Dissecting the biological content of host and parasite-derived extracellular vesicles in malaria

Kioko Mwikali

The Open University, Initiative to Develop African Research Leaders,

The KEMRI-Wellcome Trust Research Programme

This dissertation is submitted for the degree of Doctor of Philosophy

January 2023
Abstract

*Plasmodium falciparum* causes the deadliest form of malaria, especially in African children under five years. The parasite and its human host secrete small nanoparticles called extracellular vesicles (EVs) into their extracellular milieu. EVs contain bioactive molecules such as RNA and proteins reflecting that in the secreting cells. However, the biological contents of EVs are not well-characterized in the context of *falciparum* malaria. EVs are attractive sources of non-invasive biomarkers for diseases affecting inaccessible tissues such as the brain. Analysing the biological content of EVs will provide insights into their functions and could illuminate pathophysiological mechanisms of severe malaria syndromes, including cerebral malaria, respiratory distress, and severe malaria anaemia.

The first part of this thesis aimed to investigate the biological benefit of EVs to the secreting parasite. To achieve this purpose, I sequenced four-hourly RNA samples obtained from EVs secreted by six *P. falciparum* isolates and compared them to the RNA within the secreting whole parasites. The data suggest parasite-derived EVs might be part of a post-transcriptional gene regulatory mechanism that maintains RNA homeostasis in *P. falciparum*. The second part of this thesis aimed to understand the pathophysiology of severe malaria complications. To this purpose, I quantified the protein and RNA content of plasma EVs obtained from acute malaria patients. I developed temporal models of disease progression by applying manifold learning to the cross-sectionally collected EV samples. In a parallel approach, I used plasma EV transcriptomes to develop a pseudo-temporal model of cerebral malaria. I determined that cerebral malaria might feature a progressive decrease in neural gene expression and an increase in glial gene expression, which can be followed using peripheral blood EVs. I conclude that EVs are treasure troves of biomarkers that can be used to eavesdrop on biological mechanisms in *P. falciparum* and its human host.
Declaration

This dissertation is the product of my work and does not include collaborative work unless expressly stated.

This thesis does not exceed the word limit of 100 000 set by the Open University.
Acknowledgements

I am grateful to my PhD advisors; Dr Abdirahman Abdi, Dr Alena Pance, Prof. Julian Rayner, and Prof. Philip Bejon, for their full support during the study. Many thanks go to Shaban Mwangi for assisting with some experiments, including collecting samples during the day while I collected them at night. I acknowledge Roben Atrobus and Dr Marcus Lee for facilitating data acquisition at the Cambridge Institute of Medical Research and the Wellcome Sanger Institute, respectively.

This work was supported by Wellcome Trust Awards: 209289/Z/17/Z (to Dr Abdirahaman Abdi), 222323/Z/21/Z, 206194/Z/17/Z (to Prof Julian Rayner) and 203077/Z/16/Z (a core grant to KEMRI-Wellcome Trust, awarded to Prof Philip Bejon). My studentship was also supported by the Wellcome Trust [DEL-15-003] and the UK Foreign, Commonwealth & Development Office, with support from the Developing Excellence in Leadership, Training and Science in Africa (DELTAS Africa) programme. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.
Table of Contents

Abstract..............................................................................................................ii

Declaration ..................................................................................................... iii

Acknowledgements.......................................................................................... iv

Table of Contents .............................................................................................. v

List of Figures ................................................................................................... xii

List of Tables ..................................................................................................... xiv

Abbreviations and acronyms............................................................................ xv

Chapter 1 Introduction........................................................................................1

1.1 Summary ................................................................................................... 2

1.2 Malaria is still a public health problem ...................................................... 2

1.3 Five species of Plasmodium infect humans ................................................. 3

1.4 The multistage life cycle of P. falciparum ................................................... 4

1.5 The P. falciparum genome ......................................................................... 6

1.6 Transcriptional control of gene expression in P. falciparum ....................... 7

1.7 The transcriptome of the asexual blood stages of P. falciparum ............... 8

1.8 Post-transcriptional gene regulation in P. falciparum ............................... 9

1.8.1 Translational repression in P. falciparum ........................................... 10

1.8.2 Decay and stabilization of RNA in P. falciparum. ............................... 11

1.8.3 Translational repression by RNA binding proteins (RBPs) in P. falciparum 12
1.14.4 PfEVs contain parasite biomolecules ................................................................. 46

1.15 Thesis objectives ..................................................................................................... 50

1.15.1 Main objective ................................................................................................. 50

1.15.2 Specific objective one ...................................................................................... 50

1.15.3 Specific objective two ....................................................................................... 50

1.15.4 Specific objective three ..................................................................................... 51

Chapter 2 Materials and Methods ............................................................................... 52

2.1 RNAseq of *P. falciparum* extracellular vesicles .............................................. 53

2.1.1 *P. falciparum* isolates cultured ......................................................................... 53

2.1.2 Preparation of EV-depleted albumax and culture medium ............................... 53

2.1.3 Culturing of *P. falciparum* asexual parasites .................................................. 53

2.1.4 Synchronization of parasites .............................................................................. 54

2.1.5 Harvesting of culture-conditioned medium for EV isolation ............................. 54

2.1.6 Enrichment of PfEVs from CCM ..................................................................... 55

2.1.7 PfEVs isolation from processed CCM ................................................................. 55

2.1.8 RNA isolation from PfEVs and whole parasites ................................................ 56

2.1.9 Preparation of cDNA libraries for sequencing .................................................. 57

2.1.10 Quantification of RNAseq reads ..................................................................... 58

2.1.11 RNAseq data normalization ............................................................................ 59

2.1.12 Estimation of rhythmic parameters .................................................................. 60

2.1.13 Functional analysis ......................................................................................... 60
2.2 Proteomic profiling and analysis of plasma-derived EVs ............................................. 61
2.2.1 Samples and design .................................................................................................. 61
2.2.2 Isolation of EVs from Plasma for proteomic profiling ........................................... 61
2.2.3 Extracellular vesicles bead-assisted flow cytometry............................................... 61
2.2.4 In-solution trypsin digestion of EV proteins .......................................................... 62
2.2.5 Searching raw peptide spectra against the human database .................................. 63
2.2.6 Processing of peptide-spectrum-matches and normalization .................................. 64
2.2.7 Differential protein abundance analysis ................................................................. 65
2.2.8 Estimation of disease progression pseudotime ..................................................... 65
2.2.9 Pathway and gene ontology analysis ......................................................................... 65
2.3 RNAseq analysis of plasma EVs from cerebral malaria patients .............................. 66
2.3.1 Samples and design ................................................................................................ 66
2.3.2 RNA extraction from plasma-derived EVs .............................................................. 66
2.3.3 cDNA library preparation from plasma EV-RNA .................................................... 66
2.3.4 Differential transcript abundance analysis .............................................................. 67
2.3.5 Selection of cerebral malaria signature genes ......................................................... 67
2.3.6 Estimation of cerebral malaria progression pseudotime ......................................... 68

Chapter 3 Extracellular vesicles might be part of a post-transcriptional regulatory mechanism that maintains RNA homeostasis in *Plasmodium falciparum* ............................................. 69

3.1 Summary ...................................................................................................................... 70
3.2 Introduction .................................................................................................................. 70
3.3 Results ................................................................................................................................. 72

3.3.1 Generation of PfEVs and whole parasite RNAseq data ............................................. 72

3.3.2 Secretion of RNA via PfEVs is conserved among parasite isolates ................. 74

3.3.3 RNA secretion via PfEVs occurs in a periodic manner ............................................ 76

3.3.4 Transcriptional phase-shift in the secretion of RNA via PfEVs .................... 77

3.3.5 Unusable RNA is enriched in PfEVs compared to the whole parasite .......... 78

3.4 Discussion ......................................................................................................................... 81

3.5 Conclusion .......................................................................................................................... 82

Chapter 4 Integrated proteomic analysis of circulating extracellular vesicles identifies pathological processes associated with severe malaria complications ................. 83

4.1 Summary ............................................................................................................................ 84

4.2 Introduction ......................................................................................................................... 86

4.3 Results ................................................................................................................................. 88

4.3.1 Clinical parameters ....................................................................................................... 88

4.3.2 Proteomic data generation ............................................................................................. 90

4.3.3 Batch adjustment of proteomic data .............................................................................. 92

4.3.4 Plasma-derived small EVs proteomes can distinguish SM and NSM ............... 94

4.3.5 Manifold learning identifies malaria disease progression pseudotime .......... 96

4.3.6 The inferred trajectory is associated with the known biology of SM .......... 99

4.3.7 Secretion of neural and glial proteins via small EVs over pseudotime .......... 101

4.3.8 Late pseudotime could be the disease recovery stage ................................. 102
| 4.3.9 | Host proteins in plasma EVs can distinguish malaria syndromes | 103 |
| 4.3.10 | Identifying markers of malaria-associated death | 107 |
| 4.4 | Discussion | 110 |
| 4.5 | Conclusion | 114 |

Chapter 5  
Brain-expressed RNA in circulating extracellular vesicles can discriminate retinopathy-positive and negative cerebral malaria | 115 |

| 5.1 | Summary | 116 |
| 5.2 | Introduction | 117 |
| 5.3 | Results | 118 |
| 5.3.1 | Patient clinical and laboratory parameters | 118 |
| 5.3.2 | Host EV-RNAs discriminate CM-R⁺ from CM-R⁻ | 120 |
| 5.3.3 | Brain-specific transcripts are enriched in circulating EVs | 122 |
| 5.3.4 | Generation of cerebral malaria gene signature | 125 |
| 5.3.5 | Estimation of neurodegeneration pseudotime in cerebral malaria | 128 |
| 5.4 | Discussion | 135 |
| 5.5 | Conclusion | 138 |

Chapter 6  
Discussion | 139 |

| 6.1 | Summary and Aims | 140 |
| 6.2 | The homeostatic role of PfEVs in *P. falciparum* | 140 |
| 6.2.1 | How *might* *P. falciparum* load RNA into PfEVs? | 140 |
6.2.2 The physiological requirement of a gene during the IDC can be inferred based on its exclusion from PfEVs ................................................................. 141

6.2.3 PfEVs released by mosquito stages could be enriched in RNA encoded by genes required during the IDC ........................................................................................................ 142

6.2.4 The mRNA within PfEVs might be translated into proteins in recipient parasite cells 142

6.3 Circulating EV cargo as biomarkers of severe malaria – future directions .......... 143

6.3.1 Immune capture methods could improve the diagnostic potential of EVs .... 143

6.3.2 Comparing the biological materials in EVs with plasma or parent cells could reveal interesting differences ................................................................. 144

6.3.3 How can the markers of disease progression in malaria be validated in future studies? 144

6.3.4 Application of EVs-based biomarkers to clinical disease ......................... 145

Chapter 7 Bibliography ....................................................................................... 147
List of Figures

Fig 1.1 The asexual and sexual blood stages of P. falciparum .................................................. 6
Fig 1.2 The concepts of post-transcriptional gene regulation in P. falciparum ......................... 10
Fig 1.3 The microenvironment of cerebral malaria ................................................................. 20
Fig 1.4 The pathophysiology of severe malaria anaemia ......................................................... 25
Fig 1.5 The pathophysiology of acidotic respiratory distress .................................................. 28
Fig 1.6 The biogenesis of extracellular vesicles ....................................................................... 32
Fig 1.7 Asexual stages of P. falciparum secrete extracellular vesicles ....................................... 40
Fig 2.1 Extracellular vesicle isolation by ultracentrifugation .................................................... 56
Fig 2.2 RNAseq data generation protocol ................................................................................. 59
Fig 3.1 RNAseq data generation and processing ...................................................................... 73
Fig 3.2 RNA secretion via PfEVs is conserved between isolates ............................................... 75
Fig 3.3 Secretion of RNA via PfEVs occurs in a periodic manner .............................................. 76
Fig 3.4 Secretion of RNA via PfEVs phase-shifted relative to the WP expression .................. 77
Fig 3.5 Unusable RNA is preferentially secreted via PfEVs ...................................................... 80
Fig 4.1 Comparison of clinical parameters between severe malaria complications ............... 89
Fig 4.2 Generated data is enriched in small EV proteins ......................................................... 91
Fig 4.3 Adjustment of batch effects ........................................................................................ 93
Fig 4.4 Quantitative plasma-derived proteomic analysis of SM vs NSM ............................... 95
Fig 4.5 Ingenuity pathway analysis of significantly altered proteins ...................................... 96
Fig 4.6 Molecular estimation of malaria pseudotime using manifold learning ..................... 98
Fig 4.7 Inferred staging reflects known malaria biology .......................................................... 100
Fig 4.8 Secretion of brain proteins over pseudotime ................................................................ 101
Fig 4.9 Early and very late pseudotime samples have similar protein profiles ..................... 103
Fig 4.10 Comparative proteomics analysis of different malaria disease syndromes ............... 106
Fig 4.11 Identifying markers of death caused by malaria .................................................... 108
Fig 4.12 Cox-regression analysis ............................................................................................ 109
Fig 5.1 Comparison of host EV transcriptomes between CM-R+ and CM-R− ...................... 121
Fig 5.2 Most of altered transcripts are specific to the brain ................................................... 123
Fig 5.3 Brain-expressed RNA is enriched in EVs obtained from CM patients ..................... 124
Fig 5.4 Functional analysis of differentially altered brain specific genes ............................. 125
Fig 5.5 Selection of top transcripts that can discriminate CM-R+ and CM-R− ...................... 126
Fig 5.6 Selected top transcripts that can discriminate CM-R+ and CM-R− .......................... 127
Fig 5.7 Estimation of disease progression pseudotime .......................................................... 129
Fig 5.8 Variation of EV transcripts over pseudotime .............................................................. 130
Fig 5.9 Functional analysis of transcripts altered over pseudotime ..................................... 132
Fig 5.10 Representative biological processes altered over pseudotime ............................... 133
Fig 5.11 Brain cell RNA abundance in EVs as a function of disease pseudotime .................. 134
List of Tables

Table 5.1 Summary descriptive table of patients by retinopathy status................................. 119
**Abbreviations and acronyms**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA1</td>
<td>ATP-binding cassette subfamily A member 1</td>
</tr>
<tr>
<td>ABCB6</td>
<td>ATP-binding cassette subfamily B member 6</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>ACS1b</td>
<td>Acyl-CoA synthetase, pseudogene</td>
</tr>
<tr>
<td>ADAM10</td>
<td>ADAM metallopeptidase domain 10</td>
</tr>
<tr>
<td>ADAMTS7</td>
<td>ADAM metallopeptidase with thrombospondin type 1 motif 7</td>
</tr>
<tr>
<td>ADAMTS13</td>
<td>ADAM metallopeptidase with thrombospondin type 1 motif 13</td>
</tr>
<tr>
<td>ADD1</td>
<td>Adducin 1</td>
</tr>
<tr>
<td>AGL</td>
<td>Amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase</td>
</tr>
<tr>
<td>AKAP12</td>
<td>A-kinase anchoring protein 12</td>
</tr>
<tr>
<td>AKAP9</td>
<td>A-kinase anchoring protein 9</td>
</tr>
<tr>
<td>ALAD</td>
<td>Aminolevulinate dehydratase</td>
</tr>
<tr>
<td>ALAS</td>
<td>Aminolevulinate synthase 1</td>
</tr>
<tr>
<td>ALAS2</td>
<td>Aminolevulinate synthase 2</td>
</tr>
<tr>
<td>Alba</td>
<td>Acetylation lowers affinity protein family</td>
</tr>
<tr>
<td>ALI</td>
<td>Acute lung injury</td>
</tr>
<tr>
<td>AMA1</td>
<td>Apical membrane protein 1</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>AP2-G</td>
<td>ApiAp2 transcription factor G</td>
</tr>
<tr>
<td>APCS</td>
<td>Amyloid P component serum</td>
</tr>
<tr>
<td>ApiAP2</td>
<td>Apicomplexa apetala 2 DNA binding domain</td>
</tr>
<tr>
<td>APOC2</td>
<td>Apolipoprotein C2</td>
</tr>
<tr>
<td>APOC3</td>
<td>Apolipoprotein C3</td>
</tr>
<tr>
<td>AQP1</td>
<td>Aquaporin 1</td>
</tr>
<tr>
<td>AQP2</td>
<td>Aquaporin 2</td>
</tr>
<tr>
<td>ARDS</td>
<td>Acute Respiratory Distress Syndrome</td>
</tr>
<tr>
<td>ARF6</td>
<td>ADP ribosylation factor 6</td>
</tr>
<tr>
<td>ASCA</td>
<td>ANOVA-simultaneous component analysis</td>
</tr>
<tr>
<td>ATG4</td>
<td>Autophagy-related protein 4</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>basoE</td>
<td>Basophilic erythroblasts</td>
</tr>
<tr>
<td>BBB</td>
<td>Blood-brain barrier</td>
</tr>
<tr>
<td>BCS</td>
<td>Blantyre coma score</td>
</tr>
<tr>
<td>BFU-E</td>
<td>Burst forming unit - erythroblasts</td>
</tr>
<tr>
<td>BH</td>
<td>Benjamin-Hochberg procedure</td>
</tr>
<tr>
<td>BSG</td>
<td>Basigin</td>
</tr>
<tr>
<td>BVE-PLS</td>
<td>Backward variable elimination – partial least squares</td>
</tr>
<tr>
<td>C1GALT2</td>
<td>Core 1 beta3-galactosyltransferase-specific molecular chaperone</td>
</tr>
<tr>
<td>C2</td>
<td>Complement C2</td>
</tr>
</tbody>
</table>
C3  Complement C3
C4BPA  C4b-binding protein alpha chain
C5  Complement C5
C6  Complement C6
C7  Complement C2
C8B  Complement C8b
CACNA1G  Voltage-dependent T-type calcium channel subunit alpha-1G
CACNA1I  Voltage-dependent T-type calcium channel subunit alpha-1I
CAF1  Chromatin assembly factor 1 p55 subunit
CC  Community control without *P. falciparum* infection
CCM  Culture conditioned medium
CCp1-4  LCCL domain-containing protein 1 to 4
CCR3  C-C chemokine receptor type 3
CCR4  C-C chemokine receptor type 4
CCR4-NOT  Carbon Catabolite Repression—Negative On TATA-less complex
CCR5  C-C chemokine receptor type 5
CCR7  C-C chemokine receptor type 7
CCR9  C-C chemokine receptor type 9
CD10  Cluster of differentiation 10
CD11a  Cluster of differentiation 11a
CD14  Cluster of differentiation 14
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD38</td>
<td>Cluster of differentiation 38</td>
</tr>
<tr>
<td>CD3E</td>
<td>Cluster of differentiation 3E</td>
</tr>
<tr>
<td>CD40LG</td>
<td>Cluster of differentiation 40-ligand</td>
</tr>
<tr>
<td>CD5</td>
<td>Cluster of differentiation 5</td>
</tr>
<tr>
<td>CD55</td>
<td>Cluster of differentiation 55</td>
</tr>
<tr>
<td>CD63</td>
<td>Cluster of differentiation 63</td>
</tr>
<tr>
<td>CD81</td>
<td>Cluster of differentiation 81</td>
</tr>
<tr>
<td>CD8A</td>
<td>Cluster of differentiation 8A</td>
</tr>
<tr>
<td>CD8B</td>
<td>Cluster of differentiation 8B</td>
</tr>
<tr>
<td>CD9</td>
<td>Cluster of differentiation 9</td>
</tr>
<tr>
<td>CDK5</td>
<td>Cyclin-dependent kinase 5</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>Cer</td>
<td>Ceramide</td>
</tr>
<tr>
<td>CFR</td>
<td>Case fatality ratio</td>
</tr>
<tr>
<td>CFU-E</td>
<td>Colony forming unit - erythroblasts</td>
</tr>
<tr>
<td>CHD7</td>
<td>Chromodomain helicase DNA binding protein 7</td>
</tr>
<tr>
<td>CHMP</td>
<td>Charged multivesicular body protein</td>
</tr>
<tr>
<td>CHMP3</td>
<td>Charged multivesicular body protein 3</td>
</tr>
<tr>
<td>CHMP4B</td>
<td>Charged multivesicular body protein 4B</td>
</tr>
<tr>
<td>CHMP6</td>
<td>Charged multivesicular body protein 6</td>
</tr>
<tr>
<td>CIMR</td>
<td>Cambridge institute of medical research</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CITH</td>
<td>Trailer hitch homolog, putative</td>
</tr>
<tr>
<td>CLAG3.1</td>
<td>Cytoadherence linked antigen 3.1</td>
</tr>
<tr>
<td>CM</td>
<td>Cerebral malaria</td>
</tr>
<tr>
<td>CM-R−</td>
<td>Retinopathy negative cerebral malaria</td>
</tr>
<tr>
<td>CM-R+</td>
<td>Retinopathy positive cerebral malaria</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>COX5B</td>
<td>Cytochrome c oxidase subunit 5B, mitochondrial</td>
</tr>
<tr>
<td>CPM</td>
<td>Counts per million</td>
</tr>
<tr>
<td>CPOX</td>
<td>Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial</td>
</tr>
<tr>
<td>CR1</td>
<td>Complement receptor type 1</td>
</tr>
<tr>
<td>CREBBP</td>
<td>CREB-binding protein</td>
</tr>
<tr>
<td>CRMP1-4</td>
<td>Cysteine repeat modular protein 1 to 4</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>CRT</td>
<td>Chloroquine transporter</td>
</tr>
<tr>
<td>CSP</td>
<td>Circumsporozoite protein</td>
</tr>
<tr>
<td>CTRP</td>
<td>Circumsporozoite- and TRAP-related protein</td>
</tr>
<tr>
<td>CTSC</td>
<td>Cathepsin C</td>
</tr>
<tr>
<td>CTSG</td>
<td>Cathepsin G</td>
</tr>
<tr>
<td>CVC</td>
<td>Caveola-Vesicle Complexes</td>
</tr>
<tr>
<td>CXCR3</td>
<td>C-X-C chemokine receptor type 3</td>
</tr>
<tr>
<td>CXCR4</td>
<td>C-X-C chemokine receptor type 4</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>DAI</td>
<td>Diffuse alveolar inflammation</td>
</tr>
<tr>
<td>DCP1-DCP2</td>
<td>Decapping protein1-decapping protein 2 complex</td>
</tr>
<tr>
<td>Dd2</td>
<td><em>Plasmodium falciparum</em> Dd2 strain</td>
</tr>
<tr>
<td>DEFA1B</td>
<td>Neutrophil defensin alpha 1b</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>Deoxythymidine triphosphate</td>
</tr>
<tr>
<td>dUTP</td>
<td>Deoxyuridine triphosphate</td>
</tr>
<tr>
<td>EBA165</td>
<td>Erythrocyte binding antigen 165</td>
</tr>
<tr>
<td>EBA175</td>
<td>Erythrocyte binding antigen 175</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr Virus</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>EDNRB</td>
<td>Endothelin receptor type B</td>
</tr>
<tr>
<td>ELF2</td>
<td>E74-like factor 2</td>
</tr>
<tr>
<td>ENO</td>
<td>Enolase</td>
</tr>
<tr>
<td>EP300</td>
<td>E1A-associated protein p300</td>
</tr>
<tr>
<td>EPB41</td>
<td>Erythrocyte membrane protein band 4.1</td>
</tr>
<tr>
<td>EPB41L2</td>
<td>Erythrocyte membrane protein band 4.1-like 2</td>
</tr>
<tr>
<td>EPB42</td>
<td>Erythrocyte membrane protein band 4.2</td>
</tr>
<tr>
<td>EPCR</td>
<td>Endothelial protein C receptor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>EPHB4</td>
<td>Ephrin type-B receptor 4</td>
</tr>
<tr>
<td>ESCRT</td>
<td>Endosomal sorting complex required for transport</td>
</tr>
<tr>
<td>ETRAMP11.2</td>
<td>Early transcribed membrane protein 11.2</td>
</tr>
<tr>
<td>ETRAMP2</td>
<td>Early transcribed membrane protein 2</td>
</tr>
<tr>
<td>EVs</td>
<td>Extracellular vesicles</td>
</tr>
<tr>
<td>EXP2</td>
<td>Exportin 2</td>
</tr>
<tr>
<td>F2R</td>
<td>Thrombin receptor</td>
</tr>
<tr>
<td>F7</td>
<td>Blood coagulation factor VII</td>
</tr>
<tr>
<td>FBN2</td>
<td>Fibrillin 2</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FECH</td>
<td>Ferochelatase</td>
</tr>
<tr>
<td>FET</td>
<td>Fisher’s exact test</td>
</tr>
<tr>
<td>FGA</td>
<td>Fibrinogen A</td>
</tr>
<tr>
<td>FGB</td>
<td>Fibrinogen B</td>
</tr>
<tr>
<td>FGG</td>
<td>Fibrinogen G</td>
</tr>
<tr>
<td>FOXO3</td>
<td>Forkhead box protein O3</td>
</tr>
<tr>
<td>FTH1</td>
<td>Ferritin heavy chain 1</td>
</tr>
<tr>
<td>FTL</td>
<td>Ferritin light chain</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde phosphate dehydrogenase</td>
</tr>
<tr>
<td>GATA1</td>
<td>Erythroid transcription factor</td>
</tr>
<tr>
<td>GC</td>
<td>Guanine - cytosine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>GDV1</td>
<td>Gametocyte development protein 1</td>
</tr>
<tr>
<td>GLYPA</td>
<td>Glycophorin A</td>
</tr>
<tr>
<td>GLYPB</td>
<td>Glycophorin B</td>
</tr>
<tr>
<td>GM3</td>
<td>Monosialodihexosylganglioside</td>
</tr>
<tr>
<td>GP6</td>
<td>Glypican 6</td>
</tr>
<tr>
<td>GPX4</td>
<td>Glutathione peroxidase 4</td>
</tr>
<tr>
<td>GRK2</td>
<td>G protein-coupled receptor kinase 2</td>
</tr>
<tr>
<td>GSDMB</td>
<td>Gasdermin-B</td>
</tr>
<tr>
<td>GSDMD</td>
<td>Gasdermin-D</td>
</tr>
<tr>
<td>H1</td>
<td>Histone 1</td>
</tr>
<tr>
<td>H2A</td>
<td>Histone 2A</td>
</tr>
<tr>
<td>H2B</td>
<td>Histone 2B</td>
</tr>
<tr>
<td>H3</td>
<td>Histone 3</td>
</tr>
<tr>
<td>H4</td>
<td>Histone 4</td>
</tr>
<tr>
<td>HB</td>
<td>Haemoglobin</td>
</tr>
<tr>
<td>HBA1</td>
<td>Haemoglobin A1</td>
</tr>
<tr>
<td>HBB</td>
<td>Haemoglobin B</td>
</tr>
<tr>
<td>HBD</td>
<td>Haemoglobin D</td>
</tr>
<tr>
<td>HGF</td>
<td>Hepatocyte growth factor</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HK</td>
<td>Hexose kinase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HnRNPA2B1</td>
<td>Heterogeneous nuclear ribonucleoproteins A2/B1</td>
</tr>
<tr>
<td>HP</td>
<td>Heterochromatin protein</td>
</tr>
<tr>
<td>HPA</td>
<td>Human Protein Atlas</td>
</tr>
<tr>
<td>HRP2</td>
<td>Histidine rich protein 2</td>
</tr>
<tr>
<td>HRP3</td>
<td>Histidine rich protein 3</td>
</tr>
<tr>
<td>ICAM1</td>
<td>Intercellular adhesion molecule 1</td>
</tr>
<tr>
<td>IDC</td>
<td>Intraerythrocytic developmental cycle of <em>P. falciparum</em></td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IL10</td>
<td>Interleukin 10</td>
</tr>
<tr>
<td>IL12A</td>
<td>Interleukin 12 A</td>
</tr>
<tr>
<td>IL17RA</td>
<td>Interleukin 17 receptor agonist</td>
</tr>
<tr>
<td>IL17α</td>
<td>Interleukin 12 alpha</td>
</tr>
<tr>
<td>IL18</td>
<td>Interleukin 18</td>
</tr>
<tr>
<td>IL18R1</td>
<td>Interleukin 18 receptor 1</td>
</tr>
<tr>
<td>IL1R1</td>
<td>Interleukin 1 receptor 1</td>
</tr>
<tr>
<td>IL6</td>
<td>Interleukin 6</td>
</tr>
<tr>
<td>IL6ST</td>
<td>Interleukin 6 signal transducer</td>
</tr>
<tr>
<td>IL8</td>
<td>Interleukin 8</td>
</tr>
<tr>
<td>ILVs</td>
<td>Intraluminal vesicles</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile range</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IRAK1</td>
<td>Interleukin-1 receptor-associated kinase 1</td>
</tr>
<tr>
<td>IRAK4</td>
<td>Interleukin-1 receptor-associated kinase 4</td>
</tr>
<tr>
<td>IRS1</td>
<td>Insulin receptor substrate 1</td>
</tr>
<tr>
<td>IRS2</td>
<td>Insulin receptor substrate 2</td>
</tr>
<tr>
<td>ITGA7</td>
<td>Integrin alpha 7</td>
</tr>
<tr>
<td>ITGB7</td>
<td>Integrin beta 7</td>
</tr>
<tr>
<td>ITGBL1</td>
<td>Integrin beta-like protein 1</td>
</tr>
<tr>
<td>K13C580Y</td>
<td>Plasmodium falciparum Kelch 13 C580Y mutation</td>
</tr>
<tr>
<td>KAHRP</td>
<td>Knob associated histidine rich protein 2</td>
</tr>
<tr>
<td>KASIII</td>
<td>beta-ketoacyl-ACP synthase III</td>
</tr>
<tr>
<td>KCNA1</td>
<td>Potassium voltage-gated channel subfamily A member 1,</td>
</tr>
<tr>
<td>KE01</td>
<td>Plasmodium falciparum KE01 strain</td>
</tr>
<tr>
<td>KE02</td>
<td>Plasmodium falciparum KE02 strain</td>
</tr>
<tr>
<td>KE04</td>
<td>Plasmodium falciparum KE04 strain</td>
</tr>
<tr>
<td>KE06</td>
<td>Plasmodium falciparum KE06 strain</td>
</tr>
<tr>
<td>KLF1</td>
<td>Krueppel-like zinc finger transcription factor 1</td>
</tr>
<tr>
<td>KWTRP</td>
<td>KEMRI-Wellcome Trust Research Programme</td>
</tr>
<tr>
<td>LaeCer</td>
<td>Lactoceramide</td>
</tr>
<tr>
<td>LBP</td>
<td>Lipopolysaccharide binding protein</td>
</tr>
<tr>
<td>LFA-1</td>
<td>Lymphocyte function-associated antigen 1</td>
</tr>
<tr>
<td>LIG4</td>
<td>DNA ligase 4</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>LIP5</td>
<td>Lipase 5</td>
</tr>
<tr>
<td>LMP1</td>
<td>Latent membrane protein 1</td>
</tr>
<tr>
<td>log2FC</td>
<td>Log2 fold-change</td>
</tr>
<tr>
<td>LRP1</td>
<td>Low Density Lipoprotein Receptor-related Protein</td>
</tr>
<tr>
<td>LTF</td>
<td>Lactotransferrin</td>
</tr>
<tr>
<td>LXR/RXR</td>
<td>Liver X receptor /retinol X receptor</td>
</tr>
<tr>
<td>MAHRP1</td>
<td>Membrane associated histidine rich protein1</td>
</tr>
<tr>
<td>MAP1LC3A</td>
<td>Microtubule-associated proteins 1A/1B light chain 3A</td>
</tr>
<tr>
<td>MAPK1</td>
<td>Mitogen-activated protein kinase</td>
</tr>
<tr>
<td>MBL2</td>
<td>Mannose-binding lectin-2</td>
</tr>
<tr>
<td>MC</td>
<td>Maurer’s cleft</td>
</tr>
<tr>
<td>MCP1</td>
<td>Monocyte chemoattractant protein -1</td>
</tr>
<tr>
<td>MCV</td>
<td>Mean corpuscular volume</td>
</tr>
<tr>
<td>MDH</td>
<td>Malate dehydrogenase</td>
</tr>
<tr>
<td>MDN1</td>
<td>Midasin</td>
</tr>
<tr>
<td>MDS</td>
<td>Multidimensional scaling</td>
</tr>
<tr>
<td>mEVs</td>
<td>Medium extracellular vesicles</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>miRNAs</td>
<td>microRNA</td>
</tr>
<tr>
<td>MLKL</td>
<td>Mixed lineage kinase domain-like protein</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MRSA</td>
<td>Methicillin resistant Staphylococcus aureus</td>
</tr>
<tr>
<td>MSP1</td>
<td>Merozoite surface protein 1</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4</td>
</tr>
<tr>
<td>MVBs</td>
<td>Multivesicular bodies</td>
</tr>
<tr>
<td>NCAN</td>
<td>Neurocan</td>
</tr>
<tr>
<td>NDNF</td>
<td>Neuron-derived neurotrophic factor</td>
</tr>
<tr>
<td>NDUFA8</td>
<td>NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8</td>
</tr>
<tr>
<td>NDUFB2</td>
<td>NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 2</td>
</tr>
<tr>
<td>NEDD4</td>
<td>E3 ubiquitin-protein ligase NEDD4</td>
</tr>
<tr>
<td>NLRC3</td>
<td>NLR family CARD domain-containing protein 3</td>
</tr>
<tr>
<td>NLRC4</td>
<td>NLR family CARD domain-containing protein 4</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>Neurogenic locus notch homolog protein 1</td>
</tr>
<tr>
<td>NSM</td>
<td>Non-severe malaria</td>
</tr>
<tr>
<td>orthoE</td>
<td>Orthochromatic erythroblasts</td>
</tr>
<tr>
<td>p53</td>
<td>Cellular tumor antigen p53</td>
</tr>
<tr>
<td>PAX5</td>
<td>Paired box protein Pax-5</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidyl-choline</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCC</td>
<td>Pearson correlation coefficient</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PDGFA</td>
<td>Platelet-derived growth factor subunit A</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PER2</td>
<td>Period circadian protein homolog 2</td>
</tr>
<tr>
<td>PfEMP1</td>
<td><em>P. falciparum</em> erythrocyte membrane protein 1</td>
</tr>
<tr>
<td>PfEVs</td>
<td><em>P. falciparum</em> derived extracellular vesicles</td>
</tr>
<tr>
<td>PfHRP2</td>
<td><em>P. falciparum</em> histidine rich protein 2</td>
</tr>
<tr>
<td>PfLDH</td>
<td><em>P. falciparum</em> lactate dehydrogenase</td>
</tr>
<tr>
<td>PfMC-TM</td>
<td><em>P. falciparum</em> maurer’s cleft-transmembrane protein</td>
</tr>
<tr>
<td>PHKG2</td>
<td>Phosphorylase kinase subunit gamma-2</td>
</tr>
<tr>
<td>PI-3P</td>
<td>Phosphatidylinositol-3-phosphate</td>
</tr>
<tr>
<td>PICs</td>
<td>Preinitiation complex</td>
</tr>
<tr>
<td>PLXNB2</td>
<td>Plexin B2</td>
</tr>
<tr>
<td>PLXNC1</td>
<td>Plexin C1</td>
</tr>
<tr>
<td>PMI-III</td>
<td>Plasmepsin I to III</td>
</tr>
<tr>
<td>polyE</td>
<td>Polychromatic erythroblasts</td>
</tr>
<tr>
<td>PPARα</td>
<td>Peroxisome proliferator-activated receptor α</td>
</tr>
<tr>
<td>PPO</td>
<td><em>P. falciparum</em> protoporphyrinogen oxidase</td>
</tr>
<tr>
<td>PPOX</td>
<td>Protoporphyrinogen oxidase</td>
</tr>
<tr>
<td>PPP1R3A</td>
<td>Protein phosphatase 1 regulatory subunit 3A</td>
</tr>
<tr>
<td>PRKN</td>
<td>E3 ubiquitin-protein ligase parkin</td>
</tr>
<tr>
<td>proE</td>
<td>Proerythroblasts</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidyl-serine</td>
</tr>
<tr>
<td>PSM</td>
<td>Peptide spectrum match</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensin homolog</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Prostaglandin-endoperoxidase synthase 2</td>
</tr>
<tr>
<td>PTP1</td>
<td>P/EMP1 trafficking protein 1</td>
</tr>
<tr>
<td>PTPRK</td>
<td>Receptor-type tyrosine-protein phosphatase kappa</td>
</tr>
<tr>
<td>PUF2</td>
<td>Pumilio homology domain family member 2</td>
</tr>
<tr>
<td>PURA</td>
<td>Purine-rich single-stranded DNA-binding protein alpha</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcription – polymerase chain reaction</td>
</tr>
<tr>
<td>RAB41</td>
<td>Ras-related protein Rab-41</td>
</tr>
<tr>
<td>RANBP2</td>
<td>Ran-binding protein 2</td>
</tr>
<tr>
<td>RAP1</td>
<td>Rhoptry associated protein 1</td>
</tr>
<tr>
<td>RAP2</td>
<td>Rhoptry associated protein 2</td>
</tr>
<tr>
<td>RASGRP1</td>
<td>RAS guanyl-releasing protein 1</td>
</tr>
<tr>
<td>RBCs</td>
<td>Red blood cells</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>RD</td>
<td>Respiratory distress</td>
</tr>
<tr>
<td>RESA</td>
<td>Ring erythrocyte surface protein</td>
</tr>
<tr>
<td>REX</td>
<td>Ring exported protein</td>
</tr>
<tr>
<td>RH3</td>
<td>Reticulocyte binding protein homologue 3</td>
</tr>
<tr>
<td>RH5</td>
<td>Reticulocyte binding protein homologue 5</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RHOA</td>
<td>Ras homolog family member A</td>
</tr>
<tr>
<td>RhopH2</td>
<td>High molecular weight rhoptry protein 2</td>
</tr>
<tr>
<td>RIG-I</td>
<td>Retinoic acid-inducible gene 1 protein</td>
</tr>
<tr>
<td>RLE</td>
<td>Relative log expression</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNaseq</td>
<td>RNA sequencing</td>
</tr>
<tr>
<td>RNases</td>
<td>Ribonucleases</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic</td>
</tr>
<tr>
<td>RON5</td>
<td>Rhopty neck protein 5</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RRR</td>
<td>Relative risk ratio</td>
</tr>
<tr>
<td>RSP2</td>
<td>Rhopty surface protein 2</td>
</tr>
<tr>
<td>RT</td>
<td>Retention time</td>
</tr>
<tr>
<td>RTS,S</td>
<td>Malaria vaccine</td>
</tr>
<tr>
<td>S100A8</td>
<td>S100 calcium-binding protein A8</td>
</tr>
<tr>
<td>S100A9</td>
<td>S100 calcium-binding protein A9</td>
</tr>
<tr>
<td>SAGE</td>
<td>Serial analysis of gene expression</td>
</tr>
<tr>
<td>SBP1</td>
<td>Skeleton binding protein 1</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel protein type 5 subunit alpha</td>
</tr>
<tr>
<td>SCS-beta</td>
<td>Succinate--CoA ligase [ADP-forming] subunit beta</td>
</tr>
<tr>
<td>SEMA5C</td>
<td>Semaphorin 5C</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SERPINA1</td>
<td>Alpha-1-antiproteinase, Serpin A1</td>
</tr>
<tr>
<td>SERPINF2</td>
<td>Alpha-2-antiplasmin, Serpin F2</td>
</tr>
<tr>
<td>sEVs</td>
<td>Small extracellular vesicles</td>
</tr>
<tr>
<td>sKE01</td>
<td>HBEC selected <em>P. falciparum</em> KE01 strain</td>
</tr>
<tr>
<td>SLC40A1</td>
<td>Solute carrier family 40 member 1, Ferroportin-1</td>
</tr>
<tr>
<td>SM</td>
<td>Severe malaria</td>
</tr>
<tr>
<td>SMA</td>
<td>Severe malaria anaemia</td>
</tr>
<tr>
<td>SNARE</td>
<td>Soluble N-ethylmaleimide-sensitive factor activating protein receptor</td>
</tr>
<tr>
<td>SPTA1</td>
<td>Spectrin alpha chain</td>
</tr>
<tr>
<td>SPTB</td>
<td>Spectrin beta chain, erythrocytic</td>
</tr>
<tr>
<td>STAT1</td>
<td>Signal transducer and activator of transcription 1</td>
</tr>
<tr>
<td>STEVOR</td>
<td>Subtelomeric variable open reading frame protein family</td>
</tr>
<tr>
<td>STING</td>
<td>Stimulator of interferon genes</td>
</tr>
<tr>
<td>STYXL2</td>
<td>Serine/threonine/tyrosine-interacting-like protein 2</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small ubiquitin-related modifier</td>
</tr>
<tr>
<td>TEAB</td>
<td>Triethylammonium bicarbonate</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>TFR2</td>
<td>Transferrin receptor protein 2</td>
</tr>
<tr>
<td>TFRC</td>
<td>Transferrin receptor protein 1</td>
</tr>
<tr>
<td>TGFB1</td>
<td>Transforming growth factor beta 1</td>
</tr>
<tr>
<td>THEMIS2</td>
<td>Thymocyte-expressed molecule involved in selection protein 2</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>TIGIT</td>
<td>T-cell immunoreceptor with Ig and ITIM domains</td>
</tr>
<tr>
<td>TLR6</td>
<td>Toll-like receptor 6</td>
</tr>
<tr>
<td>TLR7</td>
<td>Toll-like receptor 7</td>
</tr>
<tr>
<td>TMEM121</td>
<td>Transmembrane protein 121</td>
</tr>
<tr>
<td>TMEM135</td>
<td>Transmembrane protein 135</td>
</tr>
<tr>
<td>TMF1</td>
<td>TATA element modulatory factor</td>
</tr>
<tr>
<td>TMT</td>
<td>Tandem mass tag</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>TNFS14</td>
<td>Tumor necrosis factor superfamily member 14</td>
</tr>
<tr>
<td>TNFSF4</td>
<td>Tumor necrosis factor ligand superfamily member 4</td>
</tr>
<tr>
<td>TORC1</td>
<td>Target of Rapamycin Complex 1</td>
</tr>
<tr>
<td>TPM</td>
<td>Transcript per million</td>
</tr>
<tr>
<td>TRAF3IP2</td>
<td>TRAF3-interacting protein 2</td>
</tr>
<tr>
<td>TRAIL</td>
<td>TNF-related apoptosis-inducing ligand</td>
</tr>
<tr>
<td>TRPC1</td>
<td>Short transient receptor potential channel 1</td>
</tr>
<tr>
<td>TSG101</td>
<td>Tumor susceptibility gene 101</td>
</tr>
<tr>
<td>UBL3</td>
<td>Ubiquitin-like protein 3</td>
</tr>
<tr>
<td>UBR2</td>
<td>E3 ubiquitin-protein ligase UBR2</td>
</tr>
<tr>
<td>UBR4</td>
<td>E3 ubiquitin-protein ligase UBR4</td>
</tr>
<tr>
<td>UNC5C</td>
<td>Protein unc-5 homolog C</td>
</tr>
<tr>
<td>UROD</td>
<td>Uroporphyrinogen decarboxylase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UROS</td>
<td>Uroporphyrinogen-III synthase</td>
</tr>
<tr>
<td>USP1</td>
<td>Ubiquitin specific protein 1</td>
</tr>
<tr>
<td>USP4</td>
<td>Ubiquitin specific protein 4</td>
</tr>
<tr>
<td>USP7</td>
<td>Ubiquitin specific protein 7</td>
</tr>
<tr>
<td>USP20</td>
<td>Ubiquitin specific protein 20</td>
</tr>
<tr>
<td>USP31</td>
<td>Ubiquitin specific protein 31</td>
</tr>
<tr>
<td>USP32</td>
<td>Ubiquitin specific protein 32</td>
</tr>
<tr>
<td>VAV1</td>
<td>Proto-oncogene vav</td>
</tr>
<tr>
<td>VCAM1</td>
<td>Vascular cell adhesion protein 1</td>
</tr>
<tr>
<td>VSP</td>
<td>Vacuolar sorting protein</td>
</tr>
<tr>
<td>VSP4</td>
<td>Vacuolar sorting protein 4</td>
</tr>
<tr>
<td>VSP4A</td>
<td>Vacuolar sorting protein 4A</td>
</tr>
<tr>
<td>VWF</td>
<td>von willebrand factor</td>
</tr>
<tr>
<td>WDR81</td>
<td>WD repeat-containing protein 81</td>
</tr>
<tr>
<td>WHO</td>
<td>World health organisation</td>
</tr>
<tr>
<td>WP</td>
<td>Whole parasites</td>
</tr>
<tr>
<td>WSI</td>
<td>Wellcome Sanger Institute</td>
</tr>
<tr>
<td>ZC3H8</td>
<td>Zinc finger CCCH domain-containing protein 8</td>
</tr>
</tbody>
</table>
Chapter 1  Introduction
1.1 Summary

