Defining retinal phenotypes in cerebral malaria using unsupervised clustering

Kyle J. Wilson [1,2,3], Owen K. Banda [1,3], Tusekile Phiri [4], Harold Nkume [4], Terrie E. Taylor [4,5], Nicole O'Brien [4,6], Karl B. Seydel [4,5], Christopher A. Moxon [2,3], Nicholas A. V. Beare [1,3]

1. University of Liverpool 2. University of Glasgow 3. Malawi-Liverpool-Wellcome Programme 4. Blantyre Malaria Project 5. Michigan State University 6. Ohio State University

Background

Cerebral malaria (CM) is a severe neurological complication of malaria infection characterized by coma in the presence of confirmed malaria infection.[1] It is associated with a specific retinopathy, which has both diagnostic and prognostic significance.[2,3,4]

Malarial retinopathy (MR) is characterized by haemorrhages, ischaemic whitening and vessel discolouration (see Fig 1A and B, below).[5] However, not all malarial retinopathy is equal – retinal appearances can differ between individuals (Fig1C and D).

Our group aimed to:
a) Describe novel retinal phenotypes in malarial retinopathy
b) Determine if retinal phenotypes correspond to underlying differing underlying immunopathogeneses

Methods

We performed unsupervised clustering on 1027 MR+CM patients across a range of tuning parameters and rationalised the solutions to a single solution using a combination of pre-defined criteria and meta-clustering. Next, we trained a Decision Tree classifier to improve interpretability. Finally, we quantified 372 unique inflammatory plasma proteins in a subset of 144 patients with both retinal data and plasma samples stored as part of two ongoing prospective observational cohort studies. All analyses were performed in R (v4.5.2) and Python (v3.11.5).
This study was approved by the ethics committees at the University of Liverpool and Kamuzu University of Health Sciences.

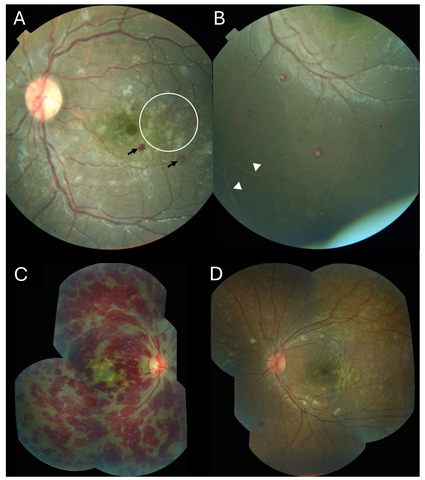

Fig 1. Malarial retinopathy. A) Retinal haemorrhages (black arrows) and ischaemic whitening (white circle and elsewhere). B) Vessel discolouration (white arrowheads). C) CM patient with highly haemorrhagic retinal presentation. D) CM patient with few haemorrhages but considerable ischaemic change.

1) WHO. World Malaria Report. World Health Organisation. Licence: CC BY-NC-SA 3.0 IGO; 2024. 2) Taylor TE, Fu WJ, Carr RA, Whitten RO, Mueller JG, Fosiko NG, et al. Differentiating the pathologies of cerebral malaria by postmortem parasite counts. Nat Med. 2004;10: 143-145. doi:10.1038/nm986 3) Wilson KJ, Liomba AM, Seydel KB, Banda CK, Moxon CA, MacCormick IJC, et al. Re-evaluating malarial retinopathy to improve its diagnostic accuracy in paediatric cerebral malaria: A retrospective study. PLOS Medicine. 2025;22: e1004727. doi:10.1371/journal.pmed.1004727 4) Beare N, Southern C, Chalira C, Taylor T, Molyneux M, Harding S. Prognostic significance and course of retinopathy in children with severe malaria. Archives of ophthalmology (Chicago, Ill : 1960). 2004;122. doi:10.1001/archopht.122.8.1141 5) Wilson KJ, Dhalla A, Meng Y, Tu Z, Zheng Y, Mhango P, et al. Retinal imaging technologies in cerebral malaria: a systematic review. Malaria Journal. 2023;22: 139. doi:10.1186/s12936-023-04566-7

Results

Defining MR phenotypes – see Fig 2 and caption.

Creating an interpretable Decision Tree – see Fig 3 and caption.

Understanding immune differences between MR phenotypes in CM – see Fig 4 and caption.

