CHAPTER FOUR: RESULTS

4.1 *Plasmodium falciparum* infected RBCS release exosome-like vesicles

Extracellular vesicles were isolated from processed parasite conditioned medium using a combination of ultrafiltration and preparative ultracentrifugation, as described in the Methods section. An examination (by David Goulding, Wellcome Sanger Institute) of the purified material using sectioned transmission electron microscopy (TEM) revealed that they had a size range of 40 – 200 nm and a dense lumen surrounded by a lipid bilayer as shown by **Figure 6**. This morphology matches that described for cell exosomes reported in literature [127], but these vesicles can only be referred to as exosome-like vesicles (ELV) because their biogenesis has not been proven. The sample examined microscopically contained a mixture of vesicles isolated from the entire intraerythrocytic parasite cycle, in recognition that the parasite stage in which PfEVs are released is not yet resolved.

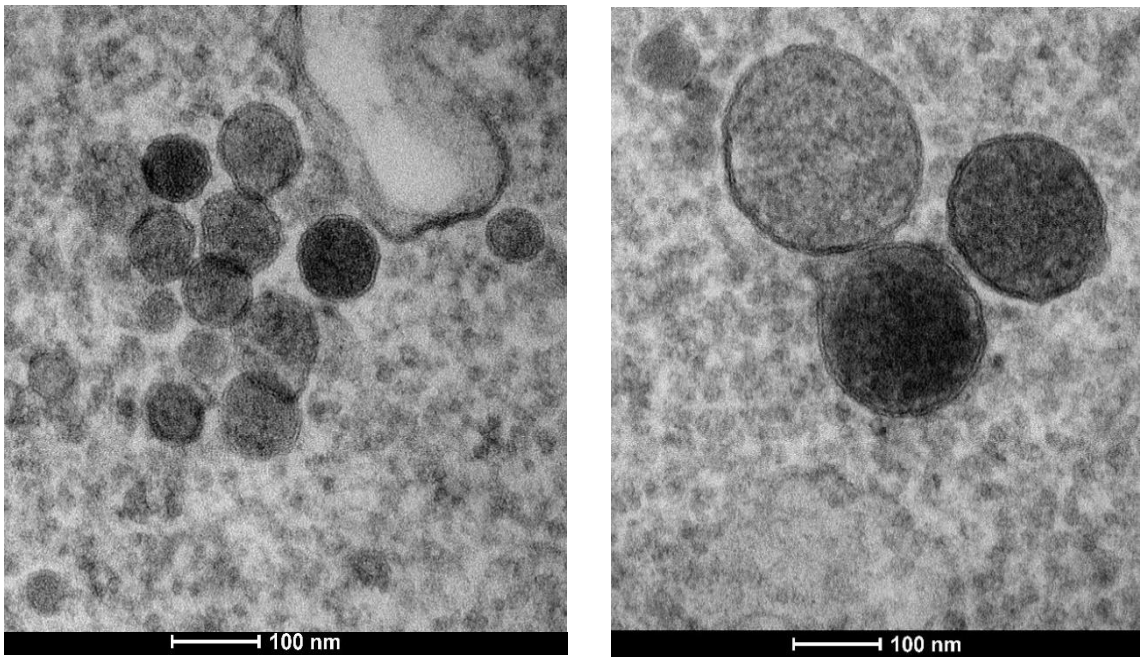

Figure 6: Sectioned transmission electron microscopy images of *Plasmodium falciparum* extracellular vesicles.

4.2 Sequencing of asexual parasite PfEV-RNA time course samples

In order to characterise transcripts enriched within PfEVs, three parasite strains comprising of the long-term adapted laboratory isolate, NF54, and two recent clinical isolates (named 9775 and 11019 as per the KEMRI-Wellcome Trust Research Programme in-house naming system) were cultured, and samples of 300 ml medium collected in four different time windows across the intraerythocytic development cycle. Parasites were synchronised at 12 h post-invasion and conditioned medium collected after 24 h, 36 h, 48 h and 60 h post-invasion in real time. As the intraerythocytic cycle is around 48 h, the 60h sample would in fact represent a time-point 12 h after a second round of invasion. It was therefore termed as the 12 h time point. The time course was repeated for all three strains to provide biological duplicate samples; in total therefore 24 samples were collected, comprising 2x4 time points from each of the three strains.

After purification of PfEVs, the pellets were treated with RNase A and DNAse I to remove non-vesicular RNA and genomic DNA respectively. The RNA content of PfEVs was subsequently purified using the guanidium lysis method. Bioanalyzer analysis revealed that PfEVs contain RNA in the size range of 25 - 4000 bp, peaking at 500 bp (**Figure 7**). The RNA integrity number (RIN) was generally less than 3 while that of the whole parasite RNA ranged between 6 and 8.5. This was expected because RIN is based on the rRNA content of the sample and PfEV-RNA contain little or no traces of ribosomal RNA. The observation was in line with a previous report of PfEV-RNA [128].