Digital dissection of the RNA and protein content of membrane-bound vesicular structures (termed extracellular vesicles) secreted by the malaria parasite *P. falciparum* and its human host was the overall goal of the work described in this thesis. The first objective was to compare the RNA content of extracellular vesicles secreted by asexual cultures *P. falciparum* (*P*EVs) to the transcriptome of the secreting whole parasites (WP). In the second objective, I aimed to analyse the protein and RNA content of extracellular vesicles secreted by the host into plasma during acute malaria infection. In this introductory Chapter, I provide background about malaria and extracellular vesicles. I start by discussing malaria with a specific focus on post-transcriptional gene regulation and the pathogenesis of severe disease syndromes. In the second section, I describe extracellular vesicles and their role in host-parasite interaction.

1.2 Malaria is still a public health problem

Malaria is a tropical disease endemic in low- and middle-income countries and is caused by protozoan parasites of the genus *Plasmodium* (Phillips et al., 2017). According to WHO 2021 malaria report, there were more than 241 million malaria cases and 627 000 malaria deaths in 2020 alone, the vast majority of which occurred in sub-Saharan African countries (WHO, 2021). The report also stated that there were approximately 14 million fewer malaria cases in 2019 compared to 2020 and 69 000 more malaria fatalities, primarily due to the diversion of resources and attention towards the Covid-19 pandemic at the expense of other infectious diseases (WHO, 2021).

The only licenced malaria vaccine, RTS, S, performed poorly (30 – 50% effectiveness) in clinical trials, and the improved version, R21, is still undergoing further testing (Datoo et al., 2022; Datoo et al., 2021; Hill, 2011; Kazmin et al., 2017; White et al., 2015). Parasite resistance to the frontline drug, artemisinin, was observed in South East Asia as early as 2009, and drug-resistant mutants are starting to emerge in sub-Saharan Africa (Balikagala et
al., 2021; Kayiba et al., 2021; Uwimana et al., 2020; WHO, 2020). The unavailability of a proven malaria vaccine combined with reports of artemisinin-resistant mutants in Africa reinforces the threat of malaria to public health.

1.3 Five species of *Plasmodium* infect humans

The *Plasmodium* species belong to the *Apicomplexa* phylum, named for the characteristic apical complex essential for host cell invasion. Over 250 documented species of *Plasmodium* infect a wide range of host species, including mammals, rodents, birds, snakes, and lizards (Ayala, 1978; Garnham, 1966; Telford, 2009; Valkiūnas, 2004). Five of these *Plasmodium* species infect humans, including *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi* (Antinori et al., 2012; Golgi, 1886; Oddoux et al., 2011; Snowden, 2008; White, 2008). Of these, *P. falciparum* causes most malaria-associated cases and deaths in sub-Saharan Africa and Asia (Alao et al., 2013; Snow, 2015). *P. vivax* is found in temperate climates, mainly in South America, Southeast Asia, and Ethiopia (Lo et al., 2015; Petersen et al., 2013). *P. vivax* preferentially invades Duffy-positive reticulocytes. Therefore, researchers have long thought that most individuals living in the sub-Saharan region were Duffy-negative and hence totally resistant to *P. vivax* malaria (Miller et al., 1976). However, recent studies have found that *P. vivax* can infect Duffy-negative individuals, although the prevalence of vivax among these populations is appropriately low (Gunalan et al., 2016; Lo et al., 2015; Ménard et al., 2010; Niangaly et al., 2017; Russo et al., 2017).

*P. ovale* and *P. malariae*, which cause the mildest clinical disease, are prevalent in Asia and West African countries (Doderer-Lang et al., 2014; Mueller et al., 2007). *P. knowlesi* was initially considered a non-human primate parasite, but zoonotic human infections have been witnessed primarily in Malaysia (Oddoux et al., 2011; White, 2008). This PhD thesis focuses on *P. falciparum*, the most lethal *Plasmodium* species.
1.4 The multistage life cycle of *P. falciparum*

The life cycle of the *P. falciparum* parasite is complex. It involves the mosquito vector and a human host (Fig. 1.1). The parasite has evolved developmental stages that allow its survival in different ecosystems within the host, including the human peripheral circulation, erythrocytes and hepatocytes, as well as the mosquito midgut, basal lamina and salivary glands (Aly et al., 2009; Phillips et al., 2017).

*P. falciparum* infects humans when a female *Anopheles* mosquito species takes a blood meal and inoculates elongated motile forms of the parasites called sporozoites acquired from a previous blood meal from an infected host. In less than 60 min (based on studies using rodent models of malaria) after injection into the blood, most of the sporozoites travel via the venous circulation to the liver, where they invade the hepatocytes (Amino et al., 2006a; Amino et al., 2006b; Sidjanski and Vanderberg, 1997; Yamauchi et al., 2007). Sporozoites invade liver cells by first transmigrating through the Kuppfer cells through a process called sporozoite cell traversal, followed by the formation of parasite-hepatocyte cell junctions and entry into the hepatocyte (called cell invasion) (Amino et al., 2008; Frevert et al., 2005; Hollingdale et al., 1981; Ishino et al., 2004; Mota et al., 2001).

Inside the hepatocytes, the sporozoites divide and differentiate into a cyst-like structure called hepatic schizonts that contain daughter parasites (Meis et al., 1985). The host does not experience any symptoms during this hepatic schizogony process. After approximately five days from inoculation, the hepatic schizonts burst and release the daughter parasites called merozoites into the blood circulation, where they invade erythrocytes in less than a minute (Meis et al., 1985; Sturm et al., 2006). While inside the erythrocytes, the merozoites transform within 48 h from ring-like stages to larger forms called trophozoites and then to multinucleated erythrocytic schizonts, which finally rupture, releasing daughter merozoites into the blood circulation. Most of the released merozoites invade new erythrocytes to continue the asexual blood stage of the parasite (Grassi, 1900; Sakaguchi et al., 2016).
thesis focuses on the intraerythrocytic developmental cycle (IDC) of \textit{P. falciparum} (Fig. 1.1A), which is the stage that causes malaria clinical disease.

In response to stress, a small fraction of the erythrocytic parasites undergo cell cycle arrest to form gametocytes, the mosquito stages transmitted from humans to the mosquito vector (Kafsack et al., 2014). Gametocytes migrate to the bone marrow and undergo five stages of maturation from immature roundish forms to mature crescent-shaped forms (Fig. 1.1B) (Aguilar et al., 2014; Joice et al., 2014; Thomson and Robertson, 1935). Mature gametocytes migrate to the skin capillaries, where a female \textit{Anopheles} mosquito takes them during a blood meal (Chardome and Janssen, 1952; Thomson and Robertson, 1935; Van Den Berghe et al., 1952).

Xanthurenic acid or the change of temperature inside the mosquito midgut activates male gametocytes to undergo three mitotic divisions to generate eight motile microgametes with flagella, while the female gametocytes mature into macrogametes (Billker et al., 2004; Billker et al., 1998; Nijhout, 1979; Raabe et al., 2009). This process is called gametogenesis. The male microgametes seek the female macrogametes, where they fuse and form zygotes, the only diploid stage of \textit{Plasmodium} parasites (Liu et al., 2008; van Dijk et al., 2001). The zygotes differentiate into elongated motile forms called ookinetes, which combine gliding motility and chitinases to invade the chitin-containing epithelial cells and enter the mosquito’s abdomen, where they grow into oocysts (Huber et al., 1991; Vlachou et al., 2004). The oocysts replicate to form sporozoites, which excyst and migrate from the basal lamina to the salivary glands. The mosquito is infectious 7 - 10 days after siphoning the mature gametocytes from the human cutaneous capillaries (Pimenta et al., 1994; Sinden, 1974).
1.5 The *P. falciparum* genome

The *P. falciparum* genome is haploid and consists of more than 23 million base pairs and 14 chromosomes with an extraordinarily high AT content averaging 80% genome-wide but reaching as high as 90% in intergenic and intronic regions (Gardner et al., 2002). During this project's last update of the *Plasmodium* genome database (PlasmoDB), 5725 *P. falciparum* genes were annotated (https://plasmodb.org/plasmo/app).
The parasite genome contains most of the genes required to maintain an organism, but approximately half of these genes have unknown functions. The mechanism of gene regulation is also not apparent from the DNA sequence; therefore, post-genomic studies have played a significant role in providing *Plasmodium* gene regulatory models (Bahl et al., 2003; Gardner et al., 2002).

### 1.6 Transcriptional control of gene expression in *P. falciparum*

The first step of gene expression is the conformation and remodelling of the chromatin, given by nucleosomes. Nucleosomes form from complexes of histones, including octamers containing two of each histone 2A (H2A), histone 2B (H2B), histone 3 (H3), and histone 4 (H4), that are wrapped by DNA 1.6 times and the complex is held together by histone 1 (H1). Posttranslational modifications of the histone side chains, mainly H3, determine the formation of transcriptionally-amenable euchromatin or heterochromatin (Clark-Adams et al., 1988; Han and Grunstein, 1988; Lorch et al., 1987; Olins and Olins, 1974). Although the *P. falciparum* possesses other genes required for chromatin formation, such as H2A, H2B, H3, and H4, the gene that codes for the essential H1 is lacking in all *Plasmodium* species (Aurrecoechea et al., 2016; Gardner et al., 2002). There is a possibility that *Plasmodium* species have a wildly divergent H1 histone that is difficult to predict or that they employ different gene regulatory mechanisms from those used by other eukaryotes.

The binding affinity of histones is lower in high AT-rich genomes than in high GC-rich genomes. Thus, *P. falciparum* histones bind poorly to the AT-rich *P. falciparum* genome, which contributes to increased accessibility of the genome. (Silberhorn et al., 2016). The organization of the *Plasmodium* chromatin architecture is also very unstable during the asexual blood stage (Ponts et al., 2010). Specifically, there is more histone-free DNA than histone-bound DNA during the asexual blood cycle because the chromatin loosens immediately after erythrocyte invasion and rapidly reorganizes during schizogony before the next parasite cycle (Ponts et al., 2010). In addition, except for a few genes, such as the *var*
family, the promoter and terminator regions of *P. falciparum* genes are histone-free (Ponts et al., 2010), which makes them readily accessible to the transcription machinery. Thus, the canonical transcriptional gene regulatory mechanism may not be active in *P. falciparum* during the IDC, which opens the floodgates for investigators to explore other models of gene regulation.

1.7 The transcriptome of the asexual blood stages of *P. falciparum*

Even before the *P. falciparum* genome was sequenced, investigators in the pre-genomic era used microarray technology to identify transcriptional differences between the asexual blood stages (Ben Mamoun et al., 2001) and between gametocytes and trophozoites (Hayward et al., 2000).

Once the genome was available, the DeRisi lab analysed the complete transcriptome of the asexual blood stages of *P. falciparum* at a one-hour scale resolution using microarray (Bozdech et al., 2003; Llinás et al., 2006). To their surprise, they found that transcription of the majority of parasite genes (> 85%) occurs during the asexual life cycle (Bozdech et al., 2003; Llinás et al., 2006), even those required by mosquito stages such as sporozoites, gametocytes, and ookinetes. They also demonstrated that the abundance of the parasite mRNAs transcribed from more than 75% of the *P. falciparum* genes is rhythmic, consisting of a single peak and a single trough (Bozdech et al., 2003; Llinás et al., 2006). These seminal findings were consistent with the histone-free DNA phenomenon described earlier and suggested that *P. falciparum* employs non-transcriptional mechanisms to terminate or switch-off gene expression. The results reported by DeRisi lab were successfully recapitulated by other research groups using DNA microarray (Hoo et al., 2016; Le Roch et al., 2003) and RNAseq (López-Barragán et al., 2011; Otto et al., 2010; Painter et al., 2018; Rijo-Ferreira et al., 2020; Smith et al., 2020), and therefore it is acknowledged within the malaria community that *P. falciparum* asexual blood stages have a rhythmic transcriptional cascade.
CHAPTER 1

In other eukaryotes, transcription factors play central roles in the activation and deactivation of genes (Babu et al., 2004). The 27 members of the Apicomplexan AP2 (ApiAP2) protein family are the only known transcription factors in Plasmodium species that regulate gene expression in all parasite developmental stages (Balaji et al., 2005; Painter et al., 2011; Painter et al., 2018). However, it may not be possible for a fully functional organism to regulate over 5700 genes using such a small set of transcription factors (Painter et al., 2018). Therefore post-transcriptional control could be the primary gene regulatory mechanism in P. falciparum.

1.8 Post-transcriptional gene regulation in P. falciparum

Post-transcriptional gene regulation encompasses all processes that control gene expression at the RNA level (Gebauer and Hentze, 2004; Vembar et al., 2016). Like other organisms, the life of RNA in P. falciparum starts in the nucleus, where the transcriptional machinery synthesizes pre-mRNA. Pre-mRNA undergoes subsequent processing such as splicing, the addition of 5’-end m7GpppN cap and 3’-end poly (A) tail structures resulting in a mature RNA, which is ready for export to the cytoplasm (Moore, 2005; St Johnston, 2005). Once the mRNA is in the cytoplasm, post-transcriptional gene control in P. falciparum can occur via four mechanisms: a) delay between transcription and translation /ribosomal stalling (Bec et al., 2019; Bunnik et al., 2013; Caro et al., 2014; Vembar et al., 2016); b) mRNA degradation/decay (Balu et al., 2011; Vembar et al., 2016; Zhang et al., 2014); c) repression of translation by RNA binding proteins (RBPs) (Bunnik et al., 2016; Reddy et al., 2015; Vembar et al., 2016); and d) antisense RNA binding to the corresponding mRNA, which silences its translation (Fig. 1.2) (Broadbent et al., 2015; Filarsky et al., 2018; López-Barragán et al., 2011). Attainment of equilibrium amongst these post-transcriptional gene control processes is crucial to prevent the mistiming of protein production.
1.8.1 Translational repression in *P. falciparum*

Ribosome-mRNA association is used as a proxy for active protein production. Under stressful conditions, the ribosome stalls at the elongation step and the stalled mRNA is temporarily stored in stress granules, but if stress persists, mRNA decay pathways are activated (Kimball et al., 2003; Mokas et al., 2009). Genome-wide studies have used this premise to advance our knowledge of translational gene control in *P. falciparum* (Bunnik et al., 2013; Caro et al., 2014). One of these studies by Bunnik et al. compared the polysome-associated mRNA to steady-state mRNA at three-time points during the IDC (Bunnik et al., 2013). They observed a delay in the abundance of polysomal-bound mRNA compared to steady-state mRNA for 1280 genes. This delay suggests that at least 20% of the *P. falciparum* mRNAs were stored in a translationally inactive state until their need arose. These translationally inactivated mRNAs were encoded by genes involved in the invasion and remodelling of the erythrocyte. However, more than 43% of the genes had matching
CHAPTER 1

transcription and translation patterns, indicating that delayed translation/ribosomal stalling was not involved in their regulation.

The second study by Caro also compared the RNA bound to the ribosome to the steady-state mRNA and found that translation and transcription in *P. falciparum* were tightly coupled in at least 90% of the genes (Caro et al., 2014). These slightly contrasting results between Bunnik and Caro et al. can be explained by: a) the fact that Bunnik et al. analysed polysomes while Caro et al. analysed both monosomes and polysomes (Bunnik et al., 2013; Caro et al., 2014), b) Bunnik analysed three IDC stages (0, 18, 36 hours post-invasion [hpi]) while Caro analysed five stages (merozoites, 11, 21, 31, 45 hpi) (Bunnik et al., 2013; Caro et al., 2014). However, both studies suggest that only a small proportion of genes are translationally inactivated in *P. falciparum*.

1.8.2 Decay and stabilization of RNA in *P. falciparum*.

In almost every organism, RNA decay is a critical process that prevents aberrant protein production by ensuring that translation only occurs when the protein is required. Decay of mRNA occurs in both 5’-3’ and 3’-5’ directions and involves deadenylation (removal of poly(A)-tail), decapping (removal of the 5’-cap), and degradation by ribonucleases (RNases). Deadenylated mRNA is either decapped by DCP1-DCP2 (decapping protein 1-decapping protein 2) complex and degraded by exonuclease Xrn1 from the 5’-end or degraded by exoribonucleases (exoRNases) from either the 5’-end or the 3’-end (Gardner et al., 2002; Garneau et al., 2007; Houseley and Tollervey, 2009). *P. falciparum* encodes genes required for RNA decay, including CCR4-NOT deadenylase complex (CAF1 and CCR4 deadenylase subunits) as well as RNase D family (RRP6) and exoRNaseII-related family (exoRNases DIS3/RRP44) (Balu et al., 2011; Zhang et al., 2014).

Disruption of the CAF1 deadenylase subunit in *P. falciparum* alters the expression of more than 1000 genes, leading to mistimed translation of these genes, most of which are involved in erythrocyte invasion and egress (Balu et al., 2011). Furthermore, *Pf*RNaseII
CHAPTER 1

post-transcriptionally regulates about 200 genes, including those encoding noncoding RNAs and a subset of var (Zhang et al., 2014). Using biosynthetic RNA labelling, Painter and colleagues observed that active transcription generally occurs throughout the IDC, with notable instances of post-transcriptional regulatory bursts during the late stages of the parasites (Painter et al., 2018). Nonetheless, the mechanisms regulating genome-wide mRNA decay rates in P. falciparum are poorly understood.

1.8.3 Translational repression by RNA binding proteins (RBPs) in P. falciparum

RNA repression is the sequestration of mRNA within RBP to prevent unwanted translation or decay and is crucial during the parasite transition from one stage to the next (Bunnik et al., 2016; Vembar et al., 2016). Like mRNA stalled in ribosomes, mRNA-RBPs complexes in P. falciparum are also temporarily sequestered in stress granules (P-bodies) (Kamenska et al., 2016). Analysis of the P28 (previously called P21) antigen in P. berghei in the early 1990s led to the first corroboration that translational repression occurs amongst malaria parasites. Specifically, two studies found that late gametocytes abundantly transcribed P28, although protein production only occurred in zygotes and ookinetes following transmission to the mosquitoes (Paton et al., 1993; Shaw et al., 1996). After that, an integrated ‘omic study found a similar regulatory pattern for two other mosquito-stage antigens, P25 and P7 (Hall et al., 2005). However, the mechanism behind this gene regulatory mechanism remained unknown until it was observed that a complex of two RBPs, named the DOZI/CITH complex (composed of Development of Zygote Inhibited and its interacting partner CAR-I/Trailer Hitch Homolog), binds RNA encoded by 731 genes including P25 and P28 (Guerreiro et al., 2014; Mair et al., 2006; Mair et al., 2010). Depletion of the DOZI/CITH complex in gametocytes leads to reduced RNA abundance of 211 of the 731 genes indicating that the complex prevents degradation of RNA encoded by these genes (Guerreiro et al., 2014; Mair et al., 2006; Mair et al., 2010). Adding further to the complexity, the Puf (Pumilio and fem-3 binding factor homolog) RBP family was also shown to repress
gametocytogenesis by binding mRNA from P25 and P28 (Cui et al., 2002; Fan et al., 2004; Miao et al., 2013a; Miao et al., 2010).

The Acetylation lowers affinity (Alba) protein family (n = 6) are the only RBPs that have been experimentally implicated in translational gene regulation during the IDC stages of malaria parasites (Vembar et al., 2015). In P. falciparum, PfAlba 1 - 4 coprecipitate with polysomes of the IDC stages (Bunnik et al., 2013), and PfAlba1 binds to mRNA transcribed from 1193 genes, 13 of which encode erythrocyte invasion proteins such as apical membrane antigen 1 (AMA1), high molecular weight rhoptry protein 3 (RhoPH3), calcium-dependent protein kinase 1 (CDPK1) and rhoptry-associated protein 1 (RAP1) (Vembar et al., 2015). However, other than PfAlba1, the role of RBPs in translational repression during the IDC remains a black box.

Two genome-wide in silico surveys have identified putative RBPs in P. falciparum (Bunnik et al., 2016; Reddy et al., 2015). The study by Reddy and collaborators used a retrieval strategy to identify 189 RBPs (3.5% of all annotated genes) in P. falciparum which falls under six categories: Zinc finger, DEAD/H-box RNA helicases, Alba, K homology, and Puf gene families (Reddy et al., 2015). About 24% of the identified RBPs were expressed at higher levels in asexual stages compared to sexual stages, indicating that they could regulate gene expression during the IDC. The other bioinformatic study performed hidden-Markov searches to identify close to 1000 RBPs (18% of all annotated genes), of which 199 interact with mRNA during the IDC (Bunnik et al., 2016). While the authors observed a high abundance of several RNA-binding domains amongst the identified RBPs, there was a substantial underrepresentation of RNA-binding domains linked to RNA degradation (Bunnik et al., 2016). This observation implies that the prominent role of RBPs in P. falciparum might be translational repression rather than RNA degradation.
1.8.4 Translational silencing by antisense RNA

Antisense RNA is another post-transcriptional regulatory mechanism in *P. falciparum*, which involves the synthesis of *cis* or *trans* RNA complementary to the mRNA. The antisense RNAs act as translational repressors by binding to their complementary mRNAs. Serial analysis of gene expression (SAGE) provided the first evidence of an antisense silencing mechanism in *P. falciparum* by revealing that 12 - 17% of the annotated genes encode antisense transcripts (Gunasekera et al., 2004; Patankar et al., 2001). It was later shown that antisense transcription in *P. falciparum* is catalysed by RNA polymerase II (Militello et al., 2005).

