Chapter 4
Functional consequences of perturbing vitamin B6 biosynthesis in *P. falciparum* parasites.

4.1 Introduction

4.1.1 Transcriptomics in the malaria parasite

Integrated systems biology is becoming increasingly essential in describing the functional organisation of drug-induced changes by integrating X-omics approaches; genomics, transcriptomics, proteomics, metabolomics and bioinformatics [222, 223]. Transcriptomics aims to describe the dynamics of transcript expression and co-expression, as well as quantities of the transcripts, which consists of mRNA, rRNA and non-coding RNA components [223]. Currently transcriptomics provides the most comprehensive view of changes within the cell, and provides insight into the functions and dynamics of the expressed genome [224]. Microarrays, serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS), quantitative reverse transcriptase (RT)-PCR (qRT-PCR) and whole transcriptome shotgun sequencing (WTSS or RNA-seq) are some transcript profiling methodologies [224]. SAGE, MPSS, as well as RNA-seq are methodologies that do not require prior knowledge of sequence information, and are useful approaches on non-model organisms which have not been sequenced. Microarrays are more affordable, high throughput and can interrogate transcriptional changes in the whole genome, however rely on prior sequence information to design oligos [224]. The genome of *P. falciparum* has been sequenced and revealed that it consisted of 5400 protein-encoding genes, of which 60% were found to be transcriptionally active during the IDC [225]. This has led to the implementation of a large number of microarray profiling studies to mine information regarding the *P. falciparum* transcriptome.

Microarray-based transcriptomic studies on *P. falciparum* parasites have aided in understanding differential transcript expression in diverse life-stage developmental forms [225-228], in distinct parasites strains [229], as well as transcript copy number variations and polymorphisms in different laboratory isolates [230]. Different pathogenicity factors in field isolates [230, 231], as well expression of transcripts potentially involved in drug transport in drug-resistant strains [232] have been elucidated using microarray studies. Microarrays have been used to correlate *in vivo* parasite proliferation to *in vitro* culturing conditions [233], and the effect of different environmental factors on *in vitro* cultured parasites [234, 235]. Some important features identified include monocistronic transcript expression patterns of most chromosomal genes during the IDC [225]. These transcripts

reach maximal expression only once during the 48 h cycle, whereas apicoplast gene expression patterns were shown to be polycistronic [225].

A large number of microarray studies on *P. falciparum* have focused on elucidating transcriptomic responses due to drug pressure as well as drug resistance mechanisms [207, 236-241]. With this information the compensatory transcriptome adaptations due to chemical knock-down of the target could be elucidated, which should aid in creating more effective drugs, as well as drug combinations. Decreased transcript expression during drug treatment could suggest that these processes potentiate the activity of a compound. A hypothetical example is decreased transcription of membrane transporters that facilitate import of the active compound. Increased expression of transcripts could facilitate removal of the drug, as has been demonstrated for mefloquine resistant parasites which have increased the *P. falciparum* MDR transporter gene copy numbers [77].

Studies have shown that the *P. falciparum* transcriptome is dynamic, and that different small molecules have profoundly different effects on transcript expression patterns. Hu *et al.* constructed interaction networks or gene-function prediction networks based on global expression patterns from a wide range of different small molecule growth inhibitors [242]. These different molecules stimulated distinctly different sub-populations of transcripts, and provided insights into the MOA of similar classes of inhibitors. Interesting small molecules such as apicidin and trichostatin A – both histone deactelyase inhibitors (HDAC inhibitors) – with different chemical scaffolds were shown to stimulate very similar transcript expression profiles. Other molecules such as quinine and CQ induced different transcription patterns compared to the former compounds, but were similar compared to one another. It was also elegantly shown with microarrays that the mechanism of drug resistance against DHA involves increased *pfmdr*1 transcript expression, confirming that this transporter is involved in conferring resistance to the drug [238]. The delayed cytotoxic responses of tetracyclines such as doxycycline on parasites were elucidated using microarrays and attributed to impaired apicoplast gene transcription [207]. A microarray study aimed at determining the MOA of a potent piperazine compound ACT-213615 showed that the drug induced differential expression of around 550 genes in late schizonts [237]. Even though the definitive MOA could not be determined, it was shown that the compound has a distinctive transcriptomic responses apart from artemisinin or CQ [237].

In some cases, it has been shown that the transcriptome of parasites is relatively unresponsive due to drug treatment. The antifolate WR99210 drug which specifically targets dihydrofolate reductase-thymidylate synthase (DHFR-TS) did not affect the target transcriptional expression [241]. The folate cofactor, tetrahydrofolate (THF), which is recycled by DHFR-TS, is essential for production

of nucleobases and inhibition of this process could have resulted in a paralysed state. In CQ-treated parasites 600 transcripts were differentially expressed, of which a only small percentage (~3%) had fold change (FC) values above 4 [240]. The authors failed to identify clear CQ-responsive networks linked to low level transcripts expression, however suggested that small changes in transcript abundances could have profound synergistic effects on cell physiology. The lack of differential transcript expression of the known targets in each case led the authors to speculate that the transcriptome is rigid and unresponsive. In contrast, this could provide clues into the mechanism of action of CQ; the drug might affect a master transcription factor and thereby rendering parasites unable to differentially express transcripts, or the toxic accumulation of heme induces a similar arrest in transcription.

In this chapter, a functional genomic study was undertaken to determine the global effects of treatment of the parasite with 4PEHz, and whether inhibition of *Pf*Pdx1 affects processes related to PLP metabolism. The compound was shown to have some specificity for *Pf*Pdx1, and inhibition of protein activity is expected to disrupt metabolic processes which could lead to alterations in transcript expression patterns. Additionally, proteins from 4PEHz-treated parasites were extracted to determine the extent of translational and post-translational effects caused by the anticipated PLP-depleted state in the parasites.

4.2 Methods

4.2.1 Parasite culturing and drug treatment

P. falciparum (3D7) parasites were maintained under aseptic conditions in continuous culture according to Trager *et al.* [201], as previously indicated by Smit *et al.*[243]. The RMPI 1640 culture medium contained 25 mM HEPES, 10 mM glucose, 20 mM sodium bicarbonate, 25 mM hypoxanthine and 0.5% w/v AlbuMAX II (Invitrogen) at pH 7.4. Parasite cultures were maintained at 5% haematocrit using human O^+ erythrocytes in 75 cm^3 flasks (Nunc) with shaking at 200 rpm at 37°C. Before drug treatments, the synchronicity of the parasite life-stages was controlled by three consecutive D-sorbitol (5% w/v in 1x PBS) treatments, each 8 h apart in early and late ring-stage parasite cultures. Briefly, parasite cultures (50 mL, 1-5% parasitaemia, ring-stage parasites) were collected by centrifugation at 3000*g* for 5 min, supernatants discarded and pellets re-suspended in 5 volumes (10 mL) of 5% D-sorbitol and incubated at RT for 10 min. The cultures were centrifuged as before and the pellets were re-suspended in complete culture media. Flasks were deoxygenated using 5% CO_2, 5% O_2 and 90% N_2 and kept in a shaking incubator (120 rpm) at 37°C.

