RESEARCH ARTICLE

REVISED Cerebrospinal fluid markers to distinguish bacterial meningitis from cerebral malaria in children [version 2; peer review: 2 approved]

James M. Njunge [1], Ian N. Oyaro [1,2], Nelson K. Kibinge [1], Martin K. Rono[1,3], Symon M. Kariuki [1], Charles R. Newton [1,4], James A. Berkley [1,5], Evelyn N. Gitau [1,6]

[1]KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research Coast, Kilifi, Kenya
[2]University of Nairobi, Nairobi, Kenya
[3]Pwani University Health and Research Institute, Pwani University, Kilifi, Kenya
[4]Department of Psychiatry, Medical Sciences Division, University of Oxford, Oxford, OX3 7JX, UK
[5]Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, OX3 7FZ, UK
[6]Alliance for Accelerating Excellence in Science in Africa (AESA), Nairobi, Kenya

v2 **First published:** 03 Jul 2017, **2**:47 (https://doi.org/10.12688/wellcomeopenres.11958.1)
Latest published: 26 Sep 2017, **2**:47 (https://doi.org/10.12688/wellcomeopenres.11958.2)

Abstract

Background. Few hospitals in high malaria endemic countries in Africa have the diagnostic capacity for clinically distinguishing acute bacterial meningitis (ABM) from cerebral malaria (CM). As a result, empirical use of antibiotics is necessary. A biochemical marker of ABM would facilitate precise clinical diagnosis and management of these infections and enable rational use of antibiotics.
Methods. We used label-free protein quantification by mass spectrometry to identify cerebrospinal fluid (CSF) markers that distinguish ABM (n=37) from CM (n=22) in Kenyan children. Fold change (FC) and false discovery rates (FDR) were used to identify differentially expressed proteins. Subsequently, potential biomarkers were assessed for their ability to discriminate between ABM and CM using receiver operating characteristic (ROC) curves.
Results. The host CSF proteome response to ABM (*Haemophilus influenza* and *Streptococcus pneumoniae*) is significantly different to CM. Fifty two proteins were differentially expressed (FDR<0.01, Log FC≥2), of which 83% (43/52) were upregulated in ABM compared to CM. Myeloperoxidase and lactotransferrin were present in 37 (100%) and 36 (97%) of ABM cases, respectively, but absent in CM (n=22). Area under the ROC curve (AUC), sensitivity, and specificity were assessed for myeloperoxidase (1, 1, and 1; 95% CI, 1-1) and lactotransferrin (0.98, 0.97, and 1; 95% CI, 0.96-1).
Conclusion. Myeloperoxidase and lactotransferrin have a high potential to distinguish ABM from CM and thereby improve clinical management. Their validation requires a larger cohort of samples that includes other bacterial aetiologies of ABM.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
REVISED **version 2** published 26 Sep 2017	✓ report	✓ report
version 1 published 03 Jul 2017	? report	? report

1 **Wilson L. Mandala**, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
Malawi University of Science and Technology, Thyolo, Malawi

2 **Clarissa Valim**, Michigan State University, East Lansing, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Biomarkers, Acute Bacterial Meningitis, Cerebral Malaria, CSF, proteomics, Myeloperoxidase, Lactotransferrin

This article is included in the KEMRI | Wellcome Trust gateway.

Corresponding authors: James M. Njunge (jnjunge@kemri-wellcome.org), Evelyn N. Gitau (e.gitau@aasciences.ac.ke)

Author roles: Njunge JM: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Oyaro IN**: Conceptualization, Investigation, Methodology; **Kibinge NK**: Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Rono MK**: Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Kariuki SM**: Resources, Writing – Review & Editing; **Newton CR**: Resources, Writing – Review & Editing; **Berkley JA**: Formal Analysis, Resources, Writing – Review & Editing; **Gitau EN**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [084538], Strategic Award; [107769], Strategic Award and Initiative to Develop African Research Leaders (IDeAL). JMN is a beneficiary of both Strategic Award and IDEAL early career post-doc, EG and MKR are beneficiaries of Strategic Award. NKK is a post-doc currently supported by IDeAL.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Njunge JM, Oyaro IN, Kibinge NK *et al.* **Cerebrospinal fluid markers to distinguish bacterial meningitis from cerebral malaria in children [version 2; peer review: 2 approved]** Wellcome Open Research 2017, **2**:47 (https://doi.org/10.12688/wellcomeopenres.11958.2)

First published: 03 Jul 2017, **2**:47 (https://doi.org/10.12688/wellcomeopenres.11958.1)

REVISED **Amendments from Version 1**

Figure S1 has been modified to exclude the CM group and the text "CM, cerebral malaria;" has been deleted from Figure S1 legend. Figure 4 has been removed since the same information is represented in the Table 3. References have been added after the following text in the discussion section "Whilst in CM the blood brain barrier is mildly impaired with few morphological changes[27–30]." The CM group has been excluded from Figure S1 as well as the text "CM, cerebral malaria;" has been deleted from Figure S1 legend.

The following text was added under limitations in the discussion section "Further, in endemic areas, the definition of CM is often challenging and CM is often over-diagnosed. The WHO definition of CM may misclassify up to 25% of cases[46] and its specificity is greatly improved by adding a clinical test for Malaria Retinopathy[47,48]. However, Retinal changes specific to CM require specialist examination techniques, are difficult to examine in conscious children, and such data was not available in this study. Samples analyzed in this study were left over specimens following normal microbiology and chemistry laboratory procedures. There is lacking data for the samples on the time taken from collection during lumbar puncture to storage at -80°C and therefore there is possibility that proteome changes occurred during that period[49]. Label-free proteomics are cost-efficient, offer higher proteome coverage and a higher dynamic range. However, as every sample is handled separately, variations that can bias the quantitative analysis may be introduced." The percent of sample for which each protein was quantified in either ABM or CM has been added to Table S2.

See referee reports

Introduction

Acute non-traumatic coma is an important cause of morbidity and mortality among paediatric hospital admissions in malaria endemic areas of Africa. This is commonly caused by acute bacterial meningitis (ABM) and cerebral malaria (CM), although viruses, fungi, and other infectious and non-infectious causes may occur. The clinical features associated with and used for diagnosis of ABM may overlap with those of CM. The World Health Organization defines CM as coma that persists >1 h after a seizure once hypoglycaemia is corrected with no other cause to explain the coma, and the presence of asexual parasites in peripheral blood[1]. Abnormal retinoscopy is associated with cerebral parasite sequestration and increases the specificity of the diagnosis for CM[2,3]. Histidine-rich protein 2, a parasite protein used to estimate the total body parasite biomass, is also considered a potential marker that increases specificity for CM[4,5], but does not exclude ABM. In practice, children with clinical signs and a positive malaria slide or rapid diagnostic test (RDT) are treated for malaria. However, in malaria-endemic regions, asymptomatic malaria parasitemia can be common and the presence of parasites may mean that a patient fulfils the diagnostic criteria for CM, when in fact another cause exists.

The diagnosis of bacterial meningitis is often more difficult. Cerebrospinal fluid (CSF) culture is the gold standard. Other surrogate markers include CSF pleocytosis with neutrophil predominance, low CSF glucose, and increased total CSF protein concentration. However, all these require laboratory facilities. CSF culture takes almost 48 hours, and although highly specific, has about 80% sensitivity, which is reduced when antibiotics have been given prior to sampling[6,7]. Thus, distinguishing patients with ABM can be difficult in malaria-endemic areas[8–10]. Consequences of failing to adequately treat ABM are increased risks of death and severe neuro-disability[11]. On the other hand, unnecessary use of antibiotics risks escalating antimicrobial resistance[12]. Therefore, a fast and reliable biochemical marker that could be developed into a point of care test with sufficient specificity and sensitivity would facilitate clinical diagnosis and appropriate management of CNS infections. Markers of the host response to CNS infection may offer the opportunity to distinguish infection aetiology to identify ABM.

Proteomic analysis allows quantitation of a large number of proteins present in biological fluids, such as plasma and CSF, providing an opportunity for unbiased discovery of biomarkers associated with clinical phenotypes[13–17]. Examples include aetiology-specific host response signatures that distinguish pneumococcal, meningococcal, and enteroviral meningitis[18]; mortality risk in pneumococcal meningitis[16]; and CM compared to other encephalopathies[17]. The latter approach enhances proteome coverage, but generally precludes quantification of identified proteins.

In this study, we aimed to determine components of the CSF protein expression profiles of children with ABM that distinguish from those with CM with a high degree of specificity and sensitivity, which could be developed into point-of-care tests.