Bioinformatic surveys have also been used to predict antisense transcripts using RNAseq data, but most of them have not been experimentally tested (Broadbent et al., 2015; López-Barragán et al., 2011). One of these surveys by Broadbent and others predicted that gametocyte development gene 1 (GDV1), a gene essential for sexual commitment, was potentially regulated by an antisense RNA that overlaps with the GDV1 start codon and initiates downstream of the GDV1 locus (Broadbent et al., 2015). After that, it was experimentally established using the loss of function mutants that the predicted GDV1-antisense RNA inhibits sexual differentiation during the IDC through translational silencing of GDV1 (Filarsky et al., 2018). The cytoadherence-linked antigen 9 (CLA9), which mediates the binding of infected erythrocytes to the cluster of differentiation 36 (CD36), is also translationally repressed by its antisense RNA (Barnwell et al., 1989; Gardiner et al., 2000). There is evidence of antisense long non-coding RNA transcribed from the intron of some *var* genes, but rather than translational repression; *var* antisense transcription activates the homologous *var* gene (Amit-Avraham et al., 2015; Jing et al., 2018).

As I conclude this topic of post-transcriptional gene regulation in *P. falciparum*, I would like to point out outstanding questions that remain to be answered. They include: 1) Is there a single genome-wide regulatory mechanism responsible for the sinusoidal patterns of *P.
CHAPTER 1

*P. falciparum* gene expression during the IDC, or is the fine-tuned cascade of gene expression due to the contribution of several regulatory mechanisms, including those described above? 2) How does the parasite handle the RNA transcribed from mosquito-stage specific genes not required during the IDC? These two questions motivated the study reported in Chapter 3 of this thesis.

1.9 Severe malaria manifests in different clinical syndromes

Diagnosis of severe malaria is performed using multiple case definitions, including cerebral malaria (peripheral parasitaemia with coma), acute respiratory distress, and severe malaria anaemia (English et al., 1997; Marsh et al., 1995; WHO, 2000). Other clinical presentations of severe malaria include metabolic acidosis, hypoglycaemia, pulmonary oedema, acute kidney injury, haemorrhage, jaundice, hyperparasitaemia, and complications of pregnancy, and these have been reviewed here (Burté et al., 2012; Murphy and Breman, 2001; Phillips et al., 2017).

1.9.1 Cerebral malaria

1.9.1.1 Definition and diagnosis of cerebral malaria

Cerebral malaria is a neurological syndrome caused by *P. falciparum*. Patients are diagnosed with cerebral malaria if they have asexual stages of *P. falciparum* and are still comatose one hour after the correction of hypoglycaemia or termination of a seizure (WHO, 2000, 2015). However, this definition is non-specific as other pathogens (viruses and bacteria) or neurological abnormalities can also cause coma with incidental peripheral parasitaemia (Idro et al., 2010). Brain autopsy studies of patients diagnosed with cerebral malaria revealed that more than 20% succumbed to other neurological conditions besides malaria (Taylor et al., 2004; White, 2011). The retina undergoes pathological changes similar to the brain and is currently used to differentiate actual cerebral malaria cases from atypical encephalopathies with incidental parasitaemia (Beare et al., 2006; White et al., 2009).
1.9.1.2 Epidemiology of cerebral malaria

Cerebral malaria is the most severe syndrome caused by *P. falciparum* (Rénia et al., 2012). In the 1990s, the incidence rate of cerebral malaria in sub-Saharan Africa was 7 and 3 cases per 1000 malaria patients aged less than five and 5 to 9 years, respectively (Brewster et al., 1990; Murphy and Breman, 2001). According to WHO, the incidence of cerebral malaria in 2020 was highest in lower malaria transmission settings (1 per 10 cases in areas with less than 5% prevalence). Still, the absolute numbers were expectantly higher in areas with a prevalence higher than 10% (WHO, 2021). Even with appropriate treatment, ~20% of children with cerebral malaria die, but mortality in adults, although slightly higher (30%), can be lowered by treatment with intravenous artesunate (Dondorp et al., 2005; Dondorp et al., 2010; Murphy and Breman, 2001). Although most cerebral malaria patients fully recover, 15 – 20% of survivors develop post-discharge neurological sequelae, including hemiplegia, aphasia, cortical blindness, and ataxia (Birbeck et al., 2010; Idro et al., 2006).

1.9.1.3 Mechanisms of brain damage in cerebral malaria

The exact pathogenesis of cerebral malaria is still controversial. Mechanisms that have been explored include adhesion and sequestration of infected erythrocytes in the brain vasculature, infiltration of activated immune cells into the brain, activation of the brain endothelial cells, and disruption of the blood-brain barrier (Newbold et al., 1999; Riggle et al., 2020; Turner et al., 2013). The sections below will discuss these pathophysiological processes in detail, followed by a descriptive illustration (Fig. 1.3).

**Sequestration of parasitized erythrocytes in the brain:** Adhesion and sequestration of parasite-infected erythrocytes in the brain microvasculature is theoretically the central cause of neural dysfunction in cerebral malaria (Silamut et al., 1999; Storm et al., 2019; Taylor et al., 2004). The parasite employs a family of virulence proteins called *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1) to bind to receptors on brain endothelial cells such as ICAM1, VCAM1, and EPCR. The adherent erythrocytes agglutinate with other
infected erythrocytes, which further increases the sequestered parasite biomass, binds to platelets and white blood cells, and forms rosettes with uninfected erythrocytes. In addition to protecting infected erythrocytes from splenic clearance, sequestration causes brain injury via vascular obstruction, reducing the supply of metabolites such as oxygen and glucose to the brain cells (Dondorp et al., 2004; Newbold et al., 1999). While there is a significant correlation between coma and sequestration, the latter is unlikely to cause tissue necrosis in the brain because malaria-associated coma is rapidly reversed by antimalaria treatment (Çizmeci et al., 2016; Ponsford et al., 2012; Silamut et al., 1999; Storm et al., 2019). Note also that sequestration alone cannot explain the pathogenesis of cerebral malaria as it also occurs among patients with other malaria syndromes (MacPherson et al., 1985).

**Infiltration of immune cells into the brain:** Inflammatory products released by sequestered *P. falciparum*-infected erythrocytes also take part in the pathophysiology of cerebral malaria. During the early stages of *P. falciparum* infection, the host mounts a protective response, including activating CD4+ and CD8+ T cells via antigen-presenting cells. Although almost any immune cell can infiltrate the injured brain, CD8+ T cells are the only ones that have received significant attention as crucial players in cerebral malaria pathogenesis. Both mouse and human studies have demonstrated that the migration, accumulation, and interaction of CD8+ T cells with endothelial cells in the peripheral brain vasculature contributes to cerebral malaria (Barrera et al., 2019; Belnoue et al., 2002; Riggle et al., 2020; Swanson et al., 2016). One of these studies by Riggle and others found that 69% of children with cerebral malaria have CD8+ T cell infiltration in the brain vasculature (Riggle et al., 2020). Using experimental models of cerebral malaria, Swanson observed that parasite-specific CD8+ T cells interact with brain microvasculature causing severe brainstem pathology (Swanson et al., 2016). Activated CD8+ T cells might also upregulate the expression of C-X-C motif chemokine receptor 3 (CXCR3) and CC motif chemokine receptor 5 (CCR5), which subsequently bind to chemokines secreted by brain endothelial
cells, further promoting T cell chemotaxis and infiltration into the brain. Lymphocyte function-associated antigen 1 (LFA-1/CD11a) expressed on the surface of CD8+ T cells can also mediate binding to ICAM-1 on endothelial cells as anti-LFA1 therapy prevents fatal disease in murine models (Howland et al., 2015; Swanson et al., 2016). Once activated, CD8+ T cells secrete chemoattractant biomolecules such as chemokines, perforin and granzyme-B which further recruit other immune cells such as natural killer cells and macrophages into the brain (Belnoue et al., 2002; Haque et al., 2011; Riggle et al., 2020).

**Activation of brain endothelial cells**: Activation of the vascular endothelial cells in the brain is a feature of cerebral malaria resulting mainly from the sequestration of parasitized erythrocytes and the secretion of pro-inflammatory cytokines. Activated vascular endothelial cells upregulate endothelial protein C receptor (EPCR) expression, which aggravates the sequestration of infected erythrocytes (Turner et al., 2013). Variants of PfEMP1 (named DC8 and DC13) preferentially bind to EPCR and have been associated with severe malaria (Argy et al., 2017; Avril et al., 2013; Ortolan et al., 2022). However, the primary function of EPCR is to increase the activation of protein C by binding and presenting it to the thrombin-thrombomodulin complex (Avril et al., 2013; Shabani et al., 2017). Activated protein C still binds to EPCR and initiates a signalling cascade that results in the downregulation of inflammatory cytokines, including IL-6 and TNF, inhibiting the coagulation pathway (Esmon, 2003; Mohan Rao et al., 2014). Thus, the binding of infected erythrocytes to EPCR blocks the available EPCR binding sites and prevents activation of protein C, resulting in increased release of inflammatory cytokines by brain endothelial cells and induction of the coagulation pathway (Shabani et al., 2017). Platelets are the central effector cells of the coagulation system, and they possess adhesion receptors (P-selectin, CD36, and ICAM-1) that mediate interaction with infected erythrocytes (agglutination) and immune cells (Cox and McConkey, 2010; Newbold et al., 1999). However, human observational studies have shown that endothelial activation and sequestration are
independently associated with severe disease and mortality (Abdi et al., 2015; Hanson et al., 2015).

**Disruption of the blood-brain barrier:** The blood-brain barrier (BBB) is a semipermeable membrane that forms the border between cerebral parenchyma and peripheral blood. It comprises a wall of endothelial cells, an end-feet ensheathment of astrocytes, and a basement of pericytes (Ballabh et al., 2004; Daneman and Prat, 2015). The BBB is a selective membrane that allows the diffusion of hydrophobic (oxygen and hormones) and small non-polar molecules while restricting the entry of pathogens, peripheral immune factors, and large hydrophilic molecules into the brain. Metabolites like glucose cross the BBB into the brain parenchyma via active transport.

The binding of infected erythrocytes to vascular endothelial receptors (ICAM-1, VCAM-1, and EPCR) induces signalling cascades involving Rho GTPases, calcium, kinases, VE-cadherin and reactive oxygen species that lead to disruption of the BBB molecular structure causing it to leak (Wittchen, 2009). Other mechanisms through which BBB disruption occurs in cerebral malaria include irreversible damage resulting from the interaction of brain endothelial cells with infected erythrocytes (Storm et al., 2019), TNF-α-mediated upregulation of miRR-155 in endothelial cells, which leads to alteration of tight junctions (Barker et al., 2017) and induction of endothelial cell apoptosis by perforins and granzyme-B released by CD8⁺ T cells (Haque et al., 2011; Schmid et al., 2017). Disruption of the BBB allows cytokines and immune cells to enter the brain and activate glial and neuronal cells resulting in brain injury and neurological sequelae (Song et al., 2022).
1.9.2 Severe malaria anaemia

1.9.2.1 Definition and diagnosis of severe malaria anaemia

A decrease in haemoglobin defines anaemia and depends on age, gender, physiological condition, and place of residence. In developing countries, malaria patients are diagnosed with severe malarial anaemia if their peripheral parasitaemia (>10,000 parasites/µl) is accompanied by haemoglobin concentrations of less than 5g/dl or a haematocrit of less than 15% (Newton et al., 1997; WHO, 2015). However, other causes of anaemia in malaria-endemic regions, including malnutrition, haemoglobinopathies, bacteraemia and hookworms, could also be accompanied by incidental parasitaemia, which makes the differential diagnosis of severe malarial anaemia very difficult.

Fig 1.3 The microenvironment of cerebral malaria

Cerebral malaria develops in overlapping mechanisms including parasite sequestration in the brain vasculature, infiltration of immune cells and factors (e.g. activated CD8 T cells and cytokines), activation of the brain endothelial cells and disruption of the blood brain barrier.
1.9.2.2 Epidemiology of severe malaria anaemia

Severe malarial anaemia is the most common complication of *P. falciparum* infection, contributing to more than 50% of severe malaria cases in East Africa in 2020 (WHO, 2021). Paediatric deaths in referral hospitals in developing countries range from 3% - 46% (English et al., 2004). About 7.5% – 34% of African children with malaria infection also have severe malarial anaemia (Taylor et al., 2006). In the early 1990s, the incidence of severe malarial anaemia in Malawi was 54.7 cases per 1000 children under five years old (Slutsker et al., 1994). Still, half a decade later, Snow reported that the incidence of severe malarial anaemia among children less than five years old was 7.6/1000 per year (Snow et al., 1999). However, more recent multicentre studies in developing countries indicate that the overall prevalence of severe malarial anaemia is about 20%, while the case fatality rate (CFR) is 8% (Perkins et al., 2011; Taylor et al., 2006).

1.9.2.3 The pathogenesis of severe malaria anaemia

The pathogenesis of severe malarial anaemia is not fully understood. Mechanisms that have been proposed include impaired erythropoiesis, reduced erythrocyte deformability, antibody and complement binding, augmented splenic clearance of abnormal erythrocytes, iron deficiency and coinfection with bacteraemia, hookworms, and HIV (Perkins et al., 2011). These mechanisms are discussed below, followed by an illustration (Fig. 1.4).

**Dyserythropoiesis:** Impaired or defective erythropoiesis is arguably the primary cause of low haemoglobin levels in children with severe malarial anaemia, but it remains unproven. Impaired erythropoietic response translates to the inability to replenish erythrocytes lost due to parasite-driven factors such as haemolysis and immune clearance. Small studies in the 1980s observed a reduced rate of erythroblast proliferation among patients with acute malaria, which was suggestive of damage to the bone marrow parenchyma (Dörmer et al., 1983; Wickramasinghe et al., 1982). Dyserythropoiesis in severe malaria patients is often accompanied by erythroid hyperplasia and coincides with reduced reticulocytosis (Abdalla
et al., 1980; Abdalla, 1990; Were et al., 2006). A marked increase in reticulocyte numbers when patients are treated with antimalarial drugs indicates impaired erythropoiesis as the cause of severe malaria anaemia (Wickramasinghe et al., 1982).

Severe malarial anaemia is not due to a deficiency of erythropoietin, as anaemic children with malaria have increased concentrations of erythropoietin compared to anaemic children without malaria (Casals-Pascual et al., 2006; Newton et al., 1997; Nussenblatt et al., 2001; Verhoef et al., 2002) and erythropoietin levels were found to be negatively correlated ($r = -0.88$, $p < 0.05$) with haemoglobin levels in acute malaria patients (Nallandhighal et al., 2019). However, the exact cause of dyserythropoiesis in patients with severe malaria anaemia is somewhat elusive. Still, possible mechanisms include suppression of erythroid proliferation by pro-inflammatory cytokines (Perkins et al., 2011) and inhibition of erythropoiesis by hemozoin-containing monocytes (Awandare et al., 2007; Casals-Pascual et al., 2006).

**Removal of rigidified uninfected erythrocytes by the spleen**: Once infected, the entire erythrocyte stiffens and becomes less deformable as the parasite matures. As mentioned earlier, while it is understandable how infected erythrocytes lose their deformability, uninfected erythrocytes also rigidify (Dondorp et al., 1999; Dondorp et al., 2002). The mechanisms causing stiffness of uninfected erythrocytes have not been adequately dissected but potentially include oxidative damage of erythrocyte membranes (Griffiths et al., 2001) and transfer of parasite-derived molecules to uninfected erythrocytes (Douki et al., 2003; Naumann et al., 1991; Uyoga et al., 2012). The spleen removes rigid uninfected cells from circulation, contributing to anaemia. Indeed, early studies observed a positive association of reduced deformability with decreased haemoglobin concentrations (Dondorp et al., 1999). A murine model also reported a progressive loss of uninfected erythrocytes in *P. berghei*-infected animals with low parasitaemia (Evans et al., 2006). However, reduced uninfected
erythrocyte deformability fails to explain severe malarial anaemia as it also occurs in patients with other syndromes (Dondorp et al., 2002; Dondorp et al., 2004).

**Increased phagocytosis and lysis of uninfected erythrocytes:** Complement activation and erythropagocytosis are involved in the pathophysiology of severe malaria anaemia (Goka et al., 2001). Circulating erythrocytes from severe malaria anaemia patients have increased IgG and immune complexes on their surface and deficiencies in complement inhibitory proteins CD55 and CR1 (Stoute et al., 2003; Waitumbi et al., 2000). Two mechanisms leading to increased deposition of antibodies have been demonstrated. In one such mechanism, parasite antigens such as RAP2/RSP2 adhere on the surface of both infected and uninfected erythrocytes, allowing specific antibodies to opsonize the antigen-erythrocyte complexes and accelerate complement-mediated phagocytosis or lysis (Layez et al., 2005). In the second mechanism, autoantibodies against phosphatidyl-serine (PS) promote the clearance of uninfected erythrocytes (Fernandez-Arias et al., 2016). Phosphatidyl-serine resides within the inner leaflet of the cell membrane, but due to inflammatory or oxidative stress caused by the parasite, it flips to the outer leaflet triggering the immune system to produce anti-phosphatidyl-serine antibodies. The levels of atypical B-cells that secrete anti-phosphatidyl-serine antibodies correlate to the progression of anaemia in patients (Fendel et al., 2010; Fernandez-Arias et al., 2016; Rivera-Correa et al., 2019; Totino et al., 2010). Nonetheless, a study in Thailand did not find increased binding of IgG to erythrocytes from malaria patients, while another one in Cameroon found that all sera recognized RAP2/RSP2 regardless of whether patients were anaemic or not and unexpectedly sera from malarial anaemia patients activated complement to a lesser extent compared to that from non-anaemic patients (Awah et al., 2011; Merry et al., 1986).

**Iron deficiency and malaria anaemia:** There are two schools of thought regarding the relationship between iron and malaria. Proponents of malaria as a cause of iron deficiency anaemia argue that the inflammatory response observed among acute malaria patients could
reduce iron uptake by increasing the expression of iron regulatory protein hepcidin (Drakesmith and Prentice, 2012). Studies supporting this theory have found that iron deficiency or anaemia declines following the interruption of malaria transmission (Frosch et al., 2014; Noland et al., 2012). Mendelian randomization analyses suggest that malaria can cause iron deficiency (Muriuki et al., 2021). However, iron deficiency is generally prevalent in malaria-endemic regions and could be coincidentally reported among malaria patients. Additionally, the mean corpuscular volume (MCV) in severe malaria anaemia patients is higher than that of healthy children within the community (Nallandhighal et al., 2019). Iron deficiency anaemia is usually associated with severe microcytosis; therefore, the higher than typical MCV values observed in severe malaria anaemia patients suggest that iron deficiency may not be involved (Massey, 1992; Nallandhighal et al., 2019).

Iron deficiency has been found to protect patients from severe *falciparum* malaria and mortality in some settings (Gwamaka et al., 2012). In others, iron supplementation promotes recovery from malaria anaemia (Schellenberg et al., 2004; van Hensbroek et al., 1995). Contrastingly, iron interventional studies, including a Pemba island prospective study that was stopped prematurely (Sazawal et al., 2006), demonstrated that iron supplementation increased malaria morbidity and mortality (Sazawal et al., 2006; Soofi et al., 2013; Veenemans et al., 2011; Zlotkin et al., 2013). Thus, it is unclear whether malaria causes iron deficiency, and although the WHO recommends iron supplementation in malaria, it is essentially not implemented (WHO, 2016).
1.9.3 Malaria respiratory distress, hypoglycaemia, and metabolic acidosis

1.9.3.1 Definition and diagnosis of malaria respiratory distress

Malaria-induced respiratory distress, hypoglycaemia, and metabolic acidosis have interlinked pathological mechanisms. The WHO defines respiratory stress as a malaria syndrome featuring deep acidotic breathing, noncardiogenic pulmonary oedema, pneumonia or any other manifestations of dysfunctional respiration. Hypoglycaemia is determined by blood glucose levels of less than 40 mg/dl and is usually accompanied by lactic acidosis (Marsh et al., 1995; Murphy and Breman, 2001; WHO, 2015).
CHAPTER 1

1.9.3.2 Epidemiology of malaria respiratory distress

Severe malaria patients occasionally present with respiratory distress. In a study by Marsh, the prevalence of respiratory distress was 13.7% (n=1,833), while that for hypoglycaemia was 13.2% (n=698). Although respiratory distress can present alone, it is often accompanied by other malaria syndromes, including cerebral pathology and anaemia. The respiratory distress CFR is estimated to be 1.2/1000 children less than five years old, while the hypoglycaemic CFR is somewhere between 1.2 to 2.8 deaths per 1000 children under five years of age (Marsh et al., 1995; Murphy and Breman, 2001; Solomon et al., 1994; Taylor et al., 1988). The prevalence of respiratory distress in adults ranges from 2 to 20% and depends on the Plasmodium species. Below is a discussion and an illustration of the pathophysiology of respiratory distress in malaria patients (Fig. 1.5).

1.9.3.3 Pathogenesis of respiratory distress, hypoglycaemia, and acidosis

Respiratory distress in children with malaria is tightly linked to systemic acidosis and is sometimes referred to as malarial hyperpnea or hyperventilation. Malaria respiratory distress is associated with lung injury and pulmonary oedema rather than metabolic acidosis in adults. Hypoglycaemia is linked to renal failure in adults, but this is less frequent in children (Arem, 1989; Murphy and Breman, 2001).

Metabolic acidosis as a cause of respiratory distress in children: Deep breathing in children with respiratory distress is an aftermath of systemic metabolic acidosis and is primarily due to the accumulation of lactic acid (English et al., 1997). Three theories explain acidotic respiratory distress in children with severe malaria: 1) the asexual blood parasites, which are incapable of oxidative phosphorylation, consume and deplete host glucose with concomitant excretion of large quantities of lactic acid to avoid toxicity (Mehta et al., 2006; Vander Jagt et al., 1990; Zuzarte-Luís and Mota, 2018); 2) peripheral tissues primarily the liver augment glucose uptake pathways (glycogenesis) while at the same time inhibiting glucose production pathways (gluconeogenesis) - hepatic gluconeogenesis is the primary
pathway of clearing and normalizing lactate in the blood (Kawo et al., 1990; van Thien et al., 2004; White et al., 1987); and 3) renal failure which is common in severe malaria could also result to insufficient excretion of lactate (English et al., 1997; Namazzi et al., 2022; Sriboonvorakul et al., 2018). Although the exact mechanism leading to hyperlactatemia is unknown, the low blood pH caused by excess lactic acid triggers the brain to increase the respiratory rate resulting in hyperpnea.

Acute lung injury as a cause of respiratory distress in malaria: Adults (and rarely in children) with severe malaria usually exhibit a type of respiratory distress called acute respiratory distress syndrome (ARDS) or acute lung injury (ALI) (Lucas et al., 1997). Unlike acidotic respiratory distress in children, ARDS is characterized by disruption of the blood-air barrier (alveolar-capillary membrane), pulmonary (alveolar) oedema, diffuse alveolar inflammation (DAI) and severe hypoxemia as described first by (Ashbaugh et al., 1967). Damage to the alveolar-capillary barrier occurs first, followed by leakage of oedema fluid into the alveoli. The engagement of hemozoin with lung tissues has been demonstrated as one mechanism through which lung damage occurs in ARDS (Deroost et al., 2013). Infiltration of CD8\(^+\) T cells into the lungs also occurs, suggesting that T cell pathological mechanisms such as secretion of perforins and granzymes might also contribute to ARDS (Claser et al., 2019; Pham et al., 2017). ARDS is usually caused by conditions concomitant in severe malaria patients, including sepsis and pneumonia but the complication can also present alone in malaria patients in the absence of any other known causes.
CHAPTER 1

1.10 Minimally invasive biomarkers of severe malaria syndromes are lacking

Severe malaria syndrome affects different remote tissues such as the brain (cerebral malaria), haematological tissue (severe malaria anaemia), and the lung (respiratory distress). However, much of what we understand about the effect of malaria syndromes in these remote tissues has been facilitated by autopsy studies because obtaining measurements from these organs from malaria patients is not practical. However, it is now widely acknowledged that all cells release tiny sacs called extracellular vesicles into their environment (Raposo and Stoorvogel, 2013; Théry et al., 2018). The biological content of extracellular vesicles resembles that of the secreting cells or tissues. Therefore extracellular vesicles are attractive sources of minimally invasive disease biomarkers, but this prospect has been poorly explored in the context of severe malaria. In chapter 4 and 5, I will demonstrate that the biological content of circulating extracellular vesicles obtained from the plasma of acute malaria

---

Fig 1.5 The pathophysiology of acidotic respiratory distress

Respiratory distress in children results from elevated blood lactate levels which results from increased synthesis by *P. falciparum* and immune cells or reduced lactate clearance by the liver and kidney

---
patients not only mirror what we know about malaria syndromes but also fuels new biological theories. To reflect the experimental goal of all my result Chapters, I provide a general description of extracellular vesicles in the next part of the introduction.

1.11 Extracellular vesicles - background

Extracellular vesicles (EVs) are a group of nano-sized particles secreted by all cells into their environment (Raposo and Stoorvogel, 2013). EVs are surrounded by a lipid bilayer and contain a wealth of biologically active molecules, including proteins, lipids, RNA, and genomic DNA, as part of their cargo. EVs transfer their cargo from donor to recipient cells, impacting the biological functions of the recipient cells. An increasing number of studies support the role of EVs in pathogenesis and as potential targets for therapies and diagnostics in cancer and infectious diseases.

1.11.1 A brief history of extracellular vesicles

Pioneering experiments with platelets were the first to hint at cell-derived vesicles' presence and biological significance. In the mid-1940s, extended high-speed ultracentrifugation (31,000 x g for 150 min) of plasma prolonged the coagulation time of the supernatant, and the clotting time shortened when the pellet was reintroduced to the supernatant showing that subcellular factors are involved in blood coagulation (Chargaff, 1945; Chargaff and West, 1946). Seventeen years later, it was demonstrated that this subcellular pellet was made up of vesicular particles with a diameter of 10-50 nm and a density of 1.020 to 1.025 g/ml (Wolf, 1967). Platelets secreted these vesicles and were thus named “platelet dust” (Wolf, 1967). After that, they were shown to contain lipids and biological cargo, including contractile proteins and ATP (Crawford, 1971).

Early electron microscopy studies also demonstrated that cells other than platelets secrete vesicular structures whose size was consistent with what we now describe as EVs. During the summer of 1966, Sun et al. found that alveolar cells secreted “vesicle-like structures” into the alveolar space (Sun, 1966). In the late 1960s, two studies described small
membranous vesicles ingrained in the hypertrophic cartilage matrix (Anderson, 1969; Bonucci, 1967). The role of multivesicular bodies (MVBs) in EV biogenesis was first described in 1974 (Nunez et al., 1974). Using the bat thyroid gland, Nunez and Gershon observed that MVBs containing intraluminal vesicles fuse with the cell membrane, releasing the vesicles to the extracellular space (Nunez et al., 1974). EVs were also found incidentally by researchers who were not explicitly searching for them. While looking for viruses in biological fluids, researchers often came across “virus-like particles” that were non-viral (Dmochowski et al., 1968; Fawcett, 1956; Levine et al., 1967; Prince and Adams, 1966; Seman et al., 1971). These “virus-like particles” were regarded as artefacts of virus isolation (Prince and Adams, 1966) until Dalton published a paper in 1975 reporting that the morphology of the “virus-like particles” resembled that of intraluminal vesicles within MVBs and that it was not right to call them “virus-like” (Dalton, 1975).

It was not until the early 1980s that the term “exosome” was conceived after the purification of similar-sized vesicular particles secreted by reticulocytes (Harding et al., 1983; Trams et al., 1981). These vesicles were enriched in the transferrin receptor protein and were thus theorized as a mechanism for eliminating the no longer required proteins as the reticulocytes transformed to mature erythrocytes (Harding et al., 1983; Trams et al., 1981). Graca Raposo first reported the role of EVs in cell-to-cell communication in 1996. In her seminal work, Raposo found that B cell-derived EVs can activate T cells in an MHC -class II-restricted manner (Raposo et al., 1996).

1.11.2 The nomenclature and biogenesis of EVs

The nomenclature of EVs has rapidly evolved and is influenced by different isolation methods and the multi-disciplinary nature of the field. A diverse range of terminologies has been used, such as dexosomes (from dendritic cells) (Nikfarjam et al., 2020), prostasomes (from prostate cancer cells) (Aalberts et al., 2014), oncosomes (from tumour cells) (Meehan et al., 2016), exosomes (Roy et al., 2018), exosome-like-particles, microvesicles and
microparticles (Raposo and Stoorvogel, 2013). To reflect the multidisciplinary nature of the EV field, the International Society of Extracellular Vesicles (ISEV) proposed two main types of vesicles based on size and biogenesis: small EVs (sEVs) and medium EVs (mEVs) (Théry et al., 2018).

Small EVs (formerly called exosomes) are 30 - 200 nm in diameter and form when late endosomes bud inwards to generate MVBs containing intraluminal vesicles (ILVs). MVBs fuse with the plasma membrane and discharge their ILVs into the environment as small EVs (Fig 1.6) (Théry et al., 2018; Wu et al., 2021). Classical surface markers of small EVs include tetraspanins (such as CD9, CD63, and CD81), heat shock proteins, and endosomal sorting complex required for transport (ESCRT) protein family, including alix and tumour susceptibility gene 101 (TSG101) (Théry et al., 2018).

Conversely, medium EVs (previously called ectosomes, microvesicles or microparticles) range from 100 - 1000 nm in diameter and form from the outward invagination of the plasma membrane (Fig 1.6) (Théry et al., 2018). Examples of canonical markers of medium EVs include annexin, phosphatidyl-serine and ADP-Ribosylation factor 6 (ARF6). The current purification protocols do not candidly distinguish between small and medium EVs. Therefore the term extracellular vesicles (EVs) accommodates all cell-derived vesicles (Théry et al., 2018).
1.12 Mechanisms of sorting cargo into EVs

1.12.1 Sorting of proteins into EVs

Ubiquitin-dependent endosomal protein sorting into EVs: Post-translational modifications mark proteins for sorting into EVs (Ageta and Tsuchida, 2019; Moreno-Gonzalo et al., 2018). Ubiquitination is the primary mechanism that targets proteins for loading into EVs (Ageta and Tsuchida, 2019). Three enzymes catalyse the classical ubiquitination reaction: a) E1:ubiquitin activating enzyme, which activates ubiquitin in an ATP-dependent manner; b) E2:ubiquitin conjugating enzyme, which is a carrier for activated...
ubiquitin; and c) E3:ubiquitin protein ligase which catalyses the conjugation of activated ubiquitin with target proteins (Deshaies and Joazeiro, 2009; Schulman and Harper, 2009; Ye and Rape, 2009). Misfolded or unwanted proteins are ubiquitinated and are either degraded by both proteasome-dependent and independent (lysosomal) pathways (Ciechanover and Schwartz, 1998; Mukhopadhyay and Riezman, 2007) or are alternatively sorted into EVs by the ESCRT machinery involved in MVB biogenesis (Ageta and Tsuchida, 2019; Moreno-Gonzalo et al., 2018).

**Ubiquitin-like endosomal protein sorting into EVs:** Some proteins such as neuronal precursor cell expressed developmentally downregulated 4 (NEDD4), ubiquitin-like sequence protein 3 (UBL3), and small ubiquitin-like modifiers (SUMO) possess ubiquitin-like sequences that modify specific proteins and target them for sorting into EVs even in the absence of classical ubiquitination (Welchman et al., 2005). For example, sumoylation was demonstrated to promote ubiquitin-independent ESCRT sorting of α-synuclein into EVs (Emmanouilidou et al., 2010; Kunadt et al., 2015), and the ubiquitin-like sequence of the UBL3 protein can modify target proteins marking them for secretion via EVs (Ageta et al., 2018; Takanashi et al., 2022).