The highly synchronous cultures (approx. 95% schizonts) were dived into treated (T) and untreated (UT) groups at 10% parasitaemia, 5% haematocrit. The treated group was treated with 20 µM 4PEHz (equivalent to 2 x IC_{50}) at the start of the ring stages (defined as 0 hours post invasion, HPI). After 24 h, the media was replaced with fresh media, maintaining drug pressure for the 4PEHz treatments. Sampling was done by harvesting infected erythrocytes at three different time points each 12 h apart (t_1 = 12 HPI, t_2 = 24 HPI and t_3 = 36 HPI) corresponding to rings, trophozoites and schizont forms of the parasite. This entailed removing 10 mL culture from the flasks and washing the harvested infected erythrocytes twice with 10 pellet volumes of cold 1x PBS as above to remove any remaining drug or culture media components. The washed infected erythrocyte pellets were flash frozen in liquid nitrogen and subsequently stored at -70°C. Additional 5 mL culture samples were harvested from both T and UT groups of each time point for proteomic analyses. Experiments were performed in duplicate, with two technical replicates for each time point.

4.2.2 Transcriptomics

4.2.2.1 RNA isolation

RNA was isolated from infected erythrocytes using Tri-reagent (Sigma-Aldrich) combined with RNeasy mini kit (Qiagen) according to the manufacturer's protocol for isolation of total RNA from animal tissues using on-column DNase-1 digestion (Qiagen), as reported previously [243, 244].

Frozen infected erythrocyte pellets were thawed at RT and gently agitated to facilitate thawing. The cells were lysed through the addition of 600 µL lysis buffer (RNAeasy kit, Qiagen) and agitated by vortexing. The lysed cell mixture was pipetted onto a QIA-shredder column, and the eluent collected after centrifugation at 13 000*g* for 2 min was separated into two equal parts in microcentrifuge tubes. Tri-reagent (Sigma-Aldrich), measuring to 600 µL, was added to each tube, followed by incubation at RT for 5 min. Tri-reagent is a RNA isolation agent that contains guanidinium thiocyanate and phenol that denatures proteins and inhibits RNAses, and following addition of chloroform allows separation of mostly RNA into upper aqueous phases, originally developed by Chomczynski *et al.* [245]. Chloroform (400 µL, Merck) was added to the tubes and these vigorously vortexed for 15 s. The chloroform-containing tubes were incubated at RT for 10 min, centrifuged at 13 000*g* for 15 min and the upper aqueous phases transferred to a clean microcentrifuge tubes. Ethanol (700 µL of 70% v/v, Merck) was added to each tube to precipitate RNA, and the contents of the previously split tubes combined into one tube. The solution (700 µL) was consecutively loaded onto a RNAeasy spin column and centrifuged for 8000*g* for 15 sec until all of the solution was passed through the column. The spin column was washed by adding 350 µL wash buffer RW1 (Qiagen) followed by centrifugation as before. On column DNase-1 digestion was performed with 80 µL DNase-1 solution in buffer RDD (Qiagen, 10 Kunitz units) at RT for 15 min. The digested genomic DNA was washed from the columns using 350 µL wash buffer RW1, followed by a final 2x washing in 500 µL buffer RPE (Qiagen) and drying of the columns by centrifugation at 8000*g* for 2 min to ensure complete removal of remaining liquid. RNA was collected by adding 30 µL of RNase-free water (RNAeasy mini kit, Qiagen) allowing incubation for 2 min at RT, followed by centrifugation at 8000*g* for 2 min, repeated twice. The collected RNA was stored at -70°C.

4.2.2.2 Assessment of RNA purity and quality

RNA was quantified using Nanodrop® ND-1000 (*NanoDrop* Technologies), which uses the UV absorption of the sample at 260 nm, together with the molecular extinction coefficient of RNA, to calculate the RNA concentration. The quality and purity of the RNA was assessed using the Experion automated electrophoresis system (Bio-Rad), which involves; separation, staining and destaining, RNA band detection, imaging and data analysis [246]. RNA (200 ng) was loaded onto the Experion™ RNA StdSens Lab-Chip microfluidic chip (Bio-Rad) according to the manufacturers recommendations. Loaded samples are labelled with a fluorescent dye enabling the detection of the RNA through laser-induced fluorescence. RNA quality is qualitatively and

quantitatively assessed, the former reported as the RNA quality indicator (RQI) which determined by analysing the integrity of the ribosomal RNA, the latter determined by comparison to an internal molecular weight standard.

4.2.2.3 cDNA synthesis

First strand cDNA synthesis was performed by mixing 4 µg RNA, 250 pmol OligodT primer ($dT_{23}VN$, Inqaba Biotechnical Industries) and 775 pmol random nonamer (N_9, Inqaba Biotechnical Industries) which was incubated at 70°C for 10 min, followed by incubation at 4°C for 10 min. A 10× aminoallyl dNTP stock was prepared with final concentrations of 5 mM 5-(3-aminoallyl)-2ˈ-deoxyuridine-5ˈ-triphosphate (aa-dUTP), 5 mM dTTP, 5 mM dGTP, 5 mM dCTP, 10 mM dATP. After incubation of the RNA-containing tubes 6 µL 5× Superscript first strand reaction buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$), 10 mM DTT, 40 U RNasin (40 U/µL, natural protein-based RNase inhibitor, Promega), 1.7 µL 10× aminoallyl dNTP stock (0.28 mM aa-dUTP, dTTP, dGTP, dCTP and 0.56 mM dATP) and 340 U Superscript III (Invitrogen) was added, to a final volume of 30 µL with RNase-free water. The reverse transcriptase reactions were incubated at 42°C for 20 h. Remaining RNA was hydrolysed by adding a final concentration of 142 mM NaOH (Fluka) and 71 mM EDTA (pH 8.0, Merck), followed by incubation at 65°C for 15 min. The synthesised cDNA was subjected to a clean-up procedure to remove hydrolysed RNA, using the PCR Clean-up kit (Macherey-Nagel). Briefly, 400 µL buffer NT was added to the cDNA-containing tubes, after which the sample was transferred to the Nucleospin extract II spin column and incubated at RT for 4 min. Under these chaotropic salt conditions the cDNA selectively binds to the silica matrix, allowing removal of any unbound material. The spin columns were centrifuged at 12 000*g* for 2 min, after which the membrane was washed using 500 µL buffer NT, followed by centrifugation as above. Columns were dried with a final centrifugation step at 12000*g* for 2 min. Pre-heated RNase-free water (30 µL) was pipetted onto the membranes, allowing incubation for 4 min at RT, followed by centrifugation at 12000*g* for 90 s. The cDNA concentration and purity was determined using a Nanodrop® ND-1000 (*NanoDrop* Technologies), which uses the UV absorption of the sample at 280 nm, together with the molecular extinction coefficient of DNA, to calculate the DNA concentration. A cDNA reference pool comprising of all the time points was created by combining 1.5 µg cDNA from each of the 24 samples, this was mixed and split into 24 aliquots.