Methods

Study participants

The study used archived CSF samples (n = 59) from paediatric admissions (2002–2011) at Kilifi County Hospital (Formerly Kilifi District Hospital), Kilifi, Kenya. All samples used in this study had been consented and approved for storage and research by the Kenya Medical Research Institute. Samples comprised two groups of children, based on clinical and laboratory findings. Acute bacterial meningitis (ABM; n=37) was defined as children who had a positive bacterial culture for CSF. Cerebral malaria (CM; n=22) was defined as children who had peripheral asexual malarial parasites >2500 parasites/µl on blood film[19], negative CSF cultures, CSF leukocyte count <10 cells/µl and no CSF biochemical feature of ABM. ABM was defined without respect to parasitemia, and therefore 5 children had *Plasmodium falciparum* coincidental infection.

Sample preparation and LC-MS/MS analysis

Aliquots of 10 µl of CSF were denatured in 50 mM ammonium bicarbonate (Fluka) containing 8 M urea (Sigma). Proteins were reduced with 20 mM dithiothreitol (Sigma) at room temperature with shaking for 1 hour (h) and subsequently alkylated in the dark for 1 h with 65 mM iodoacetamide (Sigma). Excess iodoacetamide was quenched using 65 mM dithiothreitol. Urea present in the sample was dialyzed out with 50mM ammonium bicarbonate, using 3 kDa amicon filters (Millipore). Proteins were digested with trypsin (Thermo Scientific) overnight (16 hours) and peptides obtained were desalted using C18 Spin columns (Thermo Scientific), according to manufacturer's instructions, dried in a Speedvac concentrator (Thermo Scientific), and re-suspended in 50 µl loading solvent (97.05% H_2O, 2% acetonitrile, 0.05% formic acid). Peptides (5 µl) were loaded using a Dionex Ultimate 3000 nano-flow ultra-high-pressure liquid chromatography system (Thermo

Scientific) on to a 75µm x 2 cm C18 trap column (Thermo Scientific) and separated on a 75µm x 25 cm C18 reverse-phase analytical column (Thermo Scientific). Elution was carried out with mobile phase B (80% acetonitrile with 0.1% formic acid) gradient (5 to 35 %) over 120 min. Peptides were measured using a Q Exactive Orbitrap mass spectrometer (Thermo Scientific) coupled to the chromatography system via a nano-electrospray ion source (Thermo Scientific). The ms^1 settings were: Resolution, 70000; AGC target, 3e6; scan range, 400–1800 m/z; while the ms^2 settings were: Resolution, 17500; AGC, 5e4; isolation window, 1.6 m/z. The top 15 most intense ions were selected for ms^2, which were subsequently excluded for the next 30 s.

Data preparation

Mass spectrometer files (.Raw files) were analysed by MaxQuant software version 1.5.3.30[20] by searching against the human Uniprot FASTA database (downloaded February 2014) using the Andromeda search engine[21]. Cysteine carbamidomethylation was set as a fixed modification and N-terminal acetylation and methionine oxidations as variable modifications. The false discovery rate (FDR) was set to 0.01 for both proteins and peptides with a minimum length of seven amino acids and was determined by searching a decoy database. A decoy FASTA database is generated from the target database, comprising sequences derived from the organism being studied, by switching the amino-carboxyl orientation of a protein's amino acids to generate sequences that do not exist in nature, which are then concatenated with the target FASTA database[22]. Enzyme specificity was set as C-terminal to arginine and lysine with trypsin as the protease. A maximum of two missed cleavages were allowed in the database search. Peptide identification was performed with an allowed initial precursor mass deviation of up to 7 ppm and an allowed fragment mass deviation of up to 20 ppm. The label free quantification (LFQ) algorithm in MaxQuant was used to obtain quantification intensity values.

Statistical analysis

a) Pre-processing and exploration. Study participants characteristics data was uploaded and analyzed in Stata version 13.1 and significance was tested using Wilcoxon rank-sum (Mann-Whitney) test for non-parametric variables, while two-sample t test with equal variances was used for parametric data. Chi square test was used for binary data.

Proteome data analysis was conducted in R version 3.3.2 and all samples were included in the analysis. We limited protein analysis to those identified and quantified in at least half of the samples in either ABM or CM. Proteins that were not detected in a sample were presumed to be on the lower limit of detection and their LFQ values were set at 0. Range and logarithmic normalization was performed to adjust protein quantities to a comparable scale. Unsupervised clustering using principal component analysis (PCA) and hierarchical clustering were applied to assess variation and determine group separation (i) among the ABM samples that comprised *Haemophilus influenza* (n = 12) and *Streptococcus pneumoniae* (n = 25) and (ii) between ABM and CM. Group separation among ABM samples was carried out to determine whether the host response is bacterial-specific or generic. To visualize protein clustering patterns, a heatmap was generated with the Pearson correlation coefficients as the distance metric.

b) Biomarker extraction. Fold change (FC) and FDR were used to identify differentially expressed proteins. Selection of candidate biomarkers was performed through feature-importance assignment, based on variable importance, as implemented in the random forest (RF) algorithm version 4.6-12[23]. Here, recursive feature elimination (RFE) resulted in a reduced subset of proteins whose ability to distinguish between ABM and CM groups was evaluated using the mean decrease in accuracy (MDA) scores. The Boruta R package version 5.2.0[24] is designed as a wrapper around RF facilitating RFE and MDA weight assignment. Subsequently, each protein was assessed for its ability to discriminate between ABM and CM by evaluating its receiver operating characteristic (ROC) curves. Potential biomarkers were thus identified from the differentially expressed proteins ranked according to the area under curve (AUC) and MDA scores.

Results

Characteristics of study participants

The study analysed 59 samples from two clinical groups, whose characteristics are presented in Table 1.

Table 1. Characteristics of study participants. Data are median (interquartile range), unless otherwise stated. Abbreviations: CSF, cerebrospinal fluid; iRBC, infected red blood cell; WBC, white blood cell; MUAC, mid-upper arm circumference.

Characteristic	Acute bacterial meningitis (n=37)	Cerebral malaria (n=22)	P
Age, months	35 (9 – 90)	30.5 (13 – 37)	0.61
Sex, male, n (%)	21 (56.8)	8 (36.4)	0.1
Parasite density, iRBCs $\times 10^3/\mu L$	0 (0 – 0)	230000 (100800 – 393600)	0.0001
CSF WBC count, cells/µL	3370 (288 – 5120)	2 (1 – 4)	0.0001
Total CSF protein, mg/dL	2.1 (1.3 – 2.49)	0.28 (0.22 – 0.37)	0.0001
Blood glucose, mg/dL	5.4 (4 – 7.6)	5.3 (3.2 – 7.7)	0.9
CSF glucose, mg/dL	0.6 (0.3 – 0.9)	3.1 (2.6 – 3.7)	0.0001
Ratio of CSF to blood glucose	0.1 (0.06 – 0.16)	0.69 (0.42 – 1)	0.0001
Outcome, dead, n (%)	13 (35)	0	0.001
MUAC, cm	13.95 (11.5 – 15.4)	13.95 (13.3 – 15.2)	0.3
Seizures, n (%)	9 (24)	10(46)	0.1

CSF proteomes of ABM and CM differ significantly

The LC-MS/MS analysis resulted in the quantification of a total of 708 non-redundant proteins, of which 183 proteins were commonly expressed in both ABM and CM (Figure 1A). One hundred and sixty proteins were quantified in >50% within each group and were selected for further analysis (Figure 1B). Of the 160 proteins, 32 proteins were not quantified in CM, while two proteins were not quantified in ABM, as shown in Figure 1B. Overall, ABM had a higher number of quantified proteins than CM. All quantified proteins and those included in subsequent analysis are listed in Table S1 and Table S2, respectively.

Normalized protein LFQ values revealed clear sample separation into ABM and CM groups (Figure 2). S100A8/S100A9 (calprotectin), lactotransferrin (LTF), myeloperoxidase (MPO), and myeloblastin (PRTN3) were the top five proteins driving the sample group separation observed in dimension 1 (data not visualised). The CM samples showed strong within-cluster connectivity, suggesting lower proteome variation compared to ABM samples, which showed greater cluster spread. The ABM group comprising Gram-negative *H. influenza* and Gram-positive *S. pneumoniae* did not exhibit any bacteria-specific clustering (Figure S1).

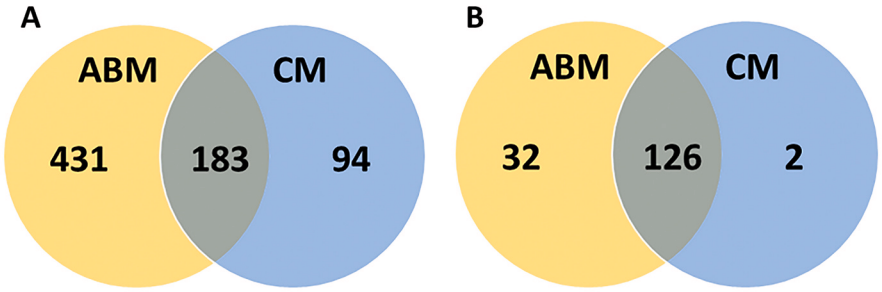

Figure 1. Cerebral spinal fluid (CSF) proteomes of acute bacterial meningitis (ABM) and cerebral malaria (CM) differ significantly. (**A**) Distribution of total proteins quantified (n = 708) between ABM and CM. (**B**) Distribution of proteins included in the biomarker analysis, where proteins had to be quantified in at least half of the samples in either ABM or CM. The CSF of ABM patients is characterized by a larger protein diversity compared to CM.