**Ubiquitin-independent endosomal protein sorting into EVs:** Some proteins belonging to the ESCRT complex, including Alix, the CHMP, and VSP protein families, are also involved in sorting proteins into EVs in a ubiquitin-independent manner. Alix is an ESCRT accessory protein that sorts proteins into EVs by binding to their cytoplasmic domains, and its depletion was found to reduce the incorporation of CD9, CD63, and CD81 into EVs (Larios et al., 2020). The CHMP proteins work together with the VSP proteins to sort proteins into EVs. For instance, the CHMP5 component of ESCRT interacts with VSP4 and lysosomal trafficking regulator-interacting protein 5 (LIP5) to sort aquaporin 2 (AQP2) into MVBs (Roche et al., 2019; Vild et al., 2015). Furthermore, the VSP4A subunit of ESCRT was found to interact with the CHMP4B subunit to sort β-catenin into EVs. Indeed,
overexpression of VSP4A increases the quantities of β-catenin in EVs without affecting the size and number of the secreted EVs, and the silencing of CHMP4B and VSP4A significantly decreases the amount of β-catenin in EVs (Han et al., 2019).

**Lipid raft-mediated protein sorting into EVs:** The lack of ubiquitination or the ESCRT complex leads to a decrease in the protein content of EVs rather than absolute depletion, suggesting that other pathways are also involved in sorting proteins into EVs (Stuffers et al., 2009). The cytosolic side of the MVBs membrane contains microdomains called lipid rafts, with high affinity to biological cargo, including proteins (de Gassart et al., 2003). The transfer of proteins associated with the lipid rafts into the lumen of EVs can occur independently of ESCRT and requires sphingolipid ceramide produced from the breakdown of sphingolipids by sphingomyelinases (Trajkovic et al., 2008). Survivin is one of the proteins whose secretion via EVs was shown to depend on lipid rafts (Valenzuela et al., 2014).

**Tetraspanin-mediated protein sorting into EVs:** Tetraspanins are membrane proteins enriched in EVs and are associated with protein sorting and biogenesis of EVs. They include CD9, CD37, CD53, CD63, CD81, and CD82 and are believed to be markers of small EVs (Escola et al., 1998). The role of CD63 in packaging latent membrane protein 1 (LMP1) into EVs has been demonstrated. LMP1 is an oncoprotein encoded by Epstein-Barr virus (EBV), and transportation of LMP1 via EVs is essential for the replication and pathogenesis of the virus. CD63-positive EVs exhibit high quantities of LMP1, and the amounts of LMP1 in EVs decrease in CD63-knockout cell lines suggesting the critical role played by CD63 in packaging LMP1 into EVs (Hurwitz et al., 2018; Hurwitz et al., 2017). The tetraspanin CD9 was shown to promote the sorting of metalloprotease CD10 into EVs. The metalloprotease CD10 secreted via EVs with the help of CD9 contributes to B-cell maturation by modifying the extracellular matrix environment (Mazurov et al., 2013).
1.12.2 Lipid sorting into EVs

The membrane bilayers of EVs are composed of lipids. The lipidic content of EVs is very different from that of the secreting parent cells. EVs contain 2-3 times higher amounts of PS, cholesterol, sphingolipids, and ceramide. At the same time, phosphatidyl-choline (PC) is lesser in exosomes compared to the parent cells, and the quantities of PE between EVs and parent cells are not different (Llorente et al., 2013; Skotland et al., 2017). Accumulating cholesterol in MVBs is essential for forming ILVs, the precursors of EVs, whereas enrichment of sphingomyelin and ceramide in EVs potentially results from the inward budding of membrane lipid rafts (Trajkovic et al., 2008).

Lipids are asymmetrically distributed within the membrane bilayer of EVs due to the action of flippases, floppases, and scramblases during EV formation (Clark, 2011; Hankins et al., 2015). For instance, there are higher levels of phosphatidyl-serine on the EV bilayer's outer leaflet than on the inner leaflet, which is abnormal because phosphatidyl-serine is preferentially located within the inner leaflet of cell membranes. Conversely, sphingolipids and PC, which are expected to be on the outer leaflet in other membranes, are enriched in the inner leaflet in EVs (Van Meer et al., 2008). Although mechanisms of sorting lipids into EVs are poorly understood, they are thought to be independent of protein sorting and largely depend on the size and quantity of EVs rather than the type of the parent cell (Haraszti et al., 2016).

1.12.3 RNA sorting into EVs

Lipid raft-mediated loading of RNA into EVs: Some RNA molecules have high affinity to the lipid raft-like regions in the outer layer of the MVBs-membrane. Cellular RNAs interact with the rafted domains of MVBs membrane using specific sorting aptamers or secondary structures and are then encapsulated into ILVs during the budding-in process mediated by ceramide (Janas et al., 2020; Trajkovic et al., 2008). The selective loading of
RNA molecules depends on their differential affinity to phospholipids within the rafted domains of the MVB membrane (Janas et al., 2006).

**RBP-mediated loading of RNA into EVs:** RBPs bind to RNA molecules to regulate gene expression post-transcriptionally. It was demonstrated that 30 RBPs in EVs form RBP-RNA complexes (Statello et al., 2018). The RBP-RNA complexes can be loaded into EVs via several mechanisms. The RBP could be sumoylated or ubiquitinated in one such mechanism, allowing the RBP-RNA complex to be sorted into EVs by the classical ESCRT machinery. It was shown that one of the most common RBPs, HnRNPA2B1, must be sumoylated before binding to EV-sorting motifs in target RNA molecules (Villarroya-Beltri et al., 2013). Some RBPs are also thought to mediate the loading of RNA into EVs by delivering the RBP-RNA complex to the lipid rafts of the MVBs membrane (Janas et al., 2015). Specifically, HnRNPA2B1 has a high affinity for lipid rafts on the surface of MVBs membrane, a property that allows it to sort miRNA into EVs (Villarroya-Beltri et al., 2013). RBPs were also found to interact with the autophagy initiator microtubule-associated protein 1A/1B-light chain 3 (MAP1LC3A/LC3), allowing the loading of the RBP-RNA complexes into EVs (Leidal and Debnath, 2020; Leidal et al., 2020).

### 1.13 The Biological Significance of EVs

EVs play myriad biological functions, summarised into three categories: 1) cellular housekeeping/homeostasis; 2) communication and exchange between cells, tissues, and organs; and 3) microenvironment seeding, including acting as decoys.

#### 1.13.1 EVs Maintain Cellular Homeostasis

EVs are used by cells to package and dispose of biomolecules if their physiological requirement has been met. Specifically, differentiating cells eliminate molecules that are no longer required via EVs. For instance, pioneering studies on EVs found that reticulocytes released the no longer required transferrin receptor as they matured to erythrocytes (Harding et al., 1983; Trams et al., 1981). Integrin α4β1 and the water channel aquaporin 1 (AQP1)
are secreted by reticulocytes via EVs (Blanc et al., 2009; Rieu et al., 2000). Other proteins that are partly secreted in reticulocyte-derived EVs include basigin, ABCB6 and CD55 (Carayon et al., 2011; Kiss et al., 2012; Rabesandratana et al., 1998), which are three receptors required by *P. falciparum* to invade host cells (Crosnier et al., 2011; Egan et al., 2015; Egan et al., 2018). On a similar note, the secretion of phosphorylated pseudokinase mixed lineage kinase domain-like (MLKL) via EVs is a mechanism used by cells to contain the necroptotic activity of this kinase (Yoon et al., 2017).

The release of nucleic acids via EVs also directly regulates their intracellular levels and is a way to avoid accumulating unwanted genetic material or to regulate gene expression at the RNA level. For example, fibroblast cells prevent unnecessary activation of DNA damage response pathways by packaging and excreting cytoplasmic DNA via EVs (Takahashi et al., 2017). During stressful cell conditions such as serum starvation, translation of specific mRNAs stops, and preinitiation complexes (PICs) containing stalled mRNA are sequestered into stress granules. Upon release of stress, translation of the mRNAs in the PICs is reinitiated, but if stress persists, the PICs and the stalled mRNAs are packaged into EVs and evacuated from cells (Bec et al., 2019).

Extracellular disposal of miRNAs via EVs is also a mechanism used by cells to keep in check the intracellular levels of miRNAs and to regulate gene expression within the secreting cells. Marabita et al. demonstrated that activated lymphocytes downregulate the intracellular levels of miR150 by secreting it via EVs. miR150 regulates mRNAs essential for lymphocyte differentiation; therefore, its disposal via EVs could be a quick way to prevent excessive lymphocyte responses (de Candia et al., 2013). This hypothesis is supported by the fact that low levels of miR150 in serum are associated with poor survival in critically ill septic patients (Roderburg et al., 2013). Cancer cells also maintain their oncogenic properties by eliminating tumour-suppressive miRNAs (e.g., miR23b and Let-7) via EVs (Kobayashi et al., 2014; Ohshima et al., 2010; Ostenfeld et al., 2014).
1.13.2 EVs transfer signalling molecules to specific targets

EVs can transfer their biological cargo from donor to recipient cells, impacting function. The specificity of EV mediated cell to cell communication is possible due to the following: 1) secreted EVs have surface transmembrane ligands such as integrins that can recognize target cells, and 2) the presence of receptors on the surface of target cells that allow binding and uptake of EVs (Stratman et al., 2022). The role of EVs in cell-cell communication has been demonstrated using EVs from antigen-presenting cells. Specifically, EVs released from B cells and dendritic cells were shown to express MHC-peptide complexes allowing the EVs to activate T cells just like the parent cells (Raposo et al., 1996; Tkach et al., 2017). Similarly, EVs released by immune cells contain TNF receptor ligands such as TRAIL, TNF, and FasL on their surface, allowing them to bind to TNF receptors and trigger apoptosis through caspase activation (Munich et al., 2012; Stenqvist et al., 2013). However, EVs can also interact with recipient cells in a non-specific manner through mechanisms such as membrane fusion mediated by Rab and SNARE proteins (Parolini et al., 2009; Prada and Meldolesi, 2016), raft-mediated endocytosis, macropinocytosis, and phagocytosis.

1.13.3 EVs seed local microenvironments and act like decoys

The seeding of local microenvironments by EVs occurs when secreted EVs contain active molecules, such as enzymes that influence the concentrations of specific signalling molecules in the target local microenvironment (Iraci et al., 2017; Stratman et al., 2022). The adipose tissue is one of the microenvironments which EVs greatly influence. Most EVs in the adipose tissue are derived from the adipocytes and can be taken up by other cells in this microenvironment, such as immune and endothelial cells. Recent studies have linked the transfer of adipocyte EVs to macrophages as a potential cause of inflammation and insulin resistance of the adipose tissue in obese patients (Kranendonk et al., 2014b; Reilly and Saltiel, 2017). There are three potential ways through which this occurs: 1) adipocyte EVs transfer lipid droplets to monocytes which stimulates the monocytes to adopt features of
adipocyte tissue macrophages (e.g., increased lysosome content, lipid accumulation, and multinucleation) as they differentiate (Flaherty et al., 2019; Kranendonk et al., 2014a), 2) Adipocyte EVs contain miR34a and miR1224 which inhibit M2-like anti-inflammatory phenotype (Pan et al., 2019; Zhang et al., 2022) and miR155 which stimulates M1-like pro-inflammatory phenotype (Zhang et al., 2016), 3) Macrophages polarized into the M1-like pro-inflammatory phenotype also release EVs that promote not only inflammation but also insulin resistance in the adipose tissue (De Silva et al., 2018; Ying et al., 2017).

Tumours also seed their local environment using EVs to allow the spread and growth of cancer cells. Melanoma EVs reprogramme bone marrow progenitor cells towards a premetastatic phenotype by transferring the oncogenic kinase, receptor tyrosine kinase MET (Peinado et al., 2012). Cancer-derived EVs also promote metastasis in organs far away from the secreting tumours by stimulating the proliferation of distantly-related cells (Costa-Silva et al., 2015; Hoshino et al., 2015). EVs also act like decoys that sequester and neutralize pathogens and their toxins. A mouse model showed that circulating host EVs promote survival from methicillin-resistant Staphylococcus aureus (MRSA) by binding and neutralizing pore-forming bacterial toxins (Keller et al., 2020). More recently, COVID-19 patients with uncomplicated clinical disease were found to have increased levels of EVs that express ACE2. The ACE2-containing EVs could bind and neutralize SARS-CoV-2 irrespective of the variant (El-Shennawy et al., 2022; Kim et al., 2022).

**Acquired immunity:** Host-EVs play a role in antigen presentation as circulating EVs derived from macrophages, B cells and dendritic cells express MHC class I and MHC class II as well as co-stimulatory molecules on their surface, which allows them to bind to pathogen molecules and directly present them to T cells (Raposo et al., 1996). Dendritic cells can also capture EVs loaded with foreign antigens and present them on their MHC class I and MHC class II molecules for subsequent activation of CD4+ or CD8+ T cells.
CHAPTER 1

1.14 Extracellular vesicles in malaria

1.14.1 *P. falciparum*-infected erythrocytes secrete extracellular vesicles

In 2013, two studies independently showed that *P. falciparum*-infected erythrocytes secrete sub-micron particles resembling EVs previously characterized in mammalian cells (Fig. 1.7) (Mantel et al., 2013; Regev-Rudzki et al., 2013). One of those studies by Mantel et al. found that erythrocytes infected with *P. falciparum* secrete 10 - 13 times more EVs than uninfected erythrocytes (Mantel et al., 2013). The authors used *in vitro* experiments to show that secretion of EVs by *P. falciparum*-infected erythrocytes increases as the parasite progresses through the IDC and peaks during schizogony (Mantel et al., 2013). The other study by Regev-Rudziki et al. contrastingly found that ring stages of *P. falciparum*-infected erythrocytes secrete small EVs ranging from 80 to 120 nm in diameter (Regev-Rudzki et al., 2013). The differences between these two studies are attributable to; a) different EV isolation protocols, with Mantel et al. using 60% sucrose cushioned centrifugation (100,000 x g, 16 h) while Regev-Rudziki et al. used 10-30% OptiPrep gradient centrifugation (250,000 x g, 18 h), b) Regev-Rudziki focusing on ring-stage parasites only (Mantel et al., 2013; Regev-Rudzki et al., 2013).

![Fig 1.7 Asexual stages of *P. falciparum* secrete extracellular vesicles](image.png)

**Fig 1.7 Asexual stages of *P. falciparum* secrete extracellular vesicles**

A Early-stage parasites release exosome-like EVs ranging in size from 80 to 120 nm as observed by (Regev-Rudzki et.al., 2013). B Late stage parasites release small and medium sized EVs as observed by (Mantel et.al., 2013).
CHAPTER 1

1.14.2 How might Plasmodium species generate and secrete PfEVs?

The biogenesis of *P. falciparum*-derived extracellular vesicles (termed PfEVs in this thesis) is incompletely understood. The current literature, however scanty, suggests that *P. falciparum* secretes a mixture of both small and medium EVs. The size of ring-stage PfEVs (80 - 120 nm) and the presence of flotilin1, a small EV marker, partially insinuates that they could be small EVs originating directly from the parasite (Regev-Rudzki et al., 2013; Sampaio et al., 2018). Recent studies further support that *P. falciparum* possesses an endocytic sorting complex required for transport (ESCRT) which plays a role in PfEVs secretion (Avalos-Padilla et al., 2021). On the other hand, the PfEVs released by schizonts are likely to be medium EVs (blebs of infected erythrocytes) due to their size (100 - 400 nm) and the presence of medium EVs markers such as ADP-ribosylation factor 6 (ARF6) and phosphatidyl-serine (Mantel et al., 2013).

Nonetheless, there is evidence that vesicular trafficking of proteins occurs in parasite-infected erythrocytes. The cytosol of infected erythrocytes is filled with a heterogeneous mixture of parasite-derived intracellular vesicles whose size matches that of PfEVs. However, it is unclear whether these cytosolic vesicular structures are nascent PfEVs. For instance, *P. falciparum* secretes two types of relatively mobile vesicles into the erythrocyte cytosol, namely the 15 - 25 nm coated vesicles (Hanssen et al., 2008; Wickert et al., 2003) and the 60 - 100 nm uncoated vesicles (Taraschi et al., 2001; Trelka et al., 2000). The smaller coated vesicles bud off from the Maurer’s clefts (MC) and later merge with the erythrocyte membrane (Hanssen et al., 2008; Wickert et al., 2003). The presence of PfEMP1 and PfEMP3 (exported parasite proteins) in the larger uncoated vesicles confirms their function in the transport of parasite material to the erythrocyte membrane (Taraschi et al., 2001; Trelka et al., 2000). It was recently revealed that artemisinin resistance mutation, K13C580Y, induces an increase in the formation of intracellular vesicles that contain PfEMP1 and phosphatidylinositol 3-phosphate (PI-3P) (Bhattacharjee et al., 2018). Since
ring-stage \textit{Pf}EVs contain PI-3P, which is essential for forming multivesicular endosomes, there could be an unexplored link between cytosolic vesicles in infected erythrocytes and \textit{Pf}EVs (Nascimbeni et al., 2017; Petiot et al., 2003; Sampaio et al., 2018).

Caveola-vesicle complex (CVC) structures of pinocytotic origin are on the plasma membrane of \textit{vivax}-infected erythrocytes. CVCs are cup-shaped invaginations of the plasma membrane called caveolae that measure 90 nm in diameter and contain small (50 nm) vesicles (Aikawa et al., 1975; Barnwell et al., 1990). \textit{P. falciparum} ring stages release small, highly mobile compartments called J dots into the host cell cytosol ranging from 0.1 to 0.2 μM in diameter (Gormley et al., 1992; Hibbs and Saul, 1994; Külzer et al., 2010). The highly mobile vesicles contain two-parasite type II HSP40 that bear a J domain, hence the name J dots, and they stain with rhodamine and acridine dyes suggesting that they contain phospholipids and nucleic acids, respectively (Gormley et al., 1992; Hibbs and Saul, 1994; Külzer et al., 2010). MC resident and transient proteins, including STEVOR, KAHRP, SBP1 and \textit{Pf}EMP1, are absent in J dots indicating that they are distinct structures (Külzer et al., 2010). Another study demonstrated the trafficking of small vesicles attached to actin filaments between the MC and the erythrocyte membrane, but this vesiculation pathway is lost in individuals with haemoglobin mutations (Cyrklaff et al., 2011). Nonetheless, actin is an essential protein in extracellular vesicle biogenesis, and further studies may reveal the similarity between these intracellular vesicles and their extracellular counterparts (Holliday et al., 2019; Lombardi et al., 2021).

Eukaryotic MVBs are enriched in cholesterol and lysophosphatidic acid (Ilnytska et al., 2021; Möbius et al., 2003). The \textit{Apicomplexan} rhoptry organelle situated at the apical part is rich in cholesterol (Coppens and Joiner, 2003; Foussard et al., 1991), and the endocytic pathway is believed to play an essential role in its formation (Bannister et al., 2000). The honeycomb appearance of the rhoptry lumen is believed to result from internal vesicles that contain phospholipid membranes bearing transmembrane proteins (Bannister et al., 2000).
Rhoptry discharge during erythrocyte invasion releases intramembranous particles. \( P/f \)EVs are enriched in rhoptry proteins suggesting that rhoptries could be prototype MVB-related organelles and that those membranous particles are indeed nascent extracellular vesicles (Abdi et al., 2017; Bannister et al., 1986; Dluzewski et al., 1989; Stewart et al., 1986). Cumulatively, all these studies indicate that vesiculation occurs in \( P. falciparum \).

1.14.3 The role of \( P/f \)EVs in malaria

Although recently discovered (Mantel et al., 2013; Regev-Rudzki et al., 2013), it is already established that \( P/f \)EVs play a role in the interaction within the parasite population and with the host. \( P/f \)EVs are higher in severe malaria compared to uncomplicated malaria (Combes et al., 2004) and have the potency to activate cells of the immune system (Mantel et al., 2013; Mwangi et al., 2021; Sampaio et al., 2018) and alter the integrity of the endothelial cells (Mantel et al., 2016). Interestingly, when added to parasite cultures, \( P/f \)EVs promote gametocytogenesis (Mantel et al., 2013; Regev-Rudzki et al., 2013). \( P/f \)EVs also promote erythrocyte invasion (Dekel et al., 2021) and can act as anti-malarial drug delivery vehicles (Borgheti-Cardoso et al., 2020). The sections below will cover a descriptive overview of these functions of PfEVs.

1.14.3.1 \( P/f \)EVs and drug resistance

\( P/f \)EVs have been implicated in drug resistance. Regev-Rudzki showed that transgenic parasite lines could transfer drug-resistance genes to recipient parasites. Plasmids containing genes for blasticidin deaminase or dihydrofolate reductase enzymes were used to confer resistance to blasticidin or WR99210, respectively. The two parasite lines died when cultured separately in the presence of both drugs. However, both survived when the isolates were cultured in the presence of the two drugs. Additional experiments were performed by treating the drug-sensitive parasites with \( P/f \)EVs from drug-resistant parasites revealing that genetic transfer of drug resistance was occurring via \( P/f \)EVs secreted during the early ring stages of the IDC (Regev-Rudzki et al., 2013). However, while these experiments proved
that genetic exchange between parasites occurs via *Pf*EVs, it is worth noting that *P. falciparum* does not naturally harbour plasmids; therefore, this mechanism of drug resistance may not happen in the wild.

1.14.3.2 *Pf*EVs regulate parasite growth.

While searching for apoptosis markers in *P. falciparum*, Correa found that *Pf*EVs isolated from high parasitaemia cultures (20-25%) activate death signals within the parasite population. Proteomic analysis revealed that *Pf*LDH was more enriched in *Pf*EVs from highly parasitized cultures than those from lowly parasitized cultures. Several assays were performed, which established that *Pf*EVs obtained from high-density cultures inhibit parasite growth in new cultures and induce apoptosis markers such as increased ROS levels, caspase activation, and phosphatidyl serine translocation. The amplification of ROS levels in treated parasites was shown to be mediated by *Pf*LDH (Correa et al., 2019).

Contrastingly, another related study found that treating uninfected erythrocytes with *Pf*EVs promotes parasite growth. Proteomic analysis of the *Pf*EVs in this study showed enrichment of human kinases and proteasome subunits, which were shown to be functional using biochemical methods. Treating uninfected erythrocytes with *Pf*EVs induced phosphorylation of cytoskeletal proteins such as β-adducin, ankyrin-1 and dematin. Finally, the researchers showed that the phosphorylated proteins serve as targets for degradation by 20S proteasome, which *Pf*EVs also delivered. Digestion of the erythrocyte cytoskeletal proteins increases erythrocyte membrane deformability which primes these cells for the forthcoming parasite invasion (Dekel et al., 2021). These two studies show that *Pf*EVs contain signalling molecules that regulate the parasite population.

1.14.3.3 *Pf*EVs promote the formation of gametocytes.

*Pf*EVs also play a role in the sexual differentiation of the parasites. When added to parasite cultures, *Pf*EVs induce the formation of gametocytes, the sexual stages of the parasite transmitted from humans to the mosquito vector (Mantel et al., 2013; Regev-Rudzki
et al., 2013). Even before these seminal findings, it was still a common practice among researchers working on sexual stages to use a parasite-conditioned medium to stimulate gametogenesis (Fivelman et al., 2007; Miao et al., 2013b). Thus, PfEVs isolated from the same conditioned medium may be the trigger to switch to sexual stages. However, the active component in PfEVs that promotes gametogenesis has not been investigated.

1.14.3.4 PfEVs modulate the immune cells

It is increasingly being established that PfEVs are virulence factors that activate the immune system in malaria. Early studies on PfEVs secreted by schizonts demonstrated that they could activate monocytes to secrete pro-inflammatory (IL-6, TNFα) and anti-inflammatory (IL-10) cytokines (Mantel et al., 2013). These results supported that parasite survival highly depends on the delicate balance between pro-inflammatory and anti-inflammatory mediators. Similarly, PfEVs secreted by ring-stage parasites contain genomic DNA, which activates a cytosolic immune sensor in monocytes called stimulator of interferon genes (STING) (Sisquella et al., 2017). Once activated, STING initiates a signalling cascade that ultimately leads to increased production of type I interferon (IFN) by monocytes (Sisquella et al., 2017; Skopelja-Gardner et al., 2022). The authors concluded that type I IFN likely promotes parasite growth by dampening the immune system (Sisquella et al., 2017). Other studies have shown that Type I IFN can induce inflammation and damage host tissues, leading to severe malaria.

PfEVs also affect granulocytes, specifically neutrophils. Using microfluidic devices, researchers have shown that PfEVs induce increased chemotaxis and probably exhaustion of neutrophils (Mantel et al., 2013). Conversely, PfEVs also suppress the ability of neutrophils to produce reactive oxygen species and cytokines (Babatunde et al., 2018b). This suppression could negatively impact the bactericidal activity of neutrophils and could partially explain the increased co-infection with bacteria among malaria cases (Scott et al., 2011). Recently, it was demonstrated that PfEVs containing RNA could fuse with and
activate natural killer (NK) cells. This activation was via the melanoma differentiation-associated protein 5 (MDA5), a retinoic acid inducible gene 1 (RIG-1) like receptor in natural killer (NK) cells involved in sensing cytosolic foreign RNA (Ye et al., 2018).

We recently showed that PfEVs isolated from ring-stage parasites have a differential ability to modulate immune cells compared to infected erythrocytes. Specifically, PfEVs preferentially induce the secretion of Th17 cytokines such as MCP1, IL-10 and IL-17α (Mwangi et al., 2021). All these studies prove that PfEVs can transfer their biological cargo to immune cells, impacting function.

1.14.3.5 PfEVs increase the permeability of host endothelial cells

The transfer of PfEVs cargo to endothelial cells has also been demonstrated. In brief, the uptake of PfEVs containing human miRNA-Argonaute-2 complexes by endothelial cells leads to the downregulation of the caveolin-1 gene. This downregulation was explicitly mediated by the binding of miR451a to its target caveolin 1. Due to the essential role of caveolin 1 in maintaining the endothelial barrier integrity, its downregulation leads to increased permeability of the endothelial cells. The uptake of PfEVs by endothelial cells also increased the secretion of cytokines and the expression of VCAM-1. Still, the exact component in PfEVs that activates endothelial cells has not been identified. The researchers who carried out this study concluded that upregulated expression of VCAM-1 (a receptor for parasite virulence protein, PfEMP1) could augment the sequestration of infected erythrocytes. At the same time, the increased secretion of cytokines could promote local inflammation (Mantel et al., 2016).

1.14.4 PfEVs contain parasite biomolecules

1.14.4.1 The lipid composition of PfEVs

Generally, EVs have a phospholipid bilayer resembling a conventional cell membrane (Raposo and Stoorvogel, 2013). Only two studies have analysed the lipid composition of PfEVs (Borgheti-Cardoso et al., 2020; Gulati et al., 2015). Comparing the lipid profile of
CHAPTER 1

PfEVs to that of the secreting whole parasites, Gulati et al. determined that phosphatidylserine and phosphatidyl-inositol were enriched in PfEVs while phosphatidyl-choline was excluded. They also found that immune modulatory lipids such as ceramide (Cer), lactosylceramide (LacCer) and monosialodihexosylganglioside (GM3) were abundant in PfEVs. However, the authors did not explain whether the PfEVs-lipids originated from the host cell or the parasite (Gulati et al., 2015). The second lipidomic study of PfEVs did not find significant qualitative differences between the lipid content of EVs from infected and uninfected RBCs. However, it was found that the most abundant lipids in both EVs were phosphatidylcholines, but small amounts of sphingomyelins were also detectable (Borgheti-Cardoso et al., 2020). However, these studies seem pretty different in that the first compares the EVs to the parasite while the second compares them to RBCs.

1.14.4.2 The protein content of PfEVs

Several studies have demonstrated that PfEVs contain host and parasite proteins (Abdi et al., 2017; Dekel et al., 2021; Mantel et al., 2013; Sampaio et al., 2018). The first study described the protein content of PfEVs released by late-stage parasites and identified over 450 human and 78 P. falciparum proteins. The parasite proteins detected in this study included Maurer’s cleft resident proteins (PfMC-TM, SBP1, MARHP1 and REX), parasitophorous vacuole membrane proteins (EXP2 and ETRAMP2), exported surface proteins (RESA and CLAG3.1) and rhoptry proteins (RAP2 and RhopH2), but surprisingly KAHRP and PfEMP1 were not detected (Mantel et al., 2013). The second study by Abdi and colleagues analysed the PfEVs secreted by a recent clinical isolate adapted to short-term in-vitro culture (~ <100 cycles). They compared their data to that of the first study by Mantel and identified 100 proteins that were only present in PfEVs released by a recent clinical isolate (Abdi et al., 2017). In 2018, another proteomic study reported that PfEVs released within the first 12 hours post-invasion (hpi) contain 178 host proteins and 101 parasite proteins, including the virulence factor PfEMP1 (Sampaio et al., 2018). In 2021, the same
research team performed another proteomic profiling of PfEVs and identified 800 human and 387 parasite proteins. The human host proteins were enriched in kinases and 20S-proteasome subunits that synergistically play a role in priming naive erythrocytes for invasion, as described above (Dekel et al., 2021).

Note that the above proteomic studies were limited in the number of biological and technical replicates and were not performed in precisely the same stages of the parasites. Hence, differences in protein number and abundance should be interpreted with caution.

1.14.4.3 The RNA content of PfEVs

Three studies have reported parasite-derived RNA's presence in PfEVs (Babatunde et al., 2018a; Sisquella et al., 2017; Ye et al., 2018). The first study sequenced PfEV-RNA obtained from the early stages of the parasite and found that PfEVs were enriched in genomic DNA, mature microRNA, and unannotated parasite RNA. In addition, they also detected parasite mRNA in PfEVs, of which they verified the full-length mRNA of at least one P. falciparum gene that encodes ETRAMP11.2 using targeted PCR (Sisquella et al., 2017).

The second study also found that PfEVs contained host miRNA species involved in biological processes such as cell cycle and cell adhesion. Other non-coding RNA species identified included snoRNA, piwi RNA, tRNA, Y-RNA, and Vault RNA. The researchers also reported the presence of mRNA in PfEVs, most of which code for exported proteins such as RESA, ETRAMP and PHIST family (Babatunde et al., 2018a). The final study used agarose gel to show that PfEVs contain capped RNA as large as 8 kb, which was not detected in EVs from uninfected erythrocytes (Ye et al., 2018).

Although these studies have provided evidence that mRNA is secreted via PfEVs, the benefit of this secretion to the secreting parasite is unknown. In addition, these studies were limited in the number of timepoint samples collected as the parasite progressed throughout the IDC and, therefore, could not capture the genome-wide RNA profiles of PfEVs. In Chapter 3, I provide more detailed information about the RNA content of PfEVs by
CHAPTER 1

comparing 4-hourly transcriptomes of PfEVs from six culture-adapted isolates to those of the secreting whole parasites.

1.14.4.4 Plasma-derived EVs are associated with severe malaria

Human and animal studies have been used to demonstrate associations between host-derived EVs and severe malaria. EV shedding in malaria mainly depends on the ATP-binding cassette transporter A1 (ABCA1), which is required for membrane vesiculation (Boadu and Francis, 2006; Hamon et al., 2000). An experimental model of cerebral malaria found that the knock-out of the ATP-binding ABCA1 gene reduced EV secretion (Combes et al., 2005). Secretion of EVs in patients with cerebral malaria and the wildtype ABCA1 gene was significantly higher than in patients with uncomplicated malaria and polymorphisms in the ABCA1 gene (Sahu et al., 2013).

A Malawian clinical study found that plasma-derived medium EVs (called microparticles/microvesicles) were elevated in cerebral malaria patients (Combes et al., 2004). There is also a correlation between TNF synthesis and the release of EV by endothelial cells in human cerebral malaria, suggesting that the secretion of EVs is a pathological effect of TNF-mediated inflammation (Combes et al., 2004; Sahu et al., 2013; Wassmer et al., 2011). Furthermore, an in vitro study demonstrated that LMP-420, which potently inhibits TNF production, also inhibits membrane vesiculation (Wassmer et al., 2005). These studies suggest that the ability to release EVs is associated with severe malaria. However, these studies focused on medium EVs, and the role of small EVs in malaria pathogenesis is underexplored.

Characterizing the protein and RNA content of plasma-derived small EVs from malaria patients would be a crucial step towards understanding their role in malaria pathology. However, such studies are lacking, and the only similar ones available focused primarily on medium-sized extracellular vesicles (Antwi-Baffour et al., 2017; Ketprasit et al., 2020). Host-derived small EVs possess advantages over traditional markers of severe malaria,
including i) the capacity to function as non-invasive biomarkers that reveal the effects of malaria infection to remote organs such as the kidney, brain and liver; ii) the ability to reflect clinical disease progression; iii) ability to protect biological cargos from the harsh physiological environment or frequent freeze/thaw cycles during storage. In Chapters 4 and 5, I demonstrate that the biological cargo of host-derived small EVs reflects the general biology of severe *falciparum* malaria.

1.15 Thesis objectives

1.15.1 Main objective

To dissect the biological cargo of small EVs secreted by *Plasmodium falciparum* and its human host using transcriptomic and proteomic approaches.

