Genomics Data 3 (2015) 41–48

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Whole blood gene expression profiling of neonates with confirmed bacterial sepsis

Paul Dickinson [a,b,*,1], Claire L. Smith [c,1], Thorsten Forster [a,b], Marie Craigon [a], Alan J. Ross [a], Mizan R. Khondoker [a,2], Alasdair Ivens [d,3], David J. Lynn [e,4], Judith Orme [c], Allan Jackson [c], Paul Lacaze [a], Katie L. Flanagan [f,5], Benjamin J. Stenson [c], Peter Ghazal [a,b,*]

[a] Division of Infection and Pathway Medicine, Edinburgh Infectious Diseases, University of Edinburgh, Edinburgh EH16 4SB, UK
[b] SynthSys—Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3JD, UK
[c] Neonatal Unit, Simpson Centre for Reproductive Health, Royal Infirmary of Edinburgh, Edinburgh EH16 4SA, UK
[d] Fios Genomics Ltd., Edinburgh BioQuater, Edinburgh EH16 4SB, UK
[e] Animal & Bioscience Research Department, AGRIC, Teagasc, Grange, Dunsany, Co. Meath, Ireland
[f] MRC Research Laboratories, Atlantic Boulevard, PO Box 273, Fajara, Gambia

ARTICLE INFO

Article history:
Received 31 October 2014
Accepted 6 November 2014
Available online 15 November 2014

Keywords:
Neonatal sepsis
Whole blood
Gene expression profiling
Microarray

ABSTRACT

Neonatal infection remains a primary cause of infant morbidity and mortality worldwide and yet our understanding of how human neonates respond to infection remains incomplete. Changes in host gene expression in response to infection may occur in any part of the body, with the continuous interaction between blood and tissues allowing blood cells to act as biosensors for the changes. In this study we have used whole blood transcriptome profiling to systematically identify signatures and the pathway biology underlying the pathogenesis of neonatal infection. Blood samples were collected from neonates at the first clinical signs of suspected sepsis alongside age matched healthy control subjects. Here we report a detailed description of the study design, including clinical data collected, experimental methods used and data analysis workflows and which correspond with data in Gene Expression Omnibus (GEO) data sets (GSE25504). Our data set has allowed identification of a patient invariant 52-gene classifier that predicts bacterial infection with high accuracy and lays the foundation for advancing diagnostic, prognostic and therapeutic strategies for neonatal sepsis.

Specifications	
Organism/cell line/tissue	Homo sapiens/whole blood
Sex	Male and female
Sequencer or array type	Illumina HT-12V3.0 Whole Human Genome microarray, CodeLink 55K Whole Human Genome microarray, Affymetrix U219 Whole Human Genome microarray and Affymetrix HG U133 Plus 2.0 Whole Human Genome microarray
Data format	Raw data (Tab delimited text files of background subtracted signals and .CEL files)
Experimental factors	Blood culture or cerebrospinal fluid positive bacterial sepsis vs. healthy control whole blood samples and culture negative suspected infected samples
Experimental features	A case–control gene expression profiling study of whole blood taken from neonates at the first clinical sign of sepsis and control healthy neonates. Study includes training and replication sets for blood culture positive samples and clinical evaluation set of blood culture negative sepsis cases. Results compared blood culture or cerebrospinal fluid positive septic neonates, blood culture negative septic neonates and healthy control neonates. Prior power calculations were based on Healthy Edinburgh neonates using the CodeLink platform and Gambian infants (9 months of age) were used for further refinement of power calculations using Illumina HT-12 platform.

* Corresponding authors at: Division of Pathway Medicine, Edinburgh Infectious Diseases, University of Edinburgh, Edinburgh EH16 4SB, UK.
E-mail addresses: paul.dickinson@ed.ac.uk (P. Dickinson), p.ghazal@ed.ac.uk (P. Ghazal).

[1] These authors contributed equally to this work.

[2] Present address: Department of Biostatistics, Institute of Psychiatry and NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust, King's College, London, UK.

[3] Present address: Centre for Infection Immunity and Evolution, King's Buildings, University of Edinburgh, Edinburgh, UK.

[4] Present address: EMBL Australia Laboratory, South Australian Health and Medical Research Institute, North Terrace, Adelaide, South Australia 5000, Australia.

