RNA sequencing analysis of *Plasmodium falciparum* extracellular vesicles

Kioko Mwikali

Wolfson College

University of Cambridge

This dissertation is submitted for the degree of Master of Philosophy

August 2018

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed word limit for the relevant Degree Committee.

Kioko Mwikali km19@sanger.ac.uk

The work reported in this thesis was carried out under the supervision by the following:

Dr Julian Rayner jr9@sanger.ac.uk

Dr Alena Pance ap9@sanger.ac.uk

Dr Abdirahman Abdi aabdi@kemri-wellcome.org

Acknowledgements

My sincere appreciation goes to Drs Alena Pance, Abdirahman Abdi and Julian Rayner whose mentorship and constant supervision of the work presented was performed. I am very grateful to David Goulding for selflessly imaging the extracellular vesicle preparations. Special thanks goes to Derek Pickard for his technical help in running the ultracentrifuge during the initial days of the project. I thank Drs Alison Kemp, Lia Chappell and Alena Pance for their guidance during the preparation of cDNA libraries. I greatly appreciate Dr Adam Reid for his valuable advice on RNAseq analysis, and Dr Daniella Murallo for technical training in running bioinformatics tools. Gratitude goes to Liam Prestwood (Lab manager) for ensuring that I had access to all lab reagents and equipment used in this study

I acknowledge my thesis committee members: Drs Pete Bull, Mara Lawniczak, Alena Pance, Julian Rayner and Prof Margret Robinson for their insightful advice and suggestions on the project. I appreciate my college tutor Dr Marina Salorio-Corbetto for her encouraging advice, when experiments sometimes seemed not to work. I also acknowledge Dr Marie Kent for recruiting me into the Wolfson College Social Rowers Team (WCST), which allowed me to have a relaxing break from work every Sunday. I acknowledge Drs Annabel Smith and Christina Herdberg-Delouka for monitoring my progress during my project. I would like to thank the Wellcome Sanger Institute and the KEMRI-Wellcome Trust Research Programme whose collaboration gave me the great opportunity to do the work presented in this thesis. Finally, I thank God for the gift of life and success of this work

Abstract

Plasmodium falciparum is the most lethal malaria parasite. Blood stage *P. falciparum* parasites secrete extracellular vesicles (EVs), which contain a variety of bioactive molecules and can transfer them to host and/or parasite recipient cells. The RNA content of *P. falciparum* EVs, which could potentially affect biological processes within the host or mediate communication within the parasite population, has not been previously characterised. In my MPhil project, I described transcripts that are preferentially enriched in PfEVs relative to the whole parasite transcriptome.

P. falciparum culture supernatants were collected at multiple developmental time windows from both a long-term laboratory isolate and two short-term lab-adapted clinical isolates. EVs were purified using preparative ultracentrifugation and imaged using sectioned transmission electron microscopy. PfEV-cDNA libraries were generated using the dUTP method and sequenced. Read mapping was done using HISAT2, while differential expression analysis and gene ontology were performed using the R/Bioconductor packages '*edgeR*' and '*ClusterProfiler*' respectively.

Sectioned transmission electron microscopy imaging showed PfEVs to have a dense lumen surrounded by a lipid bilayer. The size range of the PfEVs was 40 - 200 nm with median size of 100 nm. Analysis of the EV-RNA using an Agilent Bioanalyzer pico RNA chip showed that EV-RNA was in the size range of 25 - 4000bp, peaking at 500bp with little or no traces of ribosomal RNA. I identified 958 transcripts enriched in PfEVs secreted by late stage parasites relative to the total parasite RNA profile as a whole , and fewer than 300 for the other parasite time windows. Gene ontology showed that the transcripts enriched in PfEVs released by mature parasites expressed proteins involved in regulation of cellular processes. There were no overrepresented gene ontology terms for the early and mid-parasite time windows.

Conclusively, PfEVs are enriched in transcripts that express regulatory proteins such as RNA binding proteins and transcription factors, as well as the exported family of P. falciparum proteins. In other systems, EV-RNA is translated in recipient cells and can activate cytosolic immune sensors. Therefore, further studies are required to shed light on the role of PfEV-RNA in host-parasite interaction.