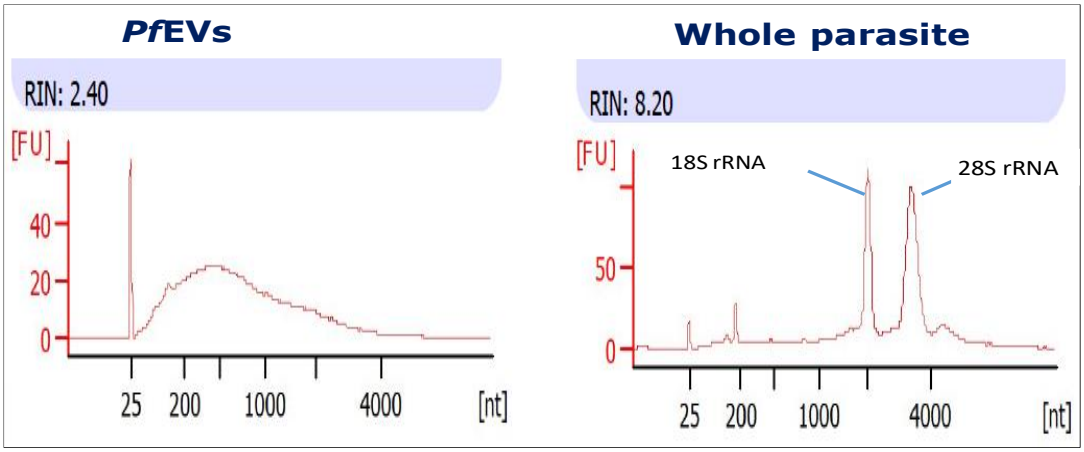

Figure 7: Examples of Bioanalyser traces of both PfEV and whole parasite RNA. The whole parasite RNA was diluted 10 fold before quantification to make its concentration within the measurement range of the Bioanalyser Pico RNA chip

Next, PfEV-cDNA libraries were prepared using the dUTP method to preserve strand specificity, and the KAPA Hifi polymerase was used to amplify the libraries in 10 cycles to boost cDNA concentration. Out of the total 24 samples collected in two batches, libraries could only be prepared from 13 of them as the remaining samples yielded too little RNA for library generation (**Table 1**). Nevertheless, at the end of the library preparation exercise, there was at least one successful library for each time window from NF54 and Pf9775. However, there were libraries for only three time points of Pf11019, as all the samples for the 48 hr time point had to be excluded due to low yield.

Strain	Parasite time window	Batch sample sequenced
Nf54	0-12h	first
	12-24h	first
	24-36h	second
	36-48h	first and second
Pf11019	0-12h	first and second
	12-24h	first
	24-36h	first
	36-48h	none
Pf9775	0-12h	first
	12-24h	first
	24-36h	first
	36-48h	first

Table 1: Table showing PfEV-RNA samples taken for sequencing

The thirteen samples were sequenced on two lanes in an Illumina HS2500 genome analyser in 158 cycles. Sequencing depth ranged between 10 to 16 million reads and a high base quality and zero overrepresented sequences was noted for 12 out of the 13 samples. However, the sample for the second time window of Pf9775 (12-24 h) had a noticeably low yield of 5 million reads per lane and a high number of overrepresented sequences.

4.3 Mapping, quantification of expression and data normalisation

Although the input PfEV-RNA contained little or no traces of rRNA as detected by Bioanalyzer (**Figure 7**), experience has it that even samples treated with ribosomal depletion kits before library preparation still contain some traces of rRNA reads. Cellular

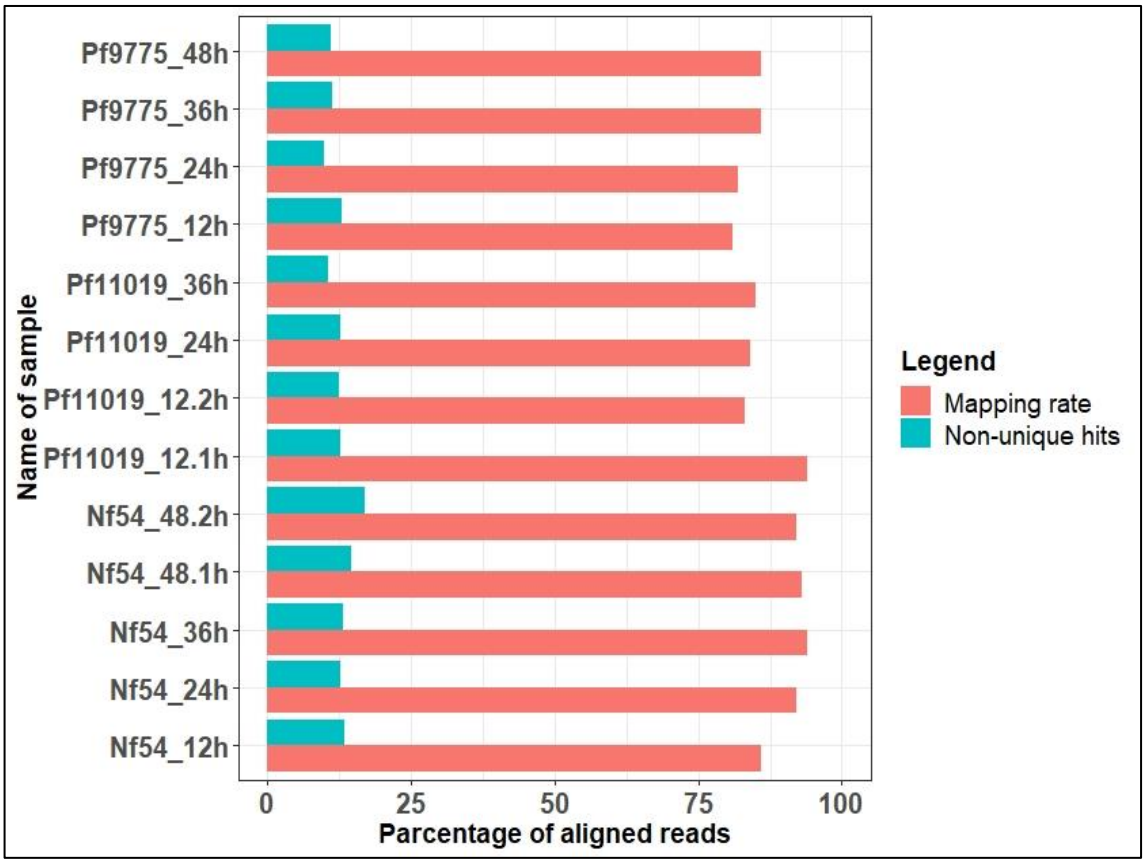

Figure 8: An overview of the alignment rate. Red bars correspond to the overall alignment rate while green bars correspond to non-primary hits

