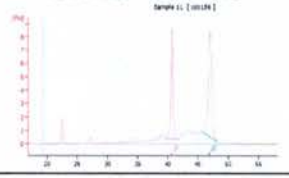
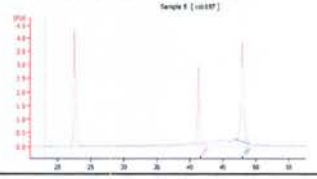
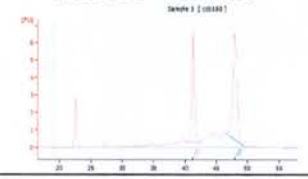
csb138 RIN 9.3 	**csb139 RIN 9.1** 	**csb140 RIN 9.2** 	**csb141 RIN 7.6** 	**csb142 RIN 8.9** 
csb143 RIN 9.6 	**csb145 RIN na** 	**csb146 RIN na** 	**csb147 RIN 8.9** 	**csb148 RIN 8.7** 
csb149 RIN 9.6 	**csb150 RIN 8.9** 	**csb151 RIN na** 	**csb152 RIN 8.7** 	**csb153 RIN na** 
csb154 RIN na 	**csb155 RIN 9.2** 	**csb156 RIN 8.9** 	**csb157 RIN 9.4** 	**csb158 RIN 9.1** 

Table 4.1.3h: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

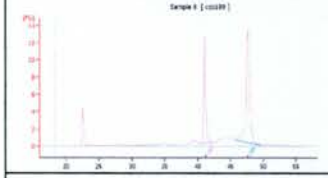
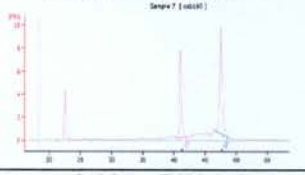
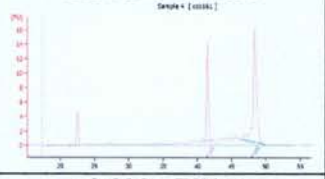
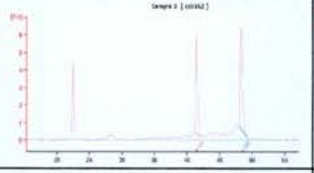
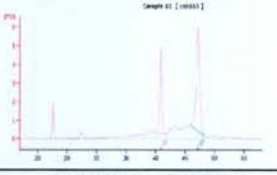
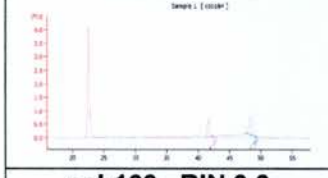
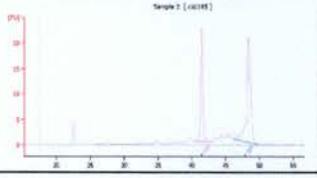
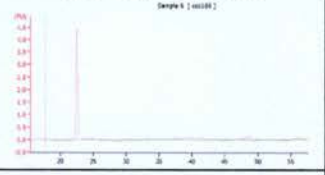
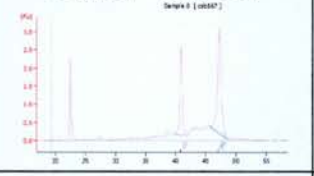
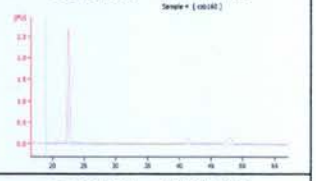
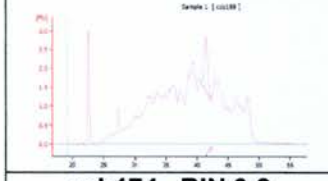
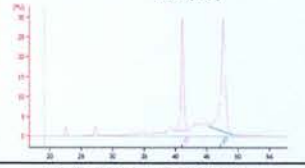
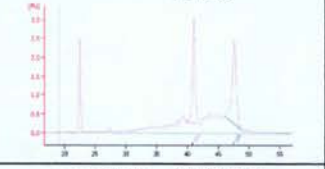
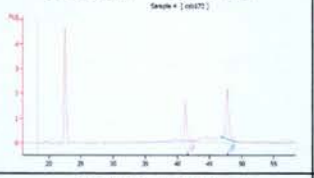
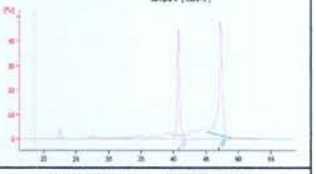
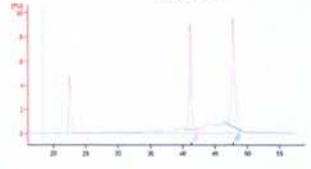
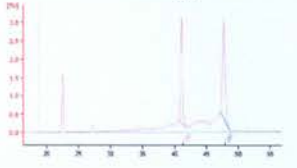
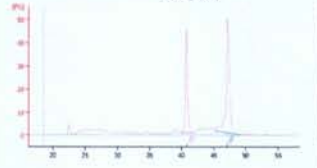
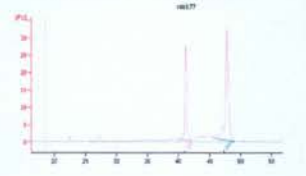
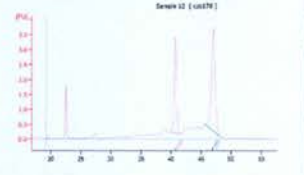
csb159 RIN 9.4	csb160 RIN 9.3	csb161 RIN 9.1	csb162 RIN 8.5	csb163 RIN 8.7
				