4.2.2.4 cDNA labelling

cDNA (1.5 µg) of each sample, together with 1.5 µg reference pool samples, were dried *in vacuo*, after which the samples were resuspended in 2.5 µL water (SABAX, Adcock-Ingram). The dyes were prepared by adding 12 µL dimethyl sulphoxide (DMSO, 99%, Sigma-Aldrich) to each Cy3 and Cy5 (GE Healtcare) dyes. Reference pool samples were labelled with Cy3, whereas Cy5 dye was used to label the individual cDNA samples. The dissolved dyes (2.5 µL) were diluted in 100 mM freshly prepared Na_2CO_3 (pH 9.0) and added to resuspended cDNA samples. The reaction mixtures were gently mixed, and incubated at RT in a desiccator for 2 h. Excess dye was removed using the QIAquick PCR purification kit (Qiagen) by adding 100 µL of buffer PBI (Proprietary, Qiagen) to the labelling mixture, after which the sample was loaded onto the QIAquick spin column and allowed to incubate at RT for 4 min. The spin columns were centrifuged at 12 000*g* for 1 min followed by three washing steps comprised of the addition of 500 µL buffer PE (Proprietary, Qiagen), followed by centrifugation at 12000*g* for 2 min, discarding the eluent in each case. Spin column membranes were thoroughly dried by centrifugation at 12 000*g* for 2 min. Labelled cDNA was eluted from the column by applying 30 µL water, allowing incubation at RT for 4 min, followed by centrifugation at 12000*g* for 90 sec, collecting the eluted liquid. Dye incorporation efficiency was determined using a Nanodrop® ND-1000 (*NanoDrop* Technologies), where the efficiency was calculated using equation 4.1, in which more than 10 labelled nucleotide per 1000 nucleotides was required to proceed.

Eq. 4.1 $$\text{Efficiency} = \frac{\text{pmol dye} \times 324.5\ \text{pg/mol}^*}{\text{ng DNA}}$$

* average mass of one dNTP

4.2.2.5 Sample fragmentation and slide oligonucleotide hybridization

Individual Cy5 labelled cDNA samples (20 pmol) were combined with 20 pmol of the Cy3 labelled reference pool cDNA sample, together with 5 µL 10× blocking buffer (Proprietary, Agilent), 1 µL 25× fragmentation buffer (Proprietary, Agilent) to a final 25 µL volume. The tubes were incubated at 60°C for 30 min, after which 25 µL 2× GE hybridization buffer was gently added to ensure no unwanted foaming occurred, and stored on ice. The arrays consisted of 8 chambers into which a total 20 pmol of each Cy5 and Cy3-labelled cDNA was loaded into each chamber. Samples were loaded into assembled hybridization chambers in the gasket slide, after which the microarray oligonucleotide slide (Agilent, with 'Agilent' side down), was carefully placed on top. The slide

chamber assembly was tightened and placed in a hybridization oven at 65°C for 18 h on rotational speed 10, making sure of liquid spreading evenly over the slide.

4.2.2.6 Hybridization clean-up

Hybridized slides were removed from assembly chambers and carefully disassembled. The slides were washed twice in wash buffer 1 (Proprietary, Agilent) for 1 min, followed by washing in preheated (37°C) wash buffer 2 (Proprietary, Agilent) for 1 min. Slides were dried by centrifugation using an Agilent desktop centrifuge, after which slides were scanned on a Axon GenePix 4000B scanner (Molecular devices).

4.2.2.7 Data processing

An improved 60-mer Agilent platform was created, partly adapted from the 70-mer Operon Array containing 8089 *P. falciparum* genes as described by Smit *et al.* [243] At the time, the most recent annotated *P. falciparum* (strain 3D7) genome from PlasmoDB 5.4 (www.plasmodb.org) was used to design the 60-mer based array. After removal of unnecessary "NULL" Operon –specific controls the remaining 7004 oligonucleotides were shorted to 60 nucleotides by using a 10 nucleotide scanning window from either the 3'- or 5'-ends, whilst ensuring that the microarray annealing temperature (Tm) was maintained close to 65°C. Equation 4.2 was used to calculate the Tm.

Eq. 4.2 $$\mathrm{Tm} = 64.9°\mathrm{C} + 41°\mathrm{C}(\mathrm{GC}\text{-}16.4)/\mathrm{N}$$

in which GC corresponds to the number of G and C nucleotides in the target sequence, where N is the total length of the sequence.

The 60-mer sequences were validated using NBLAST (www.ncbi.nlm.nih.gov/BLAST) from which all the submitted sequences had target E-values below 10^{-6}. Additionally, the 60-mer sequences were submitted, as FASTA file, into the ArrayOligoSelector tool (http://arrayoligosel.sourceforge.net/) which optimizes oligo selection based on sequence complexity, uniqueness and lack of self-complementarity. The redesigned sequences were again validated using NBLAST and submitted for printing onto Agilent slides. Arrays were interpreted using the Axon GenePix 6.0 software (Molecular devices). Automated spot detection was performed to ensure correct spot quality, saturation and signal-to-noise ratios using criteria as listed in Table 4.1.

Table 4.1: Parameters used for spot flagging and assessment in GenePix.

Parameter [a]	Function	Flag	Average percentage of flagged spots
Saturation [b]	[F532 % Sat.] > 20 Or [F635 % Sat.] > 20	Bad	2.8 – 4.6 %
Signal-to-noise (S/N) [c]	[SNR 532] < 3 And [SNR 635] < 3	Bad	29 - 32 %

[a] Flagging of spots based on circularity and intensity was not performed due to the low number of positive flagging events. [b] Spots saturation was evaluated in both foreground red channels (532 nm) and green channels (635 nm) and flagged as bad, therefore were not used, if the saturation level was above 20. [c] Spots with S/N levels below 3 were removed from arrays by this function.

For simplicity, the Cy5-red and Cy3-green intensity values from an array spot is expressed as A and M values, where M-values represents difference in average log intensity of spots in the array (Eq. 4.3). For a particular spot, if the red channel value is greater than the green, this results in a positive M-value which equates to greater levels of red-labelled cDNA, or in this case Cy5 labelled sample, compared to green-labelled Cy3 reference cDNA. The $\log_2$ ratios of the individual colour channels are used to relate expression in terms of M, as listed in equation 4.3.

Eq. 4.3 $$M = \log_2 R - \log_2 G$$

$$= \log_2 \left(\frac{R}{G}\right)$$

Where R is the intensity of in the red channel at 532 nm for the Cy5 dye and G is the spot intensity in the green channel at 635 nm for the Cy3 dye. The M-value also represents the $\log_2$ fold change ($\log_2$FC).

The A-values represents the log intensity in the colour channel for a particular spot, and is calculated by the average log intensities of both red and green channel intensities [247]. The average intensity of a spot was interpreted as the A-value using equation 4.4, which

Eq. 4.4 $$A = \frac{(\log_2 R + \log_2 G)}{2}$$

In which R is the intensity of in the red channel and G is the spot intensity in the green channel.

The LIMMA (Linear models for microarray data, [248], www.bioconductor.com) and LIMMA-GUI packages were used to analyse the gene expression microarray data. The background from all the arrays was corrected using a subset of 50 [249]. For values of M the assumption is made that the majority of genes are unaffected by the drug treatment and therefore plots of M are expected to be centred around zero [247]. With both M- and A-values, MA plot can be generated, and represent the distribution of spot data. This is useful to compare between-array M-values and determines the

extent of normalisation which is required. Within-array and between-array comparisons of both A- and M-values are necessary for accurate estimations of expression levels and consider both intrinsic array variations as well as biological differences between samples from separate arrays. Within-array normalisation methods include Loess and Robust spline normalisation, the former method fits a normalisation curve using local linear regression, the latter employs Bayes statistics together with robust fitting of regression splines to normalise the data [250, 251]. Gquantile is a between-array normalisation technique which scales the quartiles of the data [250]. Robust-spline normalisation was performed within each array, after which Gquantile normalisation was performed between the arrays which incorporates the use of the Cy3-labelled cDNA reference pool [252].