Conclusions

We described for the first time MR phenotypes in paediatric CM.

MR phenotypes can be accurately predicted using a simple clinical decision-making tool.

MR phenotypes may provide insights into the underlying immunopathogenesis per individual, thus representing a potential method for targeting adjunctive therapies for CM.

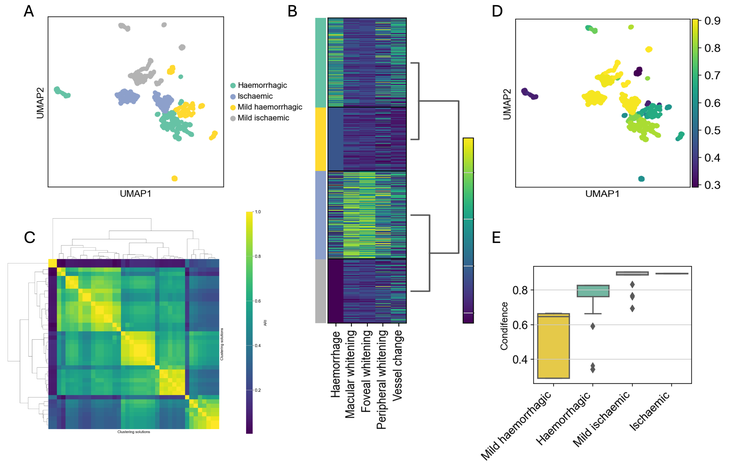

Fig 2. Unsupervised clustering of retinal data reveals novel retinal phenotypes in paediatric cerebral malaria. A) UMAP representation of the retinal data, coloured by cluster. Clusters were identified using the Leiden algorithm with a range of tuning parameters. Consensus metaclustering was performed using a spectral clustering algorithm. B) Heatmap of retinal features for clusters shown in A demonstrate clear differences between the groups. C) Heatmap of pairwise similarity between all retained clustering solutions, calculated with the adjusted Rand index. Families of internally stable but globally distinct clusterings rather than a single dominant plateau likely represent changes in number of clusters as resolution increases. D) UMAP representation of the retinal data, coloured by assignment confidence. Assignment confidence was calculated as mean within-cluster co-assignment across clustering solutions. E) Boxplot of assignment confidence by phenotype. Assignment confidence was calculated as mean within-cluster co-assignment across clustering solutions.

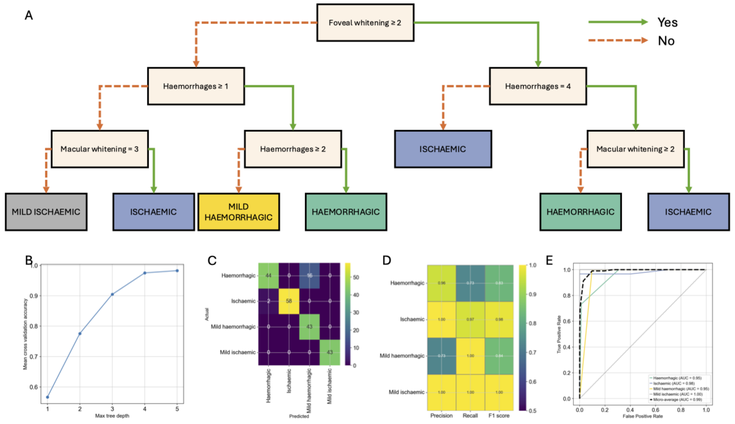

Fig 3. Predictive performance of a decision tree classifier to determine retinal phenotype from raw data. A) The simplest decision tree with good predictive performance. B) Plotting decision tree depth against average model performance in five-fold cross validation reveals a depth of three has the best balance of performance and simplicity. C) Heatmap of retinal phenotype determined by clustering versus predicted retinal phenotype using the model. D) Performance metrics for the model, stratified by retinal phenotype. E) Receiver operating characteristic curves for each retinal phenotype are plotted alongside a micro-average of each retinal phenotype. Area under the curve is uniformly high, demonstrating good predictive performance across phenotypes.