csb164 RIN na	csb165 RIN 9.2	csb166 RIN na	csb167 RIN 8.8	csb168 RIN na
				
csb169 RIN 3.3	csb170 RIN 9.1	csb171 RIN 7.9	csb172 RIN 8.7	csb173 RIN 9.2
				
csb174 RIN 9.2	csb175 RIN 7.8	csb176 RIN 8.8	csb177 RIN 9.6	csb178 RIN 8.5
				

Table 4.1.3i: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

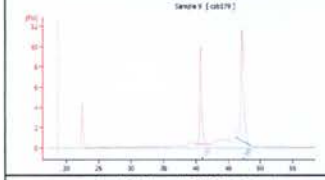
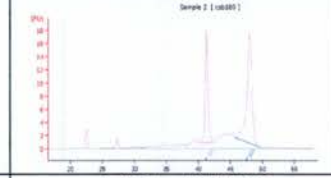
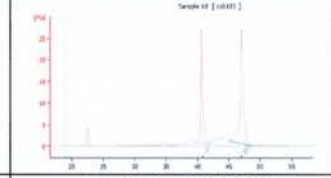
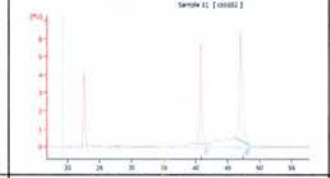
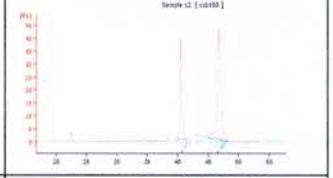
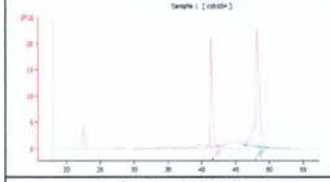
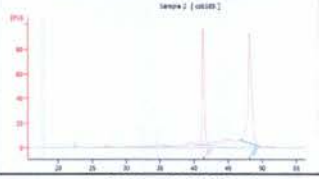
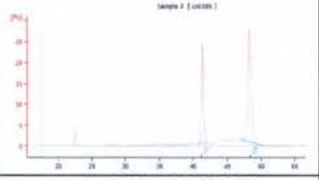
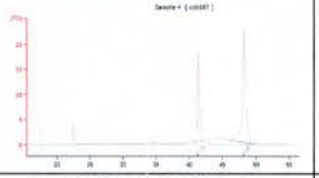
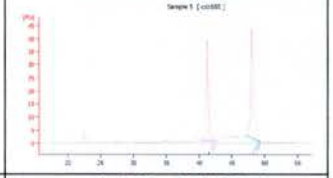
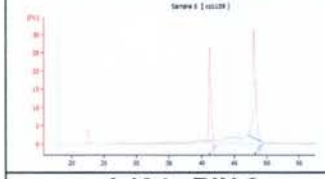
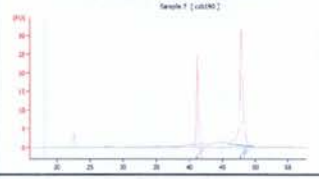
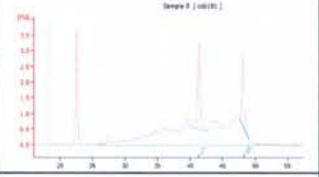
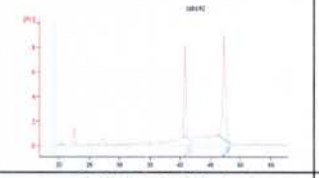
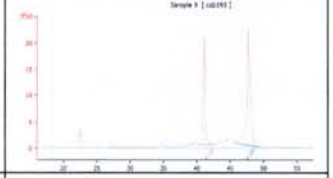
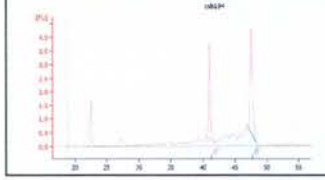
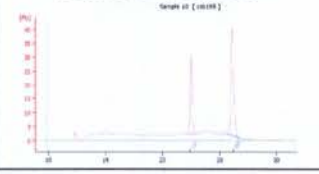
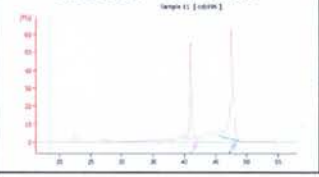
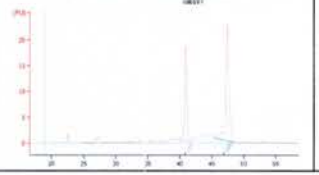
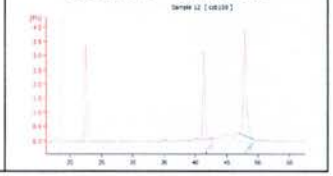
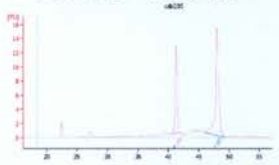
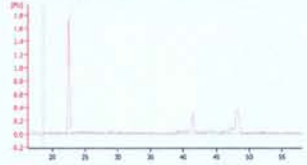
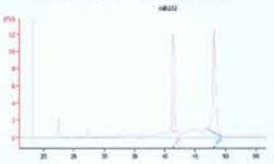
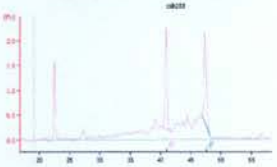
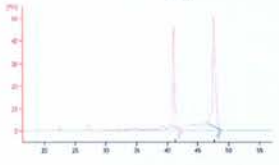
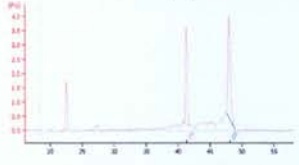
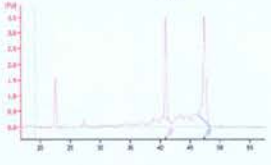
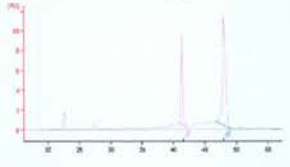
csb179 RIN 9.2 	csb180 RIN 9 	csb181 RIN 9.4 	csb182 RIN 9.4 	csb183 RIN 9.6 
csb184 RIN 9.7 	csb185 RIN 9.3 	csb186 RIN 9.6 	csb187 RIN 9.4 	csb188 RIN 9.3 
csb189 RIN 9.4 	csb190 RIN 9.9 	csb191 RIN 6.5 	csb192 RIN 8.5 	csb193 RIN 9.4 
csb194 RIN 8 	csb195 RIN 7.4 	csb196 RIN 9.2 	csb197 RIN 9.4 	csb198 RIN 9.6 

Table 4.1.3j: ***Bioanalyzer Graph Tracings and RIN Values for each Neonatal Sample.***

csb199 RIN na 	csb200 RIN 9.3 	csb201 RIN na 	csb202 RIN 8.8 	csb203 RIN 7.2 
csb204 RIN 9.3 	csb205 RIN 8.4 	csb206 RIN 8 	csb207 RIN 9.1 	-

Appendix 4: Publications and Presentations Arising From This Study

The following article was published in 2007 and it has been reproduced here by permission of The Royal Society of Chemistry and can also be found at: http://www.rsc.org/publishing/journals/AN/article.asp?doi=b707122c

Claire L Smith, Paul Dickinson, Thorsten Forster, Mizanur Khondoker, Marie Cragion, Alan Ross, Petter Storm, Stewart Burgess, Paul Lacaze, Benjamin J Stenson and Peter Ghazal. Quantitative assessment of human whole blood RNA as a potential biomarker for infectious disease. *Analyst* 2007, **132**, 1200-1209.

I presented the results of an earlier pilot of this study (9 infected and 28 control samples) at the following scientific meetings:

European Society of Paediatric Research Meeting, October 2007, Prague.

Scottish Paediatric Society, St Andrews Day Symposium, November 2007, Edinburgh.

The abstract appeared in the ESPR book of conference abstracts as:

Investigation of Host RNA Responses to Infection in Neonates: A New Avenue for Diagnosis? CL Smith, P Dickinson, T Forster, P Lacaze, BJ Stenson, P Ghazal (p8).

Paul Lacaze analyzed results from some of our earliest microarray work from my neonatal samples (3 infected, 3 control and 4 with suspected sepsis). This work was submitted as his MSc. By Research in Life Sciences Maxi-Project entitled:

"A Pathway Biology Investigation of Systemic Host Responses to Infection in Newborn Infants."

2006, University of Edinburgh.

I have subsequently presented results of the work in Chapter 5 at several scientific meetings (listed below) and work is ongoing to submit this work for publication in peer-reviewed journal(s).

Edinburgh Perinatal Festival, May 2010, Edinburgh.

Neonatal Society Summer Meeting, June 2010, Nottingham.

REaSoN Meeting, July 2010, Coventry (poster presentation).

I will also be presenting at the European Academy of Paediatric Societies Congress in October 2010, Copenhagen. The abstract submitted to this conference will appear in *Pediatric Research Journal*.

Quantitative assessment of human whole blood RNA as a potential biomarker for infectious disease†

Claire L. Smith,[ab] Paul Dickinson,[a] Thorsten Forster,[a] Mizanur Khondoker,[a] Marie Craigon,[a] Alan Ross,[a] Petter Storm,[a] Stewart Burgess,[a] Paul Lacaze,[a] Benjamin J. Stenson[b] and Peter Ghazal*[a]

Received 10th May 2007, Accepted 15th October 2007
First published as an Advance Article on the web 31st October 2007
DOI: 10.1039/b707122c

Infection remains a significant cause of morbidity and mortality especially in newborn infants. Analytical methods for diagnosing infection are severely limited in terms of sensitivity and specificity and require relatively large samples. It is proposed that stringent regulation of the human transcriptome affords a new molecular diagnostic approach based on measuring a highly specific systemic inflammatory response to infection, detectable at the RNA level. This proposition raises a number of as yet poorly characterised technical and biological variation issues that urgently need to be addressed. Here we report a quantitative assessment of methodological approaches for processing and extraction of RNA from small samples of infant whole blood and applying analysis of variation from biochip measurements. On the basis of testing and selection from a battery of assays we show that sufficient high quality RNA for analysis using multiplex array technology can be obtained from small neonatal samples. These findings formed the basis of implementing a set of robust clinical and experimental standard operating procedures for whole blood RNA samples from 58 infants. Modelling and analysis of variation between samples revealed significant sources of variation from the point of sample collection to processing and signal generation. These experiments further permitted power calculations to be run indicating the tractability and requirements of using changes in RNA expression profiles to detect different states between patient groups. Overall the results of our investigation provide an essential first step toward facilitating an alternative way for diagnosing infection from very small neonatal blood samples, providing methods and requirements for future chip-based studies.