rRNA is more abundant relative to other species and may confound the results of mRNA expression analysis. To remove any traces of rRNA sequences, *bowtie2* [113] was employed to map all reads from each sample against an artificial *P. falciparum* rRNA fasta file, and only reads not mapping to these rRNA were used in subsequent steps. Approximately 0.07 – 2% of reads in each sample mapped uniquely to rRNA sequences. Non-rRNA reads were aligned to the *P. falciparum* genome using HISAT2 [114] and mapping quality assessed using metrics first published by Wang et.al [119]. Briefly, quality metrics calculated include strand specificity, overall alignment rate, exonic rate, intronic rate and intergenic rate. The overall alignment rate ranged between 81 – 94 % as shown by **Figure 8.**

Despite a strand-specific protocol being used to prepare the cDNA libraries, strand specificity was very low as only about 55 -72% of reads mapped to the reference strand in the correct orientation. This was expected because a final amplification of the cDNA libraries was included to increase yield, but which also made the libraries unstranded. Although majority of the reads mapped uniquely to exons, about 2.5 - 5% of reads in each sample aligned to non-coding regions. Exonic reads mapped to the full length of transcripts as shown by **Figure 9** (IGV snapshot), indicating that PfEVs contain full mRNA rather than random fragments.

A popular transcriptome assembly tool called *stringtie* [121] was used to assemble the uniquely mapped reads into transcripts. *Stringtie* was given two instructions: 1) to not assemble novel transcripts as the strand specificity was not preserved during library preparation, and 2) to normalise the reads by transcript length in order to output expression data in FPKM values. To make the Lopez-Barragan et.al [116] whole parasite data comparable to the PfEV data, normalisation by sequencing depth was done by first converting the FPKM values into log2FPKM, and then using this to estimate raw counts.

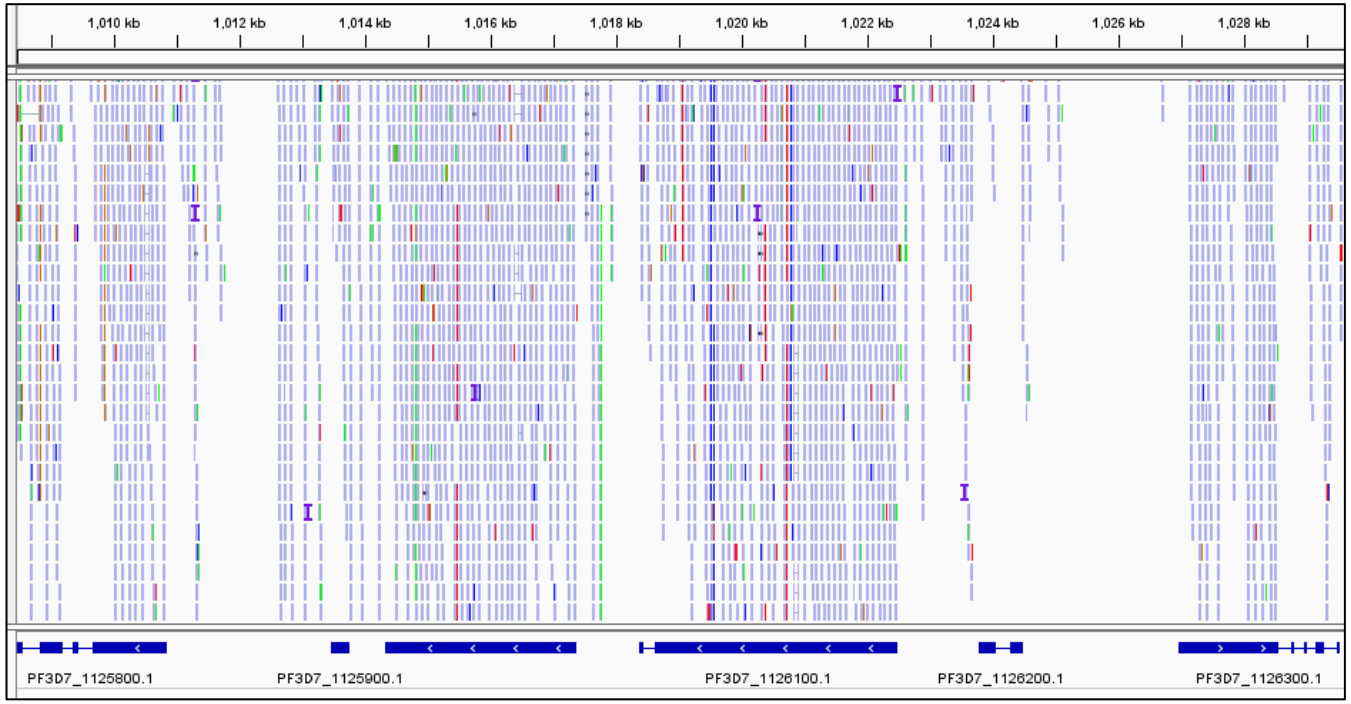

Figure 9: An example of a snapshot showing the read coverage. Reads mapped to the entire lengths of the transcripts.