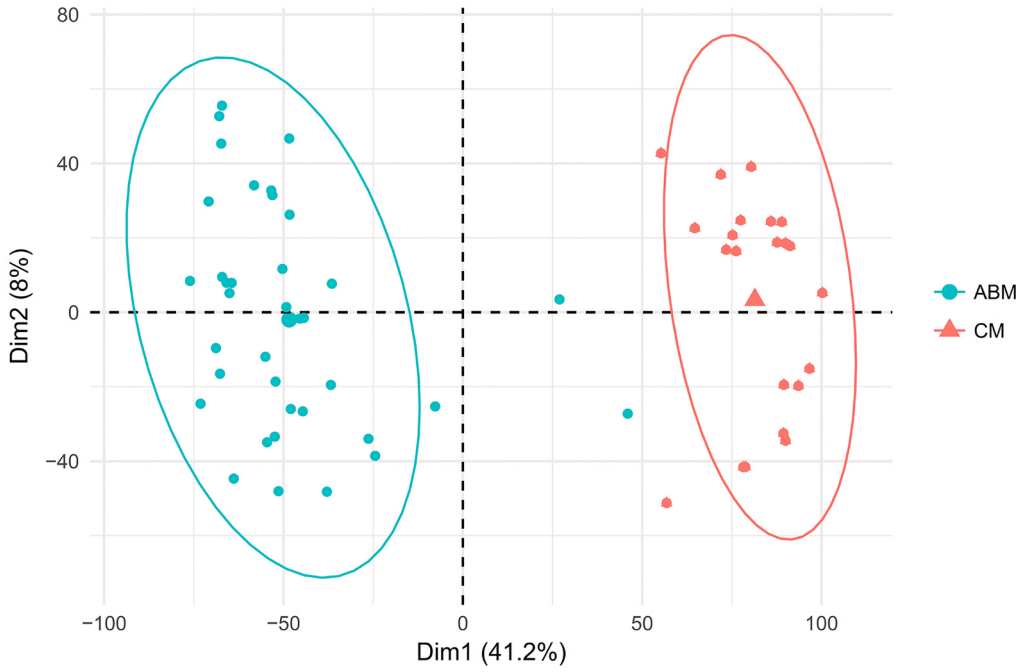

Figure 2. Host response acute bacterial meningitis (ABM) and cerebral malaria (CM) is pathogen specific. Unsupervised clustering using principal component analysis (PCA) was employed to determine clustering patterns of samples. The PCA score plot of the cerebral spinal fluid proteomes depicts clear group separation. Dimension (Dim) 1 of the PCA accounted for 41% of variation, while Dim 2 accounted for 8%.

In order to visualize clustering patterns of proteins based on normalized LFQ quantities, a heatmap with the rows, representing 160 proteins, and columns, representing 59 samples, was generated with the Pearson correlation coefficients as the distance metric (Figure 3). The protein expression profiles distinguished the two groups (Figure 3). It is notable that two ABM samples clustered with the CM group and this was similarly observed with the PCA analysis (Figure 2 and Figure 3).

Analysis revealed that 52 proteins were differentially expressed (FDR<0.01, Log FC≥2) (Table 2, Figure S2), of which 83% (43/52) were upregulated in ABM compared to CM (Figure S2). Proteins including MPO, LTF, PRTN3, PFN1, LCN2, MMP8, MMP9, RETN, PGLYRP1, and S100A8/S100A9 were among the significantly expressed proteins in ABM, while SPARC, CNTN1, CHGB, and SPARCL1 were upregulated in CM (Table 2).

MPO and LTF as the best biomarkers distinguishing ABM from CM

To identify top ranking biomarkers that distinguish ABM from CM, differentially expressed proteins were subjected to a feature-based weighting procedure where protein importance was assigned using the MDA scores. Higher scores imply increased ability of a protein to distinguish between ABM and CM (Figure S3). The proteins selected by the algorithm as the most important biomarkers are listed on Table 3, where MPO, SPARCL1, and LTF rank as the top three proteins.

Consistently, there was an overlap of differentially expressed proteins and the top ranking biomarker proteins. Additionally, ROC curves were generated independently for each of the biomarkers and the AUC determined. The top ranking biomarkers were selected based on an AUC >0.9 and MDA scores above those of

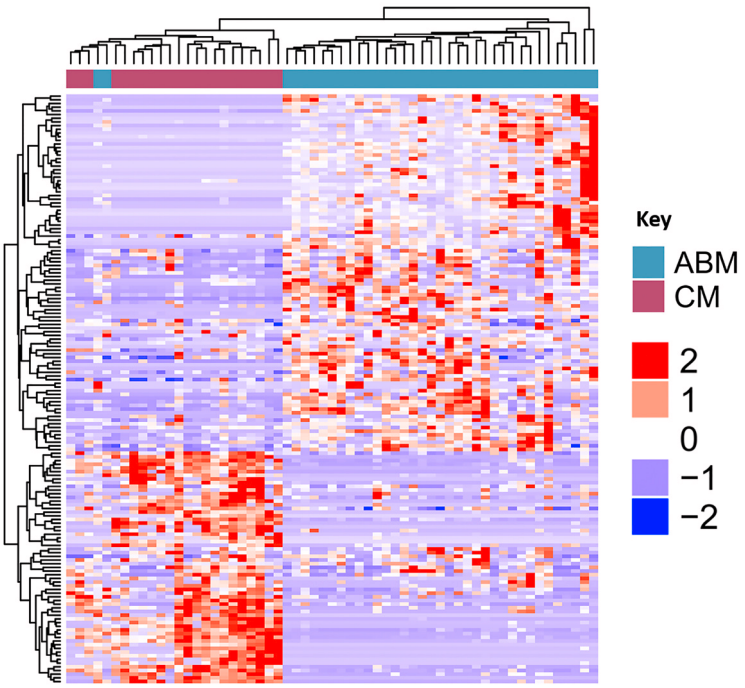

Figure 3. Heatmap demonstrating sample clustering based on protein expression profiles from acute bacterial meningitis (ABM) and cerebral malaria (CM). The heatmap was generated using hierarchical clustering based on protein expression levels calculated from normalized label free quantification values (2 and -2). The Pearson correlation coefficients were used as the distance metric. Rows represent individual proteins, while columns represents samples. Red indicates upregulation and blue indicates downregulation. Distinct sample clustering based on protein expression levels was observed with clear separation between the ABM and CM, except for two ABM samples that clustered with the CM.

Table 2. List of 52 quantified proteins that showed differential expression (FDR<0.01, and log FC≥2). FDR, false discovery rate; FC, fold change.