Introduction

Infection is an important source of morbidity and mortality in neonates and infants. In the developed world, 65% of extremely low birth weight infants develop presumed sepsis in the neonatal period.[1] With mortality rates of 10–50%, a four-fold increase in cerebral palsy and increased risk of hearing, growth and neuro-developmental impairments, the costs are great.[1] On a global perspective, infection accounts for more than half of all deaths worldwide of children younger than age 5.[2] During the first year of life, the developing human immune system encounters many challenges from both infections and vaccinations. Systemic deficiencies of both innate and adaptive immunity are thought to contribute to impaired neonatal host defences while protection through maternal antibodies, which is deficient in preterm babies, wanes after approximately six months in term infants. Early diagnosis of infection is key to providing timely and appropriate treatment. Blood is the primary source of clinical material available and when one considers that some patients may have no more than 40 ml total blood volume and that current procedures often withdraw several millilitres for various blood tests then this sets a stringent ethical and research challenge of working with extremely small quantities of blood. It is noteworthy that the standard diagnostic tool for infection is the blood culture, but this does not give a rapid result (up to 48 h), has poor sensitivity (50–80% at best but often considerably less[3]) and requires blood volumes that represent a significant proportion of an infant's circulating blood volume.

Many signaling molecules and concurrent biological pathways responsible for the initiation and propagation of an inflammatory response to infection have been identified in circulating serum – constituting what has been termed 'Systemic Inflammatory Response'. Individual molecular components identified as part of the systemic inflammatory response have been helpful toward understanding the underlying physiology of inflammation and have also shown potential diagnostic and therapeutic value, including C-reactive protein (CRP) and tumor necrosis factor-α (TNF-α).[4] From early developmental stages onward, the various activities of circulating immune cells contribute to local as well as systemic levels of cytokines and inflammatory molecules. In this context, blood serves as an integrative tissue whereby its cells and associated signaling and cytokine

[a]*Division of Pathway Medicine, The University of Edinburgh, The Chancellor's Building, 49 Little France Crescent, Edinburgh, Midlothian, UK EH16 4SB. E-mail: p.ghazal@ed.ac.uk*
[b]*Neonatal Unit, Simpson Centre for Reproductive Health, The Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh, Midlothian, UK EH16 4SA*
† This paper was published as part of the special issue on the Rapid Diagnostic Testing of Infection and Disease States.

networks relay or enhance the contribution made by sites of infection or tissue damage, to effect protection or cell injury repair responses.

Infections not only elicit but also modify, in a pathogen-specific manner, immune inflammatory responses. This occurs at two levels, the infected cell and the systemic host response levels. At the cell level, analysis of a wide range of studies looking at infection in tissue-culture experiments has defined differential and common host transcriptional responses.[5] This is due in part to the recognition of pathogen-associated molecular patterns by an array of cellular receptors to specific pathogen products. This is especially true for professional antigen presenting cells which orchestrate the selective and appropriate protective immunity. Accordingly, the nature of the systemic inflammatory response generated *in vivo* to an infective agent will also vary depending upon the specific pathogen resulting in both qualitative and quantitative aspects of the immune response being markedly influenced by the various countermeasures enacted by the pathogen. The systemic responses can be seen in altered cytokine levels, specific lymphocyte responses and can also be detected by alterations in the host RNA phenotype in response to infection.[6–10] Microarrays have proven themselves to be useful means for global analysis of gene or protein content and expression. Studies of variation in gene expression among individuals has revealed a surprising consistency, but also evidence of distinct patterns of inter-individual and temporal variation.[11] Microarray technology has shown many potential clinical applications including classification of cancer patients on the basis of disease outcome and prediction of treatment response.[12–15] In recent months there have been increasing numbers of publications reporting microarray experiments using RNA extracted from clinical whole blood samples in adults and children.[11,16–30] In this report we show for the first time that it is feasible to isolate RNA of sufficient quality and quantity from small volumes of whole blood collected from neonates in order to be able to use RNA as a potential early biomarker for infectious disease.

Experimental

Ethical consent

Ethical permission was obtained from the local research ethics committee for this study. Written informed consent was taken from the parent(s) in each case.

Sample collection

Work to determine optimal RNA extraction from neonatal blood was carried out using umbilical cord blood. Umbilical cord blood sampling took place from the cord segment still attached to the placenta after the cord was cut at delivery. The umbilical cord was cleaned with a sterile swab soaked in phosphate buffered saline and the cord cut using sterile scissors. The umbilical vein was then catheterised with a sterile 5-gauge nasogastric tube and blood aspirated into sterile syringe(s). The samples were then injected immediately into sample collection tubes.

For array analysis neonatal whole blood was used. For these samples neonatal blood sampling was performed by trained members of clinical staff. Gloves were worn during the procedure to avoid contamination. The infant's skin was cleaned with an iodine-based solution and then washed with sterile saline and dried with a sterile swab. The needle or cannula was then inserted and a blood sample of approximately 0.5 ml drawn into a syringe. The sample was injected immediately into a PAXgene™ blood RNA tube which was then inverted ten times. Samples were taken from needles, newly inserted venous cannulae or newly inserted arterial cannulae. Samples were not taken from heparinised lines. If samples were to be processed the same day they were transferred to the laboratory and incubated at room temperature for a minimum of 2 h. Otherwise samples were put directly into a −20 °C freezer located within the clinical area until they were transported to the laboratory. In each case data were gathered for each infant including the age of the infant and the mode of sampling (needle or cannula).

Blood collection tube assessment

Five different blood collection media were investigated. Blood was injected into one each of clinical blood tubes containing EDTA, Lithium Heparin and Sodium Citrate. Blood was also injected directly into a PAXgene™ blood RNA tube and into a microcentrifuge tube containing TRIZOL® LS reagent. All tubes were transferred to the laboratory on ice with the exception of the PAXgene™ tube which was transferred at room temperature. With the exception of the PAXgene™ tubes, 0.5 ml of blood from each of the blood collection tubes was processed immediately using the TRIZOL® LS extraction method followed by an on-column cleanup. The PAXgene™ tubes were processed using the PAXgene™ blood RNA extraction kits after at least 2 h incubation at room temperature.

RNA extraction assessment

Five different methods for RNA extraction were performed ranging from organic phase extraction to the use of magnetic bead technology. These were TRIZOL® LS (Invitrogen™ Corporation), QIAamp RNA Blood Mini Kit (QIAgen Ltd.), TRIZOL® LS followed by QIAamp RNA Blood Mini Kit, MagaZorb® (CorTex Biochem™, Inc., San Leandro, CA) and PAXgene™ (PreAnalytiX GmbH).

At the time of sampling for these experiments, blood was injected into PAXgene™ and EDTA tubes. RNA extraction was performed using the PAXgene™ system and using 0.5 ml of blood from the EDTA for each of the other methods except MagaZorb®. As a separate experiment later, 0.5 ml of blood processed in a PAXgene™ tube was compared to 0.2 ml of blood collected in a clinical EDTA tube and processed using the MagaZorb® magnetic bead extraction (results shown later for the MagaZorb® yield in Table 2 multiplied to give the yield equivalent to 0.5 ml).

TRIZOL® LS extraction

0.5 ml of RNase-free water was added to 0.5 ml of blood, then 3 ml of TRIZOL® LS solution was added and repetitive

pipetting was carried out to lyse the cells. RNA extraction was carried out as per TRIZOL® LS instructions with the exception of the initial centrifugation step being carried out at 4000 rpm in an Eppendorf 5810R centrifuge for 1 h.

On-column cleanup following TRIZOL® LS RNA extraction

100 μl samples obtained from the TRIZOL® LS reaction were carried into the first step of the QIAamp RNA Mini Protocol for RNA Cleanup and this protocol was followed to the end. The optional on-column DNase step, second centrifugation step and repeated elution step (to give a final elution volume of 100 μl) were incorporated.

QIAamp RNA extraction

The QIAamp RNA Mini Protocol for Isolation of Total Cellular RNA from Whole Human Blood was followed with the following variations: the blood/Buffer EL mix was incubated on ice for 20 min; after centrifugation and removal of the supernatant the pellet was incubated on ice for a further 10 min; and after further addition of Buffer EL, 600 μl (rather than 350 μl) of buffer RLT was added to the sample. The optional on-column DNase step and second centrifugation step after the addition of buffer RPE were incorporated. The sample was eluted in 100 μl of RNase-free water.