The calibrated top line indicates a section of chromosome 11 of the *P. falciparum* genome. The majority light blue bars correspond to the reads that mapped correctly to the genome in a BAM file. Mismatched bases are shown in different colours: blue (cytosine), green (adenine), brown (guanine) and red (thymine). This purple bar I is the symbol for insertions. Deletions are indicated by a blank space crossed by a line. The bottom bars correspond to *P. falciparum* genes in the annotation GTF file, while spaces in between are the intergenic regions. A gap in a gene crossed by a line indicates an intron.

4.4 *P. falciparum* extracellular vesicle contain differentially enriched genes

To achieve the experimental goal of this thesis, the quasi-likelihood F test of *edgeR* [123] was applied to identify genes that are enriched in PfEVs relative to the parasite. This test was chosen because it reflects the uncertainty of estimating dispersion for each gene especially when the number of replicates is small. The 0 -12 h and 36 - 48 h time points were first individually compared with the respective Lopez-Barragan et.al., [116] whole

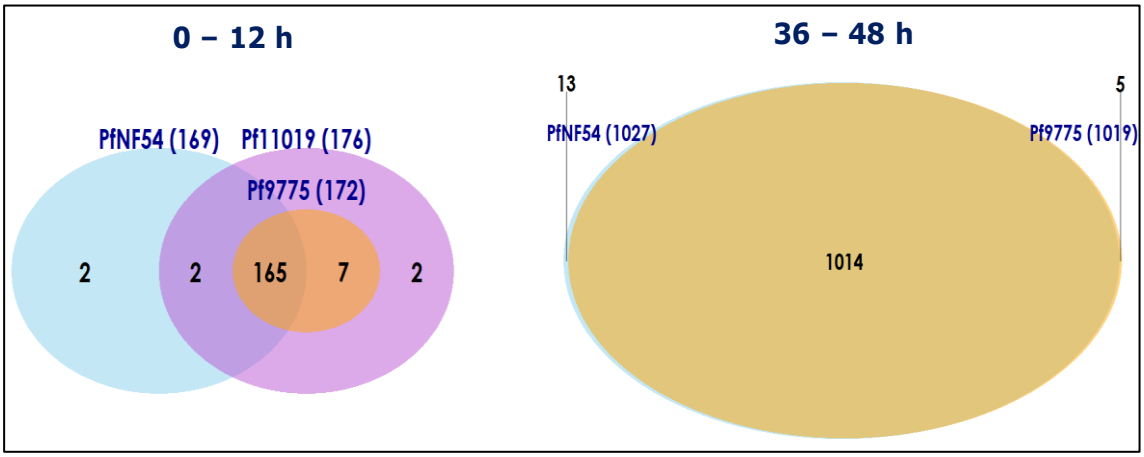

Figure 10: Parasite strain comparison of PfEV enriched genes (evDEGs) **a)** Comparison of 0 - 12 h evDEGs between strain PfNF54, Pf11019 and Pf9775 identified an overlap of 165 genes. **b)** Comparison of 36 -48 h evDEGs between PfNF54, and Pf9775 identified an overlap of 1014 genes. For each comparison, only genes with a P.adj < 0.05 and log2FC>5 were considered as evDEGs. The total number of evDEGs in each strain appears in parentheses

parasite transcriptome data using *edgeR*. This was done to provide some insights as to whether there is a difference between the transcriptome of PfEVs released by a long-term laboratory strain and more recent clinical isolates. These two time points were chosen due to the availability of at least two samples belonging to at least one strain, which is a requirement of *edgeR*. In this comparison, there was a substantial overlap of PfEV differentially enriched genes (evDEGs) across all the three isolates, as shown by the high number of transcripts within the intersection of the samples in **Figure 10**. The number of overlapping DEGs was higher in the 36 - 48 h PfEVs (1014 genes) than 0 -12 h (165 genes).

Reliable estimation of biological coefficient of variation (BCV) between conditions requires replicates [123]. Therefore, the observation that the RNA content of PfEVs from recent clinical isolates behave similarly to that of NF54, prompted the combination of samples from all strains for each individual parasite time window. This allowed me to treat the samples from a given time point as replicates, which allowed for a more for a more robust comparison. In the combined comparison, the following number of evDEGs were significantly enriched in PfEVs as compared to the parasite: 0 - 12 h (173 evDEGs), 12 -24 h (291 evDEGs), 24 - 36 h (135 evDEGs) and 36 - 48 h (958 evDEGs) as shown by **Figure 11** and **Figure 12**. This supports the notion that PfEVs released by late stage parasites contain a higher number of transcripts than those released by early stage parasites, just as late stage parasites contain more RNA than early stage parasites do. Although, there was an overlap of genes between the time points, a large proportion of evDEGs were time specific. Interestingly, only 56 genes (1014 minus 958) (**Figure 10** and **Figure 11**) detected as evDEGs during the individual comparison were not