Protein ID	Protein name	Gene name	Log FC	*P* value	FDR
P05164-2	Myeloperoxidase	MPO	-31.7	8.32E-77	1.33E-74
P02788	Lactotransferrin	LTF	-32.8	5.05E-66	4.04E-64
U3KPS2	Myeloblastin	PRTN3	-30.8	9.89E-42	5.27E-40
P07737	Profilin-1	PFN1	-30.7	1.21E-37	4.85E-36
P29401	Transketolase	TKT	-30.6	1.23E-33	3.94E-32
P08670	Vimentin	VIM	-32	1.7E-28	4.54E-27
P13796	Plastin-2	LCP1	-32	1.41E-27	3.22E-26
P12814	Alpha-actinin-1	ACTN1	-30.5	1.32E-22	2.65E-21
P31146	Coronin-1A;Coronin	CORO1A	-29.6	3.36E-22	5.98E-21
P80188	Neutrophil gelatinase-associated lipocalin	LCN2	-30.6	2.67E-19	4.27E-18
P22894	Neutrophil collagenase	MMP8	-29.4	8.23E-19	1.2E-17
P09486	SPARC	SPARC	28.5	5.41E-18	7.21E-17
Q12860	Contactin-1	CNTN1	28.6	5.91E-18	7.27E-17
Q9HD89	Resistin	RETN	-29.9	1.11E-17	1.27E-16
P06744	Glucose-6-phosphate isomerase	GPI	-29.8	1.57E-17	1.68E-16
P14780	Matrix metalloproteinase-9	MMP9	-29.4	2.95E-17	2.95E-16
P80511	Protein S100-A12	S100A12	-31.2	1.8E-15	1.7E-14
P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA	-30.1	3.59E-15	3.19E-14
P31949	Protein S100-A11	S100A11	-28.9	2.6E-14	2.19E-13
P26038	Moesin	MSN	-29.4	4.99E-13	3.99E-12
P11142	Heat shock cognate 71 kDa protein	HSPA8	-29.3	6.65E-13	5.07E-12
P06753-2	Tropomyosin alpha-3 chain	TPM3	-29.4	3.48E-12	2.53E-11
P60660-2	Myosin light polypeptide 6	MYL6	-28.9	5.27E-12	3.67E-11
P62158	Calmodulin	CALM1	-29.3	1.94E-11	1.3E-10
P08107	Heat shock 70 kDa protein 1A/1B	HSPA1A	-29.5	3.34E-11	2.14E-10
P04003	C4b-binding protein alpha chain	C4BPA	-28.5	1.32E-10	8.12E-10
O75594	Peptidoglycan recognition protein 1	PGLYRP1	-28.1	1.68E-10	9.93E-10
Q93079	Histone H2B type 1-H	HIST1H2BH	-30	2.73E-10	1.56E-09
Q01518	Adenylyl cyclase-associated protein 1	CAP1	-29.4	7.17E-10	3.95E-09
P04114	Apolipoprotein B-100; Apolipoprotein B-48	APOB	-28.3	1.07E-09	5.73E-09
P04083	Annexin A1	ANXA1	-29	3.69E-09	1.9E-08
P07900	Heat shock protein HSP 90-alpha	HSP90AA1	-28.9	4.54E-09	2.2E-08
P35579	Myosin-9	MYH9	-30.2	4.52E-09	2.2E-08
O43866	CD5 antigen-like	CD5L	-27.2	2.32E-08	1.07E-07
P00338	L-lactate dehydrogenase A chain	LDHA	-28.4	2.34E-08	1.07E-07

Protein ID	Protein name	Gene name	Log FC	*P* value	FDR
P02649	Apolipoprotein E	APOE	4.5	1.25E-05	5.56E-05
P36955	Pigment epithelium-derived factor	SERPINF1	3.5	1.45E-05	6.27E-05
P05090	Apolipoprotein D	APOD	2.2	2.54E-05	0.0001
P02766	Transthyretin	TTR	3.9	4.57E-05	0.0002
P06702	Protein S100-A9	S100A9	-9	5.58E-05	0.0002
P05109	Protein S100-A8	S100A8	-8.9	0.000161	0.0006
P23142-4	Fibulin-1	FBLN1	3.6	0.000219	0.0008
P02675	Fibrinogen beta chain	FGB	-4.9	0.000399	0.0014
P02679-2	Fibrinogen gamma chain	FGG	-4.7	0.000689	0.0025
P00738	Haptoglobin	HP	-5.2	0.000988	0.0032
P01034	Cystatin-C	CST3	3.7	0.001006	0.0032
P05060	Secretogranin-1	CHGB	7.9	0.000988	0.0032
P60709	Actin	ACTB	-5.9	0.000949	0.0032
P41222	Prostaglandin-H2 D-isomerase	PTGDS	4	0.001688	0.0052
P06733	Alpha-enolase	ENO1	-9	0.001814	0.0052
Q14515	SPARC-like protein 1	SPARCL1	8	0.00214	0.0063
P59666	Neutrophil defensin 3	DEFA3	-6.6	0.00245	0.0071

Table 3. Area under the curve (AUC), sensitivity (Sens.), specificity (Spec.), and mean decrease in accuracy (MDA) scores for the best performing biomarkers.

Biomarker	AUC (95% CI)	Sens.	Spec.	Classified correctly %	MDA
MPO	1.00 (1 to 1)	1	1	100	8.7
LTF	0.98 (0.96 to 1)	0.97	1	98	6.9
PRTN3	0.96 (0.91 to 1)	0.92	1	95	5.6
PFN1	0.96 (0.91 to 1)	0.92	1	95	5
TKT	0.95 (0.89 to 0.99)	0.89	1	93	4.1
VIM	0.91 (0.86 to 0.97)	0.83	1	90	2.98
LCP1	0.90 (0.84 to 0.97)	0.81	1	88	2.97
ACTN1	0.90 (0.84 to 0.97)	0.81	1	88	2.76
CORO1A	0.90 (0.84 to 0.97)	0.81	1	88	3.16
CSF Glucose (mg/dL) <2.4*	0.88 (0.80 to 0.99)	0.91	0.87	88	-
CSF WBC count, cells/μL >10*	0.95 (0.89 to 1)	1	0.90	97	-
Total CSF protein >0.54*	0.94 (0.87 to 1)	0.97	0.90	95	-

*indicates the best cut-off that achieved high sensitivity and specificity.

shadow proteins (Table 3). However, among the biomarkers, MPO and LTF achieved high sensitivity (≥0.98) and specificity (1), depicting their predictive potential as biomarkers (Table 3).

Discussion

Clinical differentiation of ABM and CM is important as it dictates management and prognosis. A diagnostic test based on host proteins avoids heterogeneity in pathogen proteins between infecting bacterial species. We found significant proteome difference between ABM and CM, implying that the host responds differently to bacterial and *Plasmodial* infections. This is consistent with previous work indicating a differential host response in plasma and CSF of children with a diagnosis of CM compared to those with a malaria-slide-negative ABM[17]. In a previous study, Gitau *et al*. reported differentially expressed proteins in plasma and CSF from children with CM, ABM, and nonspecific encephalopathies[17]. The approach enhanced proteome coverage through pre-fractionation, but precluded relative quantification of proteins. Biomarker discovery has been enhanced by recent developments in mass spectrometry instrumentation[25] and advanced computational and bioinformatics algorithms[20,26]. In the present study, we performed shotgun proteomics and quantitative differences in host proteins in CSF.

ABM was characterized by a higher CSF total protein concentration and higher host protein diversity. This likely results from protein infiltration following breakdown in the blood brain barrier and secretion from host cells including infiltrating neutrophils. Whilst in CM the blood brain barrier is mildly impaired with few morphological changes[27–30] and *Plasmodia* parasites are usually restricted to the vascular compartment rather than the meninges or the parenchyma of the brain, unless haemorrhage occurs.

The aetiology of ABM comprised Gram-negative *H. influenza* and Gram-positive *S. pneumoniae*. However, there was no species-specific clustering, implying that host responses to bacterial infections is largely generic. This homogeneity supports using host biomarkers for distinguishing meningitis.

A large proportion of the proteins that were not consistently quantified were excluded from the main analysis where biomarker mining followed criterion to include only proteins quantified in more than half of the samples from either of the two groups. This increased the reliability of a selected biomarker. Reducing the dataset in this way resulted in a substantial increase in proteins shared between the two groups, with ABM retaining a higher proportion of unique proteins compared to CM.

MPO and LTF were the most promising proteins for consideration as biomarkers based on their expression, sensitivity and specificity, as judged by the AUC and also ranked among the top biomarkers using RF. MPO contributes to innate host defences through microbial killing and is stored in large quantities in the azurophilic granules of neutrophils and released upon cell activation[31,32]. MPO and LTF are also part of neutrophil extracellular traps; fibrillar matrices comprised of chromatin and antimicrobial proteins, released by activated neutrophils[33]. As a biomarker, MPO has previously been shown to be higher in the CSF of patients with infectious causes compared to those of non-infectious causes, and its levels have been shown to correlate well with neutrophil counts[34]. MPO has already been developed as a diagnostic tool for cardiovascular disease risk stratification[35].

LTF is an antimicrobial polypeptide found in secondary granules of neutrophils and in human mucosal secretions, and plays a role in iron metabolism and inflammation[36,37]. LTF plays a bacteriostatic role in host defence due to its iron sequestering properties that inhibit bacterial proliferation. It is ineffective against bacteria able to acquire their iron from either LTF or transferrin[38–40]. Previously, LTF has been shown to be elevated in the CSF of patients with bacterial meningitis[41–44]. In a study in adults, CSF LTF showed diagnostic efficiency (AUC; 0.946, sensitivity; 96.6, specificity; 92.4) when distinguishing between bacterial and aseptic meningitis[45]. Such findings point to LTF as a molecular marker for ABM. Currently, LTF can be assayed using ELISA, but has the potential to be developed as a diagnostic test. It is notable that both MPO and LTF performed better than surrogate markers in the CSF, including glucose, pleocytosis, and total protein. Determination of pleocytosis requires microscopy and couldn't be a point-of-care test. Bed-side glucose rapid tests are not currently accurate enough to reliably distinguish ABM from CM.