PAXgene™ blood RNA extraction

Each of the PAXgene™ tubes was processed according to the PAXgene™ blood RNA protocol from PreAnalytix dated April 2001. Variations from the protocol were: the incubator steps were carried out in a water bath rather than a shaker–incubator (in step 5 the samples were vortexed once during the incubation), the centrifugation step was increased to 10 min, the optional on-column DNase step and the 1 min drying centrifugation were incorporated.

MagaZorb® extraction

RNA extraction was carried using 200 μl of whole umbilical cord blood according to the MagaZorb® RNA Purification Protocol (CorTex Biochem™) with Supplementary Protocol B (DNase protocol) incorporated.

Quantification and quality assessment of RNA

RNA was quantified and an $A_{260} : A_{280}$ ratio calculated for each sample after analysis on a ThermoSpectronic Biomate 5 v1.6 spectrophotometer. RNA quality was assessed running each sample on an Agilent 2100 Bioanalyser using an RNA LabChip kit. RNA quality was assessed qualitatively by looking at the electropherogram of each sample, and quantitatively by means of the RNA Integrity Number (RIN).

Microarray processing and analysis

The CodeLink™ Human Whole Genome Bioarray was comprised of approximately 55 000 30-mer probes designed to conserve exons across the transcripts of targeted genes. These 55 000 probes represent well-annotated, full length, and partial human gene sequences from major public databases. The biotin-labelled cRNA target is prepared by a linear amplification method using tailed oligo dT priming of total RNA. After second-strand cDNA synthesis, the cDNA undergoes an *in vitro* transcription (IVT) reaction to produce the target cRNA. This method produces approximately 1000-fold to 5000-fold linear amplification. Various quality control procedures are incorporated. Hybridisation is performed overnight and post-hybridisation processing includes a stringent wash to remove unbound and non-specifically hybridised target molecules and staining with Cy™5-streptavidin conjugate. Several non-stringent washes remove unbound conjugate. The bioarrays are then dried and scanned on the Agilent G2567A scanner at 5 μm resolution. Raw data were obtained from the scanned images using CodeLink™ EXPv4.1 (GE Healthcare) feature extraction software. Subsequent data validation comprising data quality control and normalisation involved the use of the statistical software package R (v 2.2.1) and Bioconductor modules for R (v 1.7). Microarray data have been deposited in the GPX MIAME compliant database at http://www.pathwaymedicine.ed.ac.uk/gpx (Accession number: GPX000071 will be made available upon publication).

Statistical methodology and analysis

For comparison of blood collection methods, sample storage and RNA extraction procedures a two-tailed paired Student's *t*-test was employed. For microarray data analysis, a simple (per-gene) analysis of variance model was employed. To define the model, suppose that the median spot intensity $X_{g(i,j,k,l,m,n)}$ corresponds to the gth gene, ith operator, jth way of taking the sample, kth freezing status, lth level of time to extraction, mth category of age, and nth sample. We assume that the median intensity $X_{g(i,j,k,l,m,n)}$ consists of a systematic component $S_{g(i,j,k,l,m)}$ and a non-systematic (random error) component $R_{g(i,j,k,l,m,n)}$ and are related by a multiplicative relationship. That is,

$$X_{g(i,j,k,l,m,n)} = S_{g(i,j,k,l,m)} \times R_{g(i,j,k,l,m,n)}, \quad (1)$$

where the random component $R_{g(i,j,k,l,m,n)}$ is distributed as log-normal with scale parameter 1 [or, $\log_e(1) = 0$] and shape parameter $\sigma_{g(i,j,k,l,m)}$. Therefore,

$$R_{g(i,j,k,l,m,n)} \sim LN(0,\sigma^2_{g(i,j,k,l,m)}). \quad (2)$$

A random variable X has the distribution $LN(\mu,\sigma^2)$ if $Y = \log_e(X)$ has the normal distribution $N(\mu,\sigma^2)$. Therefore, fitting of a log-normal distribution, in principle, reduces to fitting of a normal distribution. Now with the transformations

$$S_{g(i,j,k,l,m)} = \exp(\eta_{g(i,j,k,l,m)}), \text{ and } R_{g(i,j,k,l,m,n)} = \exp(\varepsilon_{g(i,j,k,l,m,n)}),$$

model corresponding to eqn (1) for log-intensity $Y_{g(i,j,k,l,m,n)} = \log_e(X_{g(i,j,k,l,m,n)})$ can be written as

$$Y_{g(i,j,k,l,m,n)} = \eta_{g(i,j,k,l,m)} + \varepsilon_{g(i,j,k,l,m,n)}, \quad (3)$$

where $\eta_{g(i,j,k,l,m)}$ is the log-systematic component and $\varepsilon_{g(i,j,k,l,m,n)}$ is the log-error. Assuming that the log-systematic component

$\eta_{g(i,j,k,l,m)}$ is as an additive function of the effects of the systematic sources of variation,

$$\eta_{g(i,j,k,l,m,n)} = \mu_g + \alpha_{g(i)} + \beta_{g(j)} + \gamma_{g(k)} + \delta_{g(l)} + \tau_{g(m)}, \quad (4)$$

the per-gene analysis of variance (ANOVA) model for log-intensity $Y_{g(i,j,k,l,m,n)}$ is expressed as,

$$Y_{g(i,j,k,l,m,n)} = \mu_g + \alpha_{g(i)} + \beta_{g(j)} + \gamma_{g(k)} + \delta_{g(l)} + \tau_{g(m)} + \varepsilon_{g(i,j,k,l,m,n)}, \quad (5)$$

where μ_g is the overall log-expression of gene g, $\alpha_{g(i)}$ the effect of the ith operator, $\beta_{g(j)}$ the effect of the jth way of taking blood, $\gamma_{g(k)}$ the effect of the kth freezing status, $\delta_{g(l)}$ the effect of the lth level of time to extraction, $\tau_{g(m)}$ the effect of the mth category of age, and $\varepsilon_{g(i,j,k,l,m,n)}$ is the corresponding log-error term distributed as

$$\varepsilon_{g(i,j,k,l,m,n)} \sim N(0,\sigma^2_{g(i,j,k,l,m,n)}). \quad (6)$$

The parameters of model (5) can be estimated using a least-squares method by minimising the error sum of squares given by,

$$SSE_g = \sum_{i,j,k,l,m,n} \varepsilon^2_{g(i,j,k,l,m,n)} = \sum_{i,j,k,l,m,n} (Y_{g(i,j,k,l,m,n)} - \left\{\mu_g + \alpha_{g(i)} + \beta_{g(j)} + \gamma_{g(k)} + \delta_{g(l)} + \tau_{g(m)}\right\})^2. \quad (7)$$

This is done by partially differentiating SSE_g with respect to each of the parameters, and then solving the resulting equations by setting them equal to zero. Details of the methods have previously been described[31] and are implemented in all standard statistical analysis software. We used the R program aov to fit the model (5). If the normality assumption of the log-errors (6) is true, then the least-squares estimates of the model (5) are equivalent to the maximum likelihood estimates. We employed a factorial design to analyse the data. There are five factors with a number of levels resulting in 144 factor combinations. The experiment was difficult to implement as balanced for such multi-factor analysis and therefore have cell frequencies ranging from 0 to 5. The reason for this primarily relates to lack of control on the selection of infants, ages and clinical procedures performed. Thus, while interpretation of results obtained from analysis of variance of unbalanced factorials may sometimes be less precise, the current analysis should reflect the overall behaviors of data in terms of the multiple factors, and provide guidelines for future designs and analyses using larger replication studies.

Results and discussion

Sample collection and processing

Sample collection was performed using a closely defined set of standard operating procedures as in the Experimental section. Tables 1 and 2 show results of the investigations into finding the optimal blood collection tube and extraction method (using umbilical cord blood as a surrogate for neonatal blood) comparing five different commercial RNA extraction procedures. The PAXgene™ blood RNA system consistently gave the best quality RNA while yielding sufficient quantity for microarray analysis. Using standardised protocols for sample collection and processing following identification of optimal methods means that data are as robust and reproducible as possible. Importantly our subsequent studies using neonatal samples have shown that it is possible to obtain sufficient RNA of consistently high quality from neonatal whole blood samples of 0.5 ml. For the 58 samples examined in this report, the mean yield of RNA was 8.45 μg (range 1.84–43.82), the mean RIN value was 9.1 (range 6.6–10) and the mean A_{260}/A_{280} ratio was 1.86 (range 1.24–2.47). We also show in Table 3 that storage of these PAXgene™ blood samples at −20 °C for 7 days prior to RNA extraction leads to no loss in quality of RNA.