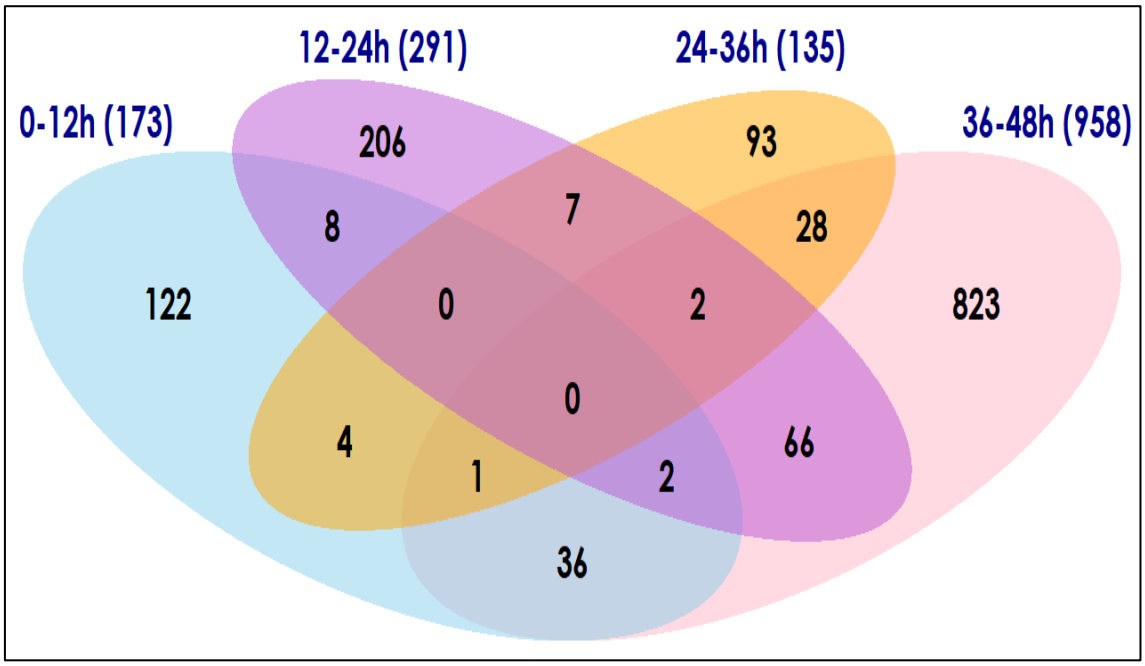

Figure 11: Overlapping of evDEGs across the asexual parasite cycle. Only genes with an adjusted P-value <0.05 and log2 (Fold Change)>5 were considered to be upregulated in PfEVs. The total number of evDEGs at each parasite time window appears in parentheses

differentially expressed when the 36 – 48 h samples were treated as replicates. This supports the theory that vesiculation could be a conserved phenomenon that may not vary much between parasite isolates.

Generally, transcripts expressing members of multi-copy families of proteins were highly upregulated in PfEVs in all time points. These include transcripts encoded by the *var, rif, phist, stevor and surfin* genes. However, there were also stage specific transcripts. For example, interesting genes enriched in 0 - 12 h PfEVs include a member of APiAP2 transcription factors (Pf3D7_1456000), sortilin and syntaxin (Qa-SNARE family). Gene products enriched in 12-24 h PfEVs include TATA-box binding protein (PF3D7_0506200), karyopherin, DNA repair protein RAD2, putative (PF3D7_0206000), kelch 13, WD repeat-containing protein 82 putative (PF3D7_1243800), and histone-lysine N-methyltransferase, H3 lysine-4 specific (PF3D7_1221000). Others include kelch 13, ABC transporter B, SNARE (SYN6) and cop-coated vesicle membrane protein p24 precursor (Pf3D7_1314500).

The PfEVs harvested between 24-36 h are rich in transcripts encoding early transcribed membrane proteins (ETRAMPs), the highly variable *Plasmodium* proteins: PfEMP1, SURFIN, RIFIN and PHISTb, as well as Maurer's cleft resident proteins (Pfmc-2TM Maurer's cleft two transmembrane proteins). Interestingly, an mRNA for a protein expressed by early sexual stages, gametocyte erythrocyte cytosolic (GECO) protein, HAD domain ookinete protein and gametocyte exported protein 20 (GEXP 20) had a higher fold change in 24 -36 h PfEVs relative to the parasite. Other interesting mRNA enriched in 24-36 h PfEVs are two non-coding RNA; signal recognition particle RNA