Differently expressed transcripts were identified by firstly comparing UT arrays with T array using Pearson correlations, after which between-array replicate spots were correlated using linear modelling (lmFit) which makes use of empirical Bayes methods [253]. Transcripts with a *P*-value smaller than 0.05, with the $\log_2$ ratio greater or equal to 0.75 or smaller and equal to -0.75 were considered significant ($\log_2$ ratio $\geq$ 0.75 and $\log_2$ ratio $\leq$ -0.75). The $\log_2$FC for differential expression in T vs. UT comparisons was calculated by subtracting the $\log_2$ (Cy5/Cy3) in the treated sample from the $\log_2$ (Cy5/Cy3) of the untreated sample. The fold change (FC) could be calculated using the antilog$_2$ of the $\log_2$ ratio, according to Eq. 4.5. The FC significance level was therefore greater and equal to 1.68 or smaller and equal to -1.68.

Eq. 4.5 $$\mathrm{FC} = 2^{\log_2 \mathrm{ratio}} \; or \; 2^{M}$$

The log-odds value or B-statistic, a statistical descriptor generated during array data comparison, refers to the probability that the gene of interest was differentially expressed, where a B-statistic of 0 corresponds to a 50-50 chance that the gene is differently expressed. Rankings of expression data using the B-statistic results in the same order if the data are ranked according to *P*-value (or the adjusted *P*-value). Some transcripts were represented by two or more spots in the arrays, and accounts for alternatively spliced variants of the same mRNA. In the case were duplicate identifiers were present, expression values from spots with the greatest B-statistic, which corresponds to the smallest adjusted *P*-value, were used.

Gene ontology (GO) terms of the differentially expressed transcripts were obtained from PlasmoDB 8.1 (www.plasmodb.org) and the transcripts were arranged according to GO biological function term. Grouping was verified using Micro Array Data Interface for Biological Annotation (MADIBA) [254] (http://www.bi.up.ac.za/MADIBA/) and DAVID (http://david.abcc.ncifcrf.gov/). Hierarchical clustering, by average linkage, was performed using Cluster [255] (http://rana.lbl.gov/EisenSoftware.htm). Gene set enrichment analysis (GSEA) was performed

using gene set enrichment analysis, GSEA software, and Molecular Signature Database (MSigDB) (http://www.broad.mit.edu/gsea/) [256]. Normalised M-values from the Tt_3 and UTt_3 array data were compared in a gene set permutation type analysis. The number of permutation was 1000, with a weighted enrichment statistic and signal-to-noise metric for ranking genes.

4.2.2.8 Real-time quantitative RT-PCR (qRT-PCR)

Real-time qRT-PCR was used to verify gene expression levels in UT and 4PEHz-treated parasites from the microarray study. The cDNA from T and UT samples of t_3 were diluted to 1.5 ng/µL with DNase-free dddH_2O (Invitrogen). A dilution series was created from pooled UT samples corresponding to 1/10, 1/20, 1/50 and 1/100 dilutions. The housekeeping transcripts; lactate dehydrogenase (LDH,) cyclophillin and seryl-tRNA synthetase (StSyn) were used to create the standard curve. Additionally, standard curves of the six target transcripts were created using the same dilution series of UT pooled cDNA at 1.5 ng/µL starting concentration (Table 4.2). Reactions were set up in a 384-well plate (Roche) consisting of 5× KAPA SYBR Fast qPCR (Kapa Biosystems), 3.75 ng cDNA and 5 pmol of each forward and reverse primer, as listed in Table 4.2,in a final 10 µL reaction volume. The plates were amplified and analysed in a Lightcycler 480 (Roche). The plates were incubated at 95°C for 10 min, followed by 48 amplification cycles consisting of 95°C for 10 s, annealing at 53°C for 5s, and extension at 72°C for 7s. A melting curve analysis was performed to verify whether primer dimers were present. Results were analysed using QBase Plus (Biogazelle).

Table 4.2: **Real-time qRT-PCR primers used to validate microarray data**.

PlasmoDB ID	Description	5' - 3' sequence
PFE0505w Cyclophilin	CycloF	AATTCTTTGACCATCTTAATCATTC
	CycloR	CAAAACAATTTTACTTCCTTGGGTTA
PF13_0141 Lactate dehydrogenase (LDH)	LDHF	GATTTGGCTGGAGCAGATATGTA
	LDHR	CAACAATAATAAAAGCATTTGGACAA
PF07_0073 Seryl-tRNA synthetase (StSyn)	StSynF	TTCGGCACATTCTTCCATAA
	StSynR	AAGTAGCAGGTCATCGTGGT
PF14_0224 (serine/threonine protein phosphatase)	PF14_0224F	AGTGCAGATACAGCTAATTAT
	PF14_0224R	TTGCCTTGACAAAATCGTTAT
PFB0695c (acyl-CoA synthetase, PfACS8)	PFB0695cF	GTCAAAAGGAACGCACTAA
	PFB0695cR	TTTTCTTGAGAACGTGATT
PFI1370c phosphatidylserine decarboxylase	PFI1370cF	AGCCATCAGTGCATATAAT
	PFI1370cR	ATTATGGATGAGCCTACTTTA
PF11_0256 pyruvate dehydrogenase E1 alpha subunit	PF11_0256F	TTACAACAAGCAGAATTAGAT
	PF11_0256R	TCATTAGAAATGTCGTCAAAT
PF14_0155 serine C-palmitoyltransferase, putative	PF14_0155F	AATGCAATACAAGCTGCTATG
	PF14_0155R	CAAACAGCCAAAATAGAATCA
PF08_0066 lipoamide dehydrogenase	PF08_0066F	AAATAATTTGAGTAAACCCATA
	PF08_0066R	AATTTCAACACCGACATTATC

4.2.3 Proteomics

4.2.3.1 Protein extraction

Additional UTt_3 and Tt_3 parasite samples (5 mL) taken at the same timepoint as for the transcirptome analyses were used for proteomic analyses. The infected erythrocyte pellets (250 µL) were thawed at RT, and re-suspended in 5 mL PBS. Parasites were isolated from the erythrocytes using saponin lysis by adding 5 µL 10% w/v saponin (Merck) followed by incubation on ice for 5 min. The samples were centrifuged at 4500*g* for 15 min at 4°C and the red supernatant was discarded. The remaining pellets were washed two consecutive times using 1 mL PBS to re-suspend the pellet and centrifuging at 15 000*g* for 2 min at 4°C. The remaining pellets were lysed using 500 µL lysis buffer consisting of 8 M urea, 2 M thiourea, 2% CHAPS, 0.5% w/v freshly prepared DTT. Samples were sonified seven times using a sonifier (Branson 450) with microtip at 20% duty cycle for 1 min with pulsing, in each case immediately halting once sample frothing was observed, with 1 min resting intervals on ice. Samples were clarified by centrifugation at 16 000*g* for 30 min at 4°C, and the remaining supernatant containing the soluble protein fraction was used in downstream

processes. The 2-D quant kit (GE Healthcare) was used to quantify the parasite protein samples, according to manufacturer's instructions.