Chip-based measurements – contending with systematic and non-systematic variation

Next we implemented the defined clinical and RNA extraction standard operating procedures for collection and processing of

Table 1 Comparison of blood collection tubes

Blood tube	n	Mean (range) RNA per 0.5 ml blood/μg	Mean (range) $A_{260:280}$	Mean (range) RIN values	RIN p-value *cf.* PAXgene™
EDTA	8	9.80 (3.51–24.91)	1.73 (1.01–2.10)	5.0 (1.0–7.6)	0.002
Li heparin	7	14.33 (4.35–30.60)	1.99 (1.79–2.17)	7.9 (7.1–8.9)	0.0004
Na citrate	8	10.20 (2.34–20.23)	1.85 (1.62–2.0)	5.5 (1.1–8.3)	0.01
TRIzol® LS	9	18.21 (5.18–39.96)	1.86 (1.63–2.04)	8.2 (6.8–9.2)	0.01
PAXgene™	8	8.16 (3.33–13.17)	2.03 (1.93–2.2)	9.5 (8.3–10)	

Table 2 Comparison of RNA extraction methods

RNA extraction method	n	Mean (range) RNA per 0.5 ml blood/μg	Mean (range) $A_{260:280}$	Mean (range) RIN values	RIN p-value *cf.* PAXgene™
MagaZorb®	7	15.92 (9.98–21.85)	1.71 (1.53–1.85)	1.9 (1.0–2.9)	0.008
QIAamp	7	2.91 (1.34–4.18)	2.77 (0–7)	4.7 (1.0–8.2)	0.1
TRIzol® LS	7	26.35 (11.88–44.48)	1.80 (1.65–1.92)	2.2 (0.0–6.4)	0.00009
TRIzol® LS and QIAamp	7	16.29 (9.03–25.41)	1.91 (1.5–2.02)	6.0 (0.0–7.8)	0.2
PAXgene™	7	9.25 (4.41–16.66)	1.98 (1.95–2.01)	7.5 (5.5–8.9)	

Table 3 Comparison of storage conditions

Sample storage	n	Mean (range) RNA/μg	Mean (range) $A_{260:280}$	Mean (range) RIN values	RIN p-value
Not frozen	6	4.16 (1.72–5.81)	1.86 (1.55–2.06)	7.8 (1.4–9.6)	0.29
Frozen at −20 °C for 7 days	6	5.61 (3.56–7.92)	1.81 (1.64–2.00)	9.4 (8.4–9.9)	

58 neonatal whole blood samples. Whole blood RNA samples are comprised of RNA from a range of blood cells including reticulocytes, and the presence of high levels of globin mRNA from these cells has led to the use of globin reduction protocols for whole blood samples.[29] We evaluated the use of a globin reduction protocol for our samples (data not shown) but observed limited improvement on sensitivity and specificity of chips. For this reason we processed our samples directly without globin reduction. Gene expression data were determined for the 58 neonatal whole blood samples and analysed for the purpose of investigating the sources and magnitude of systematic and non-systematic (random) variation and to explore an appropriate error model for such data. Specifically, five sources of variation were examined: age of patient (<5 days, 5–10 days or >10 days), blood sample collection method (needle/cannula), freezing status of sample (frozen/not frozen), time to extraction (<3 h, 3–4 h or >4 h), and technical operator (one of four operators). The data were comprised of 58 samples, corresponding to RNA extracted from neonatal whole blood of 58 infants and these samples were analysed using CodeLink™ Human Whole Genome Bioarrays (GE Healthcare), providing expression profiling of *ca.* 55 000 human gene targets in a single array. The first step of our analysis was to look for any significant operator variation in the data. We employed a simple (per-gene) analysis of variance model as outlined in the Experimental section [eqn (5)]. Results of applying the model to the first row of data (g = 1) are summarised in Table 4 and show that the factor 'time to extraction' has the highest level of variation, but none of these sources of variation are statistically significant. A summary of these results is graphically presented in Fig. 1. The top panel, Fig. 1A shows the number of genes with a mean squared error (MSE) greater than or equal to a certain level plotted against that level of variation. The bottom panel, Fig. 1B shows the number of genes having a significance level less than or equal to a certain value plotted against the corresponding level of p-values. Mean squared error (MSE) due to all the systematic sources of variation and error, and the p-values for the significance of systematic sources of variation, are computed on the basis of the per-gene ANOVA model (5). These analyses show that a significant source of variation can be attributed to the 'operator'. Another significant source appears to be the factor 'blood draw method'. That is whether a needle or a cannula was used for taking blood. We see that up to a certain level of variation (MSE = 1.4) the number of genes exceeding a certain threshold of operator variation is higher than that exceeding the same level of variation due to other sources. However, after the level 1.4, 'blood draw method' takes over the 'operator'. Testing if other confounding variables also contribute to these particular variations will require further investigation. Nevertheless, the proportion of genes showing significant variation at p = 0.05 for all five sources of variation considered in this study are shown in Table 5.

Table 4 Results of applying the analysis of variance model to the first row of data (g = 1)

Source of variation	Mean squared error (MSE)	p-Value
Operator	0.42	0.48
Blood draw method	0.19	0.55
Freezing status	0.78	0.22
Time to extraction	1.21	0.10
Age	0.21	0.67
Random error	0.51	

Error models and power calculations required for RNA biomarker identification

We assumed in our basic model (1) that the error associated with the untransformed spot intensity $X_{g(i,j,k,l,m,n)}$ is $R_{g(i,j,k,l,m,n)}$ and follows a log-normal distribution, $R_{g(i,j,k,l,m,n)} \sim LN(0,\sigma^2_{g(i,j,k,l,m,n)})$. Equivalently, the distribution of log-error, which is the error associated with the log-spot-intensity $Y_{g(i,j,k,l,m,n)} = \log_e(X_{g(i,j,k,l,m,n)})$ is denoted by $\varepsilon_{g(i,j,k,l,m,n)}$, and follows a normal distribution

$$\varepsilon_{g(i,j,k,l,m,n)} \sim N(0,\sigma^2_{g(i,j,k,l,m)}).$$

So, fitting a log-normal distribution to errors $R_{g(i,j,k,l,m,n)}$ is equivalent to fitting a normal distribution to log-errors $\varepsilon_{g(i,j,k,l,m,n)}$. The second option is more commonly used and convenient in many ways. Our per-gene ANOVA model (5) actually fits a normal distribution to the log-errors. Therefore, if log-normal is an appropriate choice for errors $R_{g(i,j,k,l,m,n)}$, residuals from the fitted model (5), given by,

$$e_{g(i,j,k,l,m,n)} = Y_{g(i,j,k,l,m,n)} - \left\{\hat{\mu}_g + \hat{\alpha}_{g(i)} + \hat{\beta}_{g(j)} + \hat{\gamma}_{g(k)} + \hat{\delta}_{g(l)} + \hat{\tau}_{g(m)}\right\}$$

should be approximately normally distributed. Empirical distributions and the corresponding fitted normal distributions to the residuals (log-errors) corresponding to ten randomly selected genes are plotted in Fig. 2. It is seen that except for the minor multimodal features of the empirical distributions, normal distributions provide a reasonable fit to the data. General conclusion from Fig. 2 may be that model (5) with normally distributed log-errors provides a reasonable fit to the data. It would, however, be worth investigating if the multimodality as seen in Fig. 2A may be captured with a more appropriate distribution.