1.15.2 Specific objective one

To describe the RNA content of small extracellular vesicles secreted by *P. falciparum* (PfEVs) infected erythrocytes in the context of the following key issues:

1) Transcriptional conservation: Is the RNA content of small PfEVs transcriptionally conserved across isolates just like the intracellular RNA?

2) Periodicity: Does secretion of RNA via small PfEVs occur stochastically like the steady-state RNA within the secreting parasites?

3) Relative abundance: Are RNAs from specific subsets of *P. falciparum* genes enriched in small PfEVs to the whole parasites?

1.15.3 Specific objective two

To describe the protein content of EVs obtained from acute malaria patients while focusing on the following:

1) Which proteins and pathways are enriched in host small EVs from severe malaria compared to non-severe malaria patients?

2) Can the host small EV proteomes predict the progression of severe malaria?
3) Are there biomarkers in small EVs capable of distinguishing between severe malarial anaemia, respiratory distress and cerebral malaria?

1.15.4 Specific objective three

To compare the host-RNA content of small EVs obtained from retinopathy positive and negative cerebral malaria while addressing the following questions:

1) Which host-RNA content of plasma-derived small EVs distinguish retinopathy positive and negative cerebral malaria?

2) Are RNAs encoded by brain-specific transcripts enriched in plasma-derived small EVs from retinopathy-positive or negative cerebral malaria cases?

3) Is retinopathy-negative cerebral malaria a “milder” form of the disease compared to retinopathy-positive cerebral malaria, or do the two evolve in distinct disease trajectories?
Chapter 2

Materials and Methods
CHAPTER 2

2.1 RNAseq of *P. falciparum* extracellular vesicles

2.1.1 *P. falciparum* isolates cultured

I cultured six *P. falciparum* isolates to obtain EVs secreted into the culture medium. Five of these were obtained from malaria cases in Kenya, have been in culture for less than 100 asexual cycles and were named KE01, sKE01, KE02, KE04 and KE06 based on the KEMRI-Wellcome Trust Research Programme (KWTRP) parasite naming system. Isolate sKE01 is a product of selecting KE01 using HBEC cells in nine passages. KE01 and KE04 were obtained from cerebral malaria cases, while KE02 and KE06 were from uncomplicated malaria cases. KE01, KE04, and KE06 were hospitalized cases, while KE02 was not. The sixth isolate was a long-term laboratory isolate called Dd2, whose parent was initially obtained from South Asia (Laotian malaria patient).

2.1.2 Preparation of EV-depleted albumax and culture medium

Five percent (w/v) Albumax II (Gibco) was dissolved in RPMI 1640 medium and centrifuged at 150,000 x g for 2 h at 4 °C (Beckmann Coulter) to deplete vesicles. The EV-depleted Albumax II was then sterilized by filtering using 0.22 µm (Millipore). The processed Albumax II was aliquoted into 50 ml portions and stored at -20 °C until use. An incomplete medium was prepared using the following: RPMI1640 (10.43 g/L), glucose (0.2%v/v), L-glutamine (2 mM), hypoxanthine (50 µg/ml), gentamicin (25 µg/ml) and HEPES (37.5 mM) (all from Gibco) under sterile conditions. To make 500 ml of incomplete medium, I added EV-depleted Albumax II. The incomplete medium was also used for the routine washing of parasite cultures, while the complete medium was used for growing the parasites.

2.1.3 Culturing of *P. falciparum* asexual parasites

I cultured asexual blood parasites in 50 ml culture flasks kept at 2-4% parasitaemia and 1-2% haematocrit in fresh (less than one week old) human O+ RBCs using the complete
CHAPTER 2

medium prepared in section 2.1.2 above. The culture-conditioned medium was changed daily, and the culture flasks were gassed using a mixture of 5% O₂, 5% CO₂ and 90% N₂. The parasite cultures were incubated at 37°C, and the temperature of the incubators was checked twice a day. Giemsa smears were prepared every day to monitor parasitaemia. The parasites were then expanded to six 50 ml culture flasks per isolate with occasional synchronization using 5% sorbitol which kills the mature stages of the parasites.

2.1.4 Synchronization of parasites

Once the six flasks of parasite cultures reached 5 - 8% parasitemia at 1.5% haematocrit, I synchronized the parasites using Percoll as previously described (Miao and Cui, 2011). In brief, parasite cultures having late trophozoites (35 - 40 hpi) were overlaid on a 65% Percoll and centrifuged at 2500 × g for 10 min with low brake. The supernatant and the RBC pellet (having mostly ring stages) were discarded while the middle layer having the mature parasites, was washed twice in the incomplete medium at 800 × g for 5min at room temperature. Giemsa smearing was performed to confirm that the asexual stage of the parasites was primarily late trophozoite. The cultures were then resuspended in fresh medium while at the same time adjusting the parasitaemia and hematocrit to 4% and 1.5%, respectively. The flasks were gassed and incubated, and Giemsa smearing was performed after 15 hours. Suppose most (>80 %) of the parasites were at the early ring stage. In that case, synchronization using 5% D-sorbitol was performed to kill any mature parasites which were yet to progress to the new cycle. The parasite cultures were resuspended in the complete medium, gassed, and incubated at 37°C.

2.1.5 Harvesting of culture-conditioned medium for EV isolation

I collected four-hour samples of culture-conditioned medium (CCM) containing parasite-derived EVs into 50 ml Falcon tubes (totalling 300 ml per timepoint) and stored them temporarily at 4 °C. A sample (~20 µL) of the whole parasite (WP) pellet was also collected,
lysed using 5 mL of 0.02% saponin, centrifuged at 8000 x g for 5 min and washed once in 1x phosphate-buffered saline (PBS) to exclude the red cell material and the parasite pellet was resuspended in RNA lysis buffer which is part of the Isolate II RNA Min Kit (Bioline). The extracted whole parasite pellets were kept at -80 °C until use, while the rest of the culture was resuspended in fresh medium, gassed, and returned to incubation at 37 °C.

2.1.6 Enrichment of PfEVs from CCM

CCM processing and PfEVs isolation were performed using an established protocol in our laboratory (Fig. 2.1) (Abdi et al., 2017). Succinctly, the collected CCM samples were centrifuged at 2000 x g for 10 min at 4 °C to remove any remaining cells and large culture debris. The supernatant was then transferred to a new 50 ml Falcon tube and centrifuged at 3200 x g for 30 min, 4 °C, with the pellet again discarded. Final centrifugation at 15000 x g for 20 min at 4 °C was performed to remove medium-sized EVs (microvesicles/microparticles). The debris-free CCM was passed through 0.22 µm (Millipore) to remove particles larger than 220 nm and stored at -80 °C until use.

2.1.7 PfEVs isolation from processed CCM

The processed CCM was thawed, transferred to 15 ml Millipore concentration columns (50kDa cut-off) and centrifuged at 3200 x g to decrease the volume by approximately 30 times. The concentrate was transferred to 13.5 ml Quick seal ultracentrifuge tubes (Beckmann) and centrifuged at 150,000 x g, for 2 h at 4 °C without brakes using a preparative ultracentrifuge (Beckmann Coulter, 70 Ti rotor). The supernatant was discarded, and pellets were resuspended in 500 µL of pre-filtered PBS and treated with 10 µL of RNase A for 15 min at room temperature. The treated pellets were washed once with 13 ml of PBS and centrifuged for 2 h at 150,000 x g, 4 °C, with the supernatant being discarded. Sectioned transmission electron microscopy confirmed that the pellets contained extracellular vesicles.
For RNA extraction, the \( P/\)EVs pellets were resuspended in RNA lysis buffer, which is part of the Isolate II RNA Min Kit (Bioline), and kept at -80 °C until use.

![Extracellular Vesicles (EV)](image)

**Fig 2.1 Extracellular vesicle isolation by ultracentrifugation**

A Culture conditioned medium containing extracellular vesicles. B Processing of culture conditioned medium by differential centrifugation. C Filtration of culture conditioned medium using 220nm filter. D Concentration of culture conditioned medium using 50KDa filter columns. E Isolation of extracellular vesicles by ultracentrifugation and treatment of EV pellets with RNAse A to digest non-vesicular RNA.

### 2.1.8 RNA isolation from \( P/\)EVs and whole parasites

The \( P/\)EVs and whole parasite-lysis buffer suspensions were thawed while in ice. RNA was isolated using the Isolate II RNA Min Kit (Bioline) following the manufacturer’s instructions. Once on the binding column, the RNA was treated with DNAse I to digest genomic DNA. The RNA was washed and eluted in 14 \( \mu \)L of nuclease-free water. The nature of extracted total RNA was assessed using Bioanlyzer Pico RNA chips (Agilent). Messenger
CHAPTER 2

RNA (mRNA) was enriched from the isolated whole parasite total RNA using NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) while following the manufacturer's instructions.

2.1.9 Preparation of cDNA libraries for sequencing

Researchers at the Wellcome Sanger Institute developed a protocol to prepare the cDNA libraries, which was used with slight modifications (Chappell et al., 2020). Synthesis of first strand cDNA was performed using Superscript III reverse transcriptase (Invitrogen) with the following cycling parameters: 25 °C for 10 min, 42 °C for 60 min and a final hold of 4 °C. Random hexamers and oligo d(T) from Qiagen were used to prime first strand cDNA synthesis while dithiothreitol (DTT) was included in the reaction to stabilize the reverse transcriptase. Actinomycin D and RNAse inhibitor were added to reduce antisense artefacts and prevent RNA digestion. The first strand cDNA reaction was cleaned using 1.8 volumes of RNAcleanXP beads solution and washed twice in 85 % ethanol while on the magnetic stand (Invitrogen). First-strand cDNA was eluted in 20.5 µL of Qiagen elution buffer (10mM Tris-HCl pH 8.0), but the beads were left in the wells to be used when cleaning the second strand cDNA reaction. Unless stated otherwise, the Qiagen elution buffer was used to elute cDNA in all downstream steps.

The second strand was synthesised using NEBNext® RNA Second Strand Synthesis Module (New England Biolabs) (parameters: 16 °C for 2.5 h and a final hold of 4 °C). In the second strand synthesis reaction while dUTP rather than dTTP was used. To clean the second strand cDNA, the binding properties of the beads already in the reaction were rejuvenated by adding 1.8 volumes of 20% polyethylene glycol (PEG) and 2.5 M NaCl solution. The beads were washed twice in 85% ethanol while still on the magnetic stand, and the cDNA was eluted using Qiagen elution buffer.
cDNA libraries were prepared using the NEBNext® Ultra II DNA Library Prep Kit (New England, Biolabs) following the manufacturer’s instructions. First, double-stranded cDNA was enzymatically shredded to achieve an average length of 500 bp and then ligated to NEXTflex adapter oligos with barcodes. The cDNA libraries were cleaned using 0.8 volumes of RINAcleanXP beads solution followed by two successive washes in 85% ethanol while still in the magnet. The cDNA libraries were treated with Uracil-Specific Excision Reagent (USER enzymes, Biolabs) at 37 °C for 15 min, followed by a DNA denaturation step at 95 °C for 10 min to digest the second strand. The cDNA libraries were amplified in 19 cycles using KAPA HiFi HotStart ReadyMix PCR kit (KAPABIOSYSTEMS) and P5 and P7 Illumina primers to increase yield. The PCR products were cleaned using 0.8 volumes of RINAcleanXP beads solution as described above and quantified using qRT-PCR. cDNA library samples were pooled in equimolar batches of 24 samples each and sequenced using the Illumina Hiseq 4000 genome analyser at the Wellcome Sanger Institute (WSI).

2.1.10 Quantification of RNAseq reads

Fastq files were downloaded from the Sanger sequencing pipeline, and quality parameters such as Phred scores, percentage GC content and presence of overrepresented sequences were assessed using the tool FastQC. Assembling and summation of gene expression were performed using a super-fast aligner called Kallisto using default settings (Bray et al., 2016). The Plasmodium falciparum 3D7 transcriptome fasta file was downloaded from PlasmoDB and used as the reference. There is a consensus that the expression hypervariable genes such as var, rif and stevor cannot accurately be compared across P. falciparum isolates due to a high degree of polymorphism; therefore these were filtered from the dataset. In addition, because mRNA was enriched from the whole parasite RNA, all non-polyadenylated genes were excluded from further analysis.
CHAPTER 2

2.1.11 RNAseq data normalization

Normalization of the RNAseq data before modelling in downstream steps was done to minimize systematic technical variation. First, a single data matrix containing the read counts was formed by merging the expression data from the individual samples using the tool *tximport*. The count data was then normalized for sequencing depth and gene length by converting to transcripts per million (TPM). TPM values were transformed using *vsn* and smoothed using *loess* algorithms in R. The tool *ARSyn* (Nueda et al., 2012), based on the ASCA model, was used to adjust the data to reduce the inherent effects of asynchronous parasite isolates. Multidimensional scaling and Pearson correlation were used to assess the data after the normalization procedure.
2.1.12 Estimation of rhythmic parameters

The normalized expression matrix was split into small datasets containing whole parasite or PfEVs data from each isolate. Then, the RNA profile of each parasite gene was modelled as a sinusoid, as previously described (Bozdech et al., 2003). The following parameters were estimated from the rhythmic model; 1) the phase, which is the timing of gene expression in whole parasites (or secretion via PfEVs) within the 48 h cycle of asexual blood stage parasites, 2) the mesor, which is the rhythm adjusted mean around which the RNA abundance oscillates (Fig. 2.3). Differences between WP and PfEVs regarding the rhythmic parameters were also statistically estimated and tested. All the chronobiological analysis was performed using CircaCompare tool in R. Phaseograms were created using ComplexHeatmap, and genes were ordered along the y-axis based on their phase.

![Fig 2.3 Rhythmic parameters estimated from cosinor regression models](image)

2.1.13 Functional analysis

Gene ontology terms were retrieved from the malaria Bioconductor database. Functional enrichment analysis was carried out to identify which sort of genes were over-represented among those enriched in PfEVs relative to the whole parasite. The level of significance was tested using a one-tailed Fisher’s exact test (FET). In addition, the Seurat tool was used to
assign *Plasmodium falciparum* genes as expressed by either asexual or mosquito parasites stages (gametocytes, sporozoites and ookinetes) by reanalysing the Malaria Cell Atlas data (Howick et al., 2019; Real et al., 2021; Reid et al., 2018). Enrichment of asexual and mosquito stage genes in *Pf* EVs was assessed using bar plots and correlation of log2FC (asexual vs mosquito stage) to delta mesor (*Pf*EV minus WP).

### 2.2 Proteomic profiling and analysis of plasma-derived EVs

#### 2.2.1 Samples and design

 Archived plasma samples obtained from acute malaria patients were used in this study. Sample size (N=216) was pragmatically determined based on available samples, resources and clinicopathological data.

#### 2.2.2 Isolation of EVs from Plasma for proteomic profiling

 EVs were isolated using nanofiltration followed by ultracentrifugation. In summary, 200 µL of plasma was transferred to 15 ml falcon tube and mixed with 13.5 ml of prefILTERED PBS. The diluted plasma was filtered using 0.22 µm (Millipore) to exclude traces of cellular debris and centrifuged at 150,000 x g for 2 h at 4 °C with no brakes. The supernatants were discarded while the pellets were washed once in 13.5 ml and centrifuged at 150,000 x g for 2 h at 4 °C. The supernatants were again discarded while the pellets containing EVs were lysed using 100 µL 6 M urea in 100 mM Triethylammonium bicarbonate (TEAB) and stored at -80 °C until when required. EV pellets for non-proteomic experiments such as flow cytometry were resuspended in 200 µL PBS.

#### 2.2.3 Extracellular vesicles bead-assisted flow cytometry

 Qualitative analysis of the isolated EVs was done using bead-assisted flow cytometry as described previously (Suárez et al., 2017). 50 µL of EVs in PBS were mixed with 1 µL of aldehyde/sulfate-latex beads (Invitrogen). 1 ml of PBS was added, and the mixture was incubated for 12 h at room temperature on a rotary wheel. Blocking was done by adding 110
µL of 1M glycine followed by incubation at room temperature for 30 min. The beads were pelleted by centrifugation at 2000 x g for 5 min, and the supernatant was removed. The pellet was washed once in 1 ml of PBS, resuspended in PBS supplemented with 0.5% fetal bovine serum (PBS+0.5%FBS) and stained using anti-CD9-APC (BD, Bioscience) and anti-CD63-PE (BD, Bioscience) for 30 min at 4 °C. Beads included with unstained EV sample, with antibodies only or with isotype antibodies were used as negative controls. Washing was done twice with 500 µL PBS+0.5%FBS, and the stained beads were pelleted by centrifugation at 2000 x g for 10 min. Data acquisition was performed using Fortessa flow cytometry (BD, Bioscience) and visualized using scatter plots.

2.2.4 In-solution trypsin digestion of EV proteins

The TMT 10-plex isobaric Mass labelling kit (Thermo Fisher Scientific, Catalog number 90113) was used to prepare the samples for mass spectrometry-based proteomic quantification while following the manufacturer’s instructions.

2.2.4.1 Reduction, alkylation, and precipitation

The EV protein lysate (100 µL) was mixed with 5 µL of the 0.5M tris(2-carboxyethyl)phosphine (TCEP) and incubated at 55 °C for 1 h. Next, I alkylated the proteins by adding 5 µL of 375mM iodoacetamide to reduce the solution and incubated for 30 min at room temperature while protecting from light. Methanol-Chloroform precipitation method was used to pellet the alkylated proteins. To the reduction-alkylation solution, 400 µL methanol was added and vortexed thoroughly, followed by 100 µL chloroform. 300 µL of mass spectrometry grade water was added, making the mixture cloudy with precipitate. Centrifugation was then performed at 14000 x g for 1 min. The result was three layers: a small chloroform layer at the bottom, a circular protein flake at the interphase and a large top aqueous layer. The aqueous layer on top was carefully removed without disturbing the protein flake, and 400 µL of methanol was added. This was followed by centrifugation at
CHAPTER 2

20,000 \times g \text{ for } 5 \text{ min}, \text{ which slammed the protein precipitate against the tube wall. The methanol was discarded while the protein precipitate was dried under a vacuum for } 5 \text{ min.}

2.2.4.2 Protein digestion and peptide labelling

Two enzymes were used to digest the EV proteins: lysyl endopeptidase and trypsin. First, the precipitated protein pellet was resuspended in 100 \mu L of 100mM TEAB, mixed with 2.5 \mu g of lysl endopeptidase, and incubated at 37 °C for 12 h. The solution was then mixed with 2.5 \mu g of trypsin and incubated for 18 h at 37 °C. The digested peptides were labelled using TMT 10-plex isobaric tags (Thermo Fisher Scientific). A reference sample was prepared by pooling 5 \mu L of each of the individual samples and was added as a control in each pool to be used to remove batch effects from the proteomic datasets as described later.

The 0.8 mg TMT reagents were reconstituted in 40 \mu L of acetonitrile (ACN), and 10 \mu L of the reagent was added to 25 \mu L of the 100 \mu L sample. The solution was mixed by pipetting and incubated for 1 h at room temperature. The labelling reaction was quenched by adding 2 \mu L of 5% hydroxylamine and incubating at room temperature for 15 min. Samples were pooled into 24 tubes, each containing 9 individual patient samples and one reference sample. Subsequently, samples were completely dried using a vacuum centrifuge and cleaned using C18 Ziptip peptide desalting columns (Millipore). Desalted TMT labelled peptides were eluted in 15 \mu L mass spectrometry sample loading buffer (2% ACN, 98% H\textsubscript{2}O) and shipped for MS3-based data acquisition at Cambridge Institute of Medical Research (CIMR) in the United Kingdom.

2.2.5 Searching raw peptide spectra against the human database

All 24 raw instrument files were processed in a single search using Maxquant version 2.0.1.0. Searching of MS3-spectra was performed using the Andromenda search engine in Maxquant against the Homo sapiens proteome fasta file downloaded from the Uniprot database (Organism ID: 9606, Proteome ID: UP000005640). The “Reporter ion MS3”
quantification setting was specified with a 0.003 Da mass tolerance while configuring TMT on lysine and n-terminus residues as a static protein modification. Trypsin specificity was specified in the Maxiquant run while allowing for two missed cleavages, N-terminal acetylation (+42.01056 Da), and methionine oxidation (+15.99492 Da) as dynamic protein modifications. In addition to TMT, fixed carbamidomethyl (+57.021 Da) modifications on cysteine protein residues were also configured.

To obtain as many peptide-spectrum-matches (PSM) as possible for subsequent determination of match confidence using retention time, the FDR cut-off in Maxiquant, which is based on the less sensitive target-decoy method, was set at 100%. The PSM confidence scores were determined using the DART-ID (Chen et al., 2019). DART-ID is a recent python tool developed by Slavov lab (Budnik et al., 2018; Chen et al., 2019) which uses Bayesian statistics to align global retention times (RT) and applies the estimates in determining the confidence of the peptide-spectrum-matches (Chen et al., 2019). The cut-off of peptide-spectrum-match confidence was set at q-value < 0.01.

2.2.6 Processing of peptide-spectrum-matches and normalization

The proteus R package was used to collate peptide-spectrum-matches into proteins. First, the median was used to summarise reporter intensities of spectra that matched the same peptide. Peptide intensities were then aggregated into proteins using the high-flyer method (Silva et al., 2006). Only proteins matched by at least three peptides and having not more than 50% missingness were used for subsequent analysis. A reference sample-based signal drift correction procedure was performed using custom R scripts based on the ruv algorithm (Jacob et al., 2015) to minimise the effects of batches. Assessment of the data before and after correction of signal drift was performed using MDS plots and pairwise Pearson correlation of the reference samples. Sample-wise normalization of the post-corrected dataset was done using cyclic loess in limma (Ritchie et al., 2015).
CHAPTER 2

2.2.7 Differential protein abundance analysis

*Limma* R/Bioconductor package (Ritchie et al., 2015) was used to perform differential protein expression analysis. *Limma* fits a linear model to each feature in the expression data to obtain coefficients that describe the design matrix. The test of significance was performed by applying the Empirical Bayes method, which shrinks the standard errors towards a common value followed by a t-test moderated using F-statistics. FDR adjustment of the raw p-values was performed using the Benjamini-Hochberg (BH) method. Results of differential expression analyses were presented as volcano plots, heatmaps and MDS plots.

2.2.8 Estimation of disease progression pseudotime

The EV proteome samples were ordered along the disease trajectory using Slingshot (Street et al., 2018). In brief, PCA embeddings for each sample were calculated and used as input for ordering the samples based on their similarity. Slingshot was then used to estimate pseudotime values for each EV sample. Linear models were fitted for each protein in *limma* while using pseudotime as the explanatory variable to identify proteins differentially altered as a function of the inferred trajectory. Similarly, proteins having different dynamic effects between two clinical groups of samples were determined using covariate-pseudotime interaction linear models. Canberra distance measure applied in the *Complexheatmap* package was used to cluster the differentially altered proteins while ordering the samples based on pseudotime.

2.2.9 Pathway and gene ontology analysis

The Ingenuity Pathway Analysis (IPA) tool (Krämer et al., 2014) was used to perform overrepresentation tests and upstream regulator analysis. The Fisher’s exact overrepresentation z-scores were visualized using custom R scripts. Weighted one-tailed FET implemented in the ViseaGO (Brionne et al., 2019) algorithm was used to characterize
differentially altered proteins based on gene ontology annotations. The threshold of significance was set at \( P \text{-value} < 0.05 \).

### 2.3 RNAseq analysis of plasma EVs from cerebral malaria patients

#### 2.3.1 Samples and design

Archived plasma samples from cerebral malaria cases that had been assessed for retinopathy were used. The sample size (\( n=40 \)) was pragmatically determined based on available samples, clinical data, and resources. Specifically, RNAseq data was generated from plasma samples from 12 retinopathy-positive and 28 retinopathy-negative cerebral malaria. The aim was to have approximately equal samples in the two phenotypes. However, only 12 retinopathy-positive samples were available after assessing the archived samples, resulting in an unbalanced design.

#### 2.3.2 RNA extraction from plasma-derived EVs

The isolation of plasma EVs for RNA isolation was generally similar to that described in section 2.2.2 with slight modifications. Succinctly, plasma was diluted in 13.5 ml of PBS and passed through a 0.22 µm filter. The filtrate was transferred into ultracentrifuge tubes and spun at 150,000 x g for 2 h at 4°C. The supernatant was discarded while the pellet was resuspended in 300 µL of PBS and treated with RNase A at 37 °C to digest non-vesicular RNA. After 15 min, the mixture was transferred to 13.5 ml ultracentrifuge tubes (Beckman). The tubes were filled using PBS and centrifuged at 150,000 x g for 2 h at 4 °C. The final pellets were digested using 250 µl of RNA lysis buffer (Bioline) and kept at -80 °C until when required. RNA was isolated using the Isolate II RNA Min Kit (Bioline) while following the manufacturer’s instructions.

#### 2.3.3 cDNA library preparation from plasma EV-RNA

The dUTP method developed by Chappell and others (Chappell et al., 2020) and fully described earlier in section 2.1.8 was used to prepare the cDNA libraries for sequencing.
CHAPTER 2

Briefly, total EV-RNA is used to generate the first strand. Before second strand synthesis, the samples were cleaned using RNaCleanXP beads to remove traces of dNTPs. During the synthesis of the second strand, dTTP was replaced with dUTP. Double-stranded cDNA was then enzymatically shredded and ligated to NEXTflex adapters. The cDNA was treated with uracil glycosylase, which digests dUTPs to make the libraries strand-specific and amplified in 15 cycles to increase yield. Sequencing was done using the Illumina Hiseq 4000 genome analyser at the Wellcome Sanger Institute.

2.3.4 Differential transcript abundance analysis

RNAseq fastq files were downloaded from the Sanger sequencing pipeline, and transcript read estimates were obtained by aligning the data to the human and P. falciparum 3D7 transcriptomes using Kallisto and tximport (Bray et al., 2016; Soneson et al., 2015). Comparison between retinopathy positive and negative cerebral malaria was performed using a popular differential gene expression analysis algorithm called edgeR (Ritchie et al., 2015; Robinson et al., 2010). In edgeR, the raw read counts were internally normalised using relative log expression (RLE) and the likelihood ratio test was chosen in this case. Adjusted P-values were calculated using the Benjamini - Hochberg statistical procedure. FDR of less than 5 % was set as the cut-off for significance. I used the Human Protein Atlas (HPA) (Pontén et al., 2008) single-cell RNAseq to provide insights into the origin of human transcripts detected as differentially altered between retinopathy-positive and negative cerebral malaria patients. Functional enrichment analysis was done as described in section 2.2.9.

2.3.5 Selection of cerebral malaria signature genes

I used algorithms implemented in the DaMiRseq (Chiesa et al., 2018) package to find the most informative biomarkers of retinopathy-positive cerebral malaria. In brief, I excluded non-informative features by performing the backward variable elimination – partial least
squares (BVE-PLS) method. A pairwise correlation was then performed, and if two genes had a correlation coefficient of more than 0.85, only the one with the least mean absolute correlation was retained. The \textit{RReliefF} feature selection algorithm based on the random forest was used to rank the genes based on their ability to distinguish retinopathy-positive and negative cerebral malaria. Genes with an importance z-score higher than one were chosen. The discriminatory ability of the selected signature was assessed using an MDS plot and ROC curve analysis.

\subsection*{2.3.6 Estimation of cerebral malaria progression pseudotime}

I used deep learning to construct pseudotime trajectories of the EV transcriptome samples. First, two outlier samples were filtered from the dataset. The count data were normalized by gene length and sequencing depth and converted to counts per million (CPM) units. A pseudotime trajectory was inferred from the logCPM data using \textit{PhenoPath} \textit{(Campbell and Yau, 2018; Lee and Choi, 2022)} R package. \textit{PhenoPath} is a tool that employs Bayesian statistics to model latent expression of each gene. \textit{PhenoPath} was chosen because it has been used before to analyse bulk RNAseq \textit{(Campbell and Yau, 2018; Lee and Choi, 2022)}. A non-linear model was fitted to determine transcripts altered as a function of the inferred disease pseudotime. Functional analysis was performed as described in section 2.2.9.
Chapter 3  Extracellular vesicles might be part of a post-transcriptional regulatory mechanism that maintains RNA homeostasis in *Plasmodium falciparum*
CHAPTER 3

3.1 Summary

Extracellular vesicles secreted by the *Plasmodium falciparum* contain RNA (Babatunde et al., 2018a; Sisquella et al., 2017; Ye et al., 2018). However, previous studies did not compare the RNA content of *Pf*EVs to that of the secreting whole parasites. It is also well-established that *Plasmodium falciparum* transcriptomes are rhythmic (Bozdech et al., 2003; Llinás et al., 2006), but whether RNA secretion via *Pf*EVs also adopts a similar stochastic behaviour remains to be studied. A more detailed comparison of the RNA content of *Pf*EVs to that of the parasite could reveal the benefits of RNA secretion to the secreting whole parasite.

In this Chapter, I expanded our knowledge about RNA secretion via *Pf*EVs by examining the secreted and intracellular RNA from tightly synchronized asexual blood-stage parasites. I found that secretion of RNA via *Pf*EVs is conserved among *Plasmodium falciparum* isolates and adopts the well-known tightly rhythmic behaviour of malaria transcriptomes. The data further revealed that the timing of RNA secretion via *Pf*EVs is shifted compared to the secreting whole parasite. I also demonstrated that unusable RNA in asexual parasites such as that encoded by mosquito and liver stages is preferentially secreted via extracellular vesicles. These findings suggest that *Pf*EVs are a housekeeping cellular mechanism that maintains RNA homeostasis in *Plasmodium falciparum*.

3.2 Introduction

A highly periodic transcriptional cascade drives the intraerythrocytic developmental (IDC) stages of *Plasmodium falciparum*, with most genes peaking only when their transcribed products are required (Bozdech et al., 2003; Llinás et al., 2006). This high rate of transcriptional switching generates much molecular waste that the parasite must eliminate. In addition, the vast majority of the more than 5700 *Plasmodium falciparum* genes are transcribed by the asexual blood stage parasites, including pseudogenes and genes whose
products are only required by other parasite stages such as sporozoites, gametocytes, ookinetes and oocysts (Bozdech et al., 2003; Llinás et al., 2006). However, the fate of the RNA transcribed from genes not essential at all for the growth of asexual blood-stage parasites is not known. Therefore, I hypothesized that *Plasmodium falciparum* possesses a waste management system that excretes unused or unusable molecular waste such as RNA.

This proposition is supported by recent studies reporting that *Plasmodium falciparum* secretes EVs packaged with biological cargos such as proteins, lipids, and genetic materials (Babatunde et al., 2018a; Mantel et al., 2016; Mantel et al., 2013; Regev-Rudzki et al., 2013; Sisquella et al., 2017; Ye et al., 2018). In other systems, EVs maintain cell homeostasis by containing and excreting unwanted or toxic materials from cells (Takahashi et al., 2017). Previous studies did not capture the role of *Plasmodium falciparum*-derived EVs in parasite waste management because they did not compare the RNA content of *Pf*EVs to that of the secreting whole parasites (Babatunde et al., 2018a; Sisquella et al., 2017; Ye et al., 2018). Previous studies on RNA secretion via *Pf*EVs did not investigate the potential rhythmicity of RNA secretion because they either focused on a single timepoint within the IDC (Babatunde et al., 2018a; Sisquella et al., 2017) or analysed RNA extracted from mixed parasite cultures (Ye et al., 2018). Therefore, it is reasonable to say that the complete transcriptome of *Pf*EVs across the IDC has not been appropriately dissected. To fill these gaps, I compared the RNA content of *Pf*EVs obtained from *in vitro* cultures of asexual blood parasites to that of the secreting whole parasites (WP).
CHAPTER 3

3.3 Results

3.3.1 Generation of PfEVs and whole parasite RNAseq data

To generate genome-wide transcriptomes of PfEVs and WP, I sequenced samples collected from tightly synchronized in vitro cultures of six asexual blood isolates (named KE01, sKE01, KE02, KE04, KE06 and Dd2) with diverse origins. A detailed overview of these Plasmodium falciparum isolates is available in the methods section 2.1.1. I adopted a large-scale method using 300 ml per isolate to obtain adequate amounts of PfEV-RNA. Four-hourly 300 ml culture conditioned medium (CCM) and whole parasite pellet were sampled beginning 2-4 hours post-invasion while at the same time monitoring culture synchronization by Giemsa staining. Only asexual blood forms of the Plasmodium falciparum were observed as the parasites progressed through the cycle. Next, I isolated PfEVs from the CCM using an established protocol in our lab that involves nanofiltration, concentration of the total 150 ml to a manageable 10 ml, and ultracentrifugation.

