CHAPTER THREE: METHODS

3.1 *Plasmodium falciparum* culture

3.1.1 Preparation of EV-depleted culture medium

Five percent (w/v) Albumax II (Life Technologies) was dissolved in water and depleted of vesicles by centrifugation at 150,000 x g for 2 h at 4 ℃ (Beckmann Coulter, 70 Ti rotor). The EV-depleted Albumax was sterilised by filtration through 0.22 µM (Millipore). Depleted Albumax was aliquoted into 12.5 ml portions and stored at – 20 ℃ until use. A litre of incomplete medium was prepared as follows: 10.43 g RPMI 1640 Life Technologies 51800019 (with L-Gln, no NaHCO3), 7.15g HEPES (Sigma), 2 g glucose (Sigma), 1 ml of 50 mg/ml hypoxanthine (Sigma), 0.5 ml of 50 mg/ml gentamicin (Sigma) and 900 ml Milli-Q H_2O and pH adjusted to 7.2. The incomplete medium was filtered using a 0.22 µM filter and stored at 4 ℃. To 1 L incomplete medium, 7.5% NaHCO3 (Sigma) and 5 g/l (12.5 ml) of EV-depleted Albumax II (both pre-filtered) were added to make up the complete culture medium, which was stored at 4 ℃ until use.

3.1.2 *P. falciparum* isolates used in this study

Three *Plasmodium falciparum* strains were used for this study. These included *P. falciparum* NF54, which is a long term-adapted laboratory strain that is known to produce gametocytes, and two Kenyan clinical isolates that have more recently been adapted to culture (approximately <100 cycles). The two Kenyan isolates are named as 9775 and 11019 based on KEMRI-Wellcome Trust parasite isolate naming system. The genomes of the two isolates have been sequenced at the Wellcome Sanger Institute and have been re-named as PfKE11 and PfKE10 respectively.

3.1.3 *Plasmodium falciparum culturing for* EV isolation

The parasites were maintained at 2% parasitaemia and 2% haematocrit in fresh (less than two weeks old) human O^+ RBCs (NHS Blood and Transplant, Cambridge, UK; informed consent from donors was obtained by NHSBT) using complete medium as prepared above, gassed using a mixture of 5% O2, 5% CO2 and 90% N2, and incubated at 37 ℃. Giemsa smears were performed on a daily basis to monitor parasitaemia. When required, the parasites were synchronized using 5% sorbitol, which kills mature stages (trophozoites and schizonts) and expanded into three 50 ml culture flasks per parasite strain for EV production.

3.1.4 Production of PfEVs

Once parasites reached optimal growth, the three 50 ml flasks were synchronized for the second time using 5% sorbitol at around 12 h post-invasion (small rings) and expanded to six 50 ml flasks per strain (5% parasitaemia, 2% haematocrit). After 12 hours (24 hrs after invasion), a Giemsa – stained smear was prepared to confirm the asexual parasite stage of the cell culture, which was predominantly early trophozoites. The cultures were suspended by pipetting up and down and transferred to 50 ml tubes. The 50 ml tubes were centrifuged (Centrifuge 5810, eppendorf) for 5 min at 600 x g with a low brake. The supernatant, referred to as culture conditioned medium (CCM), which contained PfEVs released between 12 - 24 h after invasion, was collected into 50 ml Falcon tubes and stored temporary at 4 ℃. The iRBC pellets were resuspended in fresh medium and returned to incubation at 37 ℃. The processing of the CCM was done on the same day once iRBCs pellets had been resuspended in fresh media. This process of 12 - hourly collection of CCM was repeated twice more, resulting in collection of conditioned

medium from three successive parasite development time windows: 12 - 24 h, 24 - 36 h, 36 - 48 h (in the same cycle as 12 - 24 h) (**Figure 5**). After collection of the 36 - 48 h sample, the cultures were diluted to 2% parasitaemia and the 0 -12 h sample was collected in the next cycle. The entire process was later repeated once to obtain a duplicate CCM samples for PfEV isolation from each strain.