4.2.3.2 Polyacrylamide gel electrophoresis

SDS-PAGE was performed as previously established by Laemmli *et al.* with modifications [198]. Electrophoresis was performed in Biometra Minigel protein electrophoresis housings (Biometra) with glass assembly plates. Protein samples were denatured in sample loading buffer (200 mM Tris-Cl, pH 6.8, 8% SDS, 40% glycerine, 0.02% v/v β-mercaptoethanol, 0.05% w/v bromophenol blue) by adding 2 - 18 µL of sample loading dye and boiling samples for 5 min. A 10% separating gel was prepared by adding 1.5 M Tris-Cl pH 8.7, bis-acrylamide (29.1% acrylamide, 0.9% bis-acrylamide), 0.1% SDS, 10% APS, and 33.5 nM (5µL) TEMED. The stacking gel component at 5% consisted of Tris-Cl, pH 6.7, Bis-acrylamide, 0.1% SDS, 10% APS, 33.5 nM (5 µL) TEMED. Electrophoresis buffer consisted of 25 mM Tris-Cl, 192 mM Glycine, and 0.1% w/v SDS. Loaded protein samples were run at 120V. SDS-PAGE gels were stained with Coomassie, containing 0.3 mM Coomassie R-250, 40% v/v MeOH, 7% v/v glacial acetic acid. Destaining was performed using a solution containing 40% v/v MeOH and 7% v/v glacial acetic acid.

4.2.3.3 MS Proteomics

Excised lanes from the SDS-PAGE gel was submitted to Dr Salome Smit at the Proteomics laboratory, Central Analytical Facility, Tygerberg Campus, University of Stellenbosch. Briefly, the proteins were extracted from the acrylamide gel and digested with trypsin. The digested protein samples were subjected to mass spectrometric analysis using Thermo Scientific EASY-nLC II connected to a LTQ OrbitrapVelos mass spectrometer (Thermo Scientific) with nano-electropsray source. The Thermo Proteome Discoverer 1.3 software program (PD 1.3, Thermo Scientific) was used to identify proteins via automated database searching using Mascot (Matrix Science) and Sequest (Yates lab, University of Washington) of all tandem mass spectra against the PlasmoDB 7.2 and SwissProt_57.15 databases. Proteins were considered positive matches at least 2 tryptic peptides per protein was identified with a Mascot score threshold of 20 and Sequest score threshold of 1.5.

4.3 Results

4.3.1 Transcriptomics

4.3.1.1 RNA integrity

RNA is enzymatically degraded within the cell by RNases and suffers from chemical instability at higher temperatures. The integrity and the quality of RNA extracted from treated (T) and untreated (UT) infected erythrocytes was assessed using automated electrophoresis. Traditional methods for assessing RNA quality compare the quantity of smaller and larger ribosomal RNAs (rRNA), where the 28S/18S ratio greater than 1.8 is considered as relativity intact RNA [257]. The RNA quality indicator (RQI) was developed using a series of increasingly degraded RNA samples, which were used to construct an indicator value ranging from 1 to 10, where 1 is completely degraded RNA [258]. The RQI algorithm interprets both the rRNA 28S and 18S regions from electrophoregrams, and also takes into account the pre-18S regions in which degraded RNA can occur [258]. Electropherograms of RNA extracted from both treated and untreated infected erythrocytes indicated the various time points had narrow and resolved fluorescence peaks visible for 18S and 28S ribosomal RNA with little or no shorter RNA fragments reflected as low baseline visible at the early part of the electropherogram (Figure 4.1). The RQI values for all the RNA extracts were above 7 (results not shown), indicative of good quality RNA.

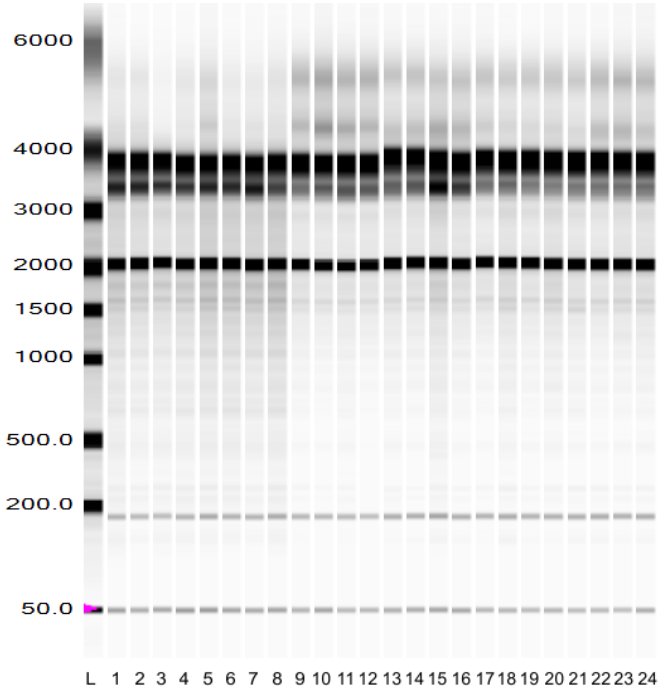

Figure 4.1: Electropherogram of RNA isolated from untreated and 4PEHz-treated parasites. A molecular weight marker, denominated in base-pairs (BP), is indicated on the left of the figure. Samples from 1 to 24 represent 200 ng RNA extracted from UTt_1 (1-4) and Tt_1 (5-8), UTt_2 (9-12) and Tt_2 (13-16) and UTt_3 (17-20) and Tt_3 (21-24), respectively. The 18S and 28S rRNA is visible at approximately 2000 and 3500 BP, respectively. Low levels of smearing visible between 2000 and 50 BP indicated that almost no degraded RNA was present.

4.3.1.2 Hybridization and raw image processing

The Agilent microarray platform, representing ~15K (15 744) oligonucleotides of which 7004 were unique identifiers, was used to determine the transcriptomic effects of 4PEHz treatment on the intraerythrocytic development of *P. falciparum* parasites. Positive control spots in each individual array included bright and dull spots at the corner of the array (Figure 4.2). Microarray spots representing Cy3 (green) labelled reference pool cDNA or Cy5 (red) labelled cDNA oligonucleotides were correlated to mapped grids containing the target information. A common reference-designed microarray was used in which the Cy3 reference pool consisted of equal amounts of pooled cDNA from all T- and UT-extracts.