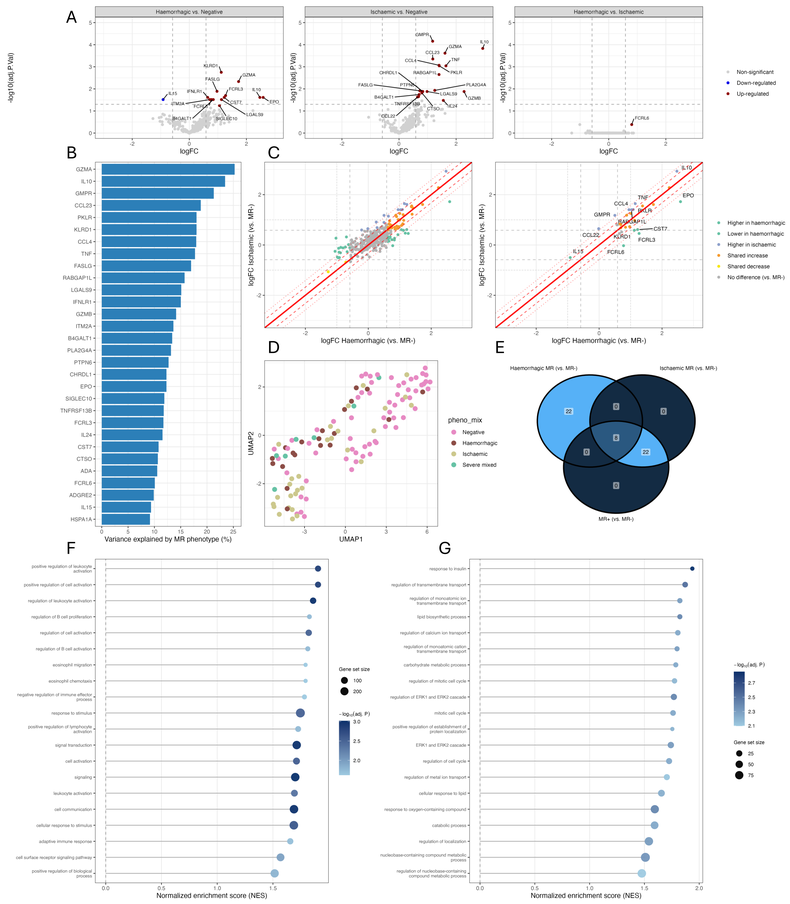

Fig 4. Inflammatory profiles differ between retinal phenotypes. A) Volcano plots of differentially abundant proteins by contrast. Proteins identified as significant after stage-wise testing are highlighted. Vertical dashed lines indicate 50% increase in protein abundance (log2(1.5)). Horizontal dashed line indicates adjusted P value 0.05 following blanket correction (not using stage-wise approach). Because proteins that were not significant in an omnibus test (F-test) would not move into the second stage of stage-wise testing a volcano can only be represented using conventionally-adjusted values. B) Per-protein variance explained by MR phenotype. Values generated using *variancePartition* and adjusting for age and sex. C) Scatter plots of fold-changes comparing haemorrhagic MR (vs. MR-) against ischaemic MR (vs. MR-). Left: Proteins with at least 50% increase in protein abundance (log2(1.5)) against reference (MR-) and a difference in protein abundance of at least 25% (log2(1.25)) are highlighted. Right: Same plot, filtered to show only the proteins that met the criteria for statistical significance during stage-wise testing in one of the two comparisons displayed. D) UMAP embedding of differentially abundant proteins coloured by retinal phenotype. Severe mixed group have both severe haemorrhagic and ischaemic features but were originally designated ischaemic. Partial separation within the embedding suggests that the inflammatory environment influences retinal phenotype. E) Venn diagram showing overlap of top 30 GO terms for haemorrhagic vs. MR-, ischaemic vs. MR- and MR+ vs. MR-. GO terms derived using GSEA with log2 fold-change as the ranking variable. There is little overlap in the most significantly enriched pathways between haemorrhagic and ischaemic MR. The heavier overlap between ischaemic MR and MR+ than haemorrhagic MR and MR+ is probably because there are fewer haemorrhagic cases than ischaemic cases. F) Top 20 significantly enriched pathways identified in the haemorrhagic vs. MR- contrast but not present in the ischaemic vs. MR- contrast. Pathways exclusive to the haemorrhagic phenotype are predominantly related to immune activation, including activation of B cells. G) Top 20 significantly enriched pathways identified in the ischaemic vs. MR- contrast but not present in the haemorrhagic vs. MR- contrast. Pathways exclusive to the ischaemic phenotype are predominantly related to metabolism, particularly of lipids and calcium.