These analyses support the possibility of performing microarray experiments on neonatal whole blood and raise the question of whether or not patient group variance is amenable to profiling RNA biomarkers. Fig. 3A shows the

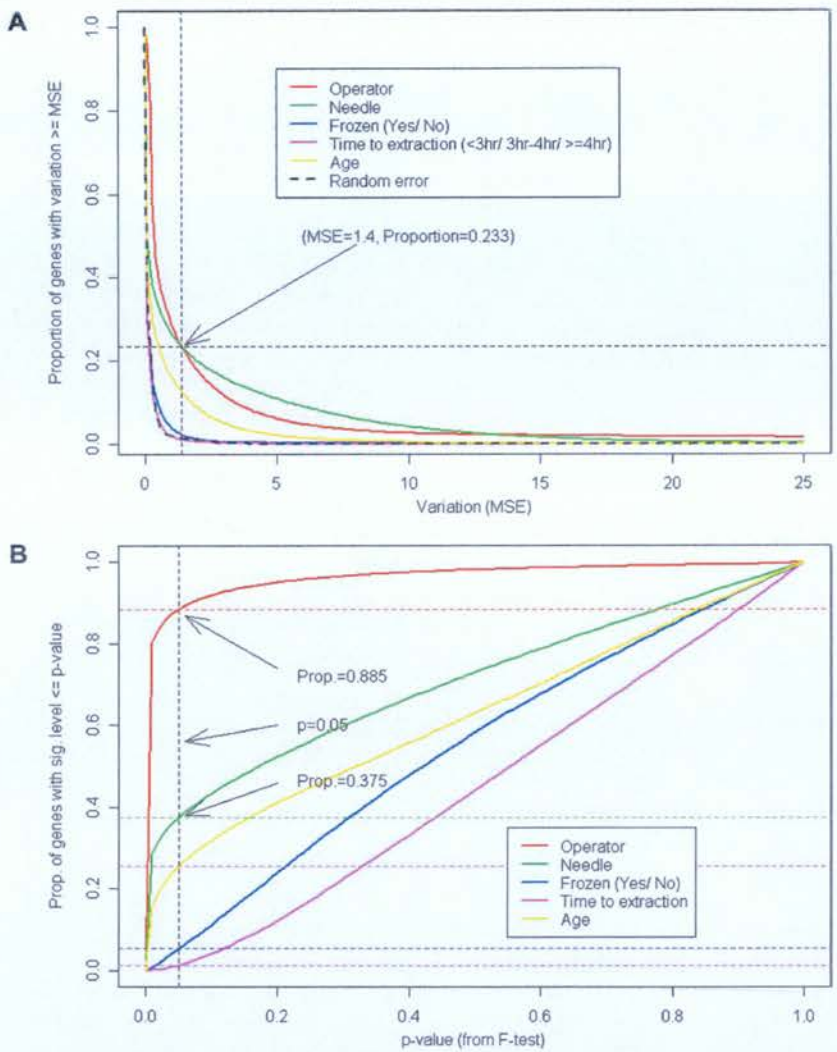


Fig. 1 Investigation of systematic and non-systematic variation in microarray data generated from neonatal whole blood samples. The number of genes having mean squared error (MSE) greater than or equal to a certain level plotted against the level of variation (**A**), and number of genes having a significance level less than or equal to a certain value plotted against the level of *p*-values (**B**). Mean squared error (MSE) due to the systematic sources and random error, and the *p*-values for the significance of systematic variations are computed on the basis of the ANOVA model (5).

Table 5 Proportion of genes having significant variation corresponding to all five systematic sources considered in this study ($p = 0.05$)

	Proportion of genes showing significant variation at $p = 0.05$				
Sources	Operator	Needle/ cannula	Freezing	Time to extraction	Age
Proportion	0.89	0.38	0.05	0.01	0.25

level of variation seen in the expression data generated from the preliminary clinical data. Plots of coefficient of variation of patient samples show acceptable levels of variation. Some exploratory plots of the data are shown in Fig. 3A. If we ignore the outliers in the mean *vs.* CV plot a non-linear trend can be seen in the mean–CV relationship. Determining sample size per experimental condition for a given level of confidence in inferring differential expressions is an important issue, and needs to be decided as an essential first step. The multiple number of RNA markers estimated for each sample makes it difficult to apply traditional sample size calculation techniques and has left most practitioners to rely on rule-of-thumb techniques. A method for computing the sample size for microarray experiments for a given pre-determined level of confidence (power) in inferring differential expressions has been described.[32] The method is based on the assumption that the microarray is set up to compare gene expressions between one treatment group and one experimental group. It is further assumed that the data have been normalised and transformed so that the data for each gene are sufficiently close to a normal distribution so that a standard two-sample pooled variance *t*-test will reliably detect differentially expressed genes. Here we compute the sample size separately for each gene according to the standard formula for the

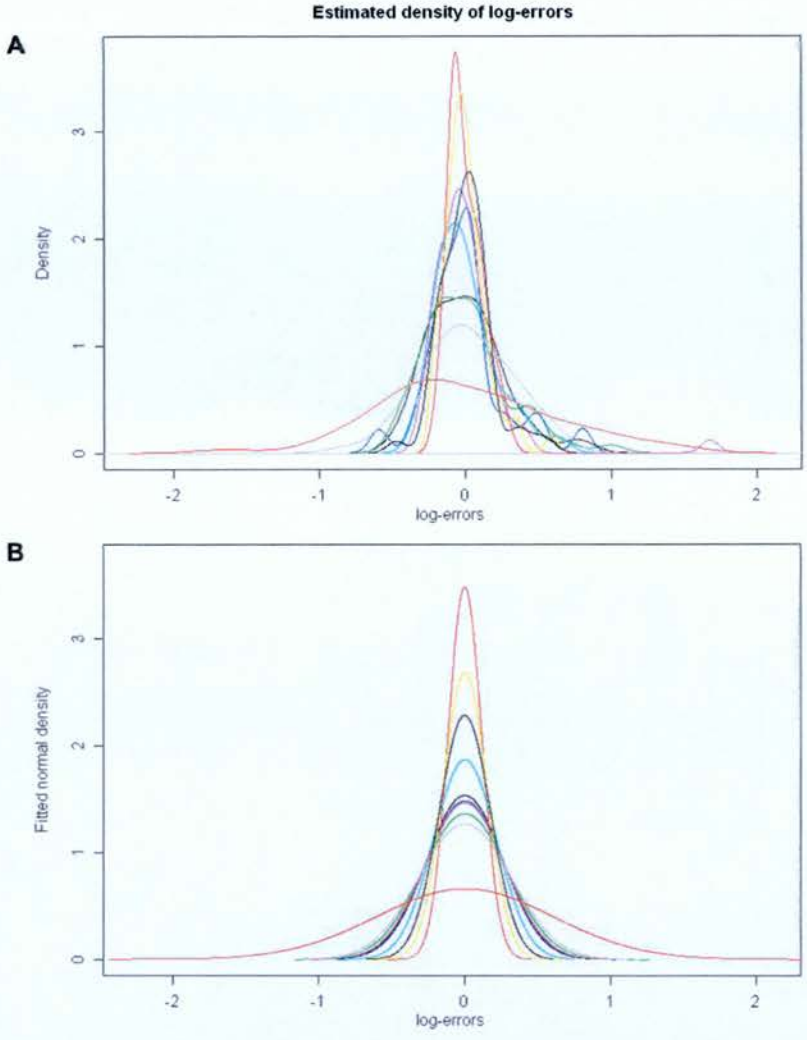


Fig. 2 Empirical distributions (**A**) and the corresponding fitted normal distributions (**B**) to the residuals (log-errors) corresponding to ten randomly selected genes.

two-sample t-test:

$$1-\beta = 1 - T_{n_1+n_2-2}\left(t_{\alpha_G/2,n_1+n_2-2} \middle| \frac{\Delta}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right) + T_{n_1+n_2-2}\left(-t_{\alpha_G/2,n_1+n_2-2} \middle| \frac{\Delta}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right)$$

where $T_d(\cdot|\theta)$ is the cumulative distribution function for non-central t-distribution with d degrees of freedom and the non-centrality parameter θ.

We apply the above method to calculate the sample size to achieve 90% power on the basis of the standard deviations computed from the quantile normalised data of 58 control samples of the neonatal whole blood gene expression study. The results are plotted in Fig. 3B representing the sample size required to achieve 90% power for a given proportion of genes on the arrays. Power calculations and sample size estimates based on $n = 58$ samples with the same type of sample and on the same microarray platform suggest that 100 samples per group will be required to detect two-fold differential expression with 90% power for at least 90% of the probes on the array, at a significance level of $\alpha = 0.001$ (corrected for multiple testing by the Bonferroni method[33]). We conclude from these analyses that identifying RNA biomarkers is tractable with a case control group size of 100 patients.