In resource limited settings in Africa, point-of-care diagnostics could considerably help in diagnosis. In malarious areas, a rapid test sensitive and specific for ABM could help in identification and timely management of comatose children suffering from ABM even though they may have a positive malaria slide or RDT. Such an approach would contribute to antibiotic stewardship in the face of increasing resistance and improve resource use. The markers identified in this study await validation and development as point-of-care diagnostics.

Our study had several limitations. For ethical reasons, we lacked CSF control samples without disease, as previously described[17]. The sample size used was relatively small, which could limit comparison between groups of syndromes, and so the results should be validated using an independent cohort of study participants. Further, in endemic areas, the definition of CM is often challenging and CM is often over-diagnosed. The WHO definition of CM may misclassify up to 25% of cases[46] and its specificity is greatly improved by adding a clinical test for Malaria Retinopathy[47,48]. However, Retinal changes specific to CM require specialist examination techniques, are difficult to examine in conscious children, and such data was not available in this study. Samples analysed in this study were left over specimens following normal microbiology and chemistry laboratory procedures. There is lacking data for the samples on the time taken from collection during lumbar puncture to storage at -80°C and therefore there is possibility that proteome changes occurred during that period[49].

Label-free proteomics are cost-efficient, offer higher proteome coverage and a higher dynamic range. However, as every sample is handled separately, variations that can bias the quantitative analysis may be introduced. The results for ABM need to be validated with other aetiologies of ABM and other age groups, such as neonates, since the two pathogens are now uncommon due to the introduction of conjugate vaccines[50].

Conclusions

Children with ABM and CM have different CSF host proteomes[17]. In the present study, two neutrophil proteins, MPO and LTF, were found to be the best biomarkers to distinguish ABM and CM. They have the potential to be developed as point-of-care diagnostics following validation in an independent cohort.

Ethical statement

The samples used in this study had been collected under a previous study examining the mechanisms of neurological damage in Kenyan children with cerebral malaria and acute bacterial meningitis with the approval of KEMRI Scientific Steering Committee (protocol No. 480; by Prof. Charles R. Newton). The parents or guardians of all study participants had provided written informed consent for sample use for future study during sample collection.

Data availability

The mass spectrometry raw files generated and analysed in the current study have been deposited to the ProteomeXchange Consortium[51] (PXD006357), via the MassIVE partner repository (MSV000080979), under the following title: Cerebrospinal fluid markers to distinguish bacterial meningitis from cerebral malaria in children.

The FTP for the dataset is available here.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by the Wellcome Trust [084538], Strategic Award, [107769]; Strategic Award and Initiative to Develop African Research Leaders (IDeAL). JMN is a beneficiary of both Strategic Award and IDEAL early career post-doc, EG and MKR are beneficiaries of Strategic Award. NKK is a post-doc currently supported by IDeAL.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank the study participants, their families, and the clinical staff at the KEMRI-Wellcome Trust Programme. We would like to thank Prof. Philip Bejon and Dr. Abdirahman Abdi for critical reviews of the manuscript.

This paper is published with the permission of the Director of the Kenya Medical Research Institute.

Supplementary material

Table S1. List of quantified proteins.

Click here to access the data.

Table S2. List of proteins included in the biomarker analysis quantified proteins.

Click here to access the data.

Figure S1. Host response to *Haemophilus influenza* and *Streptococcus pneumoniae* is generic. The principal component analysis score plot depicts overlap of the cerebral spinal fluid proteomes of *H. influenza* and *S. pneumoniae.* HIN, *H. influenza*; SPN, *S. pneumoniae*.

Click here to access the data.

Figure S2. A volcano plot showing differential protein expression between acute bacterial meningitis and cerebral malaria, with the latter as the reference. FDR < 0.01 (y axis) and Log FC≥2 (x axis) were set as the significance threshold values as shown with blue lines. FDR, false discovery rate; FC, fold change.

Click here to access the data.

Figure S3. Weight assignment of protein biomarkers using the Random Forests mean decrease in accuracy scores. “Important biomarkers” were extracted using a recursive feature elimination (Boruta wrapper package). Each box plot corresponds to mean decrease in accuracy (MDA) scores for 1000 iterations. The proteins are ranked according to their mean importance score. Briefly, MDA scores for actual proteins were compared to those of shadow proteins (randomly shuffled quantities).

Click here to access the data.