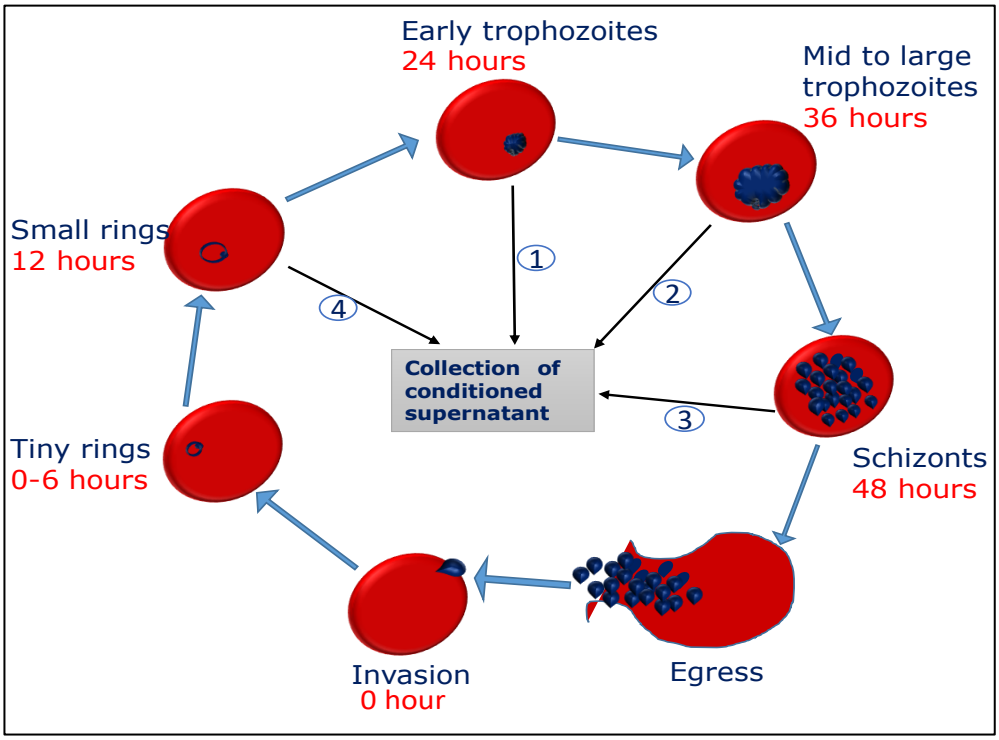

Figure 5: Illustration of time series collection of parasite conditioned medium. Parasite were synchronized and suspended in fresh medium at "small ring-stage". Samples were collected as indicated: 1; (12-24 h), 2; (24-36 h), 3; (36-48 h) and 4; (0-12 h).

3.2 Processing of CCM and purification of PfEVs

Purification of PfEVs was done as previously described [3]. The collected CCM samples briefly stored at 4°C (**Section 2.1.3**) were centrifuged at 2000 x g in 50 ml tubes for 10 min at 4 °C to remove traces of cells and debris. The supernatant was transferred to a new 50 ml tube and spun at 3200 x g for 30 min, 4 °C with low brake, with the pellet being again discarded. The supernatants from each time point of the parasite life cycle

were pooled together and passed through 0.22 µM (Millipore) to exclude vesicles larger than 220 nm majority of which are microvesicles. The collected sterile filtrate was aliquoted into 50 ml tubes, temporarily stored at -20 ℃ and later transferred to -80 ℃ until use.

The processed culture supernatants were thawed and centrifuged for 20 min at 15000 x g, 4 ℃. The supernatant was transferred to 15 ml concentration columns (50kDa cut-off, Millipore) to decrease the supernatant volume by a factor of 3 - 4. The concentrate was transferred to 13.5 ml Quick seal ultracentrifuge tubes (Beckmann) and spun for 2 h at 150,000 x g, 4 ℃ in a preparative ultracentrifuge (Beckmann Coulter, 70 Ti rotor). The supernatant was discarded and the pellets from the same time course pooled, washed once with 13.5 ml of PBS and centrifuged for 2 h at 150,000 x g, 4 ℃. The presence of extracellular vesicles in both the pellet and the PfEV suspension was confirmed by sectioned and negative stained transmission electron microscopy respectively.