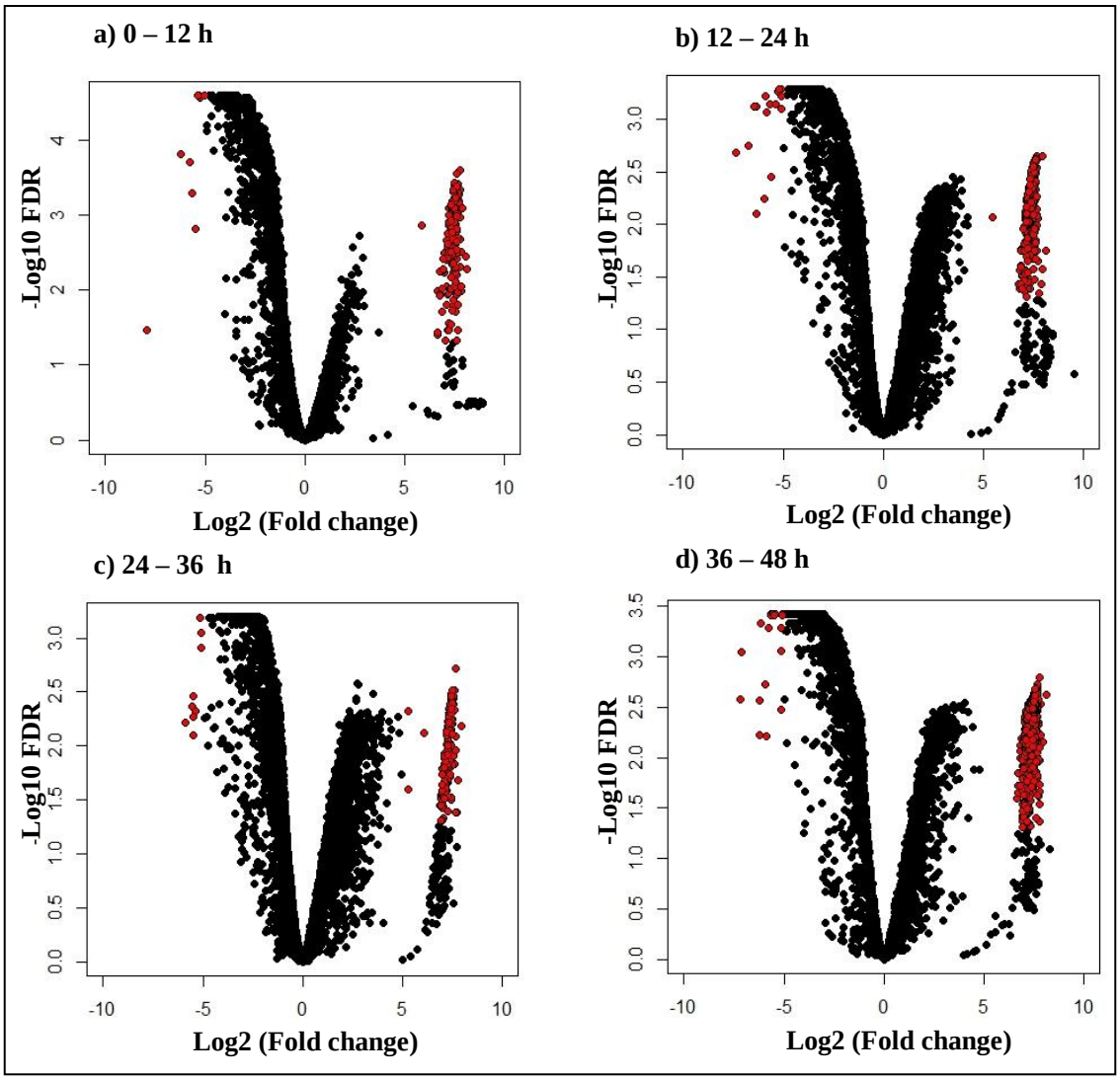

Figure 12: Volcano plots showing differentially enriched genes across the asexual parasite cycle. y-axis: negative log10 of FDR adjusted P value. x-axis: log2 (Fold change). Red shows significant genes while black shows non-significant genes. Genes with adjusted P value < 0.05 and an absolute log2Fold change threshold of 5 were termed significant.

and a member of *Plasmodium* RNA of unknown function (Pf3D7_0420800) as well as members of the exported protein family. Also enriched in 24-36 h PfEVs was multidrug resistance protein 1 (Pf3D7_0112200).

The gene products enriched in the 36-48 h PfEVs were mainly transcription factors (and associated enzymes) such as AP2 domain containing transcription factors (AP2-G, PF3D7_0420300, PF3D7_0802100, PF3D7_1222400, PF3D7_1456000, PF3D7_1449500), histone deacetylase (PF3D7_1472200), histone acetyltransferase

(PF3D7_0416400), repressor of RNA polymerase III transcription MAF1 (PF3D7_0416500), mRNA-binding protein PUF1 (PF3D7_0518700), translation initiation factors (PF3D7_0111800 and PF3D7_0716800), HCNGP-like protein, cyclins, zinc finger proteins and conserved *Plasmodium* proteins whose exact functions are not yet known. Other gametocyte-associated transcripts included: gamete antigen 27/25, GEXP10, GEXP17, gamete egress and traversal protein (GEST) and gametocytogenesis implicated protein (GIG). Another class of abundant gene products in 36-48 h PfEVs were transcripts that express the retromer complex commonly referred to as vacuolar protein sorting associated (VPS) proteins. Specifically, VPS2, VPS3 VPS4 and VPS9 were identified as highly abundant transcripts in 36-48 h PfEVs. Small nucleolar RNAs (snoR3, snoR6, snoR23, snoR24 and snoR27) and survival motor neuron-like protein (SMN) were also detected as 36-48 h evDEGs.

4.5 Gene ontology and network analysis

The 36-48 h PfEV transcriptome was highly enriched in mRNA encoding products involved in regulation of cellular processes and were assigned GO terms such as “regulation of macromolecule biosynthesis”, “regulation of gene expression”, “regulation of RNA biosynthesis”, among others (**Figure 13**). There were no overrepresented GO terms for the other three time windows. To explore the interaction between the GO terms, an enrichment map (**Figure 14**) and a category gene network plot (CGNP) (**Figure 15)** were constructed. These enrichment functional visualizations enabled mutually overlapping GO terms to cluster together. The upregulated GO terms for the 36 - 48 h comparison were organized in an intricate network involving all the 21 GO terms identified (**Figure 14**). This points out that parasites seem to pack functionally related gene products into PfEVs to perform a particular task.