References

1. WHO: **Severe falciparum malaria. World Health Organization, Communicable Diseases Cluster.** *Trans R Soc Trop Med Hyg.* 2000; **94**(Suppl 1): S1–90.
PubMed Abstract
2. Milner DA Jr, Whitten RO, Kamiza S, *et al.*: **The systemic pathology of cerebral malaria in African children.** *Front Cell Infect Microbiol.* 2014; **4**: 104.
PubMed Abstract | Publisher Full Text | Free Full Text
3. MacCormick IJ, Beare NA, Taylor TE, *et al.*: **Cerebral malaria in children: using the retina to study the brain.** *Brain.* 2014; **137**(Pt 8): 2119–2142.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Kariuki SM, Gitau E, Gwer S, *et al.*: **Value of *Plasmodium falciparum* Histidine-Rich Protein 2 level and malaria retinopathy in distinguishing cerebral malaria from other acute encephalopathies in Kenyan children.** *J Infect Dis.* 2014; **209**(4): 600–609.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Seydel KB, Fox LL, Glover SJ, *et al.*: **Plasma concentrations of parasite Histidine-Rich Protein 2 distinguish between retinopathy-positive and retinopathy-negative cerebral malaria in Malawian children.** *J Infect Dis.* 2012; **206**(3): 309–318.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Nigrovic LE, Malley R, Macias CG, *et al.*: **Effect of antibiotic pretreatment on cerebrospinal fluid profiles of children with bacterial meningitis.** *Pediatrics.* 2008; **122**(4): 726–730.
PubMed Abstract | Publisher Full Text
7. Bohr V, Rasmussen N, Hansen B, *et al.*: **875 Cases of bacterial meningitis: Diagnostic procedures and the impact of preadmission antibiotic therapy Part III of a three-part series.** *J Infect.* 1983; **7**(3): 193–202.
PubMed Abstract | Publisher Full Text
8. Berkley JA, Mwangi I, Ngetsa CJ, *et al.*: **Diagnosis of acute bacterial meningitis in children at a district hospital in sub-Saharan Africa.** *Lancet.* 2001; **357**(9270): 1753–1757.
PubMed Abstract | Publisher Full Text
9. Berkley JA, Versteeg AC, Mwangi I, *et al.*: **Indicators of acute bacterial meningitis in children at a rural Kenyan district hospital.** *Pediatrics.* 2004; **114**(6): e713–719.
PubMed Abstract | Publisher Full Text
10. Berkley JA, Mwangi I, Mellington F, *et al.*: **Cerebral malaria versus bacterial meningitis in children with impaired consciousness.** *QJM.* 1999; **92**(3): 151–157.
PubMed Abstract | Publisher Full Text
11. Molyneux EM, Walsh AL, Forsyth H, *et al.*: **Dexamethasone treatment in childhood bacterial meningitis in Malawi: a randomised controlled trial.** *Lancet.* 2002; **360**(9328): 211–218.
PubMed Abstract | Publisher Full Text
12. Carlet J, Jarlier V, Harbarth S, *et al.*: **Ready for a world without antibiotics? The Pensières Antibiotic Resistance Call to Action.** *Antimicrob Resist Infect Control.* 2012; **1**(1): 11.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Yuan X, Desiderio DM: **Proteomics analysis of human cerebrospinal fluid.** *J Chromatogr B Analyt Technol Biomed Life Sci.* 2005; **815**(1–2): 179–189.
PubMed Abstract | Publisher Full Text
14. Jesse S, Steinacker P, Lehnert S, *et al.*: **A proteomic approach for the diagnosis of bacterial meningitis.** *PLoS One.* 2010; **5**(4): e10079.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Gitau EN, Kokwaro GO, Newton CR, *et al.*: **Global proteomic analysis of plasma from mice infected with *Plasmodium berghei* ANKA using two dimensional gel electrophoresis and matrix assisted laser desorption ionization-time of flight mass spectrometry.** *Malar J.* 2011; **10**(1): 205.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Goonetilleke UR, Scarborough M, Ward SA, *et al.*: **Proteomic analysis of cerebrospinal fluid in pneumococcal meningitis reveals potential biomarkers associated with survival.** *J Infect Dis.* 2010; **202**(4): 542–550.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Gitau EN, Kokwaro GO, Karanja H, *et al.*: **Plasma and cerebrospinal proteomes from children with cerebral malaria differ from those of children with other encephalopathies.** *J Infect Dis.* 2013; **208**(9): 1494–1503.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Cordeiro AP, Silva Pereira RA, Chapeaurouge A, *et al.*: **Comparative proteomics of cerebrospinal fluid reveals a predictive model for differential diagnosis of pneumococcal, meningococcal, and enteroviral meningitis, and novel putative therapeutic targets.** *BMC Genomics.* 2015; **16**(Suppl 5): S11.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Bejon P, Berkley JA, Mwangi T, *et al.*: **Defining childhood severe falciparum malaria for intervention studies.** *PLoS Med.* 2007; **4**(8): e251.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol.* 2008; **26**(12): 1367–1372.
PubMed Abstract | Publisher Full Text
21. Cox J, Neuhauser N, Michalski A, *et al.*: **Andromeda: A peptide search engine integrated into the MaxQuant environment.** *J Proteome Res.* 2011; **10**(4): 1794–1805.
PubMed Abstract | Publisher Full Text
22. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.** *Nat Methods.* 2007; **4**(3): 207–214.
PubMed Abstract | Publisher Full Text
23. Breiman L: **Random Forests.** *Mach Learn.* 2001; **45**(1): 5–32.
Publisher Full Text
24. Kursa MB, Rudnicki WR: **Feature selection with the Boruta package.** *J Stat Softw.* 2010; **36**(11): 1–13.
Publisher Full Text
25. Michalski A, Damoc E, Hauschild JP, *et al.*: **Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.** *Mol Cell Proteomics.* 2011; **10**(9): M111.011015.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Cox J, Hein MY, Luber CA, *et al.*: **Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.** *Mol Cell Proteomics.* 2014; **13**(9): 2513–2526.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Mauro P: **Blood-Brain Barrier and Cerebral Malaria: The Road So Far.** *J Bacteriol Parasitol.* 2014; **5**: 2155–9597.
Publisher Full Text
28. Polimeni M, Prato M: **Host matrix metalloproteinases in cerebral malaria: new kids on the block against blood–brain barrier integrity?** *Fluids Barriers CNS.* 2014; **11**(1): 1.
PubMed Abstract | Publisher Full Text | Free Full Text
29. Rénia L, Howland SW, Claser C, *et al.*: **Cerebral malaria: mysteries at the blood-brain barrier.** *Virulence.* 2012; **3**(2): 193–201.
PubMed Abstract | Publisher Full Text | Free Full Text
30. Gitau EN, Newton CR: **Review Article: Blood–brain barrier in falciparum malaria.** *Trop Med Int Health.* 2005; **10**(3): 285–292.
PubMed Abstract | Publisher Full Text
31. Klebanoff SJ: **Myeloperoxidase.** *Proc Assoc Am Physicians.* 1999; **111**(5): 383–389.
PubMed Abstract
32. Rausch PG, Pryzwansky KB, Spitznagel JK: **Immunocytochemical Identification of Azurophilic and Specific Granule Markers in the Giant Granules of Chediak-Higashi Neutrophils.** *N Engl J Med.* 1978; **298**(13): 693–698.
PubMed Abstract | Publisher Full Text
33. Brinkmann V, Reichard U, Goosmann C, *et al.*: **Neutrophil extracellular traps kill bacteria.** *Science.* 2004; **303**(5663): 1532–1535.
PubMed Abstract | Publisher Full Text
34. Michael BD, Griffiths MJ, Granerod J, *et al.*: **Characteristic Cytokine and Chemokine Profiles in Encephalitis of Infectious, Immune-Mediated, and Unknown Aetiology.** *PLoS One.* 2016; **11**(1): e0146288.
PubMed Abstract | Publisher Full Text | Free Full Text
35. Schindhelm RK, van der Zwan LP, Teerlink T, *et al.*: **Myeloperoxidase: a useful biomarker for cardiovascular disease risk stratification?** *Clin Chem.* 2009; **55**(8): 1462–1470.
PubMed Abstract | Publisher Full Text
36. Metz-Boutigue MH, Jollès J, Mazurier J, *et al.*: **Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins.** *Eur J Biochem.* 1984; **145**(3): 659–676.
PubMed Abstract | Publisher Full Text
37. Levay PF, Viljoen M: **Lactoferrin: a general review.** *Haematologica.* 1995; **80**(3): 252–267.
PubMed Abstract
38. Ratcliff B: **Iron and Infection. Molecular, Physiological and Clinical Aspects.** *Yale J Biol Med.* 1988; **61**(6): 560–561.
Free Full Text
39. Weinberg ED: **Iron withholding: a defense against infection and neoplasia.** *Physiol Rev.* 1984; **64**(1): 65–102.
PubMed Abstract
40. Arnold RR, Cole MF, McGhee JR: **A bactericidal effect for human lactoferrin.** *Science.* 1977; **197**(4300): 263–265.
PubMed Abstract | Publisher Full Text
41. Maffei FA, Heine RP, Whalen MJ, *et al.*: **Levels of antimicrobial molecules defensin and lactoferrin are elevated in the cerebrospinal fluid of children with meningitis.** *Pediatrics.* 1999; **103**(5 Pt 1): 987–992.
PubMed Abstract
42. Terent A, Hällgren R, Venge P, *et al.*: **Lactoferrin, lysozyme, and beta 2-microglobulin in cerebrospinal fluid. Elevated levels in patients with acute cerebrovascular lesions as indices of inflammation.** *Stroke.* 1981; **12**(1): 40–46.
PubMed Abstract | Publisher Full Text
43. Gutteberg TJ, Flaegstad T, Jørgensen T: **Lactoferrin, C-reactive protein, alpha-1-antitrypsin and immunoglobulin GA in cerebrospinal fluid in meningitis.** *Acta Paediatr Scand.* 1986; **75**(4): 569–572.
PubMed Abstract | Publisher Full Text
44. Lykova OF, Zakharova EV, Konysheva TV, *et al.*: **[Level of lactoferrin in serum and cerebrospinal fluid of patients with meningitis].** *Zh Mikrobiol Epidemiol Immunobiol.* 2007; (2): 80–84.
PubMed Abstract
45. Dastych M, Gottwaldová J, Čermáková Z: **Calprotectin and lactoferrin in the**

cerebrospinal fluid; biomarkers utilisable for differential diagnostics of bacterial and aseptic meningitis? *Clin Chem Lab Med.* 2015; **53**(4): 599–603.
PubMed Abstract | Publisher Full Text

46. Taylor TE, Fu WJ, Carr RA, *et al.*: **Differentiating the pathologies of cerebral malaria by *postmortem* parasite counts.** *Nat Med.* 2004; **10**(2): 143–145.
PubMed Abstract | Publisher Full Text

47. Beare NA, Taylor TE, Harding SP, *et al.*: **Malarial retinopathy: a newly established diagnostic sign in severe malaria.** *Am J Trop Med Hyg.* 2006; **75**(5): 790–797.
PubMed Abstract | Free Full Text

48. Beare NA, Lewallen S, Taylor TE, *et al.*: **Redefining cerebral malaria by including malaria retinopathy.** *Future Microbiol.* 2011; **6**(3): 349–355.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Ranganathan S, Polshyna A, Nicholl G, *et al.*: **Assessment of Protein Stability in Cerebrospinal Fluid Using Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Protein Profiling.** *Clin Proteomics.* 2006; **2**(1–2): 91–101.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Brouwer MC, Tunkel AR, van de Beek D: **Epidemiology, diagnosis, and antimicrobial treatment of acute bacterial meningitis.** *Clin Microbiol Rev.* 2010; **23**(3): 467–492.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Vizcaíno JA, Deutsch EW, Wang R, *et al.*: **ProteomeXchange provides globally coordinated proteomics data submission and dissemination.** *Nat Biotechnol.* 2014; **32**(3): 223–226.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 10 November 2017

https://doi.org/10.21956/wellcomeopenres.13799.r26405

✔ **Clarissa Valim**
Department of Osteopathic Medical Specialties, Michigan State University, East Lansing, MI, USA

The authors response and revisions addressed all my concerns. I have no further comments.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 27 September 2017

https://doi.org/10.21956/wellcomeopenres.13799.r26406

✔ **Wilson L. Mandala**
[1] Malaria Immunology Group, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
[2] Academy of Medical Sciences, Malawi University of Science and Technology, Thyolo, Malawi

I am satisfied with the changes made after my initial review. All concerns have been addressed appropriately and satisfactorily. I now approve this manuscript for publication without any reservations.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Malaria Immunology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 14 August 2017

https://doi.org/10.21956/wellcomeopenres.12925.r24112

?

Clarissa Valim
Department of Osteopathic Medical Specialties, Michigan State University, East Lansing, MI, USA

The article by Njunge *et al*. is a very well written report of a sound investigation about biomarkers to differentiate bacterial encephalitis from cerebral malaria in children. The findings of this study, set keystones in the research agenda to develop a so needed point-of-care test for pediatric encephalitis in Africa. We have two major comments and some minor ones.

Major

1. The study was conducted with samples collected during 9 years. During these years, were samples from only 59 patients collected and archived? Given the importance of selection bias in biomarker studies, it would be important for the authors to provide details of the criteria used to select patients/samples for this study. Perhaps they could also provide a flowchart of enrollment in the supplementary material.