3.3 PfEV-RNA isolation

The PfEV pellet was suspended in 400 µl of PBS and treated with 2 units (1 µl) of RNase A (Biolab) for 15 min at room temperature to remove non-vesicular RNA. The treated PfEV suspension was digested with 500 µl of lysis buffer (Isolate II RNA Min Kit, Bioline) and stored at -80 ℃. RNA was isolated using the Isolate II RNA Min Kit, (Bioline) following manufacturer's instructions and eluted in 100 µl of nuclease free water. Purified RNA was quantified using Bioanalyzer Pico RNA chips (make) and stored at -80 ℃, until library preparation.

3.3 Preparation of cDNA libraries

An established protocol [111] was used to prepare the cDNA libraries with some minor modifications. Succinctly, RNAcleanXP beads (Biolabs) solution was used to concentrate the total PfEV-RNA five times. The concentrated RNA was eluted from the beads using a Qiagen elution buffer (10 mM Tris-HCl pH 8.0). All RNA elution in downstream steps was done using this buffer. First strand cDNA was synthesised using Superscript III reverse transcriptase (Invitrogen) with the following cycling parameters: 25 ℃ for 10 minutes, 42 ℃ for 60 minutes and a final hold of 4 ℃. Both oligo d(T) and random hexamers from Qiagen were used in the first strand cDNA synthesis reaction. The first strand reaction was cleaned using 1.8 volumes of RNAcleanXP beads solution followed by two successive washes with 80% ethanol while on the magnetic stand. RNA was eluted from the beads with 41 µl of Tris-HCl, but the beads were left in the wells for later use in cleaning the second strand reaction. Second strand cDNA was synthesised using NEBNext RNA Second Strand Synthesis Module (Biolabs) (parameters: 16 ℃ for 2.5 hours and a final hold of 4 ℃) replacing dTTP with dUTP. Double stranded cDNA was cleaned using the RNAcleanXP beads already in the wells. The binding properties of the second strand cDNA to the beads was rejuvenated by addition of 1.8 volumes of 20 % polyethylene glycol (PEG), 2.5 M NaCl solution. Two successive washes of the beads were performed using 80% ethanol while on the magnetic stand. The cDNA was eluted, quantified using a Bioanalyzer high sensitivity DNA chip and then the beads were discarded. Cleaned cDNA in 50 µl elution buffer was processed in AFA Covaris for 30 sec (parameters: duty factor 5, peak power 175, cycles 200) to obtain cDNA fragments. Quality control of the fragmented cDNA was performed using the Agilent Bioanalyzer high sensivity DNA chip. End repair and dA tailing of the cDNA fragments were done using NEBNext Ultra II DNA Library Prep Kit. The blunt-ended, polyA-tailed cDNA

was purified using 1.8 volumes of AMPureXP beads in 20% PEG, 2.5M NaCl solution, washed twice using 80% ethanol while still in the magnetic stand and eluted in 50 µl elution buffer. The end-repaired cDNA was ligated to 5 µl of 0.1 µM NETflex adapter oligo mix using Quick DNA ligase and incubation at 20 ℃ in a thermal cycler for 15 min. The ligated cDNA was purified using 0.8 volumes of 20% PEG, 2.5 M NaCl AMPureXp bead solution, washed twice using 80% ethanol while still in the magnetic stand. The clean ligated cDNA was eluted in 24 µl elution buffer (10 mM Tris-HCl pH 8.0) and quality control was performed using Agilent Bioanalyzer HS DNA chip. The libraries were digested with 0.1 volumes of Uracil-Specific Excision Reagent (USER enzymes, Biolabs) at 37 ℃ for 15 min followed by a DNA denaturation step at 95 ℃ for 10 min. The single strand ligated cDNA was amplified (in 10 cycles) using KAPA HiFi HotStart ReadyMix PCR kit (KAPABIOSYSTEMS) and the PCR primers P5 and P7 (Illumina) following the manufacturer's recommended cycling parameters. PCR products were quantified using qRT-PCR and again cleaned using 0.8 volumes AMPureXP beads in 20% PEG, 2.5 M NaCl solution. All the cDNA library samples were pooled in equimolar concentrations into one tube and sequenced using the Illumina HS2500 genome analyser in 158 cycles.