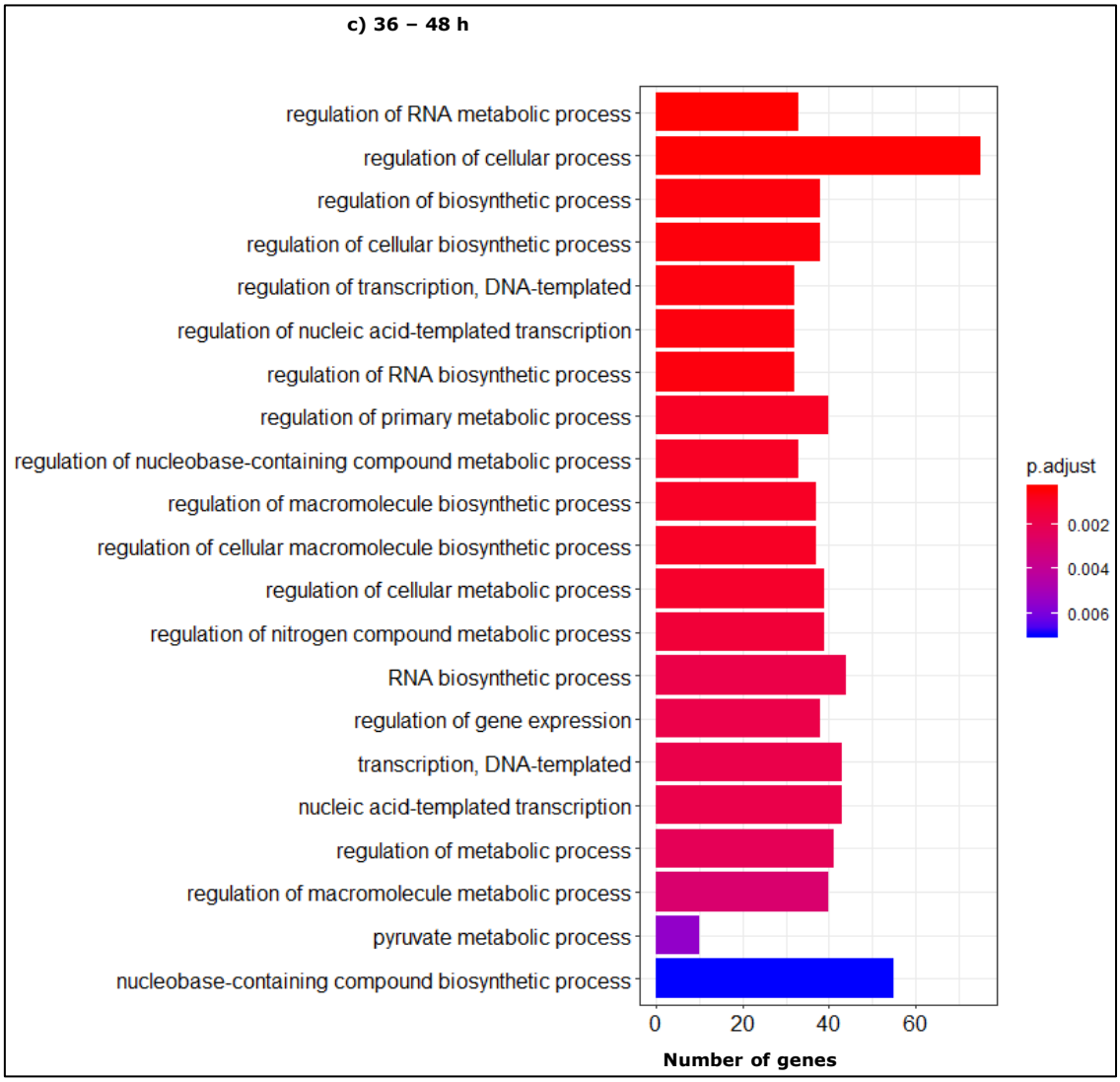

Figure 13: Enriched GO term analyses for the 36-48 h time window. GO terms with a P.adj value < 0.05 were considered significant. The red colour indicates GO terms with the smallest P.adj value with the increasing blue colour corresponding to larger adjusted P values