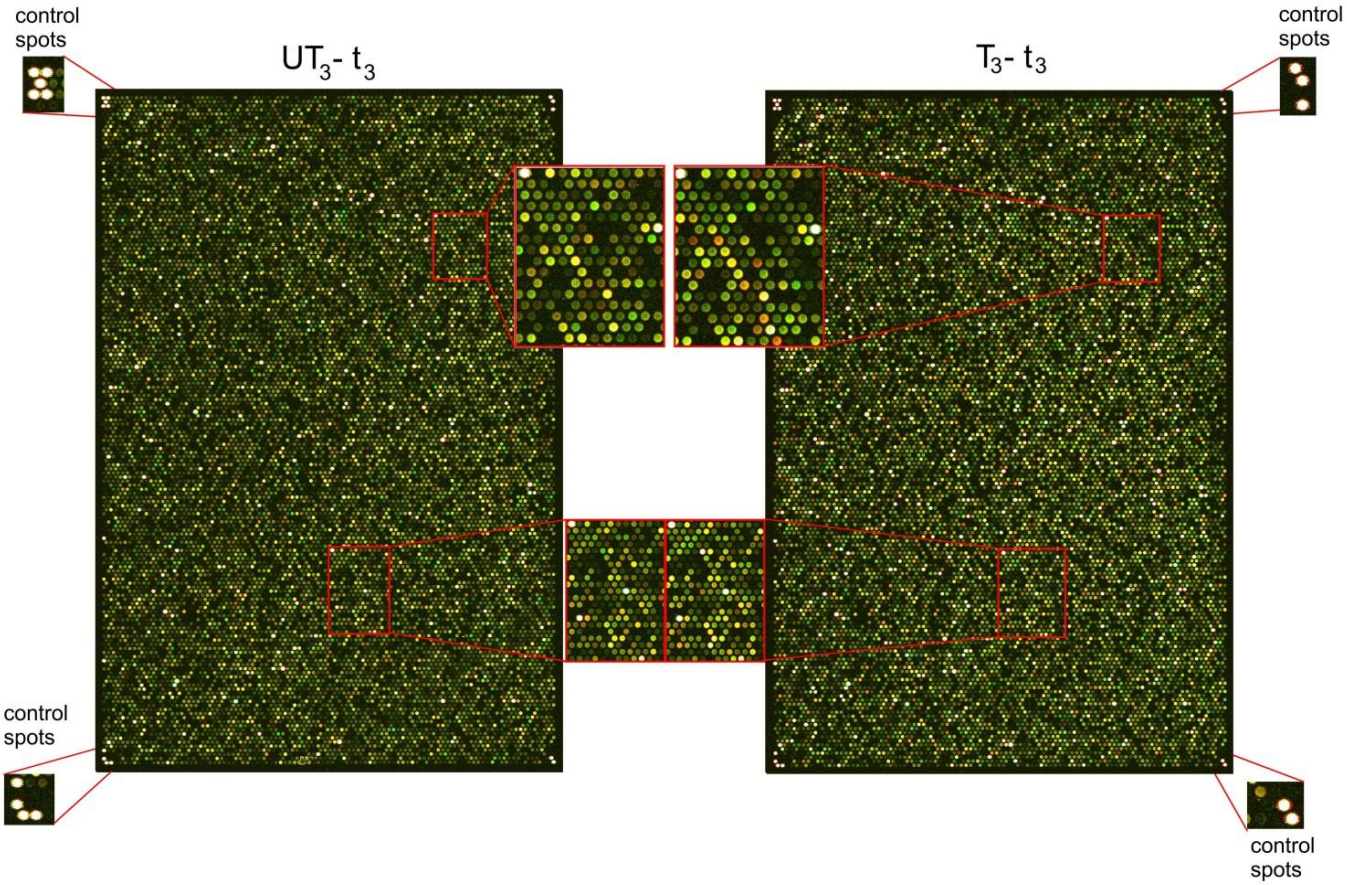

Figure 4.2: Individual arrays from a single UTt_3 and Tt_3 sample. The expanded regions in the centre compare the same region of spots from different arrays, and as can be seen the Tt_3 array some spots have higher red intensities, which indicate greater abundance of those particular transcripts in the array. The corner regions contained bright and dark control spots which are used to determine whether cDNA hydridization occurred successfully.

Spots from the arrays were subjected to evaluations based on signal intensity, spot saturation and signal to noise levels as indicated in the methods section. When comparing two individual arrays (UTt_3 and Tt_3), differences in spot brightness, intensity and colour reflects are noticeable (Figure

4.2). The spots, of either red or green colour channel, were interpreted with imaging. Spots from all of the slides were subjected to spot filtering procedures which entailed removing low signal-to-noise spots, as well as spots which were over saturated. On average between all the slides 2.8 – 4.6% of spots were removed based on over saturation. Between 29 – 32% of spots were removed from further interpretations based on the low (< 3) signal-to-noise ratios. On average more than 11K spots in each slide conformed to the quality filtering parameters and were further subjected to spot and array normalization procedures.

4.3.1.3 Normalisation of microarray image data

Background or ambient spot signal is caused by residual material left on the arrays after washing or by the non-specific binding of labelled samples [249]. Background correction of the individual red- and green-colour channels is necessary to remove biased interpretation and accounts for array-based variance including differences in labelling efficiency and spatial-associated differences with the array, ensuring correct interpretation of biological variation in terms of gene expression [249, 252]. In experiments performed the corrected adaptive background method (*normexp*) was used, with an offset of 50, thereby also removing unnecessary control spots thereby only retaining important information regarding target spots [248]. Background processing using the *normexp+offset* normalisation, instead of traditional subtraction of local background, gives lower false discovery rates [249].

Before normalisation, the log intensity (A) values of with the arrays had median values of between 3 and 4 with outlier values ranging as low as -1 (Figure 4.3 A). The $\log_2$ ratio of expression (M) values of the data had median values between -1 and 0 (Fig 4.5 B). MA plots, pre- and post-normalisation, generated from arrays of t_2 and t_3 in the 4PEHz study are shown in the Appendix section (Fig 4.A1- 4.A3). MA plots of array 7 – 12, as shown in Fig 4.A1, showed M values centred at 0. Loess normalisation aims to produce median centred data, and the quartiles are forced to have the same values using Gquantile between-array normalisation. Plots of Loess and Gquantile normalised data were median centred (Figure 4.3 C, D), however the differences were still visible in the quartile values of the data, suggesting ineffective Gquantile scaling of all the data. Array 4 and 14 had visibly smaller 25^{th} percentile values in A-plot (Figure 4.3 C), with some outliers as well, which is reflected in M-plots where array 4 and 14 can be seen with lowered median values (Figure 4.3 D). Robust spline normalisation together with Gquantile scaling resulted in improved median correction for A-values (Figure 4.3 E), in which the quartile values were effectively scaled to each other. The M-values appeared more evenly centred to 0 compared to Loess normalisation (Figure 4.3 F).