Stringency of transcriptome and variation of RNA phenotype

The above studies indicate that while there is a significant level of variation there is nevertheless a relatively stable, well-correlated RNA phenotype that can be used to identify a specific set of RNAs as potential biomarkers. Overall, this observation is indicative of homeostatic mechanisms underpinning a stringently regulated genome. In support there are a number of well-documented genetic diseases, mainly of non-protein coding mutations, that lead to an alteration of very subtle changes (up to two-fold or less) in gene expression in comparison with normal individuals and which result in

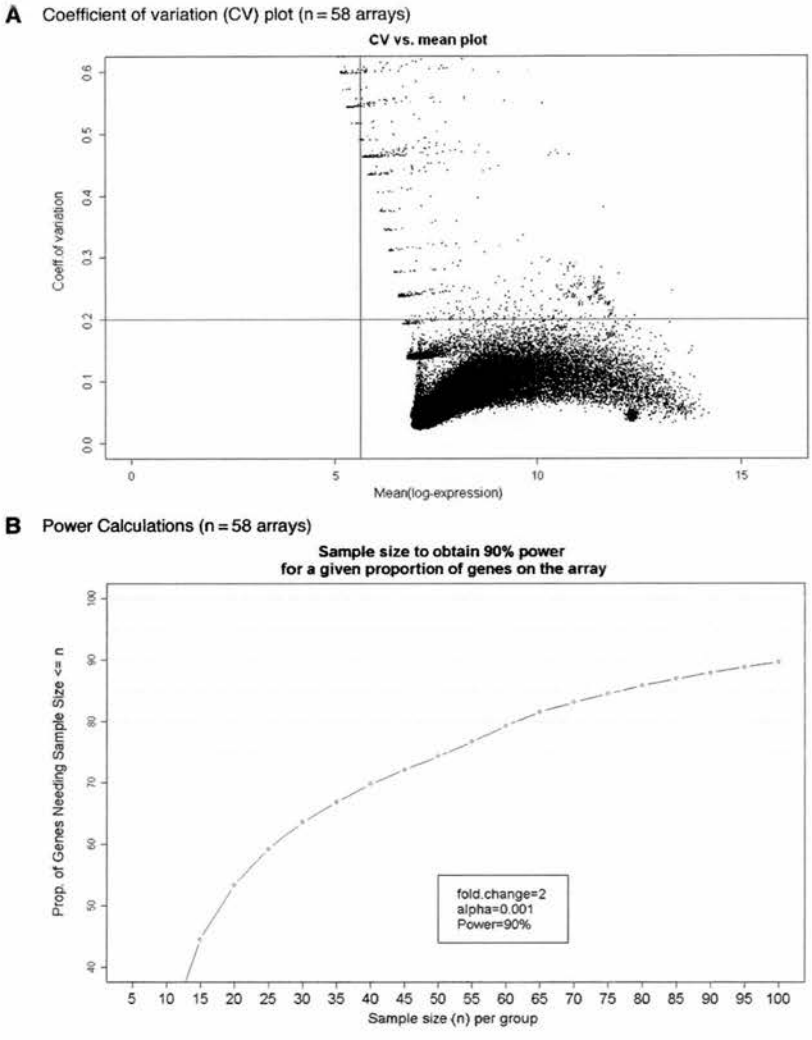


Fig. 3 Scatter plots of inter-patient variance and power calculations.

marked clinical phenotypes.[34–37] These studies indicate that gene expression is tightly regulated and relatively intolerant of dramatic variation. In support, microarray studies of variation in expression among individuals have revealed a surprising level of overall consistency but also evidence of distinct patterns of inter-individual and temporal variation.[11] A recent chip study shows the possibility of even detecting slight alterations in gene expression due to allelic variation.[38] Our recent experience in performing a range of clinical molecular profiling studies from intestinal, mammary, adult blood, endometrial, ovarian and testicular biopsies supports the view for a remarkable stringency in the inter-individual regulation of the transcriptome (see ref. 39 and unpublished observations). To date, neonatal whole blood shows the highest level of variability in our experience. This could be due to greater levels of complexity of procedural, sample handling and age differences as well as a degree of biological variation. Nevertheless, even with these samples our microarray profiling observations show an excellent correlation between individuals, further indicating the human transcriptome to be under stringent homeostatic regulation. Accordingly, even without standardisation, early microarray studies have shown potential clinical applications including classification of cancer patients on the basis of disease outcome and prediction of treatment response.[12–15]

RNA phenotype of systemic host response

One particularly exciting application of microarrays could be investigating infection by detecting alteration in host RNA phenotype in response to infection.[6–9] This would be particularly useful if unique host signatures could identify individual pathogens.[10] Recently, there have been a few small/pilot studies looking at gene expression profiling in response to infection.[17,19,23,27] Also, there have been an increasing number of publications reporting microarray experiments using RNA from clinical whole blood samples in adults and children.[11,16–30,40] These reports are at this time limited not only in terms of statistical power but also to the gene-analytic approach applied. However, one study has attempted to explore networks of inter-related genes.[41] In this connection it is worth noting that biological pathways provide a central level of physiological organisation and, to date, a pathway-centric approach is markedly absent.

Perspective: bio-chip platforms for point-of-care

Biochip platforms based on pathogen detection both at the nucleic acid[42] and protein[43] level are seen as key in the accurate diagnosis of infection. However, the multi-parameter testing of changes in whole blood RNA expression has the potential to use extremely small quantities of material which does not require the presence of the infective agent in the sample. It is possible to envision that micro-devices would have the capacity not only to process and extract RNA but also to detect directly the presence of specific host pathway responses to infection. It is also likely that these pathway responses may also be detected at the protein level based on predictions from the RNA phenotype. Biochips incorporating a combination of both pathogen detection and host response would give the widest possible coverage to detect signatures diagnostic of infection. Nevertheless, regardless of the platform technology used to detect such signatures it is of fundamental importance that a clear understanding of the levels and contributions of systematic and non-systematic variation are fully appreciated. In this report we have shown that the operator and the point of collection can provide a significant source of changes in gene expression. Therefore if RNA is to fulfil a future role as a potential biomarker it is essential that appropriately powered studies are performed which account for known systematic variation and error models developed to account for non-systematic variation.

Conclusion

In conclusion we show that multi-parameter testing of changes in RNA expression offers innovative potential and an amenable means for measuring an RNA phenotype using relatively small quantities of whole blood. We provide optimal methods and procedures and attendant clinical and experimental SOPs for applying a rigorous, chip-based investigation. Even with such methods, significant challenges and limitations remain, related to both systematic and non-systematic variations. From an analysis of the RNA phenotypes from 58 samples of neonatal whole blood considered in this study we identify among the systematic sources of variation as potentially contributing variability: laboratory operator, the way the blood is drawn (by needle/cannula), and the age of the infants. Further case studies are required to validate these findings. We show that a linear additive analysis of variance model for log-transformed data with Gaussian-distributed log-errors seems reasonable to adjust the data for systematic variation. Except for the minor multimodal features of the empirical distributions, log-normal distributions for errors or, equivalently, normal distributions for log-errors provide a reasonable fit to the data accounting for non-systematic variation. A sample size of about 100 per-group seems reasonable to achieve 90% power for 90% of markers on the chip.

Acknowledgements

We would like to acknowledge funding by the INFOBIOMED Network of excellence (European Commission FP6 Programme: contract number IST-2002-507585), Wellcome Trust, Scottish Enterprise, and Chiesi Pharmaceuticals, and help from all of our colleagues in the Division of Pathway Medicine and clinical staff.