2. The authors may want to provide minimal details of the collection and storage of their study samples, since Mass Spec proteomics can be very sensitive to details in the collection process including elapsed time to centrifuge and refrigerate samples. If not many details are known, given the age of these samples, it would be helpful to add to the discussion a limitation specifying that some proteins could have been degraded and thus, could not be properly quantified.

   To better understand the importance of selection bias to the author's results, it would also be helpful to see a distribution of samples by the year they were collected. Hopefully, it can be shown that bacteria samples were collected contemporaneous to malaria samples.

3. The authors state that proteins with more than 50% of samples missing quantification were excluded from the analysis but those that could be quantified in up to 50% of samples a value set to 0 when they could not be quantified. That is fine but it would be helpful to know what was the percent of samples that could not be quantified in each of the two comparison groups for the most important markers (perhaps those included in Table S2). Have the authors considered that an alternative explanation for the non-quantification was that the samples were degraded?

Minor

1. Given the importance of the case classification criteria in this study, the authors may want to provide a few more details of the non-ABM group. They were classified as cerebral malaria but how strong was the evidence of the presence of cerebral malaria in this group? Could they have viral encephalitis, for instance, or a non-detected ABM? Could the authors provide their rational to state that this group represented patients with CM? Regardless of the presence of CM in the group without ABM, the results of the study are relevant since they show biomarkers that identified ABM and non-ABM (or unlikely ABM) patients.

2. The authors may want to include in Table 3, the optimal cut-off value of the proteins that achieve the reported sensitivity, specificity. It would also be helpful for the top 10 or so proteins to see the boxplots comparing the two diagnosis groups to have a better sense of the variability.

3. The authors may want to reconsider and exclude the ROC curve from Figure 4. The curve of an AUC-ROC of approximately 1 is not very illustrative. Also, although the actual accuracy of the identified proteins is probably very high, the AUC-ROC estimated by the authors was probably influenced by the fact that proteins were selected and had their accuracy estimated in the same samples. The curve in this paper is probably giving more weight to the estimated accuracy than it would be necessary.

4. We have a few statistical suggestions for a future study conducted by the authors. However, the authors shall feel free to ignore those suggestions in this study. These different analysis would be unlikely to change conclusions and we believe the results of the study shall be made public and not wait for any additional analysis. The suggestions are:

   a) Wilcoxon and t-tests assuming equal variances were used to compare protein in the two comparison groups. Often the variance of proteins is very different across the two groups and none of these two tests are adequate when the variances vary across comparison groups.

   b) If the number of samples with zeros were substantial, a more appropriate approach would be using Zero inflated or Zero truncated models to compare proteins.

   c) In the future, the authors may want to explore using multiple imputation if missingness is substantial

   d) In the future the authors may want to explore additional feature selection procedures. Random forests are excellent algorithms to create classification signatures but may not be the most appropriate algorithm to select markers. Moreover, conducting a few different algorithms and evaluating consensus may help to decrease a bit the impact of overfitting in each one of these algorithms.

   e) In future studies the authors may explore using re-sampling to estimate accuracy, for instance, using cross-validation. That may help to discount the overestimation in accuracy that occurs when there is neither a validation nor a test set.

5. We could not find any “phenotype” or metadata associated with the protein data. Without that, it would not be possible to reproduce the analysis presented in this paper. I do not consider any action is necessary to address this comment since the authors made available what nearly all

investigators do.

6. Alternative explanations for the extremely high estimated AUC-ROC could be inserted in the discussion, such as issues with sample storage and selection that could have differentially affected some proteins (that have different degradation rates), and the impact of misclassification of patients in the CM group.

7. My understanding is that the study used a label-free Mass Spec. To prevent problems with the Mass Spec community, the authors may want to add a comment to the discussion about limitations of the technology used as opposed to a labelled quantitative approach. This is not terribly relevant and it is fine if the authors choose to ignore this comment.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biomarkers in bacterial infection, Epidemiology of malaria, Biostatistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Sep 2017

James Njunge, Centre for Geographic Medicine Research Coast, Kilifi, Kenya

Major

1. The study was conducted with samples collected during 9 years. During these years, were samples from only 59 patients collected and archived? Given the importance of selection bias in biomarker studies, it would be important for the authors to provide details of the criteria used to select patients/samples for this study. Perhaps they could also provide a flowchart of enrollment in the supplementary material.

We provide below bar plots of sample distribution per year. This was a discovery study for biomarkers and we randomly selected samples that had left over CSF samples and fitting study inclusion. We selected samples from a database of approximately 3,420 CSF samples that includes ABM, CM, and unknown aetiology as well as others. Since the proteomics approach used was label free protein quantification, the number of sample that could be used for the discovery study was limited. The study used samples collected over 9 years, although the distribution of these samples for CM was in a span of 4 years (2003 – 2006; Figure 1). For ABM, samples included mainly originated from the years (2002 – 2005 and 2007) (Figure 1.).

2. The authors may want to provide minimal details of the collection and storage of their study samples, since Mass Spec proteomics can be very sensitive to details in the collection process including elapsed time to centrifuge and refrigerate samples. If not many details are known, given the age of these samples, it would be helpful to add to the discussion a limitation specifying that some proteins could have been degraded and thus, could not be properly quantified.

In the limitation section, the following statement has been included "Samples analyzed in this study were left over specimens following normal microbiology and chemistry laboratory procedures. There is lacking data for the samples on the time taken from collection during lumbar puncture to storage at -80°C and therefore there is possibility that proteome changes occurred during that period [1]"

To better understand the importance of selection bias to the author's results, it would also be helpful to see a distribution of samples by the year they were collected. Hopefully, it can be shown that bacteria samples were collected contemporaneous to malaria samples.

Figure 1: Yearly distribution of ABM and CM samples included in the study

3. The authors state that proteins with more than 50% of samples missing quantification were excluded from the analysis but those that could be quantified in up to 50% of samples a value set to 0 when they could not be quantified. That is fine but it would be helpful to know what was the percent of samples that could not be quantified in each of the two comparison groups for the most important markers (perhaps those included in Table S2). Have the authors considered that an alternative explanation for the non-quantification was that the samples were degraded? *The percentages of proteins quantified per group has been provided in Table S2 as requested.*

In the limitation section, the following statement has been included "Samples analyzed in this study were left over specimens following normal microbiology and chemistry laboratory procedures. There is lacking data for the samples on the time taken from collection during lumbar puncture to storage at -80°C and therefore there is possibility that proteome changes occurred during that period [1]".

Minor

1. Given the importance of the case classification criteria in this study, the authors may want to provide a few more details of the non-ABM group. They were classified as cerebral malaria but how strong was the evidence of the presence of cerebral malaria in this group? Could they have viral encephalitis, for instance, or a non-detected ABM? Could the authors provide their rational to

state that this group represented patients with CM? Regardless of the presence of CM in the group without ABM, the results of the study are relevant since they show biomarkers that identified ABM and non-ABM (or unlikely ABM) patients.

The following text has been included in the limitation section. "Further, in endemic areas, the definition of CM is often challenging and CM is often over-diagnosed. The WHO definition of CM may misclassify up to 25% of cases [2] and its specificity is greatly improved by adding a clinical test for Malaria Retinopathy [3, 4]. However, Retinal changes specific to CM require specialist examination techniques, are difficult to examine in conscious children, and such data was not available in this study."

2. The authors may want to include in Table 3, the optimal cut-off value of the proteins that achieve the reported sensitivity, specificity. It would also be helpful for the top 10 or so proteins to see the boxplots comparing the two diagnosis groups to have a better sense of the variability.

Bar plots have been generated as below for review purposes

Figure 2: SPARC
Figure 3: PRTN3
Figure 4: PFN1
Figure 5: MPO
Figure 6: LTF
Figure 7: LCP1
Figure 8: CORO1A
Figure 9: ACTN1
Figure 10: SPARCL1

3. The authors may want to reconsider and exclude the ROC curve from Figure 4. The curve of an AUC-ROC of approximately 1 is not very illustrative. Also, although the actual accuracy of the identified proteins is probably very high, the AUC-ROC estimated by the authors was probably influenced by the fact that proteins were selected and had their accuracy estimated in the same samples. The curve in this paper is probably giving more weight to the estimated accuracy than it would be necessary.