Sectioned transmission electron microscopy (TEM) revealed that the pellet obtained from this process contained vesicular particles whose size (100 nm) matched that described for small EVs (Théry et al., 2018) (Fig. 3.1A). Consistent with established knowledge (Jenjaroenpun et al., 2013; Valadi et al., 2007), the RNA isolated from extracellular vesicles lacked the ribosomal peaks in WP RNA (Fig. 3.1B). RNAseq data were generated from the PfEVs and WP, and gene-level abundances were estimated using Kallisto (Bray et al., 2016). Visual inspection of the normalized datasets using multidimensional scaling revealed that both PfEVs and WP transcriptomes were strikingly cyclic, with samples of the same time points clustering independently of isolate (Fig. 3.1C-D).
Fig 3.1 RNAseq data generation and processing

A Sectioned transmission electron microscopy showing that the *Plasmodium falciparum* extracellular vesicles. B Bioanalyser traces showing that PfEVs lack the ribosomal RNA peaks found in whole parasites. C and D Multidimensional scaling plots of both whole parasite (WP) and *Plasmodium falciparum* extracellular vesicle (PfEVs).
3.3.2 Secretion of RNA via PfEVs is conserved among parasite isolates

To interrogate the degree of transcriptional conservation between the six isolates, I used Pearson Correlation Coefficients (PCC) to compare the five Kenyan isolates to Dd2 (KE01 vs Dd2, sKE01 vs Dd2, KE02 vs Dd2, KE04 vs Dd2, KE06 vs Dd2). Consistent with what is known about *Plasmodium falciparum* transcriptomes (Bozdech et al., 2003; Llinás et al., 2006), the whole parasites showed high conservation between isolates (Fig. 3.2A, top panel) with an overlap of 3970 genes (75%) exhibiting PCC greater than 0, and 4427 genes (85%) showing PCCs greater than 0 when considering any four of the five correlation pairs (Fig. 3.2B).

Conversely, I observed that PfEVs transcriptomes were also transcriptionally conserved (Fig. 3.2A, bottom panel) with an overlap of 2407 (44%) positively correlated genes between the five comparisons and 3702 genes (71%) when considering overlaps between any four of them (Fig. 3.2C). When I correlated the average RNA abundance in whole parasites to that of the corresponding PfEVs, I observed high concordance (rho>0.72, p-value<0.0001) between the two (Fig. 3.2D). These observations suggested that the mechanism of loading transcripts into PfEVs was broadly conserved across isolates and could also be under temporal control.
Fig 3.2 RNA secretion via *PfEVs* is conserved between isolates

**A** Histograms showing Pearson correlation coefficients (PCCs) between KE01, sKE01, KE02, KE04 and KE06 using Dd2 as the reference, WP=whole parasites, *PfEVs*=*Plasmodium falciparum*-derived extracellular vesicles. 
**B** and **C** Venn diagram analysis; **B** whole parasites and **C** extracellular vesicles gene profiles with PCC scores > 0 in the five comparisons of Kenyan isolates with Dd2. 
**D** The RNA content of *PfEVs* and WP is positively correlated.
3.3.3 RNA secretion via PfEVs occurs in a periodic manner

I wondered whether the secretion of individual *Plasmodium falciparum* transcripts via PfEVs adopted a stochastic pattern like that described previously in WP (Bozdech et al., 2003; Llinás et al., 2006). To this end, I modelled the transcriptional profiles as sinusoidal curves to define rhythmic parameters of RNA abundance in PfEVs and within the secreting WP. The parameters included the phase (timing of expression), mesor (rhythm-adjusted mean) and amplitude (distance between mesor and phase). Separate phaseograms of WP and PfEVs transcriptomes were created by ordering the transcripts using the estimated phase (Fig. 3.3A-B). The RNA abundance in WP was highly periodic, with most expressed transcripts having a single peak and a single trough (Fig. 3.3A), as witnessed by others (Bozdech et al., 2003; Llinás et al., 2006). Secretion of RNA via PfEVs also showed striking periodic patterns, with 4585 to 4868 genes exhibiting significant rhythmicity (p-value < 0.05) among the six isolates (Fig. 3.3B).

**Fig 3.3** Secretion of RNA via PfEVs occurs in a periodic manner

Phaseograms of blood-stage transcriptomes of A whole parasites and B extracellular vesicles obtained from the six *Plasmodium falciparum* isolates (KE01, sKE01, KE02, KE04, KE06 and Dd2). The phaseograms were generated from the mean centered Fourier transformed logTPM and the genes are ordered along the y-axis based on their phase. The numbers on the left of each phaseogram represent the total number of significantly rhythmic genes at p-value < 0.05.
3.3.4 Transcriptional phase-shift in the secretion of RNA via PfEVs

Next, I used the profiles of transcripts that were significantly rhythmic in both PfEVs and WP (p-value < 0.05) to create combined phaseograms using the same gene ordering for the two sets of data in each of the two compartments (Fig. 3.4A). This analysis showed that the phases of RNA abundance in PfEVs were shifted compared to those of WP (Fig. 3.4A), with

**Fig 3.4 Secretion of RNA via PfEVs phase-shifted relative to the WP expression**

A Phaseograms showing that the peak of WP-RNA expression for the vast majority *P. falciparum* genes corresponds to the trough of PfEVs-RNA secretion. Only genes detected as significantly rhythmic in both WP and PfEVs were used to construct the phaseograms and the order of genes in WP and EV samples is the same in each isolate. Mean z-scoring of WP and PfEVs datasets was performed separately. B and C Phase-shift demonstration using two well-known genes; skeleton binding protein 1 (SBP1) and merozoite surface protein 1 (MSP1), B separate plots C combined plots.
CHAPTER 3

the peak of RNA in WP roughly matching the trough of secretion via PfEVs (Fig. 3.4B-C). These findings signify that RNA is maximally secreted via PfEVs when not required by the asexual blood parasites.

3.3.5 Unusable RNA is enriched in PfEVs compared to the whole parasite

To test whether specific subsets of genes were preferentially enriched in PfEVs, I calculated the differences in rhythmic means (mesor) between PfEVs and WP while using the WP as the reference. The genes were then grouped into six clusters (denoted c1 to c6) based on the delta mesor and consistency of enrichment across the six isolates (Fig. 3.5A). Genes in clusters c1 (n=1072) and c2 (n=623) were consistently enriched in PfEVs across all isolates while genes in clusters c5 (n=771) and c6 (n=1625) were preferentially excluded. Clusters c3 and c4 represented a relatively smaller number of genes (623 and 544, respectively) whose delta change mesor between PfEVs and parasites was inconsistent among isolates (Fig. 3.5A).

Cluster c1 and c2 genes belonged to a wide range of biological functions primarily active during the mosquito stages of the parasite, including heme (Ke et al., 2014) (e.g. PPO, UROD and ALAS) and fatty acids (Vaughan et al., 2009) (e.g. KASIII) synthesis, the citric acid cycle (van Dooren et al., 2006) (e.g. MDH and SCS-beta), formation of gametocytes (Filarsky et al., 2018; Kafsack et al., 2014) (e.g. GDV1, AP2-G, PUF2), the crystalloid organelle in ookinete and young oocyst (CCp1-4), sporozoite host cell traversal and invasion (e.g. CTRP, CSP and CRMP1-4) (Fig. 3.5A-B). In contrast, clusters c5 and c6 included genes belonging to well-characterized functions active at the IDC stages of the parasites, including invasion of red blood cells (Weiss et al., 2015) (e.g. RH5, MSP1, EBA175, RON5), the digestive food vacuole (Lamarque et al., 2008) (e.g. HRP2, HRP3, CRT, PMI-III), glycolysis (Oyelade et al., 2016) (e.g. ENO, GAPDH, HK), and protein processing and
export to the red blood cell membrane (Mundwiler-Pachlatko and Beck, 2013) (e.g. KAHRP, MAHRP1, SBP1, PTP1) (Fig. 3.5A-B).

The above observations signified that unusable RNA not required at all during the IDC was preferentially eliminated from the parasite via PfEVs. To validate this observation, I grouped 4279 P. falciparum genes as either expressed during IDC or non-IDC stages by analysing single-cell RNAseq dataset generated by the Malaria Cell Atlas studies (Howick et al., 2019; Real et al., 2021; Reid et al., 2018). First, I observed that clusters c1 and c2 had more genes belonging to non-IDC stages than IDC, while clusters c5 and c6 consisted mainly of IDC-stage specific genes (Fig. 3.5C). Second, the correlation between the log2FC (non-IDC vs IDC, Malaria Cell Atlas RNA-seq) and the median delta mesor (PfEVs vs WP) revealed that the relative gene expression in non-IDC stages was positively correlated (R = 0.44, P-value < 0.001) to the relative enrichment in PfEVs (Fig. 3.5D).

In addition, out of the 81 P. falciparum pseudogenes (e.g. EBA165, RH3 and ACS1b) analysed, 52 belonged to clusters c1 and c2 while only 11 belonged to c5 and c6 (Fig. 3.5E). These findings suggest that the secretion of RNA via PfEVs is essential to maintain cellular homeostasis by excreting unwanted RNA from the parasite cells.
Fig 3.5 Unusable RNA is preferentially secreted via PfEVs

A Heatmap of delta mesor obtained by comparing the RNA profiles of PfEVs to those of whole parasites. B Example abundance profiles of CSP (sporozoite marker) and MAHRP (IDC stage marker). C Bar plots showing the number of parasite stage specific genes enriched in cluster c1 to c6. D The relative enrichment of RNA in PfEVs (x-axis) positively correlates with the relative expression in non-IDC stages (y-axis). E Majority of pseudogenes are highly enriched in PfEVs (41 belong to cluster c1) which is in line with the argument that unusable RNA is preferentially secreted via PfEVs.
3.4 Discussion

Here, I suggest for the first time a homeostatic role of PfEVs during the intraerythrocytic cycle of *Plasmodium falciparum* by comparing the RNA content of PfEVs obtained from six biological replicates to that of the secreting WP (intracellular RNA). The data in this Chapter provides a novel model in which PfEVs are part of a control mechanism that selectively excretes obsolete or unusable RNA from the parasite. Excretion of dispensable cargo was the first function of extracellular vesicles described in 1983, where two research teams separately observed that reticulocytes excreted transferrin receptors via extracellular vesicles as part of the maturation process to form erythrocytes (Harding et al., 1983; Pan and Johnstone, 1983).

Consistent with the transcriptional behaviours of malaria parasites (Bozdech et al., 2003; Hoo et al., 2016; Llinás et al., 2006; Rijo-Ferreira et al., 2020; Smith et al., 2020), my findings reveal that RNA from the vast majority *Plasmodium falciparum* genes is periodically secreted via PfEVs during the asexual blood stage of the parasites. The secretion patterns are also correlated among isolates with different backgrounds indicating that the process of RNA excretion via PfEVs is a transcriptionally conserved mechanism essential for *Plasmodium falciparum* to survive. Interestingly, PfEVs transcriptomes are phase-shifted relative to the secreting whole parasites indicating that the “just in time gene expression” pattern in the whole parasites described first by DeRisi lab (Bozdech et al., 2003; Llinás et al., 2006) is followed by “a just in time RNA secretion” to the extracellular space. This hysteretic relationship between gene expression and excretion of RNA via PfEVs could be crucial in preventing the build-up of untranslated RNA within the cytoplasm of the highly transforming malaria parasite.

It is established that *Plasmodium falciparum* transcribes all its 5700 plus genes (Bozdech et al., 2003) during the asexual stages. Still, little is known about how the parasite eliminates
uneval RNA encoded by mosquito stage-specific genes and pseudogenes. In this Chapter, I showed that *Pf*EVs contained relatively higher amounts of unusable RNA expressed by mosquito stage-specific genes and pseudogenes compared to the secreting whole parasites. We suggest that *Pf*EVs secretion preserves cellular parasite homeostasis by blocking the accumulation of unwanted RNA required only in mosquito and liver parasite stages, thus preventing the aberrant production of inoperative proteins. These findings are consistent with other biological scenarios that have implicated extracellular vesicles as a convenient way of post-transcriptional silencing (Desdín-Micó and Mittelbrunn, 2017). For instance, cancer cells have been shown to secrete the tumour-suppressive let-7 miRNAs into their extracellular environment to preserve their oncogenic properties (Ohshima et al., 2010).

In summary, the data fits nicely with a paradigm in which secretion of *Pf*EVs maintains parasite cellular homeostasis by excreting unused or unusable molecules from the parasite if their physiological need has been met or if they are not physiologically required at all during the IDC. Perturbation of this molecular waste management process might be detrimental to *P. falciparum*, and therefore our findings unfold new prospects to target the parasite.

### 3.5 Conclusion

In this Chapter, I compared the RNA content of *Pf*EVs to those of the secreting WP. I found that the secretion of RNA via *Pf*EVs is conserved and periodic, like the intracellular RNA within the parasite. I further showed that the timing of RNA secretion via *Pf*EVs is shifted compared to that of expression in the WP. Finally, the data revealed that unusable RNAs encoded by physiologically unwanted genes, such as those specific to mosquito and liver stages, were preferentially secreted via *Pf*EVs. I conclude that *Pf*EVs are part of a post-transcriptional regulatory mechanism that maintains RNA homeostasis in *P. falciparum*. 
Chapter 4 Integrated proteomic analysis of circulating extracellular vesicles identifies pathological processes associated with severe malaria complications.
CHAPTER 4

4.1 Summary

Severe malaria in children presents as overlapping clinical syndromes, which include cerebral malaria (CM), respiratory distress (RD), and severe malaria anaemia (SMA) (English et al., 2004; Marsh et al., 1995). The pathophysiological processes underlying each of these syndromes are poorly understood. Circulating extracellular vesicles (EVs) contain proteins and RNAs derived from the secreting cells. Therefore, their contents can be used to understand pathological processes occurring in inaccessible organs affected by severe malaria, such as the brain, heart and liver (Raposo and Stoorvogel, 2013).

The work described in this Chapter aimed at identifying human proteins in circulating EVs and using them to infer organism-wide pathological processes caused by \textit{P. falciparum} infection. To achieve this, I compared the host proteins in EVs from severe and non-severe malaria patients (SM vs NSM). Standard biological processes dysregulated in severe malaria include acute phase response, coagulation, lipid and glucose utilization pathways, and complement cascade activation. Next, I applied manifold learning to the EV proteomic datasets to obtain a latent variable representing disease state for each sample (also called pseudotime or infection clock). Most samples with late-stage pseudotime signatures were obtained from SM patients, and disease pseudotime was positively correlated with parasite burden (\textit{P}fHRP2) and negatively correlated to haemoglobin (HB). Unfavourable biological processes such as acute inflammation, coagulation, late complement activation (downstream of C5), increased glucose uptake and cell death pathways increased with disease progression. In contrast, favourable processes such as early complement activation (upstream of C5), TORC1 signalling, and cell production (differentiation and proliferation) pathways decreased over disease pseudotime.

In a parallel approach, I compared the proteins of EVs from pure malaria syndromes (SMA, RD and CM) to NSM to identify dysregulated pathways between the three
syndromes. SMA was associated with an increment in proteins involved in inflammatory response (e.g. CRP and NLRC4) and a decrease in proteins involved in TGFB1 signalling (e.g. TGFB1) and erythrocyte development (e.g. SPTB and HBB). Inflammatory markers were also increased in CM but to a lesser extent compared to SMA. RD patients did not exhibit an increment of inflammatory markers, but proteins involved in glycogen metabolism (e.g. IRS1 and IRS2) were significantly increased in this group of patients. These findings suggest that: 1) SMA is an erythropoietic problem caused by excessive inflammation and deficiency of TGFB1; 2) RD is not an inflammatory problem but is instead a metabolic one; 3) the processes that distinguish CM from the other malaria syndromes could not be conclusively defined using the current dataset.

Finally, I compared the proteomes of patients who succumbed to malaria to those who did not and identified 32 death markers. The most predictive marker was the STYXL2 protein (Relative Risk Ratio > 2.5; Hazard Ratio > 3; P < 0.01), which is highly expressed in heart cardiomyocytes (Pontén et al., 2008). This observation signals that heart complications could be the primary cause of malaria deaths. Furthermore, proteins involved in acute phase response, such as CRP, FTL, S100A8, and LBP, were positively associated (P < 0.01) with malaria-related deaths. These results suggest that EVs have considerable potential as sources of biomarkers that explain the pathophysiological processes associated with severe malaria.
4.2 Introduction

Severe cases of childhood *falciparum* malaria manifest in different clinical phenotypes (English et al., 2004; Marsh et al., 1995; Phillips et al., 2017). CM, SMA and RD are the most frequently encountered severe malaria syndromes in African children under five (Burté et al., 2012). The current criteria for diagnosis of severe malaria syndromes include a combination of peripheral parasitaemia with coma, respiratory distress, severe anaemia, metabolic acidosis and hypoglycaemia (Burté et al., 2012; English et al., 2004; English et al., 1997; Marsh et al., 1995; Phillips et al., 2017). However, although these case definitions are helpful for clinical recognition, they provide limited mechanistic information on the pathophysiological processes underlying the clinical syndromes.

Whole blood transcriptomic studies have attempted to explain the pathogenesis of severe malaria syndromes (Boldt et al., 2019; Cabantous et al., 2020; Lee et al., 2018; Nallandhighal et al., 2019). One disadvantage of whole blood expression studies is that they provide only information on the blood tissue and, therefore, cannot accurately infer the detrimental effects of severe malaria in remote tissues that cannot be accessed noninvasively, such as those of the brain, bone marrow, heart and liver. Plasma or serum can be used to capture events in remote organs, but the high abundance of resident plasma proteins often masks the detection of low-abundance biomarkers originating from the affected remote tissues (Burté et al., 2012; Kumar et al., 2020; Ray et al., 2015). While depletion of the most abundant proteins is used to increase the analytical depth of low-abundance proteins, this strategy is often inefficient and only leads to a slight increase (~25%) in the number of proteins identified (Cao et al., 2021; Millioni et al., 2011).

Extracellular vesicles (EVs) have massive potential as biomarkers because they are secreted by all cells into easily accessible biological fluids such as plasma, urine, and saliva (Raposo and Stoorvogel, 2013). The biological content of EVs, which include DNA, RNA,
metabolites, lipids, and proteins, primarily reflects the physiological state of the cell or tissue of origin (Raposo and Stoorvogel, 2013). Small EV membranes' lipid composition is also different from other membranes, with the former having a higher concentration of ceramide and sphingomyelin (Dinkins et al., 2017; Van Meer et al., 2008). This unique membrane-lipid composition makes small EVs less vulnerable to degradation and prolongs their half-life in biological fluids and during storage. Thus, the cell or tissue specificity coupled with prolonged half-life and easy accessibility from biological fluids such as plasma renders small EVs very attractive tools to use to understand the pathophysiology of severe malaria.

Host-derived EVs concentrations are augmented in severe *falciparum* malaria (Combes et al., 2005; Pankoui Mfonkeu et al., 2010; Sahu et al., 2013), and at least one previous study has characterized the proteomic content of medium-sized EVs (which they called microparticles) (Antwi-Baffour et al., 2017). However, to my knowledge, the protein content of the more biologically significant small EVs (often called exosomes) has not been studied in the context of severe malaria. In this Chapter, I hypothesised that small EVs obtained from the plasma of severe malaria cases contain proteins that may provide information from inaccessible tissues. I test whether the proteome profiles of plasma-derived small EVs can discriminate severe *falciparum* childhood malaria from non-severe patients and whether the profile changes with disease progression. I further test whether small EVs from severe malaria syndromes (SMA, RD and CM) have distinct proteome profiles that reflect their different pathogenesis mechanisms.
CHAPTER 4

4.3 Results

4.3.1 Clinical parameters

A total of 216 samples obtained from malaria-infected children were analysed. Of these, 163 were from severe malaria (SM) cases, while only 50 were non-severe malaria (NSM) cases. Of the NSM samples, 34 were from mild (not hospitalized) malaria cases, while 16 were from moderate (admitted) patients. Among the SM samples, 63 were from pure CM, 30 were from pure SMA, 14 were from pure RD, and the rest were diagnosed with more than one syndrome (Fig. 4.1A). These samples were part of a larger cohort collected from children with malaria between 1990 and 2014. This period has been defined into three transmission phases: pre-decline (1990 - 2002), decline (2003 - 2008) and post-decline (2009 - 2014). Detailed information about the entire cohort of patient samples was reported by Mogeni and others (Mogeni et al., 2016).

Children with malaria syndromes were significantly younger (P<0.05) than those with NSM. (Fig. 4.1B). There were no differences in anti-variable surface antigen (anti-VSA) titres between NSM and CM or RD, but the decrease in anti-VSA titres was more prominent (P<0.05) in SMA (Fig. 4.1C). *Pf*HRP2, a measure of parasite burden, was generally higher (P<0.001) in SM syndromes compared to NSM, but the alteration was not different between the syndromes (Fig. 4.1D). There were no differences in sex between the syndromes (Fig 4.1E). *Pf*HRP2 was not associated with transmission phases described by (Mogeni et al., 2016), while age, HB and anti-VSA titres were all positively correlated to the transmission period (Fig. 4.1F-I).
Fig 4.1 Comparison of clinical parameters between severe malaria complications

A Venn diagram showing overlap of samples obtained from patients with different malaria syndromes. NSM=non-severe malaria, SMA=severe malaria anaemia, RD=respiratory distress, CM=cerebral malaria

B-D Boxplots comparing age, anti-VSA titres and PfHRP2 between NSM, SMA, RD and CM.

E Bar plots showing gender distribution in each of the clinical syndromes.

F-I Scatter plots showing association of PfHRP2, age, anti-VSA titres and HB with transmission phases. The red numbering on top each scatterplot indicates the transmission period: i) pre-decline, ii) decline and iii) post-decline, as defined by Mogeni and colleagues (Mogeni et.al; 2016).
4.3.2 Proteomic data generation

To purify small EVs for proteomic profiling, I used a combination of nanofiltration and ultracentrifugation as described in the methods. Flow cytometry of small EVs captured on aldehyde latex beads confirmed that they contain conventional surface markers of small EVs, CD63 and CD9 (Fig. 4.2A). Protein extracts from purified small EVs were processed for mass spectrometry and multiplexed-tandem mass tag (TMT) labelling used to quantify >3400 human proteins at a q-value <0.01, with each protein identified by at least three peptides (Fig. 4.2B).

Only proteins measured in at least 50% of the samples were included for downstream analyses to reduce missingness per protein variable, resulting in a final total of 3194 proteins. Of these, 3071 have been annotated with biological functions, and 2159 (70%) of them overlapped with proteins previously identified in small EVs (previously called exosomes) by other researchers, according to the Vesiclepedia database (Kalra et al., 2012) (Fig. 4.2C). The following biological processes were abundantly represented within the protein dataset; signal transduction (n = 600), cell communication (n = 566), nucleic acid metabolism (n = 562), cell growth and maintenance (n = 305), transport (n=180) energy metabolic pathways (n = 174), protein metabolism (n = 220), immune response (n =74) and apoptosis (n = 25), while 653 proteins have not been assigned to biological process. On cellular component enrichment analysis, the identified proteins reside in the nucleus (n = 1192), plasma membrane (n = 590), mitochondrion (n = 155), exosomes (n = 441), lysosome (n = 284), endoplasmic reticulum (n = 165) and cytoskeleton (n = 108).
A Scatter plots showing that the isolated EVs pellets were enriched for conventional markers of small extracellular vesicles. 

B A histogram showing distribution of proteins q-values. The majority of proteins were quantified at a q-value < 0.01, and this was used as a cut-off for further analysis. 

C Overlap of host EVs proteome with that of Vesiclepedia database. The “exosome” (small EV) option was used to obtain the overlap.

**Fig 4.2 Generated data is enriched in small EV proteins**

A Scatter plots showing that the isolated EVs pellets were enriched for conventional markers of small extracellular vesicles. B A histogram showing distribution of proteins q-values. The majority of proteins were quantified at a q-value < 0.01, and this was used as a cut-off for further analysis. C Overlap of host EVs proteome with that of Vesiclepedia database. The “exosome” (small EV) option was used to obtain the overlap.
4.3.3 Batch adjustment of proteomic data

Although multiplexed TMT quantification has revolutionized the field of proteomics by increasing the sample throughput, this technology frequently creates sample batch effects, which, if not corrected, can confound the results of expression analysis. Unsupervised multidimensional scaling showed that the uncorrected proteomic samples were clustering by batch (Fig. 4.3A). To overcome this, reference (quality) control samples that I had included in each TMT sample pool were used to adjust the batch effects. Unsupervised batch effect adjustment improved the separation of SM and NSM cases (Fig. 4.3B), and the mean correlation of the reference samples increased from 0.56 to 0.97 (Fig. 4.3C); the effect of batch correction on the distribution of two representative markers, NOTCH4 and GAPDH (Fig. 4.3D) illustrates the extent of the impact. Therefore, using my data generation and processing pipeline, I created a dataset rich in small EV-specific proteins and minimal systematic noise.
Fig 4.3 Adjustment of batch effects

**A** PCA plot before batch correction showing that the samples cluster by batch. **B** PCA plots showing that samples cluster by biological condition after batch adjustment. The reference samples used in each pool cluster tightly as expected. **C** Boxplots comparing the correlation coefficients of reference (quality control) samples before and after batch correction. **D** Representative line plots showing NOTCH4 and GAPDH proteins before and after batch correction.
4.3.4 Plasma-derived small EVs proteomes can distinguish SM and NSM

To identify proteins that differ in abundance between EVs from SM and NSM patients, a standard moderated t-test analysis was applied to the batch-adjusted data. This analysis identified 1262 proteins significantly altered between SM and NSM at an FDR threshold of 5%. As shown in the volcano plot and heatmap (Fig. 4.4A, F), 945 proteins were significantly increased, while only 317 proteins were significantly decreased in SM. More than 65% (820) of the differentially abundant proteins remained significant after adjustment for age, sex, anti-VSA titres, PfHRP2 and transmission period, suggesting that alterations in protein abundance were mainly driven by intrinsic host response (Fig. 4.4B).

The protein most significantly increased in EVs from SM patients was RAB41 (Fig. 4.4C), a GTP binding protein highly expressed by cone photoreceptor cells (Pontén et al., 2008), while the most significantly decreased protein was C1GALT2 (Fig. 4.4D), a probable chaperone essential for normal glycosylation of mucin proteins (Ju and Cummings, 2002; Kudo et al., 2002). However, I also identified other tissue-specific proteins enriched in EVs from SM patients, including from the liver (HGF), heart (STYXL2) and brain (NDNF) (Fig. 4.4E). As expected, proteins that increase during infection, such as neutrophil proteins (myeloperoxidase [MPO], defensins [e.g. DEFA1]), C-reactive protein (CRP), lipopolysaccharide-binding protein (LBP), serum amyloid A4 (SAA4) and fibrinogens (FGA and FGB) were significantly elevated in EVs from SM compared to NSM patients. Unsupervised MDS performed on the altered proteins segregated SM from NSM (Fig. 4.4G).
Fig 4.4 Quantitative plasma-derived proteomic analysis of SM vs NSM

A Volcano plot showing differentially abundant proteins (FDR<0.05) in SM relative to NSM. Blue=significantly decreased in SM, red=significantly increased in SM and grey=not significant. B Venn diagram analysis comparing altered proteins obtained in the unadjusted model and one adjusted for PHRP2, anti-VSA titres, Age and Sex. C and D Boxplots showing the top and bottom most differentially altered proteins respectively. E Boxplots key genes expressed in the in the liver (HGF), heart (STYXL2) and brain (NDNF). F Heatmap representing abundance of the differentially altered proteins in SM and NSM at FDR<0.05. G Multidimensional scaling (MDS) plots showing that the altered proteins can segregate NSM from SM cases.
CHAPTER 4

To identify the pathological processes represented by the significantly altered proteins, I used the Ingenuity Pathway Analysis (IPA) tool. This analysis revealed that the significantly increased proteins were strongly associated with the activation of LXR/RXR metabolic pathways, vascular processes (such as coagulation, acute inflammatory response and complement activation), and cardiomyopathy (Fig. 4.5). Conversely, the topmost significantly decreased proteins were linked to cytoskeleton organization, activation of PPARα/RXRα signalling, calcium signalling, PTEN signalling and p53 activation (Fig. 4.5). Prediction of upstream regulators of the differentially altered proteins identified C5, NOTCH1 and LRP1 as activated proteins in SM.

![Ingenuity pathway analysis of significantly altered proteins](image)

**Fig 4.5** Ingenuity pathway analysis of significantly altered proteins

### 4.3.5 Manifold learning identifies malaria disease progression pseudotime

We did not have longitudinal data in this sample set; even if we did, later time points would be influenced by treatment effects. Therefore, the temporal trend in molecular perturbations as the disease progresses from NSM to SM cannot be directly observed. As an alternative, I applied manifold learning (Mukherjee et al., 2020) to the small EV-proteome data to define the molecular disease stage - often called disease progression pseudotime (infection clock) for each sample. Specifically, the proteomic samples were ordered based on their similarity in protein abundance, and this ordering was used to infer the molecular disease stage (Fig. 4.6A). This analysis showed that samples with late-stage disease pseudotimes were enriched in the SM sample set, while there was some mixing of SM and
CHAPTER 4

NSM samples possessing early-stage infection signatures (Fig. 4.6B). When I adopted a statistical approach to compare pseudotime between SM and NSM, I noted that disease progression was significantly ($P<0.0001$) at a more advanced stage in SM samples (Fig. 4.6C).

Severe malaria pseudotime was also positively correlated ($r=0.28$, $p<0.0001$) to plasma PfHRP2, but the association plateaued towards late-disease stages (Fig. 4.6D). When I excluded late-stage pseudotime samples (>0.35), the correlation between plasma PfHRP2 and pseudotime increased to 0.43 ($p<0.0001$) (Fig. 4.6E) suggesting that parasite burden only increased during the initial stages of severe malaria. As expected, the inferred disease trajectory was inversely correlated to HB concentration ($r= -0.19$, $p=0.0055$) and reached a nadir towards late-stage pseudotimes (Fig. 4.6F), consistent with the consumption of HB over time. The strength of the correlation between HB and pseudotime decreased to -0.22 when I excluded late-stage pseudotime samples (Fig. 4.6G). Similarly, transmission intensity described previously by Mogeni and colleagues (Mogeni et al., 2016) and mentioned in section 4.3.1 was negatively correlated to pseudotime (Fig. 4.6H), indicating increased access to care over time. Note that disease progression was not different between the malaria syndromes (Fig. 4.6I-J). These findings suggest that the latent temporal model is a good proxy for the actual progression of NSM to SM.
**Chapter 4**

A Distribution of host small EVs proteome samples by inferred disease stage. B and C Late stage pseudotimes are enriched for SM samples. D and E Inferred disease staging is positively correlated with parasite burden (PfHRP2). F and G Inferred staging is negatively corrected with haemoglobin (HB). In D and F all samples were used in the correlation analysis while in E and G only early pseudotime (<0.35) samples were used. H Correlation of pseudotime to the year of transmission. I and J Dimensional plot and boxplots showing that pseudotime does not differ between the malaria syndromes.

**Fig 4.6 Molecular estimation of malaria pseudotime using manifold learning.**

- **A** Distribution of host small EVs proteome samples by inferred disease stage.
- **B** and **C** Late stage pseudotimes are enriched for SM samples.
- **D** and **E** Inferred disease staging is positively correlated with parasite burden (PfHRP2).
- **F** and **G** Inferred staging is negatively corrected with haemoglobin (HB).
- In **D** and **F** all samples were used in the correlation analysis while in **E** and **G** only early pseudotime (<0.35) samples were used.
- **H** Correlation of pseudotime to the year of transmission.
- **I** and **J** Dimensional plot and boxplots showing that pseudotime does not differ between the malaria syndromes.
CHAPTER 4

4.3.6 The inferred trajectory is associated with the known biology of SM

Next, I wanted to find which biological pathways were differentially altered as the disease progressed. First, I applied a linear model to each protein in my dataset while using the inferred disease stage as the explanatory variable. I identified 1165 proteins whose abundance was strongly altered over pseudotime (FDR<0.05) (Fig. 4.7A), of which more than 50% (613) were not detected using static (SM vs NSM) differential analysis (Fig. 4.7B). 570 of the 1165 dynamic proteins significantly increased over pseudotime, while 595 significantly decreased as severe malaria progressed (Fig. 4.7A).