References

1 B. J. Stoll, N. I. Hansen, I. Adams-Chapman, A. A. Fanaroff, S. R. Hintz, B. Vohr and R. D. Higgins, *J. Am. Med. Assoc.*, 2004, **292**, 2357–2365.
2 J. Bryce, C. Boschi-Pinto, K. Shibuya and R. E. Black, *Lancet*, 2005, **365**, 1147–1152.
3 J. S. Gerdes, *Pediatr. Clin. North Am.*, 2004, **51**(4), 939–959.
4 R. Berner, C. M. Niemeyer, J. U. Leititis, A. Funke, C. Schwab, U. Rau, K. Richter, M. S. Tawfeek, A. Clad and M. Brandis, *Pediatr. Res.*, 1998, **44**, 469–477.
5 R. G. Jenner and R. A. Young, *Nat. Rev. Microbiol.*, 2005, **3**, 281–294.
6 J. C. Boldrick, A. A. Alizadeh, M. Diehn, S. Dudoit, C. L. Liu, C. E. Belcher, D. Botstein, L. M. Staudt, P. O. Brown and D. A. Relman, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 972–977.
7 C. J. Campbell and P. Ghazal, *J. Appl. Microbiol.*, 2004, **96**, 18–23.
8 C. A. Cummings and D. A. Relman, *Emerg. Infect. Dis.*, 2000, **6**, 513–525.
9 D. Yowe, W. J. Cook and J. C. Gutierrez-Ramos, *Microbes Infect.*, 2001, **3**, 813–821.
10 Q. Huang, D. Liu, P. Majewski, L. C. Schulte, J. M. Korn, R. A. Young, E. S. Lander and N. Hacohen, *Science*, 2001, **294**, 870–875.
11 A. R. Whitney, M. Diehn, S. J. Popper, A. A. Alizadeh, J. C. Boldrick, D. A. Relman and P. O. Brown, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1896–1901.
12 A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown and L. M. Staudt, *Nature*, 2000, **403**, 503–511.
13 M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts and V. Sondak, *Nature*, 2000, **406**, 536–540.
14 T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander, *Science*, 1999, **286**, 531–537.
15 C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein, *Nature*, 2000, **406**, 747–752.
16 F. Borovecki, L. Lovrecic, J. Zhou, H. Jeong, F. Then, H. D. Rosas, S. M. Hersch, P. Hogarth, B. Bouzou, R. V. Jensen and D. Krainc, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11023–11028.
17 R. J. Feezor, C. Oberholzer, H. V. Baker, D. Novick, M. Rubinstein, L. L. Moldawer, J. Pribble, S. Souza, C. A. Dinarello, W. Ertel and A. Oberholzer, *Infect. Immun.*, 2003, **71**, 5803–5813.
18 M. Ghielmetti, M. Bellis, M. O. Spycher, S. Miescher and G. Vergeres, *Mol. Immunol.*, 2006, **43**, 939–949.
19 M. J. Griffiths, M. J. Shafi, S. J. Popper, C. A. Hemingway, M. M. Kortok, A. Wathen, K. A. Rockett, R. Mott, M. Levin, C. R. Newton, K. Marsh, D. A. Relman and D. P. Kwiatkowski, *J. Infect. Dis.*, 2005, **191**, 1599–1611.
20 A. D. Hershey, Y. Tang, S. W. Powers, M. A. Kabbouche, D. L. Gilbert, T. A. Glauser and F. R. Sharp, *Headache*, 2004, **44**, 994–1004.
21 P. A. Horwitz, E. J. Tsai, M. E. Putt, J. M. Gilmore, J. J. Lepore, M. S. Parmacek, A. C. Kao, S. S. Desai, L. R. Goldberg, S. C. Brozena, M. L. Jessup, J. A. Epstein and T. P. Cappola, *Circulation*, 2004, **110**, 3815–3821.
22 C. C. Liew, J. Ma, H. C. Tang, R. Zheng and A. A. Dempsey, *J. Lab. Clin. Med.*, 2006, **147**, 126–132.

23 M. Prucha, A. Ruryk, H. Boriss, E. Moller, R. Zazula, I. Herold, R. A. Claus, K. A. Reinhart, P. Deigner and S. Russwurm, *Shock*, 2004, **22**, 29–33.
24 Q. Tan, K. Christensen, L. Christiansen, H. Frederiksen, L. Bathum, J. Dahlgaard and T. A. Kruse, *Hum. Genet.*, 2005, **117**, 267–274.
25 Y. Tang, T. A. Glauser, D. L. Gilbert, A. D. Hershey, M. D. Privitera, D. M. Ficker, J. P. Szaflarski and F. R. Sharp, *Acta Neurol. Scand.*, 2004, **109**, 159–168.
26 Y. Tang, M. B. Schapiro, D. N. Franz, B. J. Patterson, F. J. Hickey, E. K. Schorry, R. J. Hopkin, M. Wylie, T. Narayan, T. A. Glauser, D. L. Gilbert, A. D. Hershey and F. R. Sharp, *Ann. Neurol.*, 2004, **56**, 808–814.
27 D. C. Thach, B. K. Agan, C. Olsen, J. Diao, B. Lin, J. Gomez, M. Jesse, M. Jenkins, R. Rowley, E. Hanson, C. Tibbetts, D. A. Stenger and E. Walter, *Genes Immun.*, 2005, **6**, 588–595.
28 Z. Wang, D. Neuburg, C. Li, L. Su, J. Y. Kim, J. C. Chen and D. C. Christiani, *Environ. Health Perspect.*, 2005, **113**, 233–241.
29 R. J. Feezor, H. V. Baker, M. Mindrinos, D. Hayden, C. L. Tannahill, B. H. Brownstein, A. Fay, S. MacMillan, J. Laramie, W. Xiao, L. L. Moldawer, J. P. Cobb, K. Laudanski, C. L. Miller-Graziano, R. V. Maier, D. Schoenfeld, R. W. Davis and R. G. Tompkins, *Physiol. Genomics*, 2004, **19**, 247–254.
30 H. R. Wong, T. P. Shanley, B. Sakthivel, N. Cvijanovich, R. Lin, G. L. Allen, N. J. Thomas, A. Doctor, M. Kalyanaraman, N. M. Tofil, S. Penfil, M. Monaco, M. A. Tagavilla, K. Odoms, K. Dunsmore, M. Barnes and B. J. Aronow, *Physiol. Genomics*, 2007, **30**, 146–155.
31 D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons Inc., New York, 6th edn, 2004.
32 G. R. Warnes, Technical Report, University of Rochester, 2006 (http://www.urmc.rochester.edu/smd/biostat/people/faculty/manuscripts/0606_warnes_liu.pdf).
33 C. E. Bonferroni, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 1936, **8**, 3–62.
34 D. N. Cooper, *Ann. Med.*, 1992, **24**, 427–437.
35 E. Di Pierro, M. D. Cappellini, R. Mazzucchelli, V. Moriondo, D. Mologni, B. Zanone Poma and A. Riva, *Exp. Hematol. (N. Y.)*, 2005, **33**, 584–591.
36 X. M. Sun, C. Neuwirth, D. P. Wade, B. L. Knight and A. K. Soutar, *Hum. Mol. Genet.*, 1995, **4**, 2125–2129.
37 F. Tassone, R. J. Hagerman, W. D. Chamberlain and P. J. Hagerman, *Am. J. Med. Genet.*, 2000, **97**, 195–203.
38 P. V. Pant, H. Tao, E. J. Beilharz, D. G. Ballinger, D. R. Cox and K. A. Frazer, *Genome Res.*, 2006, **16**, 331–339.
39 H. O. Critchley, K. A. Robertson, T. Forster, T. A. Henderson, A. R. Williams and P. Ghazal, *Am. J. Obstet. Gynecol.*, 2006, **195**(406), e401–416.
40 O. Ramilo, W. Allman, W. Chung, A. Mejias, M. Ardura, C. Glaser, K. M. Wittkowski, B. Piqueras, J. Banchereau, A. K. Palucka and D. Chaussabel, *Blood*, 2007, **109**, 2066–2077.
41 S. E. Calvano, W. Xiao, D. R. Richards, R. M. Felciano, H. V. Baker, R. J. Cho, R. O. Chen, B. H. Brownstein, J. P. Cobb, S. K. Tschoeke, C. Miller-Graziano, L. L. Moldawer, M. N. Mindrinos, R. W. Davis, R. G. Tompkins and S. F. Lowry, *Nature*, 2005, **437**, 1032–1037.
42 S. W. Yeung, T. M. Lee, H. Cai and I. M. Hsing, *Nucleic Acids Res.*, 2006, **34**, e118.
43 X. Dubureq, C. Olivier, F. Malingue, R. Desmet, A. Bouzidi, F. Zhou, C. Auriault, H. Gras-Masse and O. Melnyk, *Bioconjugate Chem.*, 2004, **15**, 307–316.