The AUR-ROC curve has been removed

4. We have a few statistical suggestions for a future study conducted by the authors. However, the authors shall feel free to ignore those suggestions in this study. These different analysis would be unlikely to change conclusions and we believe the results of the study shall be made public and not wait for any additional analysis. The suggestions are:

a) Wilcoxon and t-tests assuming equal variances were used to compare protein in the two comparison groups. Often the variance of proteins is very different across the two groups and none of these two tests are adequate when the variances vary across comparison groups.

b) If the number of samples with zeros were substantial, a more appropriate approach would be using Zero inflated or Zero truncated models to compare proteins.

c) In the future, the authors may want to explore using multiple imputation if missingness is

substantial

d) In the future the authors may want to explore additional feature selection procedures. Random forests are excellent algorithms to create classification signatures but may not be the most appropriate algorithm to select markers. Moreover, conducting a few different algorithms and evaluating consensus may help to decrease a bit the impact of overfitting in each one of these algorithms.

e) In future studies the authors may explore using re-sampling to estimate accuracy, for instance, using cross-validation. That may help to discount the overestimation in accuracy that occurs when there is neither a validation nor a test set.

We will consider these suggestion which are very appropriate in our biomarker and related work.

5. We could not find any “phenotype” or metadata associated with the protein data. Without that, it would not be possible to reproduce the analysis presented in this paper. I do not consider any action is necessary to address this comment since the authors made available what nearly all investigators do.

6. Alternative explanations for the extremely high estimated AUC-ROC could be inserted in the discussion, such as issues with sample storage and selection that could have differentially affected some proteins (that have different degradation rates), and the impact of misclassification of patients in the CM group.

This has been addressed by excluding the AUC-ROC (Figure 4). We have included additional text in the limitation section of the discussion that address challenges (1) with the classification of CM and (2) sample storage.

7. My understanding is that the study used a label-free Mass Spec. To prevent problems with the Mass Spec community, the authors may want to add a comment to the discussion about limitations of the technology used as opposed to a labelled quantitative approach. This is not terribly relevant and it is fine if the authors choose to ignore this comment.

We have included the following additional text in the limitation section “Label-free proteomics are cost-efficient, offer higher proteome coverage and a higher dynamic range. However, as every sample is handled separately, variations that can bias the quantitative analysis may be introduced.”

Competing Interests: None

Reviewer Report 07 July 2017

https://doi.org/10.21956/wellcomeopenres.12925.r23986

Wilson L. Mandala
[1] Malaria Immunology Group, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
[2] Academy of Medical Sciences, Malawi University of Science and Technology, Thyolo, Malawi

This is a very well written paper which is probably a follow-up to what the group had previously done (Gitao *et al.*, 2013)[1]. I only have the following reservations;

Methods section

1. The study analysed archived CSF samples which had been collected over a period spanning almost nine years. The authors need to say more on how the samples were originally collected, how long it took before they were stored and at what temperature. They may also need to add this as a limitation in the Discussion section bearing in mind that the stability of CSF proteins is compromised if kept under different temperatures for some time (Ranganathan *et al.*, 2006[2]).

2. Definition of CM: although archive CSF samples were used, i have strong reservation with the criteria used in defining CM cases. They may need to include this as a limitation and cite recent papers on CM classification in addition to Milner *et al.,* 2014[3] (Beare *et al.*, 2011[4], Severe Malaria, 2014[5], Seydel *et al.*, 2015[6])

Discussion section

1. The statement "whilst in CM the blood brain barrier is MILDLY impaired....." may need to be revisited and supported by what has been covered in some recent reviews (Renia *et al.,* 2012[7], Polimeni and Prato, 2014[8], Prato, 2014[9] and what has been published based on ECM in mice models (Poh *et al.,* 2014[10]).)

2. Minor: in Figs 2 and S1, there are some points that seem to overlap (i.e. especially for CM in Fig2). Can these be explained in the Fig legends. For consistency it might also be better to keep the type of labels uniform in all Figs (i.e. if CM is red dot, then it stays that way in all Figs, easier to follow).

3. They may need to double check that in the Results section the number of Tables and Figures being referred to are correct.

References

1. Gitau EN, Kokwaro GO, Karanja H, Newton CR, Ward SA: Plasma and cerebrospinal proteomes from children with cerebral malaria differ from those of children with other encephalopathies.*J Infect Dis*. 2013; **208** (9): 1494-503 PubMed Abstract | Publisher Full Text
2. Ranganathan S, Polshyna A, Nicholl G, Lyons-Weiler J, Bowser R: Assessment of Protein Stability in Cerebrospinal Fluid Using Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Protein Profiling.*Clin Proteomics*. 2006; **2** (1-2): 91-101 PubMed Abstract | Publisher Full Text
3. Milner DA, Whitten RO, Kamiza S, Carr R, Liomba G, Dzamalala C, Seydel KB, Molyneux ME, Taylor TE: The systemic pathology of cerebral malaria in African children.*Front Cell Infect Microbiol*. 2014; **4**: 104 PubMed Abstract | Publisher Full Text
4. Beare NA, Lewallen S, Taylor TE, Molyneux ME: Redefining cerebral malaria by including malaria retinopathy.*Future Microbiol*. 2011; **6** (3): 349-55 PubMed Abstract | Publisher Full Text
5. Severe Malaria. *Trop Med Int Health*. 2014; **19** (s1): 7-131 Publisher Full Text

6. Seydel KB, Kampondeni SD, Valim C, Potchen MJ, Milner DA, Muwalo FW, Birbeck GL, Bradley WG, Fox LL, Glover SJ, Hammond CA, Heyderman RS, Chilingulo CA, Molyneux ME, Taylor TE: Brain swelling and death in children with cerebral malaria.*N Engl J Med*. 2015; **372** (12): 1126-37 PubMed Abstract | Publisher Full Text
7. Rénia L, Howland SW, Claser C, Gruner C, Suwanarusk R, Hui Teo T, Russell B, Ng LF: Cerebral malaria: mysteries at the blood-brain barrier. *virulence*. 2012; **3** (2): 193-201 PubMed Abstract | Free Full Text | Publisher Full Text
8. Polimeni M, Prato M: Host matrix metalloproteinases in cerebral malaria: new kids on the block against blood-brain barrier integrity?. *Fluids Barriers CNS*. 2014; **11** (1): 1 PubMed Abstract | Publisher Full Text
9. Prato M: Blood-Brain Barrier and Cerebral Malaria: The Road So Far. *Journal of Bacteriology & Parasitology*. 2014; **05** (02). Publisher Full Text
10. Poh CM, Howland SW, Grotenbreg GM, Rénia L: Damage to the blood-brain barrier during experimental cerebral malaria results from synergistic effects of CD8+ T cells with different specificities. *Infect Immun*. 2014; **82** (11): 4854-64 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Malaria Immunology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Sep 2017

James Njunge, Centre for Geographic Medicine Research Coast, Kilifi, Kenya

Methods section

1. The study analysed archived CSF samples which had been collected over a period spanning almost nine years. The authors need to say more on how the samples were originally collected,

how long it took before they were stored and at what temperature. They may also need to add this as a limitation in the Discussion section bearing in mind that the stability of CSF proteins is compromised if kept under different temperatures for some time (Ranganathan *et al.*, 2006[2]).

The following text has been included in the limitation section. "Samples analyzed in this study were left over specimens following normal microbiology and chemistry laboratory procedures. There is lacking data on the time taken from lumbar puncture to freezing at -80°C for the samples and therefore there is possibility for proteome changes before sample storage."

2. Definition of CM: although archive CSF samples were used, i have strong reservation with the criteria used in defining CM cases. They may need to include this as a limitation and cite recent papers on CM classification in addition to Milner *et al.,* 2014[3] (Beare *et al.*, 2011[4], Severe Malaria, 2014[5], Seydel *et al.*, 2015[6])

The following text has been included in the limitation section. "Further, in endemic areas, the definition of CM is often challenging and CM is often over-diagnosed. The WHO definition of CM may misclassify up to 25% of cases [2] and its specificity is greatly improved by adding a clinical test for Malaria Retinopathy [3, 4]. However, Retinal changes specific to CM require specialist examination techniques, are difficult to examine in conscious children, and such data was not available in this study."

Discussion section
1. The statement "whilst in CM the blood brain barrier is MILDLY impaired....." may need to be revisited and supported by what has been covered in some recent reviews (Renia *et al.,* 2012[7], Polimeni and Prato, 2014[8], Prato, 2014[9] and what has been published based on ECM in mice models (Poh *et al.,* 2014[10]).)

References that support mild impairment to the blood brain barrier during CM compared to bacterial and viral infections have been added as suggested

2. Minor: in Figs 2 and S1, there are some points that seem to overlap (i.e. especially for CM in Fig2). Can these be explained in the Fig legends. For consistency it might also be better to keep the type of labels uniform in all Figs (i.e. if CM is red dot, then it stays that way in all Figs, easier to follow).

Overlap of the figures removed in Figure S1 – the CM group has been excluded for clarity

3. They may need to double check that in the Results section the number of Tables and Figures being referred to are correct.

This has been thoroughly checked

Competing Interests: No competing interests were disclosed.