3.4 PfEV-RNA data analysis

3.4.1 Read processing and mapping

CRAM files for data generated in this study were obtained from the WSI sequencing pipeline and converted to fastq format using *samtools -1.3 view*. The tool FastQC [112] was used to confirm quality details of the data such as average Phred score, GC content, distribution of read length and overrepresented sequences. FastQC provides relatively

conservative cut-offs even for data generated using the state of the art sequencing facilities. However, attention was paid to details to ensure that the data did not have obvious flaws like a high number of duplicates or below average read quality. Next, *bowtie2* [113] was used to eliminate rRNA sequences from the data by mapping the short reads against an artificial *P. falciparum* ribosomal RNA sequence file and only reads not mapping to these rRNA sequences were used in downstream steps.

The super-fast splice aware mapping tool called HISAT2 [114] was used to align the processed reads to the parasite genome as described earlier. Succinctly, a HISAT2 index was produced from the *P. falciparum* genome sequences downloaded from GeneDB [115], using default parameters. Paired reads were aligned to the genome sequences using the following HISAT2 parameters: *-max-intron 5000 --rna-strandness RF --dta –I 75 –X 500.* These parameters specified a maximum intron size of 5000 for *P. falciparum*, a strand specific PfEV-cDNA library generated using the dUTP method and an expected fragment length of 75 – 500 bp. The *--dta* option commands HISAT2 to produce SAM files that are compatible with **d**ownstream **t**ranscript **a**ssembly tools.

For control comparisons, whole parasite data generated by Lopez-Barragan and others [116] was downloaded from GenBank (accession number SRP009370), processed and mapped to the reference genome as described above. This is a time series data collected at 8 h, 19 h, 30 h and 42 h of the *P. falciparum* asexual life cycle. This data is in sync with the time points of sample collection of the present study, and was therefore used as a background to predict transcripts highly enriched in PfEVs relative to the whole parasite transcriptome.

3.4.2 Mapping quality control and quantification of expression

Each SAM file was sorted and converted to a BAM file in a single step using *samtools-1.3 sort,* while *samtools-1.3 flagstat* was used to determine the unique alignment rate per sample [117]. Subsequently, *samtools-1.3 merge* [117] was used to merge BAM files originating from different sequencing lanes but belonging to the same sample. A *P. falciparum* GFF3 file was downloaded from GeneDB [115], and converted to GTF and BED formats using RSEM tools [118] for use in downstream steps. Next, RSeQC tool kit [119] was used to determine read distributions within exons, introns and intergenic regions, as well as strand specificity of the PfEV-cDNA library. Visualization of BAM files was done using the integrated genome browser (IGV) [120]. Assembling and summation of expression was done using *stringtie* [121] with the following parameters: *stringtie -G –e -A.* These parameters instructed *stringtie* to use the provided GTF file as a guide during assembly, to only output transcript alignments matching the GTF file (not to predict novel transcripts), and to output a file containing gene expression in FPKM (**f**ragments **p**er **k**ilobase of exons per **m**illion reads) values.

3.4.3 Data normalisation and expression analysis

Data normalisation was performed prior to linear modelling in subsequent steps. This was done to not only reduce technical variation within the PfEV data, but to also make it more comparable to the whole parasite data. First, the expression data from individual samples were pooled together to form a single data matrix. Subsequently, a R package called *Linnorm* [122] was used to perform library size/batch effect correction. *Linnorm* has a function (*linnorm.norm*) that transforms the FPKM values into log2FPKM values, estimates the mean variance relationship and uses this to calculate appropriate

observation-level weights (in this case raw counts). Expression analysis was done by modelling in *edgeR* [123] using the quasi-likelihood F test. A false discovery rate (FDR) adjusted P value of <0.05 was chosen as a cut-off for significant genes while the Log2 (Fold Change) threshold was set at absolute 5.

3.5 Gene ontology analysis of transcripts differentially enriched in PfEVs

Gene ontology [124], network analysis and visualization were performed using a R/bioconductor package called *clusterProfiler* [125]**.** The *Plasmodium falciparum* organism specific R package (*org.Pf.plasmo.db*) [126] was used to provide the mappings between gene symbols and gene ontology terms. The hypergeometric test (an over-representation test) option of *clusterProfiler* was used, with adjusted P value of 0.05 set as a cut-off. The results of gene ontology test were visualized by bar plots, enrichment maps and category-gene-network plots.