The proteins altered over pseudotime were strongly linked to the same Ingenuity pathways (Fig. 4.7C) identified using static (SM vs NSM) analysis (Fig. 4.5). However, I observed that the complement pathway, which I identified as upregulated in the static model (Fig. 4.5) was downregulated in the dynamic model (Fig. 4.7C). Examination of this discrepancy revealed that late complement proteins (e.g. C5, C6, C7, C8B) were positively associated with both SM and pseudotime, but the pseudotime model revealed that early-stage complement proteins (e.g. C2, C3) were negatively associated with disease progression (Fig. 4.7D). The higher number of early complement proteins (upstream of C5) compared to late complement proteins explains the reversed “complement system” z-score obtained using the dynamic model.
**Fig 4.7 Inferred staging reflects known malaria biology.**

A Heatmap showing alteration of plasma-derived host EV proteins over pseudotime. B Overlap between proteins identified as significantly altered in the covariate and psedutoime models. C Top 10 Ingenuity Pathways that decrease (left) and increase (right) over inferred pseudotime. D Comparison of complement C3 and C5 over pseudotime.
4.3.7 Secretion of neural and glial proteins via small EVs over pseudotime

Gene ontology analysis further revealed that the proteins that increased over pseudotime were strongly linked to programmed cell death, glucose homeostasis and neurogenesis. In contrast, those that decreased over pseudotime belonged to a diverse range of terms, including cell production processes, negative regulation of cytokine production and gliogenesis (Fig. 4.8A). The presence of nervous system-related biological processes suggested that manifold learning could be ordering the samples based on increasing neuronal protein abundance and decreasing glial protein abundance. To corroborate this hypothesis, I used the BRETIGEA tool (McKenzie et al., 2018) to estimate the relative brain cell protein content in each plasma-derived small EV sample. Next, I correlated the estimated brain cell proportion to pseudotime. I found that the secreted protein content of neurons in my dataset

![Gene ontology analysis of proteins altered over pseudotime.](image)

![Correlation of brain cell proportion to pseudotime.](image)

**Fig 4.8 Secretion of brain proteins over pseudotime**

A Gene ontology analysis of proteins altered over pseudotime.

B Correlation of brain cell proportion to pseudotime.
was positively correlated to disease pseudotime. In contrast, that of glial cells (astrocytes and microglia) were negatively correlated to pseudotime (Fig. 4.8B). This observation implies that by applying unsupervised manifold learning to the small EVs-proteomic profiles, I estimated the molecular changes occurring in the brain cells as the disease progresses.

4.3.8 Late pseudotime could be the disease recovery stage

I used a covariate-pseudotime interaction model to examine proteins whose dynamic effects differed between NSM and SM. There were 501 proteins whose pseudo-temporal behaviours differed between SM and NSM (Fig. 4.9A). These proteins were categorised into two groups; group 1 comprised 286 proteins that increased over pseudotime in both NSM and SM but decreased in SM towards very late-stage pseudotime while group 2 (n=215) proteins significantly decreased over pseudotime in both NSM and SM but increased only in SM as the disease progressed to extremely late-stages. Thirty-seven percent (n=192) of these proteins were also detected in both the static and pseudotime model, and only 122 (24%) were not significantly associated with pseudotime alone (Fig. 4.9B). This dynamic effect makes late-stage disease samples have some protein profile similarities with early-stage disease samples, signifying that late-stage pseudotime may represent the disease recovery period.
4.3.9 Host proteins in plasma EVs can distinguish malaria syndromes

An ANOVA-like differential protein abundance analysis was used to compare NSM to the pure SMA (n = 30), RD (n = 14) and CM (n = 63) samples. A total of 1914 proteins were significantly altered (FDR < 0.05) between the four groups of patients (Fig. 4.10A). MDS

Fig 4.9 Early and very late pseudotime samples have similar protein profiles

A Heatmap showing proteins whose dynamic effects are different between NSM and SM. B line plots showing the trends over pseudotime of two representative proteins. C Venn diagram analysis showing the overlap between proteins identified in SM vs NSM, pseudotime, and the SM vs NSM-pseudotime models.
plot confirmed that the significantly altered proteins could discriminate NSM, SMA and CM (Fig. 4.10B). The significantly altered proteins clustered into six groups based on their abundance and were named g1, g2, g3, g4, g5 and g6 (Fig. 4.10A). Representative proteins for clusters that define the syndromes have been shown in (Fig. 4.10C-F).

Cluster g1 consisted of significantly elevated proteins in all three syndromes compared to NSM (Fig. 4.10A). This cluster featured 7 proteins involved in lipid synthesis (APOC2, APOC3 and PTGS2), 16 proteins involved in cell-matrix adhesion (ITGA7, ITGB7, MUC4, CDK5 and ITGBL1) and 49 proteins involved in neuron development (e.g. CD38, SEMA5C, EPHB4, PLXNC1, PLXNB2 and UNC5C).

Cluster g2 consisted of proteins that were generally increased in both CM and SMA but were increased to a greater extent in SMA (Fig. 10A). These included proteins involved in inflammatory response (CRP, MBL2, APCS, IL6ST, NLRC4, EDNRB) and coagulation (F7, FGA, FGB, FGG, SERPINA1, SERPINF2) suggesting that these two vascular processes play a more prominent role in SMA than CM, and most likely no part at all in RD. This means pathological processes other than inflammation and augmented coagulation could drive hyperventilation in RD patients. Twelve proteins involved in heart contraction (e.g. CACNA1G, TRPC1, SCN5A, AKAP9 and GRK2) were also among cluster g2. However, cluster g2 proteins were not altered in RD (Fig. 4.10A), indicating that inflammatory response and heart disease were not pathological mechanisms in this group of patients.

The third cluster, g3, was specific to RD and featured proteins involved in energy reserve metabolic process (e.g. glycogen metabolism) such as IRS1, IRS2, PPP1R3A, PHKG2, PER2 and AGL) (Fig. 4.10A). This finding is consistent with the current definition of paediatric malarial RD as a metabolic disease rather than an inflammatory one. Twenty-two proteins in cluster g3 have a role in hemopoiesis, and 11 are involved in T cell differentiation.
CHAPTER 4

(e.g. CHD7, LIG4, VAV1, ZC3H8, and RASGRP1). However, these are not classic T cell
differentiation markers; hence, the direction in which this pathway is altered is inconclusive.

Cluster g4 consisted of significantly decreased proteins in all syndromes compared to
NSM (Fig. 4.10A). These featured 11 proteins involved in response to tumour necrosis factor
(e.g. STAT1, MAPK1, ADAM10, ADAMTS13, ADAMTS7, PRKN and AKAP12), while
33 proteins have roles in ubiquitin-dependent proteolysis (e.g. USP20, USP31, USP4, USP1,
USP7, USP32, CHMP6, UBR2, UBR4, WDR81 and PRKN). The NLRC3 inflammasome
(a close cousin to the NLRC4 inflammasome in cluster g2) was also among the proteins in
cluster g4.

Cluster g5 proteins were significantly decreased in SMA but not in the other syndromes.
They featured erythrocyte-specific proteins (SPTB, HBB, HBD, ADD1 and EPB41L2),
early complement factors (C2, C3, C4BPA), and nine members of the transforming growth
factor beta signalling pathway (e.g. TGFB1, CREBBP, EP300, FBN2, PTPRK). CD14 was
assigned to cluster g5, suggesting that the macrophages (or monocytes) could be
dysregulated in SMA compared to non-anaemic malaria patients (Fig. 4.10A). Deficiency
of TGFB1 and CD14 indicates that SMA could be an erythropoietic problem characterized
by a defect in TGFB1-producing M2-like macrophages, which are crucial elements of the
erthroblastic islands (EBI) as reviewed by (Chasis and Mohandas, 2008; Hom et al., 2015;
Li et al., 2021).

Proteins in the last cluster g6 were elevated most in CM patients but to a lower extent in
the RD group. Notable members of this cluster included brain-derived proteins (e.g. NDNF
and PCDH18) and neutrophil activation (S100A8 and MPO) (Fig. 4.10A), which was
consistent with the pathology of CM reported in the literature (Feintuch et al., 2016).
A Heatmap representation of abundances of differentially altered proteins in NSM, SMA and CM. Different clusters were identified as g1 (upregulated in all syndrome), g2 (overexpressed higher in SMA), g3 (overexpressed in RD), g4 (downregulated in all syndromes), g5 (underexpressed more in SMA) and g6 (overexpressed in CM) based on protein abundance. B Differentially altered proteins can separate NSM from severe malaria complications. C The NLRC4 inflammasome is upregulated most in SMA. D TGFB1 is downregulated in SMA and upregulated in CM but no effect in RD. E The IRS2 is upregulated in RD and downregulated in SMA and CM. F NDNF is elevated most in CM patients.
4.3.10 Identifying markers of malaria-associated death

The last analysis of this Chapter aimed at identifying host EV proteins that can be used to predict the risk of dying from malaria. Using differential abundance analysis, I identified 32 proteins whose relative abundances were significantly different between death cases and survivors (p < 0.05). Twenty-five were positive markers, while only seven were negative markers of malaria-associated deaths (Fig. 4.11 left panel). Furthermore, all the proteins upregulated among the death cases were also associated with malaria mortality (Relative Risk Ratio [RRR] > 1, FDR < 0.05). Similarly, the downregulated ones were all linked to survival (RRR <1, FDR < 0.05) (Fig. 4.11 right panel).

The most predictive marker, STYXL2 (Log2FC > 6, RRR > 2.5), is expressed exclusively by cardiomyocytes (heart cells) according to the Human Protein Atlas Single Cell RNAseq (Pontén et al., 2008), which suggests that malaria-induced heart pathology could be a significant cause of death among malaria patients (Fig. 4.11). However, a limitation to this interpretation is that damaged heart cells have not been shown to release more STYXL2 protein in EVs than intact cells. When I used cox-regression to determine small EV proteins associated with death on any day during admission, STYXL2 emerged second after FCN3 (activates lectin complement pathway) among the list of the most predictive death markers (Fig. 4.12). Other biologically interesting death markers were inflammatory markers (CRP, LBP and S100A8), ferritin light chain (FTL) and RAB41. The last in the list (RAB41) is highly expressed in cone photoreceptor cells, and therefore its upregulation links neuropathology to malaria mortality (Fig. 4.12).
Fig 4.11 Identifying markers of death caused by malaria

Forest plot showing log2FC (left panel) and relative risk ratios (RRR) (right panel) of proteins associated with malaria death. Notable protein are colored in red and include STYXL2 (expressed abundantly by cardiomyocytes), acute phase proteins (CRP, FTL, LBP and S100A8) and RAB41, a brain marker (according to Human Protein Atlas)
Fig 4.12 Cox-regression analysis

Forest plot showing hazard ratios of 109 proteins associated with death. STYXL2 (putative heart biomarker, according to Human Protein Atlas (HPA)) and GDF15 (protects heart muscles from damage) are highlighted in red. A few genes have been highlighted in orange based on expert knowledge. Among those highlighted in orange were FCN3 (complement pathway), PAWR and RAB4 (brain markers), FOXO1 (metabolic transcription factor) and FTL (acute phase protein), which were positively associated with death while those negatively associated with death were RTTN, NAV3 and DGKG (brain markers), CASP6, KYAT3 and NOTCH2 (vascular markers) and PPP1R9A and C1GALT (metabolic markers).
Delineating host responses in severe malaria is challenging due to the inaccessibility of most affected tissues and organs. Analysing the cargo of small EVs is a popular tool used to identify organism-wide responses triggered by an infection or a pathological condition. In this Chapter, I performed proteomic analysis of small EVs obtained from children with malaria complications and identified pathophysiological processes that mirror organism-wide pathophysiological processes of severe malaria.

First, I used standard differential expression analysis and identified proteins that discriminate between NSM and SM children. Some of the abnormalities I identified were consistent with the established pathology of severe malaria and included acute inflammatory response, late complement activation and augmented coagulation (Kumar et al., 2020). In addition, the small EV proteome data corroborated recent whole-blood expression studies (Cabantous et al., 2020; Ray et al., 2016) that observed activation of the LXR/RXR system in both *falciparum* and *vivax* malaria patients. Activating the LXR/RXR pathway could alter gene expression in adipocytes and hepatocytes, which may result in reduced glucose availability (hypoglycaemia) in the blood of SM patients (Laffitte et al., 2003). Specifically, activation of LXR in the liver is associated with the downregulation of gluconeogenic genes and upregulation of glucokinase which results in suppressed hepatic glucose synthesis and augmented hepatic utilization of peripheral glucose (Laffitte et al., 2003). In addition to hypoglycaemia, LXR-mediated suppression of hepatic gluconeogenesis could also lead to a lactic acid build-up in plasma, causing yet another complication of SM called lactic acidosis. Furthermore, studies have not found any association between hypoglycaemia and hyperinsulinemia, as hypoglycaemic patients have sufficiently low concentrations of insulin (Ali et al., 2011; White et al., 1987). White and others also noted high plasma alanine and lactic acid concentrations, which they proposed was due to suppressed hepatic gluconeogenesis (White et al., 1987). However, LXR/RXR activation also induces a
proinflammatory phenotype in monocytes (Sohrabi et al., 2020), and therefore the exact role of this pathway in SM remains to be elucidated. Consistently, the peroxisome proliferator-activated receptor (PPAR) system, which exhibits opposite properties (lipid-lowering and glucose-increasing properties) to those of LXR/RXR was associated with NSM. The protective effects of PPAR agonists against SM are currently being explored (Balachandar and Katyal, 2011). The p53 pathway, which was previously associated with the ability of malaria patients to control fever, inflammation and parasitaemia, was enriched in NSM (Tran et al., 2019).

Manifold learning has been used in other settings to estimate disease progression in cross-sectional studies (Lee and Choi, 2022; Mukherjee et al., 2020). I used this approach to infer disease progression from the cross-sectionally collected malaria samples in this study. While most SM samples were enriched in late-stage signatures, some had NSM-like molecular states, suggesting they could be disease-resistant severe cases. Inferred disease progression was concordant with clinicopathological measures (plasma PfHRP2, HB concentration, age, plasma anti-VSA titres, transmission period) of malaria. Many pathological processes detected using static analysis, such as coagulation, LXR/RXR activation, and PTEN signalling, were also associated with pseudotime. The consistency of temporal protein patterns with those obtained using clinicopathological classifications during admission implies that the inferred quantitative measure of disease represents the pathological molecular changes that differentiate SM from NSM. However, the inferred disease measure was necessary to associate early complement activation with NSM and late complement activation with SM. This indicates that pseudotime had ordered the samples based on their complement cascade status. The inferred disease stage also appears to map the trends in two opposing biological processes accurately. Proteins regulated in response to cell death processes such as DNA damage and apoptosis increase with pseudotime, while those involved in cell production processes such as proliferation, differentiation and actin
cytoskeleton signalling decrease with pseudotime. Ontological analysis revealed that neuronal proteins were positively linked to disease progression, while those involved in the differentiation of glial cells, such as astrocytes and microglia (gliogenesis), decreased over pseudotime. The correlation of deconvoluted brain cell-specific protein levels with pseudotime supported the ontological observations by revealing the following: (i) an increase in the neuronal protein content with pseudotime, and (ii) a decrease in the protein content of astrocytes and microglia as the severe disease progresses. I also observed that extremely early and late pseudotime samples had similar profiles of disease progression biomarkers. This could suggest that late pseudotime samples were obtained from patients who had progressed through the entire disease trajectory and were now in recovery. Overall, incorporating pseudotime into the analysis of cross-sectionally collected malaria samples orders patients based on pathophysiological processes and significantly improves the biological interpretation.

The small EVs proteomic data also revealed non-temporal differences in the pathogenesis of severe malaria syndromes; SMA, RD and CM. To start with, the concentrations of erythrocyte-specific proteins were not only lower in SMA but also under-expressed were proteins involved in erythrocyte cell differentiation and TGF-beta signalling, including TGFB1 (eight-fold decrease in SMA compared to NSM). At high levels, TGFB1 accelerates the differentiation of erythroid progenitors to terminal mature red blood cells while at the same time blocking the proliferative cycle of erythroid precursor cells (Krystal et al., 1994; Kuhikar et al., 2020; Zermati et al., 2000a; Zermati et al., 2000b). In addition, previous studies have reported that TGFB1 is an independent and even more potent inducer of terminal erythroid differentiation than erythropoietin (Zermati et al., 2000b) and therefore, its downregulation could lead to erythroid hyperplasia observed previously among SMA patients (Abdalla et al., 1980; Abdalla, 1990; Perkins et al., 2011). This proposition corroborates previous reports that observed inappropriately low reticulocytosis in SMA
(Chang et al., 2004). I also noted that acute phase proteins such as CRP, LBP, APCS and SERPINA1 were higher in SMA than in CM, confirming that acute inflammation contributes to anaemic stress in SMA. The alternate inflammasome, NLRC4, was upregulated in SMA compared to the other syndromes, which further supports the model that inflammation promotes anaemic stress in SMA. However, no other proteins in the NLRC4 inflammatory pathway were altered in the EV proteome dataset, which makes the role of NLRC4 in SMA inconclusive. Another interesting finding was that RD is associated with an increment in proteins involved in glycogen synthesis, with essentially no alterations in inflammatory markers. Considering that processes that promote glycogen synthesis not only deplete glucose in the blood but also inhibit gluconeogenesis (the primary lactate clearance pathway in humans), these results are consistent with the conceptualization of RD in children as a metabolic syndrome characterized by hypoglycaemia and hyperlactatemia.

Finally, I also identified biomarkers of malaria mortality, with the STYXL2, a protein expressed abundantly by cardiomyocytes (according to the Human Protein Atlas Single Cell RNAseq), being the most predictive marker (Pontén et al., 2008). Heart tissues come second after the brain in the list of favourite sites for the parasite to sequester (sequestration hierarchy is as follows: skin < kidney < liver < lung < heart < brain), and the risk of cardiovascular complications was reported to be significantly higher in severe malaria than healthy children (Akinkunmi et al., 2022; Gupta et al., 2021; Jervis et al., 1972). Of note, acute phase proteins such as FCN3, CRP, LBP and FTL were associated with malaria mortality, suggesting that increased acute inflammation could significantly cause malaria-related deaths.
CHAPTER 4

4.5 Conclusion

In this Chapter, I analysed the protein content of EVs obtained from patients with different clinical phenotypes, including NSM and SM syndromes (SMA, RD, and CM). EV proteins can distinguish NSM from SM patients, and they change with disease progression. Proteins associated with increasing disease severity were linked to pathological processes such as coagulation, acute inflammatory response, late complement activation and increased peripheral glucose uptake pathways. Comparison of the EV proteins between the syndromes revealed that SMA could be an erythropoietic problem characterized by the deficiency of TGFB1; RD is a metabolic problem marked by increased glycogenesis and reduced gluconeogenesis; and CM is a brain-related disease that also features neutrophil activation. Of the identified proteins, 32 could significantly predict death and included cardiomyocyte-specific protein, STYXL2, and inflammatory markers (e.g. CRP, LBP, S100A8).
Chapter 5  Brain-expressed RNA in circulating extracellular vesicles can discriminate retinopathy-positive and negative cerebral malaria.
A fraction of cerebral malaria (CM) patients have a retinal pathology called malaria retinopathy, which is a good predictor of parasite burden, duration of coma and death. Retinopathy-positive CM (CM-R+) patients are conceptualized to be the “true” CM, while those with retinopathy-negative CM (CM-R-) could have milder forms of disease or atypical encephalopathies with incidental parasitaemia (Beare et al., 2004; Beare et al., 2006; Lewallen et al., 1996). Although retinopathy is currently used as a surrogate marker for CM, biological markers that can discriminate CM-R+ from CM-R- are still lacking.

In this Chapter, I compared the host RNA content of plasma EVs isolated from CM-R+ (n = 12) and CM-R- (n = 28) cases. The RNAseq data revealed that some transcripts had significantly altered abundance between CM-R+ and CM-R- patients, and these can be used to segregate patients between these two categories. More than 30% of the significantly altered plasma EV transcripts are highly expressed in glial (astrocyte, microglia, and oligodendrocyte) and neuronal cells. In a downstream analysis, I identified a panel of 23 biomarkers that together could discriminate CM-R+ from CM-R- cases with high accuracy (> 95% sensitivity and > 90% specificity). The most predictive marker was transmembrane protein 1 (TMEM121), which encodes a poorly characterized mitochondrial membrane protein expressed at higher levels in the brain compared to other tissues. Other biologically interesting biomarkers were neurocan (NCAN, a positive marker of brain injury/gliosis) and potassium voltage-gated channel subfamily A member 1 (KCNA1, a negative marker of epilepsy). Application of deep learning to the data revealed that CM-R+ and CM-R- evolve in a single trajectory, with CM-R+ EVs being from late-stage disease patients. Consistent with other neurodegenerative pathologies (Mukherjee et al., 2020), disease progression pseudotime was associated with increased glial gene expression and decreased neural gene expression. In summary, this Chapter's data suggest that circulating EVs from CM patients contain brain-expressed RNA, which has a diagnostic and prognostic significance.
5.2 Introduction

CM is a neurological disease caused by *P. falciparum*. Brain injury in CM results from several pathological processes, including parasite sequestration, infiltration of immune cells and factors into the brain, activation of the endothelial cells and disruption of the blood-brain barrier (BBB) (English et al., 2004; Marsh et al., 1995). Currently, there are two perspectives regarding the cause of coma in CM. One school think that sequestration of infected erythrocytes in the brain is the leading cause of CM and a non-malarial aetiology underlies any coma without sequestered parasites in the brain. The second school believes that inflammation is more critical and malarial aetiology cannot be ruled out in cases where sequestered parasites do not accompany coma. The challenge for the first school has been that the presence of parasites in the brain could only be confirmed on autopsy, necessitating the search for markers that could be used to verify parasite sequestration in the brain in non-fatal cases.

The retina comprises brain-like tissue and shows parasite sequestration and pathology comparable to that in canonical brain tissues as revealed by brain and retinal imaging; it is therefore used as a marker for parasite sequestration in the brain. Patients with retinopathy-positive cerebral malaria (CM-R⁺) have a higher mortality rate than those with retinopathy-negative cerebral malaria (CM-R⁻). An autopsy study reported that malaria retinopathy is 90% sensitive and 85% specific for a cerebral pathology caused by *P. falciparum* (Taylor et al., 2004), indicating that a more accurate measure for detecting “true” CM cases is needed. Biological molecules are more attractive as diagnostic and prognostic markers than surrogate physiological phenotypes such as retinopathy. However, because blood-based biomarkers that can accurately differentiate “true” CM cases from comas of unknown origin are still lacking, malaria retinopathy is still used as a proxy for “true” CM. Additionally, investigating the pathophysiology of CM-R⁺ and CM-R⁻ may clarify whether the two disease phenotypes have different or the same aetiology. EVs have been used as biomarkers for
understanding disease mechanisms where safe and non-invasive testing is otherwise impossible, including glioblastoma and prostate cancer (Lane et al., 2019; Tutrone et al., 2020), but this has never been explored in the context of CM.

In this Chapter, I performed a comparative analysis of host RNA isolated from plasma-derived EVs obtained from CM-R⁺ and CM-R⁻ patients to disentangle the molecular signatures of CM. Furthermore, I demonstrated that CM-R⁺ and CM-R⁻ appear to evolve in a single trajectory, suggestive of a common aetiology with CM-R⁻ samples being from less severe forms of the disease compared to CM-R⁺.

5.3 Results

5.3.1 Patient clinical and laboratory parameters

Forty plasma samples obtained from children who fulfilled the WHO definition of CM at some point during their hospitalization and had retinopathy status determined were used in this study. The subset of children included in this study was a representation of the whole cohort, as the proportion of CM-R⁺ 12/40 (30%) and CM-R⁻ 28/40 (70%) were similar to those documented in the original studies (Abdi et al., 2015; Kariuki et al., 2013).

The clinical and laboratory parameters of the selected 40 children are summarized in Table 5.1. Briefly, RBC count and HB concentrations were significantly lower in CM-R⁺ compared to CM-R⁻ (P-value = 0.007 for both, Table 5.1). Consistent with the first studies to work on these samples (Abdi et al., 2015; Kariuki et al., 2013), plasma Pf/HRP2 levels tended to be higher in CM-R⁺ (P-value = 0.06, Table 5.1). Partial oxygen pressure (pO₂) was marginally higher in CM-R⁻ (mean 5.20, IQR = 4.11, 7.34) compared to CM-R⁺ patients (mean 7.40, IQR = 5.40,11.5) (P-value = 0.08, Table 5.1). Age, weight, gender, temperature, WBC, platelet count, and nutritional status (mid-upper arm circumference [MUAC]) did not significantly differ between CM-R⁺ and CM-R⁻ (Table 5.1).
Continuous normal (log scale) variables were compared using t-test while non-normal ones were compared using Mann-Whitney U test. Dichotomous variables were compared using chi-square test. The values reported for continuous variables are means (or medians for non-normal variables) and interquartile ranges while those for dichotomous variables are numbers of observations.

### Table 5.1 Summary descriptive table of patients by retinopathy status

<table>
<thead>
<tr>
<th></th>
<th>CM-R</th>
<th>CM-R+</th>
<th>P-value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months)</td>
<td>45.0</td>
<td>58.0</td>
<td>0.295</td>
<td>40</td>
</tr>
<tr>
<td>Deep breathing:</td>
<td></td>
<td></td>
<td>0.124</td>
<td>40</td>
</tr>
<tr>
<td>no</td>
<td>18</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>8.05 [6.85;10.0]</td>
<td>6.10 [5.02;7.62]</td>
<td>0.007</td>
<td>40</td>
</tr>
<tr>
<td>RBC</td>
<td>3.60 [2.96;3.90]</td>
<td>2.52 [1.80;3.30]</td>
<td>0.007</td>
<td>39</td>
</tr>
<tr>
<td>sICAM1 (log2)</td>
<td>9.10 [8.75;9.55]</td>
<td>9.13 [8.83;9.29]</td>
<td>0.791</td>
<td>40</td>
</tr>
<tr>
<td>Angiopoietin (log2)</td>
<td>11.9 [11.6;12.5]</td>
<td>12.0 [11.7;12.1]</td>
<td>0.658</td>
<td>40</td>
</tr>
<tr>
<td>Blood glucose</td>
<td>4.25 [0.77;5.20]</td>
<td>4.30 [3.85;5.50]</td>
<td>0.310</td>
<td>37</td>
</tr>
<tr>
<td>WBC</td>
<td>8.00 [8.00;9.60]</td>
<td>8.00 [8.00;8.00]</td>
<td>0.190</td>
<td>36</td>
</tr>
<tr>
<td>Platelet count</td>
<td>111 [48.0;188]</td>
<td>88.0 [77.0;144]</td>
<td>0.543</td>
<td>39</td>
</tr>
<tr>
<td>Blood group:</td>
<td></td>
<td></td>
<td>0.418</td>
<td>40</td>
</tr>
<tr>
<td>A</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>132 [129;138]</td>
<td>138 [136;140]</td>
<td>0.196</td>
<td>25</td>
</tr>
<tr>
<td>Potassium</td>
<td>3.90 [3.68;4.30]</td>
<td>4.50 [3.80;4.70]</td>
<td>0.172</td>
<td>25</td>
</tr>
<tr>
<td>pO2</td>
<td>7.40 [5.40;11.5]</td>
<td>5.20 [4.11;7.34]</td>
<td>0.077</td>
<td>36</td>
</tr>
<tr>
<td>pCO2</td>
<td>3.61 [3.36;4.18]</td>
<td>3.37 [2.66;3.67]</td>
<td>0.186</td>
<td>36</td>
</tr>
<tr>
<td>HCO3</td>
<td>16.2 [12.0;19.1]</td>
<td>13.7 [12.6;15.5]</td>
<td>0.250</td>
<td>36</td>
</tr>
<tr>
<td>Gender:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parasitaemia (log2)</td>
<td>17.4 [14.7;19.0]</td>
<td>16.3 [11.9;17.5]</td>
<td>0.063</td>
<td>40</td>
</tr>
<tr>
<td>PHb (log2)</td>
<td>3.14 [1.75;4.29]</td>
<td>4.96 [3.34;5.78]</td>
<td>0.063</td>
<td>39</td>
</tr>
<tr>
<td>Temperature</td>
<td>35.3 [37.5;38.9]</td>
<td>38.7 [37.6;39.1]</td>
<td>0.738</td>
<td>39</td>
</tr>
<tr>
<td>Weight</td>
<td>11.9 [8.90;15.7]</td>
<td>12.1 [10.6;16.1]</td>
<td>0.503</td>
<td>39</td>
</tr>
<tr>
<td>MUAC</td>
<td>15.0 [13.5;15.8]</td>
<td>14.7 [14.0;15.0]</td>
<td>0.529</td>
<td>35</td>
</tr>
<tr>
<td>Outcome:</td>
<td></td>
<td></td>
<td>0.654</td>
<td>39</td>
</tr>
<tr>
<td>alive</td>
<td>23</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dead</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.2 Host EV-RNAs discriminate CM-R\textsuperscript{+} from CM-R\textsuperscript{-}

To identify host EV-RNA signatures that discriminate CM-R\textsuperscript{+} from CM-R\textsuperscript{-}, I compared the RNA profiles of small EVs purified from 200 µL of acute plasma obtained from CM-R\textsuperscript{+} and CM-R\textsuperscript{-} malaria patients using nanofiltration (220 nm) followed by ultracentrifugation. Differential expression analysis using a standard edgeR pipeline found 1222 significantly altered transcripts (\(|\log_{2}\text{FC}|>1, \text{FDR} < 0.05\)) between CM-R\textsuperscript{+} and CM-R\textsuperscript{-} plasma EVs. (Fig. 5.1A). Of these, 534 genes were enriched in small EVs isolated from CM-R\textsuperscript{+} samples, while 688 genes were enriched in CM-R\textsuperscript{-} samples.

Supervised heatmap clustering of the differentially abundant genes segregated the two groups of cerebral malaria patients (Fig. 5.1B). Unsupervised multidimensional scaling performed on the significantly altered transcripts separated CM-R\textsuperscript{+} from CM-R\textsuperscript{-} patients (Fig. 5.1C). RAN binding protein 2 (RANBP2) was the most enriched transcript in EVs from CM-R\textsuperscript{+} patients, while TATA element modulator factor 1 (TMF1) was the least differentially abundant transcript (Fig. 5.1A). Other prominent transcripts enriched in EVs from CM-R\textsuperscript{+} patients included those encoded by TMEM135, PURA, MDN1 and CACNA1I (Fig. 5.1A), which are all highly expressed in the brain according to the Human Protein Atlas (Pontén et al., 2008) as well as NFKB activation markers (NFKBIA and NKF) and T cell exhaustion markers (TIGIT and CTLA4). The transcripts enriched in EVs from CM-R\textsuperscript{-} patients included those involved in B-cell differentiation (PAX5, THEMIS2), response to oxidative stress (GPX4 and HP), platelet activation and coagulation (F2R, PDGFA) and immune mediators such as calprotectin (S100A8 and S100A9), TLR7, IL17RA, IL12A and IRAK4.
Fig 5.1 Comparison of host EV transcriptomes between CM-R+ and CM-R-

A Volcano plot showing differentially abundant genes (FDR<0.05, |logFC| > 1) in CM-R+ (red) from CM-R (blue). B Heatmap representation of differentially abundant genes. C Unsupervised multidimensional scaling plots showing that the differentially altered genes can distinguish CM-R+ and CM-R patients. D Bar plots showing the top 20 differentially altered Ingenuity pathways. Positive (red) z-scores represent pathways enriched in CM-R+ while negative (blue) represent pathways enriched in CM-R-.
CHAPTER 5

I performed gene set enrichment analysis using the IPA tool to understand further the biological pathways that distinguish CM-R+ from CM-R- patients (Krämer et al., 2014). This analysis showed that transcripts enriched in EVs from CM-R+ patients code for proteins linked to CCR3 chemokine signalling, sumoylation, long-term synaptic potentiation, P13K/AKT signalling and glioma invasive signalling (Fig. 5.1D). Conversely, transcripts enriched in CM-R- mapped to signalling pathways, including thrombin signalling, ephrin B signalling, RHOA signalling, IL-8 chemokine signalling, cardiac hypertrophy, glutamate receptor signalling and neuronal semaphoring signalling (Fig. 5.1D).

5.3.3 Brain-specific transcripts are enriched in circulating EVs

Given that several transcripts differ in abundance between CM-R+ and CM-R- patients, I sought to establish whether a specific subset of the differentially altered transcripts could originate from the brain. To explore this possibility, I obtained the Human Protein Atlas single cell RNAseq (Pontén et al., 2008) and compared the RNA content of brain cells to that of other tissues. Among the differentially enriched transcripts between CM-R+ and CM-R- patients, 392 (32%) are expressed at higher levels (p-value<0.05) in brain cells (glia and neurons) compared to other cell types (Fig. 5.2, Fig. 5.3). Although the EVs were isolated from plasma, only 122 (10%) of the significantly altered transcripts are specific to blood and immune cells, the second largest cell type of origin of differentially altered genes (Fig. 5.2). Ontological analysis linked these brain-specific genes to well-characterized nervous system functions such as neurogenesis, regulation of neurotransmitter release and synapse organization (Fig. 5.4A-B).
Bar plots showing that 392 differentially altered transcripts are expressed at higher levels in the brain cells compared to other cell types.