Acronyms and Abbreviations

ABCA	ATP-binding cassette transporter 1
AMA1	Apical Membrane Antigen 1
AP2	Apetala 2 domain
APiAP2	Apicomplexa Apetala 2 domain
BAM	Binary alignment/Map
BED	Browser Extensible Data
Cav-1	Caveolin-1
CCM	Culture conditioned medium
cDNA	complementary deoxyribonucleic acid
CGNP	Category gene network plot
CSP	Circumsporozoite protein
dUTP	deoxy-uridine triphosphate
edgeR	Empirical Analysis of Digital Gene Expression Data in R
ELV	exosome like vesicles
ENA	European Nucleotide Archive
ESCRT	Endosomal sorting complex required for transport
ETRAMP	Early transcribed membrane protein
evDEGs	extracellular vesicle differentially enriched genes
EVs	extracellular vesicles
EXP	exported proten

FPKM	fragments per kilobase of exons per million reads
gDNA	genomic deoxyribonucleic acid
GECO	gametocyte erythocyte cytosolic protein
GEST	gamete egress and traversal protein
GEXP	gametocyte exported protein
GFF3	General Feature Format 3
GIG	gametocytogenesis implicated protein
GM3	monosialodihexosylganglioside
GO	Gene Ontology
GRA	dense granule
GTF	Gene transfer format
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HISAT	Hierarchical Indexing for Spliced alignment of Transcripts
HSP	heat schock protein
IFN	interferron
IL	interleukin
IMC	inner membrane complex
iRBC	infected-red blood cell
ISEV	International Society of Extracellular Vesicles
KAHRP	knob associated histidine rich protein
KEMRI	Kenya Medical Research Institute

LacCer	lactosylceramide
MAF1	repressor of RNA polymerase III transcription
MAHRP1	membrane associated histidine rich protein 1
MC-2TM	Maurer's Cleft two transmembrane proteins
MCs	Maurer's Cleft
mRNA	messenger ribonucleic acid
MSP1	Merozoite Surface protein1
MVs	Microvesicles
MVBs	Multi-vesicular bodies
ncRNA	non-coding RNA
NHSBT	National Health Insurance Blood and Transplant
PBS	Phosphate buffered saline
PC	Phosphatidylserine
PEG	Polyethylene glycol
PfEVs	*Plasmodium falciparum* extracellular vesicles
PfPTP1	PfEMP1 trafficking protein1
PHIST	*Plasmodium* helical interspersed subtelomeric protein
PI	Phosphatidylinositol
PIC	Pre-initiation complex
PLGA	poly-lactic-co-glycolic acid
PS	Phosphatidylserine

PTEX	*Plasmodium* translocon of exported proteins
PUF1	gene encoding a mRNA binding protein
PV	Parasitophorous vacoule
PVM	Parasitophorous Vacoule Membrane
Rap2	Rhoptry associated protein 2
RBCs	red blood cells
REXP1	ring exported protein 1
RhopH2	Rhoptry high molecular weight protein 2
RIFIN	repeatitive interspersed family of proteins
RIN	RNA integrity number
ROS	reactive oxygen species
RPMI 1640	Roswell Park Memorial Institute 1640 culture medium
rRNA	ribosomal ribonucleic acid
RSEM	RNA-Seq by Expectation Maximization
RTS,S,	a malaria vaccine
SAM	Sequence Alignment/Map
SBP1	Spectrin binding protein 1
SMN	Survival motor neuron like protein
SNARE	Soluble NSF Attachment Protein REceptor"
STEVOR	Subtelomeric variable open reading frame protein family
STING	Stimulator of interferon genes

TEM	Transmission electron microscopy
TNFα	Tumor necrosis factor alpha
TSG101	Tumor susceptibility gene 101
TVN	Tubulovesicular network
VCAM-1	Vascular cell adhesion protein 1
VPS	Vacoular protein sorting complex
WSI	Wellcome Sanger Institute

Table of Contents

Table of Figures

Tables