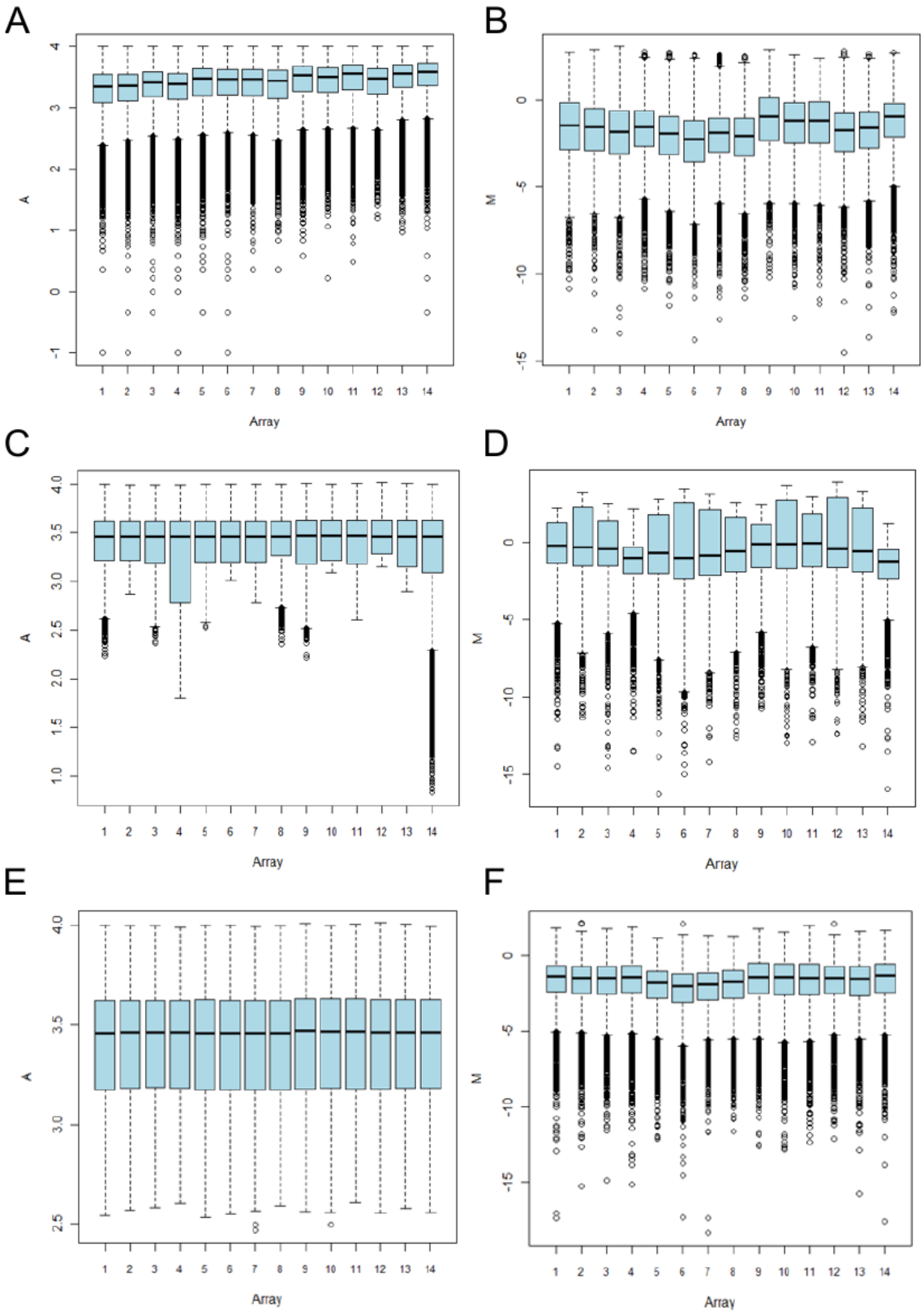

Figure 4.3: Boxplots of the log intensity (A) and log ratio expression (M) before and after array normalization. A and B) Plots of A and M before normalization, respectively. **C and D)** Plots of A and M after Loess and Gquantile normalisation**. E and F)** Boxplots of A and M after Robust spline and Gquantile normalisation. Boxplots represent the median of the data represented by a solid black line, boxed by quartiles with the top 75th percentile and bottom 25th percentile. Dotted lines represent the minimum and maximum values together with outliers marked in circles.

Density plots were used to visualise the distribution of the array data, and represents the frequency at which spot intensity values were observed, which is useful for determining array outliers [247]. Relating to the previous assumption that the M-values from an array remain unchanged or centred at 0, similarly when comparing one array to another it is expected that the overall frequency and intensity of the spots remain the same across the arrays. This is however hardly ever the case and another reason why normalisation is employed to account for between-array variation. By using a Cy3 cDNA reference pool the green channel data across all of the arrays can be normalised to one another (Figure 4.4). The green Cy3 non-normalised arrays data is depicted as green lines which have similar intensity and frequency distribution, however small between-array variations can be observed (Figure 4.4 A). After normalization, a single green representative line is observed (Figure 4.4 B and C). Loess normalisation, with Gquantile scaling, produced irregular red channel frequencies (Figure 4.4 B), compared to Robust spline normalisation with Gquantile showed markedly improved normalisation of the arrays (Figure 4.4 C). The Loess normalisation method therefore was not able to effectively normalise the data between different arrays and data from the Robust spline method was used determine differential gene expression.

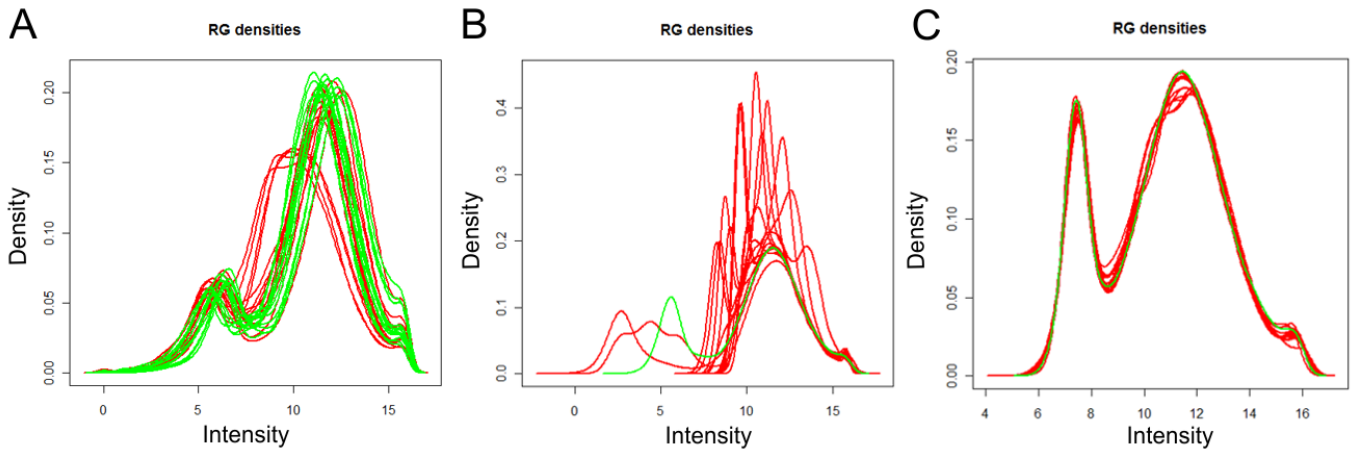

Figure 4.4: Density plots used to interpret array data before and after normalization. Density plots of array data **A)** before normalisation, **B)** after Loess and Gquantile normalisation and **C)** after Robust spline and Gquantile normalisation. The y-axis density represents the frequency of mRNA's in relationship to the x-axis intensity of the mRNA's.

After normalization all the Cy5 and Cy3 intensity values of the spots on the slides, apart from those omitted due to bad intensity or oversaturation, were used in calculations to determine the $\log_2$ FC. From these calculations only transcripts which had *P*-values smaller than 0.05, with the $\log_2$ ratio greater or equal to 0.75 or smaller and equal to -0.75 were considered significant ($\log_2$ ratio $\geq$ 0.75

and $\log_2$ ratio ≤ -0.75). Across all of these slides, on average, around 11K transcript spots in each slide were processed to derive individual spot $\log_2$ FC expression values.

4.3.2 Pearson correlations between UT and T samples

Previous microarray analyses with cytostatic therapeutics showed that Pearson correlations are necessary to ensure that, considering the 'just-in-time" nature of the transcriptome [225], correct phases of the parasite are compared to one another to observe drug-specific perturbation effects [259]. These correlations align global transcript expression patterns from T and UT parasites of different time points. As shown in the previous chapter, 4PEHz treatment of the malaria parasite did not affect the morphological composition of the parasite populations throughout a seven day period. Pearson correlations were necessary to ensure the correct T and UT-time point samples were compared to each other.