**Fig 5.2 Most of altered transcripts are specific to the brain**

Bar plots showing that 392 differentially altered transcripts are expressed at higher levels in the brain cells compared to other cell types.
**Fig 5.3** Brain-expressed RNA is enriched in EVs obtained from CM patients.

**A** Heatmap showing that 392 brain-specific genes can distinguish retinopathy positive and negative cerebral malaria. **B** Heatmap confirming that the 392 brain-specific genes are highly expressed by glial and neuronal cells (green colour). The Human Protein Atlas single cell RNA-seq data was used to create this heatmap.
5.3.4 Generation of cerebral malaria gene signature

To generate a signature composed of a small number of genes that stratifies CM-R+ and CM-R− patients, I applied the backward variable elimination-partial least squares (BVE-PLS) procedure to eliminate non-informative genes. In addition, if two transcript profiles had a correlation of greater than 0.85, the one with the largest mean absolute correlation was eliminated. These filters reduced the initial number of 1222 significantly altered genes to 166. Keeping parsimony in mind, I used a multivariate technique called RReliefF to rank the 166 selected genes by their importance, and only those having z-scores of >1 were included in the final signature. The final signature had 23 transcripts, of which 17 were enriched in EVs from CM-R+ patients while 6 were enriched in EVs from CM-R− patients (Fig. 5.5A).
The most informative gene, transmembrane protein 121 (TMEM121) (Fig. 5.5A, B), encodes a protein expressed at higher levels in the brain compared to other tissues according to the human protein Atlas (Pontén et al., 2008). I also found that neurocan (NCAN), which encodes an astrocytic proteoglycan whose expression has previously been linked to a type of brain injury called gliosis (McKeon et al., 1999), was among the selected markers (Fig. 5.5A, C). Versican (VCAN) was another brain-specific proteoglycan (Pontén et al., 2008) that was upregulated in CM-R⁺ patients (Fig. 5.5D) but was below the required level to be included in the list of the most informative genes. The most exciting signature gene whose transcripts were excluded from EVs obtained from CM-R⁺ patients was KCNA1 (Fig. 5.5A, A).

**Fig 5.5 Selection of top transcripts that can discriminate CM-R⁺ and CM-R⁻**

A BVE-PLS followed by *R*Relief feature selection performed on the differentially altered genes. Shown is a bar plot of 23 genes with *R*Relief importance z-score of greater than one. Arrows within the bars indicate the direction of change - red arrows depict enrichment in EVs from CM-R⁺ patients while blue arrows depict enrichment in EVs from CM-R⁻ patients. B-E Boxplots comparing enrichment of four markers in EVs from CM-R⁺ and CM-R⁻.
CHAPTER 5

E), a brain-specific potassium channel whose deficiency is strongly linked to epilepsy (Paulhus et al., 2020; Verdura et al., 2020).

The quantity of NCAN transcripts in EVs was positively correlated to plasma *PfHRP2* (*r*=0.59, *p*-value <0.05) among the CM-*R*+ patients, but no correlation was observed among the CM-*R*− patients (Fig. 5.6A). Furthermore, patients with the highest abundance of NCAN had the least quantity of KCNA1 (Fig. 5.6B), signifying that there could be a connection between post-traumatic gliosis and epilepsy among CM patients. MDS plot showed that the

---

**Fig 5.6 Selected top transcripts that can discriminate CM-*R*+ and CM-*R*−**

**A** Scatterplot of NCAN (marker of brain injury) and plasma *PfHRP2* (marker of parasite burden). **B** Scatterplot of NCAN and 1/KCNA1. The inverse of KCNA1 expression was used because deficiency (rather than overexpression) of KCNA1 is linked to epilepsy. **C** Unsupervised MDS plots using the 23 genes selected by the feature selection pipeline. **D** ROC curve showing that the selected 23 genes can stratify CM-*R*+ and CM-*R*− at over 90% specificity (true positive fraction) and 95% sensitivity (false positive fraction).
CHAPTER 5

23 informative genes could stratify CM-R⁺ and CM-R⁻ malaria patients (Fig 5.6C). To assess the prediction accuracy of the signature, I performed a ROC curve analysis using the first two dimensions of the MDS plots. This analysis showed that the 23 gene signature could discriminate CM-R⁺ and CM-R⁻ patients at over 90% specificity (true-positive fraction) and over 95% sensitivity (false positive fraction) (Fig. 5.6D).

5.3.5 Estimation of neurodegeneration pseudotime in cerebral malaria

To determine the progression of CM-R⁺ and CM-R⁻, I estimated the disease pseudotime using manifold learning. I found that both CM-R⁺ and CM-R⁻ evolve in a single disease trajectory, with CM-R⁺ being from the late stages of the disease (Fig. 5.7A, B). Receiver operating characteristic (ROC) analysis showed that pseudotime could distinguish CM-R⁺ and CM-R⁻ with 100% sensitivity and over 75% specificity (Fig. 5.7C). This could suggest that 25% of the CM-R⁻ samples were from “true” CM, but retinopathy failed to assign them to their correct phenotype, perhaps due to difficulty in fundoscopy assessment. Unexpectedly, the pseudotime estimated from published whole blood microarray data (Feintuch et al., 2016) was not different between CM-R⁺ and CM-R⁻ (Fig. 5.7D), indicating that EV-RNA is a better predictor of disease progression compared to whole blood RNA.

Next, I applied non-linear regression analysis to identify over 4700 transcripts that changed as a function of pseudotime (Fig. 5.8A). I noted that variation of EV-transcripts across pseudotime happens smoothly with no sharp boundaries (Fig. 5.8A). Next, I clustered the transcripts into seven modules named M1 to M7 based on the timing of enrichment in EVs. An eigentranscript (representative transcript for each module) was estimated by calculating the mean of each cluster. Then linear regression analysis was used to compare the eigentranscripts between CM-R⁺ and CM-R⁻, revealing that early pseudotime modules M1 and M2 were non-significantly enriched in CM-R⁻ while late pseudotimes modules M5 (significant) and M6 (non-significant) were increased in CM-R⁺ samples (Fig. 5.8B). Modules M3 and M7 were not different between CM-R⁺ and CM-R⁻ cases.
Fig 5.7 Estimation of disease progression pseudotime.

A Dimensional plot showing evolution of CM-R⁺ and CM-R⁻. B Boxplot showing that disease progression pseudotime estimated from the EV transcriptomes can distinguish CM-R⁺ and CM-R⁻ cases. C ROC curve analysis showing that disease pseudotime can segregate patients CM-R⁺ and CM-R⁻ at 100% sensitivity and over 75% specificity. D Boxplots showing that disease pseudotime estimated from whole blood transcriptomes (GSE72058) cannot distinguish CM-R⁺ from CM-R⁻.
I performed enrichment analysis using Fisher’s exact test to investigate the biological processes that changed as the disease progressed. Early pseudotime modules (M1, M2 and M3) mapped to biological pathways suggestive of normal response to pathogen infection such as macroautophagy (e.g. VPS4B, VPS39, CHMP7 and NEDD4), phagocytosis (TLR2, CD36, TICAM2, CLN3, and LEPR) and acute inflammatory response (TFRC, A2M, A1, Heatmap showing abundance of EV-transcripts as a function of disease progression pseudotime. Heatmap has been split into seven modules named M1 to M7 based on gene coexpression. B Fitted estimates obtained by comparing modules (M1 to M7) over pseudotime between retinopathy positive cerebral malaria (CM-R⁺) and retinopathy positive cerebral malaria (CM-R⁻) cases while using CM-R⁻ as the reference.

**Fig 5.8 Variation of EV transcripts over pseudotime**

A Heatmap showing abundance of EV-transcripts as a function of disease progression pseudotime. Heatmap has been split into seven modules named M1 to M7 based on gene coexpression. B Fitted estimates obtained by comparing modules (M1 to M7) over pseudotime between retinopathy positive cerebral malaria (CM-R⁺) and retinopathy positive cerebral malaria (CM-R⁻) cases while using CM-R⁻ as the reference.
ACVR1, TRPV1 and TREM1 (Fig. 5.9; Fig. 5.10) respectively. Transcripts from the TBX21 gene (which encodes the Th1 transcription factor, T-bet) and its target, including IL21R and IL10RA, were enriched in module M3, and their abundance in EVs assumed an inverted-U shape (Fig. 5.10B, upper panel).

Module M4 featured transcripts encoded by genes involved in the negative regulation of blood coagulation (HRG, NOS3, PDGFRA, PLG and SERPING1), activation of natural killer cells (RASGRP1, PGLYRP2, AXL and GAS6), and cellular response to interleukin-1 (CCL14, ST18, PLCB1, ZNF675 and TOLLIP) (Fig. 5.9; Fig. 5.10).

Module M5 was enriched in transcripts that code proteins involved in neuron apoptosis (FOXO3, ATF4, ITGA1, ITGAM, NUPR1, TFAP2A and CDC42), excitatory postsynaptic potential (GRIN2C, GRIK5, MECP2 and P2RX6) and granulocyte differentiation (CEBPA, FASN, GATA2 and C1QC) (Fig. 5.9; Fig. 5.10).

Module M6 transcripts were from genes involved in neutrophil-mediated immunity, such as AZU1, CTGS, and ELANE, which was consistent with a previous study that also found increased neutrophil activation among CM-R+ malaria patients (Feintuch et al., 2016) (Fig. 5.9; Fig. 5.10). Other biological processes in module M6 included T cell differentiation in the thymus (ITPKB, CARD11, STK11 and ZFP36L2), positive regulation of NF-kappaB transcription factor activity (e.g. NFKB2 and BCL-XL), cellular iron-homeostasis (e.g. PCBP1/2, FTL and FHL) and erythrocyte homeostasis (e.g. HMOX2, KLF2, and HMGB2) (Fig. 5.9; Fig. 5.10). The last module M7 was enriched in transcripts involved in the restoration of homeostases such as neuron differentiation (RHOA, FEZF2, PROX1, TCF12 and KDM4A) as well as energy metabolic pathways such as glycolysis (LDHA, PGK1, TPI1) and oxidative phosphorylation (NDUFA12, NDUFS8, SNCA, ATP5PO) (Fig. 5.9; Fig. 5.10).
Fig 5.9: Functional analysis of transcripts altered over pseudotime

Heatmap showing abundance of EV-transcripts as a function of disease progression pseudotime
Fig 5.10 Representative biological processes altered over pseudotime

**A)** Mean expression profile of selected biological processes in modules M1 to M7

**B)** Fitted estimates of genes involved in Th1 response (top panel) and neutrophil activation (bottom panel). TBX21=T-bet transcription factor, IL10RA=interleukin 10 receptor subunit alpha, IL21R=Interleukin 21 receptor, AZU1=azurocidin 1, CTSG=cathepsin G, ELANE=neutrophil elastase.
CHAPTER 5

The current understanding of neurogenerative diseases such as Alzheimer’s disease implies progressive loss and gain in neuronal and glial cell expression, respectively (De Strooper and Karran, 2016; Mukherjee et al., 2020). To test this, I deconvoluted the brain-derived cell RNA contained in each EV sample using singular variable decomposition (McKenzie et al., 2018). I then fitted a two-order polynomial model to the deconvoluted relative brain cell proportions with disease pseudotime as the dependent variable. The models showed that estimated brain cell proportions of RNA in circulating EVs vary as a function of pseudotime. I found: 1) a marked increase in glial cells (astrocytes, endothelial cells, and microglia) transcriptome, and 2) a general reduction in neuronal cell transcriptome as the disease progressed (Fig. 5.11). These observed expression patterns recapitulate the known neurodegeneration model of Alzheimer’s disease.

![Brain cell RNA abundance in EVs as a function of disease pseudotime](image)

**Fig 5.11 Brain cell RNA abundance in EVs as a function of disease pseudotime**

Cerebral malaria is characterized by increased gliosis and progressive neuronal loss. **Oligodendrocyte PC** = oligodendrocyte precursor cells
5.4 Discussion

In this Chapter, I aimed to identify brain-specific RNAs in circulating small EVs that can be used as diagnostic candidates for CM. To this purpose, I used a transcriptomic approach to compare the RNA content of small EVs from CM-R+ and CM-R- malaria cases. I found that the transcriptional profiles of plasma EVs can capture the molecular alterations occurring in the brain. To my knowledge, I compared EV transcriptomes of CM-R+ and CM-R- cases for the first time; a previous transcriptome study focused on whole blood (Feintuch et al., 2016).

Standard differential expression analysis identified transcriptional differences between EVs from CM-R+ and CM-R- patients. A surprising finding was that genes involved in platelet activation and thrombin signalling, such as VWF, PDGFA and F2R, were downregulated in CM-R+ compared to CM-R-. This finding was unexpected because CM-R- is considered a milder form of cerebral malaria; therefore, the coagulation signalling pathways are expected to be less active in CM-R- compared to CM-R+ (Taylor et al., 2004; White et al., 2009). However, the results could be indicative of either 1) thrombocytopenia in CM-R+ due to low platelet production and exhaustion and 2) the protective effects of platelet activation. The role of platelets in experimental (ECM) and human cerebral malaria has been documented (Feintuch et al., 2016; Kumar et al., 2020; Oakley et al., 2011). Parasite burdens were reported to be negatively correlated with platelet-associated killing, and purified platelets can inhibit parasite growth when added to in vitro cultures (McMorran et al., 2009; Peyron et al., 1989). However, as mentioned in the results section, platelet counts did not vary between the two types of CM, which makes it difficult to reconcile the two findings.

Furthermore, this study also suggested that retinal and intracerebral haemorrhage could cause neural damage in CM-R+ through ferroptosis. Ferroptosis is an iron-mediated cell
death caused by excessive cellular accumulation of iron or failure of the phospholipid peroxide-reducing glutathione system. In cerebral malaria, the brain is susceptible to injury from excess iron generated by RBCs lysis after intracerebral haemorrhage. HB and its oxidized by-product hemin from lysed blood cells have been shown to cause iron-mediated neuronal death (Zille et al., 2017). The primary defence mechanism against ferroptosis involves GPX4, the enzyme that can reduce lipid peroxides to non-toxic alcohols. We noted that GPX4 was excluded from EVs isolated from CM-R⁺ relative to CM-R⁻ suggesting that CM-R⁺ patients are more susceptible to iron-mediated oxidative stress and cell death. Neurodegenerative diseases with phenotypes related to CM have been linked to ferroptosis (Reichert et al., 2020). Therefore, anti-ferroptosis regimens such as ferrostatins and iron chelators currently used to treat other neurological diseases could be explored in CM treatment (Reichert et al., 2020).

Genes crucial for B cell differentiation, such as paired box protein 5 (PAX5) (Cobaleda et al., 2007; Medvedovic et al., 2011), were excluded from EVs obtained from CM-R⁺ patients. PAX5 is a transcription factor exclusively expressed by cells of the B-lineage from pro-B cells to mature germinal centre B cells. It promotes B cell commitment by activating B-specific genes while simultaneously repressing B-lymphoid inappropriate genes (Cobaleda et al., 2007). Conditional deletion of PAX5 causes dedifferentiation of mature B cells to non-committed progenitors that later develop into functional T cells in the thymus, suggesting that PAX5 is also essential for maintaining mature B cell identity (Medvedovic et al., 2011). THEMIS2 was another gene necessary for B cell selection that was significantly downregulated in CM-R⁺. THEMIS2 enhances the positive selection of germinal centre B cells and B1-cells by self and foreign antigens by lowering the B cell activation threshold of low avidity antigens (Cheng et al., 2017). Therefore, our findings showed that disturbances in B cell development and selection could be one of the ways through which *P. falciparum* causes severe disease.
Moreover, TIGIT, a coinhibitory receptor that negatively regulates activated T cells to prevent excessive harmful immune response, was augmented in EVs from CM-R+ patients (Frimpong et al., 2019). Under pathological conditions such as CM, suppression of T cell function by coinhibitory receptors could lead to exhausted (hyporesponsive) effector T cells with reduced ability to produce cytokines, stimulate B cells and clear parasites (Frimpong et al., 2019; Graham et al., 2016). The specificity of TIGIT upregulation in malaria has been previously noted where higher expression of TIGIT in CD4+ T cells was only observed in experimental and human *Plasmodium* infection but not in experimental *T. gondii* infection (Graham et al., 2016). In addition, augmented expression of TIGIT in malaria is thought to be due to chronic systemic inflammation rather than persistent stimulation of T-cells by malaria antigens (Graham et al., 2016).

Cellular EVs also contained RNA originating from distant tissues, besides transcripts derived from peripheral cells. I found that > 30 % of the significantly altered transcripts between CM-R+ and CM-R- patients were specific to the brain and mapped to actual brain functions such as neurogenesis and nerve impulse transmission. These data suggest that the transcriptome of EV circulating in the blood can be used to eavesdrop on the molecular perturbations in inaccessible tissues/ organs such as the brain.

Additionally, I identified a signature of 23 differentially abundant transcripts that could discriminate CM-R+ and CM-R- patients with high accuracy and minimal redundancy. One of the most exciting transcripts among the selected 23 was NCAN, an astrocytic proteoglycan that is upregulated upon brain insult and mediates glial scar formation (McKeon et al., 1999). This suggests that gliosis could also be involved in addition to the known CM pathological processes, such as parasite sequestration and activation of the brain endothelial cells. Similarly, the brain-specific potassium channel, KCNA1 (excluded from EVs obtained from CM-R+ patients), was the most interesting downregulated gene, and its deficiency is associated with epilepsy. Therefore, the RNA cargos of circulating EVs provide
new insights into the pathophysiology of CM and are potential non-invasive diagnostic candidates.

Currently, there are two schools of thought regarding the differences between CM-R+ and CM-R−: 1) that CM-R− is a less severe form of CM; 2) that CM-R− is an atypical disease caused by other known aetiologies of encephalopathy besides *P. falciparum*. After applying manifold learning, I found that CM-R− is most likely a milder form of the disease that has the potential to evolve into the CM-R+ phenotype. I also observed that disease pseudotime is associated with progressive neuronal loss (caused by neuron apoptosis) and increased gliosis (caused by reactive glial cells) which is characteristic of other neurodegenerative diseases such as Alzheimer’s disease (Mukherjee et al., 2020). Surprisingly, the plasma EV pseudo-temporal model of disease that separates CM-R+ and CM-R− could not be recapitulated using whole blood transcriptome. This makes circulating EVs more attractive sources of cerebral malaria biomarkers compared to whole blood.

5.5 Conclusion

In this Chapter, I compared the host RNA content of EVs obtained from CM-R+ and CM-R− malaria patients. The significantly altered transcripts mapped to both peripheral and neural biological functions. The altered peripheral processes suggested deregulated coagulation, inability to fight iron toxicity, impaired B-cell differentiation, and T-cell exhaustion among the CM-R+ compared to CM-R− patients. About 30% of the significantly altered transcripts originated from the brain. I also generated a signature matrix comprising 23 genes that can accurately distinguish CM-R+ and CM-R−. The most biologically exciting genes in the signature were NCAN (a positive marker of glial scar) and KCNA1 (a negative marker of epilepsy). Finally, I show that CM-R+ and CM-R− progress in a single lineage, with late disease pseudotime samples being those from CM-R− patients.
CHAPTER 6

6.1 Summary and Aims

Chapter 3 suggests a homeostatic role of PfEVs in *P. falciparum*. In this Chapter, I discuss future experiments that could be done to validate this observation. These range from determining the mechanism of RNA loading into parasite-derived EVs to investigating whether RNA secreted via PfEVs can be translated in recipient parasite cells. In the second part of this Chapter, I build on the work reported in Chapters 4 and 5 by describing future experiments aiming to elucidate the biological content of host-derived circulating EVs.

6.2 The homeostatic role of PfEVs in *P. falciparum*

6.2.1 How might *P. falciparum* load RNA into PfEVs?

To infer the homeostatic role of PfEVs in *P. falciparum*, I compared the RNA content of PfEVs to that of the secreting WP. While RNA has been successfully isolated from PfEVs by other studies (Babatunde et al., 2018a; Sisquella et al., 2017), the mechanism used by the parasite to load RNA into EVs is mainly unknown. RNA could be loaded into PfEVs by the lipid rafts on the MVBs limiting membrane or by RBPs. More than 15% of the *P. falciparum* genome encodes RBPs, favouring the notion that RBPs primarily mediate RNA sorting into PfEVs. Note also that studies involving mammalian systems have found that the LC3 (also called autophagy-related protein 8E [ATG8E]) conjugation machinery is involved in loading RNA-RBPs complexes into EVs (Leidal et al., 2020). The *P. falciparum* also encodes autophagy-related proteins, including an ATG8 orthologue (PF3D7-1019900) (Gardner et al., 2002). Therefore future studies can focus on establishing whether the loading of parasite RNA into PfEVs involves RBPs and secretory autophagic-related proteins. Previous studies did not find enrichment of RBPs in PfEVs, most likely because they were limited in the number of time points and replicates used. The detection of parasite proteins was also masked by the more abundant host proteins, such as proteasomes. Proteomic protocols that...
enrich parasite-derived EVs should be used to establish whether RBPs are enriched in PfEVs, as this might suggest their role in sorting RNA molecules into PfEVs.

6.2.2 The physiological requirement of a gene during the IDC can be inferred based on its exclusion from PfEVs

The biological function of more than half of all the genes encoded by *P. falciparum* remains uncharacterized (Bozdech et al., 2003; Gardner et al., 2002). Classification of these genes based on the developmental stage when their products are physiologically needed would be a crucial step towards their characterization. The use of proteomics to achieve this purpose is hindered by the low throughput nature of the technique and the fact that the IDC could require some gene products (e.g. GDV1-antisense RNA) at the RNA level (Broadbent et al., 2015; Filarsky et al., 2018; Sims and Hyde, 2006). Comparing RNA abundance encoded by the IDC to that from mosquito and liver stages can help infer genes required during each stage. However, sometimes RNA abundance is not a good proxy for protein abundance as RNA quantities do not always correlate with protein quantities in *P. falciparum*. For example, MSP1, a well-characterized gene essential during the asexual blood stage, is expressed (at RNA level) higher in mosquito stages relative to the IDC (Real et al., 2021) and reanalysis of the Malaria Cell Atlas single cell RNAseq in section 3.3.5 revealed that other merozoite genes such as P113 and AMA1 were expressed higher (at RNA level) in mosquito stages compared to the IDC (Howick et al., 2019; Real et al., 2021; Reid et al., 2018). However, in Chapter 3, I showed that RNAs encoded by P113, AMA1 and MSP1 were preferentially excluded from PfEVs, suggesting that PfEVs could differentiate genes required or not required during the IDC with higher accuracy than conventional methods. Therefore, PfEVs provide an additional molecular tool for future scientists to use to characterize *P. falciparum* genes whose function is not known.
6.2.3 \textit{Pf} EVs released by mosquito stages could be enriched in RNA encoded by genes required during the IDC

I showed that RNA encoded by genes not required at all during the asexual blood stage are secreted via \textit{Pf} EVs. This includes genes required for processes such as gametocytogenesis (GDV1 and AP2G) and gametogenesis (heme biosynthesis and fatty acid synthesis enzymes). Based on this finding, I speculate that the opposite happens during the mosquito and liver stages. During the mosquito and liver stages, the parasite could preferentially retain RNA from genes physiologically required at these stages. At the same time, those that are useful only during the IDC could be preferentially secreted via \textit{Pf} EVs. Demonstrating this (using gametocytes, for example) will reinforce my findings that the parasite secretes physiologically unwanted RNAs via \textit{Pf} EVs. This post-transcriptional regulatory role of EVs can also be explored using mammalian cells.

6.2.4 The mRNA within \textit{Pf} EVs might be translated into proteins in recipient parasite cells

The mRNA cargo of EVs was found to be translated to proteins in mammalian systems (Valadi et al., 2007). Although this remains to be demonstrated in \textit{P. falciparum}, we already know that adding \textit{Pf} EVs to asexual cultures promotes the formation of gametocytes, the sexual stages transmitted from humans to mosquitoes (Mantel et al., 2013; Regev-Rudzki et al., 2013). This suggests that molecules that induce gametocytogenesis are packed into EVs and can be transferred to recipient parasites when the \textit{Pf} EVs fuse with the plasma membrane of infected erythrocytes. In Chapter 3, I found that \textit{Pf} EVs were enriched in mRNA that encodes proteins required for gametocytogenesis, including AP2-G, GDV1 and GECO. \textit{Pf} EVs obtained from asexual cultures do not contain gametocyte proteins. Therefore the most plausible mechanism through which \textit{Pf} EVs could induce the formation of gametocytes is only if the gametocyte mRNAs within the \textit{Pf} EVs are translated into proteins in recipient
parasites (Abdi et al., 2017; Mantel et al., 2013; Regev-Rudzki et al., 2013). To explore this further, it will be worth investigating whether Pf EVs obtained from wild-type asexual parasites can induce the formation of gametocytes in AP2-G mutant parasite isolates.

6.3 Circulating EV cargo as biomarkers of severe malaria – future directions

6.3.1 Immune capture methods could improve the diagnostic potential of EVs

I have demonstrated that EVs contain proteins and RNAs originating from clinically inaccessible tissues such as the brain, making them attractive diagnostic and prognostic candidates for severe malaria. To show that EVs can be used as a type of liquid biopsy, I applied bioinformatic tools to trace back EV cargo to the parent cells theoretically. However, realizing the full potential of EVs as sources of severe malaria biomarkers will require surface markers to isolate target EVs (like those from the brain cells) from the high background of EVs from other tissues such as blood. For example, neural cell adhesion molecule L1 (L1CAM) (Faissner et al., 1984) is found on the surface of neuronal-derived EVs, and hence anti-L1CAM antibodies are widely used to isolate L1CAM+ neuronal-derived EVs from biofluids like plasma (Dagur et al., 2020; Gill et al., 2018; Goetzl et al., 2019; Guix et al., 2018; Katsu et al., 2019; Kumar et al., 2021; Madhu et al., 2019; Nogueras-Ortiz et al., 2020; Pulliam et al., 2019; Rani et al., 2019; Shi et al., 2016; Shi et al., 2014). Such immune-capture strategies can be used to separate plasma EVs originating from specific tissues affected by *P. falciparum* infection, like those of the brain, liver, and kidney. In Chapters 4 and 5, I focused more on host-derived EV proteins and RNA simply because the highly abundant host signal masked the detection of parasite materials in plasma EVs. Antibodies to *P. falciparum* proteins on the surface of plasma EVs can also be used to separate parasite-derived EVs from the high background of host EVs.
6.3.2 Comparing the biological materials in EVs with plasma or parent cells could reveal interesting differences

A study on myocardial infarction showed that proteins were robustly correlated between EVs and plasma, but some proteins, such as ADAMTS13 and DCN, were not (Gidlöf et al., 2019). Future malaria studies can compare the relative levels between the two sample types to assess whether the EV cargo differs from that in plasma. This could reveal biologically significant differences between the two sample types that will be crucial in developing EV-based diagnostic assays.

The contents of EVs in circulating biofluids are thought to broadly reflect that of the secreting cells. While this premise was generally used to interpret Chapters 4 and 5 results, I acknowledge that there are reported scenarios in the literature where the cargo in EVs is inversely correlated to that of the secreting cells (O’Grady et al., 2022). Comparing the cargo in plasma EVs from acute malaria patients to that in clinically accessible cells (e.g. immune cells) obtained from the same samples could reveal proteins and RNA that are differentially altered between the two sample types.

6.3.3 How can the markers of disease progression in malaria be validated in future studies?

There are several ways in which the work reported in Chapters 4 and 5 can be validated in future studies. For instance, future studies will need to analyse EVs from independent cohorts, preferably sampled in other African malaria-endemic regions. Plans are underway to engage our collaborators in Malawi and Uganda to acquire samples from independent malaria cohorts that will be used to validate the work reported in this thesis. Although I found pseudotime to be a good proxy for disease progression, I am aware that it is a mathematical derivative from cross-sectionally collected samples. In future, data can be generated from longitudinal samples obtained from individuals enrolled in controlled human
malaria infection (CHMI) challenge studies to validate the pseudo-temporal model of disease progression. We have archived samples from a recent CHMI study (Kapulu et al., 2018), and these could be ideal for validating what I observed using the crosssectional samples. Finally, it is worth noting that sequencing and mass spectrometry might not be the techniques used in everyday clinics in malaria-endemic areas. Therefore, the biomarkers reported in this discovery study will have to be validated using more affordable and quicker techniques such as ELISA (for protein markers) and RT-PCR (for RNA markers), which are commonly available in malaria-endemic regions.

### 6.3.4 Application of EVs-based biomarkers to clinical disease

Several EV-based liquid biopsy techniques (chips) have been developed to allow the detection of EV-based biomarkers from crude patient samples, and they include ExoScreen (Yoshioka et al., 2014), nanoplasmonic exosome (nPLEX) assay (Im et al., 2014) and EV array (Jakobsen et al., 2015). ExoScreen was first developed to diagnose colorectal cancer and could detect disease-specific (CD147) and EV-conventional (CD9) protein markers via an amplified luminescent proximity homogenous assay without the need for any EV purification steps (Yoshioka et al., 2014). A similar ExoScreen assay could be designed to detect, for instance, STYXL2 (Fig. 4.11) in malaria-derived EVs to diagnose patients who are highly likely to succumb to severe malaria. The second method, nPLEX, is a chip based on transmission surface plasmon resonance via periodic nanohole arrays (Im et al., 2014). Each of the nanoarrays in the nPLEX has antibodies to specific EV-based biomarkers that allow the profiling of intact exosomes (Im et al., 2014). An EV array, which is analogous to a protein microarray, is another technique that allows the detection of EV-biomarkers from crude samples without EV enrichment (Jakobsen et al., 2015). In an EV array, antibodies to EV surface protein markers (both disease-specific and conventional EV markers) are printed onto glass slides and used to capture disease-specific EVs. Some technologies have also been developed to detect EV-RNA molecules in crude samples. For example, the ExoDx Prostate
Test is used to perform a risk assessment for prostate cancer, which previously relied heavily on digital examination and prostate biopsy (McKiernan et al., 2020). ExoDx Prostate Test detects three RNA molecules (PCA3, ERG and SPDEF) that have been found in EVs isolated from the urines of prostate cancer patients without the need to isolate EVs. The test is also non-invasive and accurate and has saved a lot of American men from undergoing unnecessary prostate examinations or biopsies (McKiernan et al., 2020). Likewise, if future studies validate the markers reported in Chapter 5 (e.g. NCAN) in other cohorts, then an RNA chip can be designed to diagnose “true” cerebral malaria cases from other types of encephalopathies.
Chapter 7  Bibliography


CHAPTER 7


Cerebral Malaria through a Blood Transcriptomic Signature: Evidences for Erythrocyte Alteration, Immune/Inflammatory Dysregulation, and Brain Dysfunction. Mediators Inflamm 2020, 3280689.


CHAPTER 7


microparticles and are highly associated with susceptibility to severe malaria in humans. Infect Immun 81, 1287-1294.


**CHAPTER 7**


Veenemans, J., Milligan, P., Prentice, A.M., Schouten, L.R., Inja, N., van der Heijden, A.C.,
supplementation with zinc and other micronutrients on malaria in Tanzanian children: a

the malaria parasite Plasmodium spp.: systems-wide studies pave the way. Wiley Interdiscip
Rev RNA 7, 772-792.

PfAlba1 RNA-binding protein is an important regulator of translational timing in
Plasmodium falciparum blood stages. Genome Biol 16, 212.

Verdura, E., Fons, C., Schlüter, A., Ruiz, M., Fourcade, S., Casasnovas, C., Castellano, A.,
and Pujol, A. (2020). Complete loss of KCNA1 activity causes neonatal epileptic
encephalopathy and dyskinesia. Journal of Medical Genetics 57, 132-137.

Verhoef, H., West, C.E., Kraaijenhagen, R., Nzyuko, S.M., King, R., Mbandi, M.M., van
Laatum, S., Hogervorst, R., Schep, C., and Kok, F.J. (2002). Malarial anemia leads to
adequately increased erythropoiesis in asymptomatic Kenyan children. Blood 100, 3489-
3494.

the ATPase VPS4 by its cofactor LIP5 and the endosomal sorting complex required for

Villarroya-Beltri, C., Gutiérrez-Vázquez, C., Sánchez-Cabo, F., Pérez-Hernández, D.,
Vázquez, J., Martin-Cofreces, N., Martínez-Herrera, D.J., Pascual-Montano, A.,
sorting of miRNAs into exosomes through binding to specific motifs. Nat Commun 4, 2980.


analysis of data from a phase 3 randomised controlled trial. The Lancet Infectious diseases 15, 1450-1458.


Wittchen, E.S. (2009). Endothelial signaling in paracellular and transcellular leukocyte transmigration. Front Biosci (Landmark Ed) 14, 2522-2545.