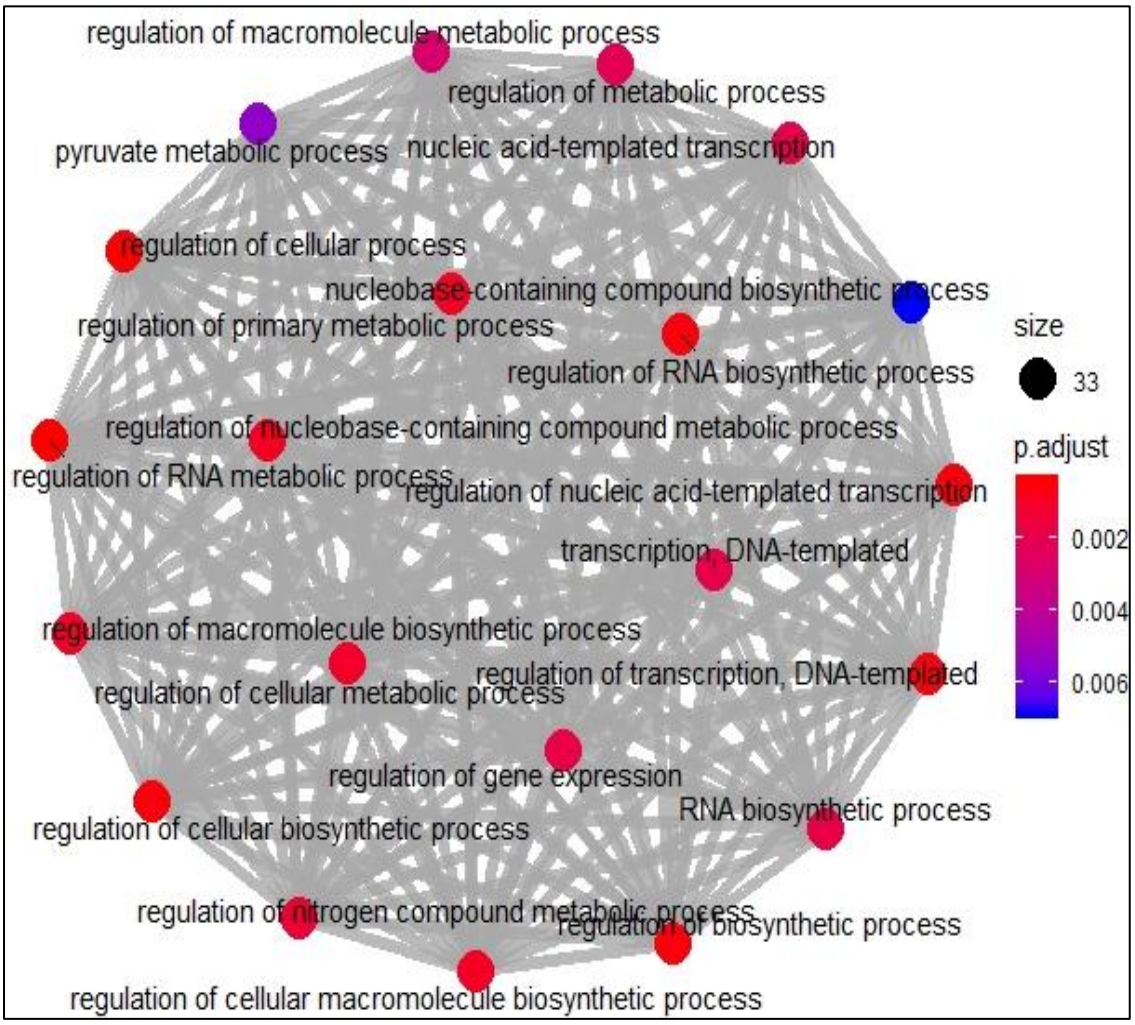

Figure 14: An enrichment map showing the interaction of GO terms in 36 - 48 h PfEVs.

The enrichment plot overcomes the problem of *P. falciparum* gene-set redundancy by grouping related GO terms into a similarity network, where nodes represent gene-sets, links represent the overlap of member genes, and node colour corresponds to the enrichment score.

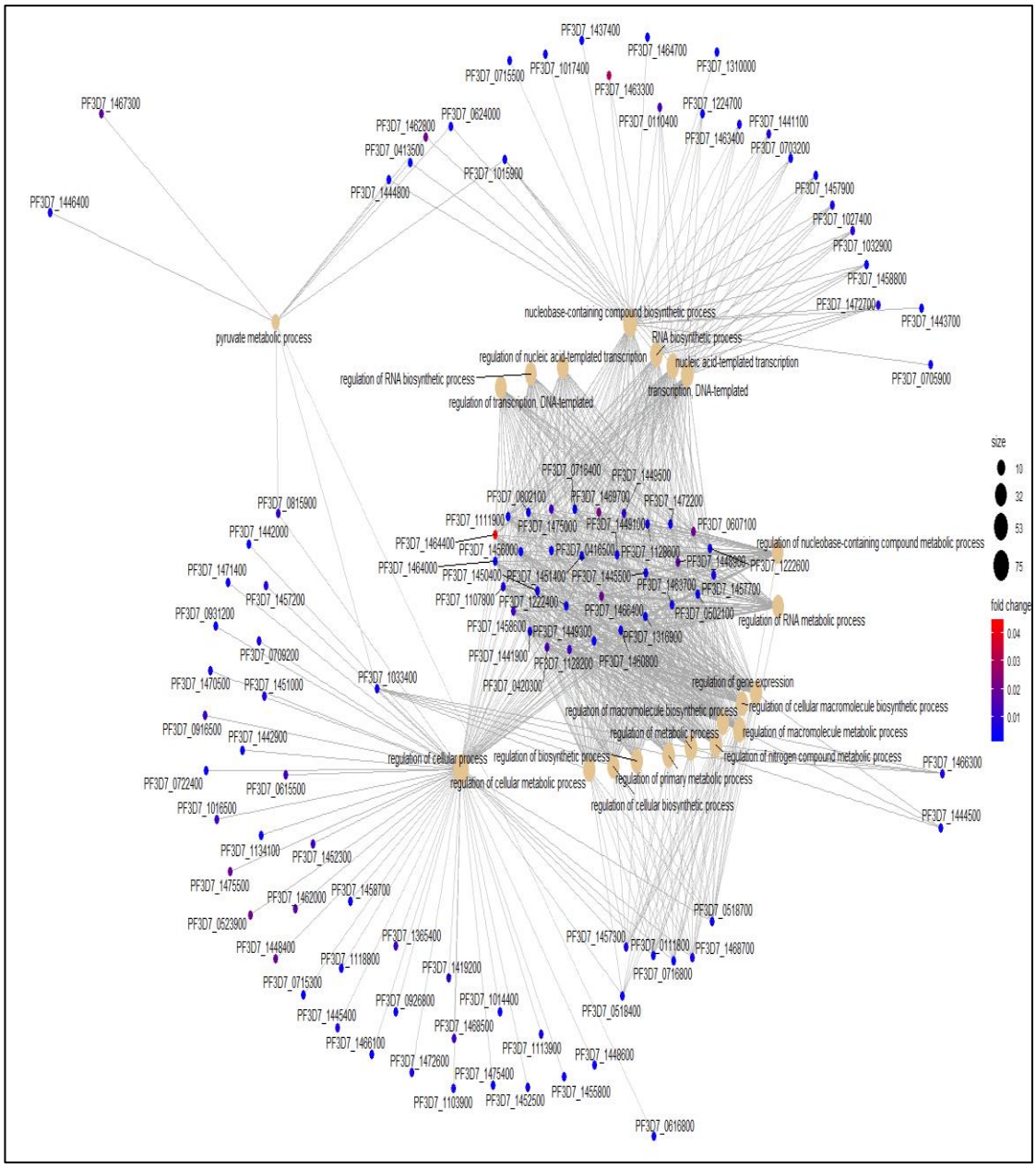

Figure 15: Category gene network for the 36-48 h comparison. The network shows that evDEGs released between the 24-36 h of the asexual parasite cycle can be grouped into three classes: **(i)**. genes exclusively involved in "Regulation of cellular processes", **(ii)**. genes exclusively involved in "nucleic acid biosynthesis" **(iii)**. and genes involved in both processes. Blue dots indicate genes assigned the smallest adjusted P-values while red dots correspond to larger adjusted P values. Yellow circles correspond to the GO terms.