[5] Present address: Department of Immunology, Monash University, Commercial Road, Prahran, Melbourne, Victoria 3181, Australia.

http://dx.doi.org/10.1016/j.gdata.2014.11.003

(*continued*)

Specifications	
Consent	Written informed consent was obtained from parents of all enrolled infants in accordance with approval granted by the Lothian Research Ethics Committee for blood samples for RNA isolation obtained at the first time of clinical signs of suspected sepsis (reference 05/s1103/3). Samples obtained from The Gambia conformed to MRC policy regarding ethical research in children and were approved by the local scientific coordinating committee (SCC), the Joint Gambia Government/MRC Ethics Committee and by the London School of Hygiene and Tropical Medicine Ethics Committee (reference SCC1085 Pilot Study 1 (L2008.63))
Sample source location	Edinburgh, UK and The Gambia

Direct link to deposited data

Deposited data are available here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25504.

Experimental design, materials and methods

Patient demographics and experimental design

The study was conducted in the Neonatal Unit, Royal Infirmary of Edinburgh and the Division of Pathway Medicine, University of Edinburgh. The patient demographics, microbial organisms isolated and reasons for blood sampling in controls for all patient sets are shown in Table 1. Infants having blood cultures taken to investigate suspected infection (Table 1B) and "well" control infants having blood taken for other clinical reasons (Table 1C) were studied. Samples taken from patients with suspected clinical infection that proved to have microbiological evidence of infection from a usually sterile body site were identified and formed the infected group. Full clinical assessment for early and late symptoms and signs of sepsis followed criteria for neonatal sepsis taken from data as detailed in Table 2, with the blood culture test used as the 'gold standard' for diagnosis of sepsis. Five infants had samples included from more than one episode of infection. 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, for example, infants with an underlying disorder causing anemia. Before embarking on this study we had previously performed a power calculation using the CodeLink chip platform [1] on neonatal samples but we also performed a power calculation using the Illumina chip platform, on an independent set of 30 infant samples at 9 months of age, before vaccination. This showed that the study design has 90% power to detect a twofold change in expression with an a of 1% (false discovery rate (FDR) corrected), for more than 99% of 35,177 gene probes present on the array [2]. A schematic of patient recruitment and sample processing workflow for the samples processed for the training, replication and validation arm of the study is shown in Fig. 1.

Sample collection and RNA extraction

For RNA isolation, blood (500 μl–1 ml) was immediately injected into PAXgene™ blood RNA tubes (PreAnalytiX BD/QIAgen) and mixed

Table 1
Patient demographics of samples used, microorganisms identified from infected patients and reasons for blood sampling in controls.

Patient demographics of samples used

Sample set	Training set		Platform test set		Validation test set	
Infection status	Infected ($n = 28$)	Control ($n = 35$)	Infected ($n = 18$)	Control ($n = 24$)	Infected ($n = 16$)	Control ($n = 10$)
Male	15 (54%)	22 (63%)	10 (56%)	15 (63%)	10 (63%)	9 (90%)
Gestation completed at birth (week): range (mean)	24–38 (28.5)	26–42 (37.9)	24–38 (28.8)	26–42 (37.3)	23–40 (28.3)	24–41 (31)
Gestation completed at sampling (week): range (mean)	26–39 (31.1)	31–44 (39.4)	26–39 (30.8)	31–44 (39.1)	25–41 (33.8)	29–42 (34.9)
Birthweight (g): range (mean)	430–3380 (1126)	650–4570 (3080)	430–3380 (1236)	650–4350 (2941)	635–3160 (1134)	800–4220 (1932)

Microorganisms identified from infected patients

Organism	Training set	Platform test set	Validation test set
Coagulase negative staphylococcus	15	8	7
Enterococcus species	4	3	1
Group B Streptococcus	2	2	1
Klebsiella species	2	1	2
Candida albicans and Klebsiella species	1	1	–
Escherichia Coli	1	1	1
Enterobacter cloacae	1	1	–
Pseudomonas aeruginosa	1	1	1
CMV	1	–	–
Listeria monocytogenes	–	–	1
Serratia marcescens	–	–	2

Reasons for blood sampling in controls

Reason	Training set	Platform test set	Validation test set
Screening test: maternal thyroid disease	17	9	–
Bilirubin check due to jaundice	5	4	1
"Routine" neonatal screening (preterms)	5	4	6
Electrolyte check: previous deranged Na	3	3	–
Screening test: pigmented scrotum	2	1	3
Blood count check: Coomb's positive	1	1	–
Screening test: newborn bloodspot	1	1	–
Neonatal encephalopathy	1	1	–

A. Patient demographics of samples used. Patient sample details are shown displaying the demographics of the population studied. B. Microorganisms identified from infected patients. Organisms detected for each infected infant are shown — these samples were taken at, or within 6 h of, the time of clinical suspicion of infection. C. Reasons for blood sampling in controls. The reasons for clinical blood sampling in the control group are shown — all of the screening tests in these infants were normal. Table 1 was adapted from Supplementary Table 3 of Smith et al. 2014 [2] by permission from Macmillan Publishers Ltd: Nature Communications [2], copyright (2014).