Table 4.3: Pearson correlations between UT and T array data of the 4PEHz treatment study.

Sample	Pearson Correlation (r)
$UTt_1 : Tt_1$	0.891
$UTt_2 : Tt_2$	0.847
$UTt_3 : Tt_3$	0.516
$UTt_1 : Tt_2$	0.093
$UTt_1 : Tt_3$	-0.521
$UTt_2 : Tt_1$	0.024
$UTt_2 : Tt_3$	0.288
$UTt_3 : Tt_1$	-0.486
$UTt_3 : Tt_2$	-0.280
$UTt_1 : UTt_2$	0.018
$UTt_1 : UTt_3$	-0.436
$UTt_2 : UTt_3$	-0.186

The highest correlation was found between the matching UT and T from each time point (Table 4.3). From the first two time points there was a high degree of correlation, which indicated preservation of global transcript expression patterns (Table 4.3). Pearson correlations between UTt_3 and Tt_3 suggested that there were greater variations in transcript expression profiles compared to the t_1 and t_2 comparisons (Table 4.3). Correlations of UT from any of the time points with any other T sample were either uncorrelated or anti-correlated (Table 4.3). Transcriptional progression is also marked by uncorrelated and anti-correlated correlations between the UTt_1 vs. UTt_2 and UTt_3 (Table 4.3), as expected due to the "just-in-time" nature of the transcriptome. From these results it was suggested that comparisons of gene expression could be made from the same time points. To

summarise, Pearson correlations suggested that the global transcript expression patterns from the first two time points were highly correlated, whereas the largest transcriptome perturbations occurred in t_3.

4.3.3 Differentially expressed transcripts

Comparison of the transcript expression ratios between UTt_1 vs. Tt_1 samples revealed only 25 unique transcripts were significantly expressed with $P < 0.05$ and had differential expression with the $\log_2$ FC > 0.75 or $\log_2$ FC < -0.75 (Figure 4.5). Of these transcripts 52% (13/25) of the transcripts had decreased expression and 48% (12/25) had increased abundance (Figure. 4.7). For the second time point (UTt_2 vs. Tt_2), 14 genes were found to be significantly expressed, all of which had decreased abundance.

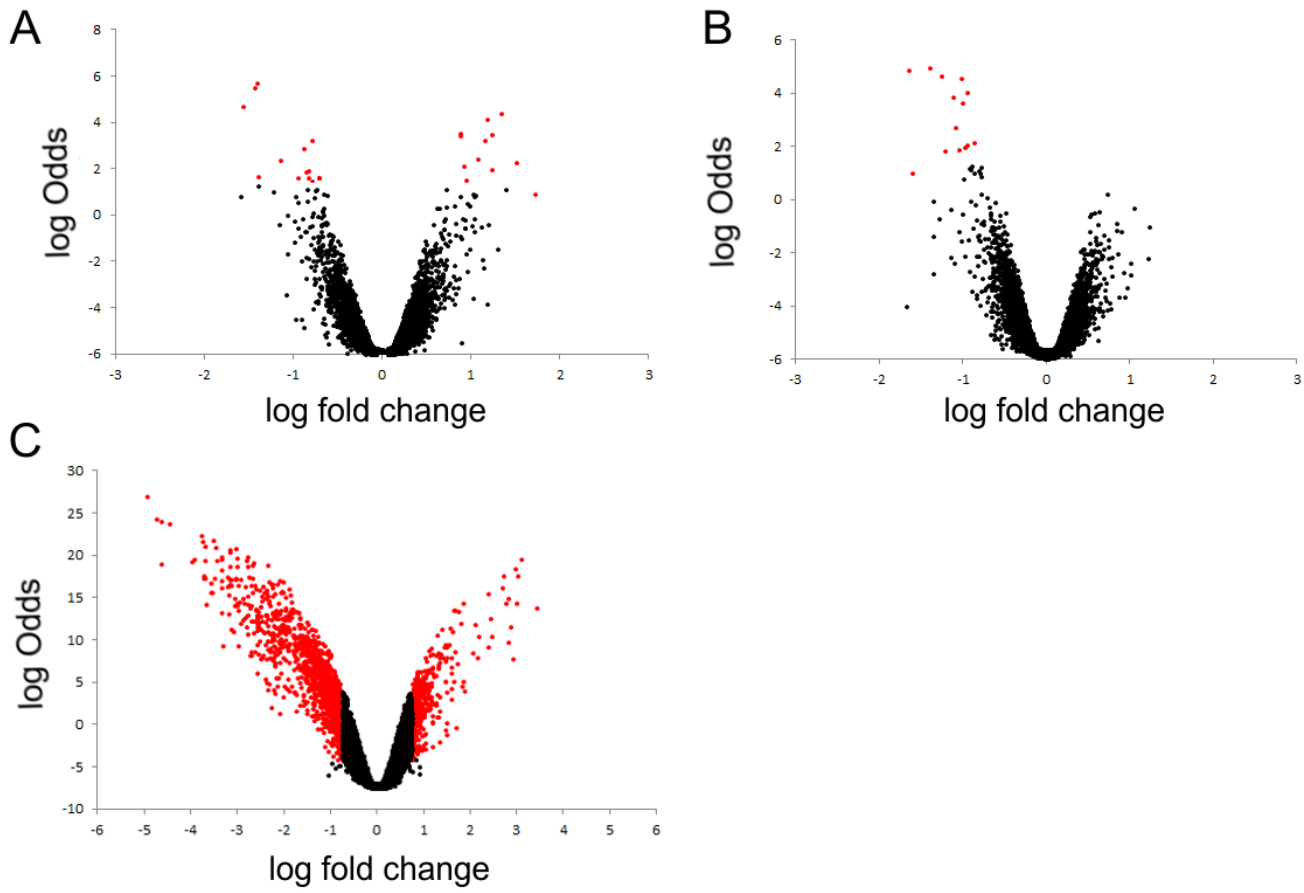

Figure 4.5: Volcanoplots generated for differently expressed transcripts affected by 4PEHz treatment. **A)** UTt_1 compared Tt_1 of which 13 unique transcripts had decreased abundance (red spots indicated $p < 0.05$, with $\log_2$FC > 0.75 or $\log_2$FC < -0.75) and 12 unique transcripts had increased expression. **B)** Volcanoplot of UTt_2 vs. Tt_2 in which 14 transcripts (from the 23 in total) were unique and that had significant decreased abundance. **C)** Volcanoplot of UTt_3 compared to Tt_3 in which 787 unique transcripts were differential expressed. Transcripts which had differential expression with the $\log_2$ FC > 0.75 or $\log_2$ FC < -0.75, and with $P < 0.05$ are marked in red.

Comparisons between the UTt_3 and Tt_3 samples showed 787 unique differently expressed transcripts. Of these 68% (535/787) of the transcripts had decreased transcript expression compared to 32% (252/787) which had increased expression. There were only a few differentially expressed transcripts in the first two time points in contrast the third time point, implicating that 4PEHz affects the parasites during later developmental stages of the IDC. The complete list of the differentially expressed transcripts from only the third t_3 time point is provided in the Appendix 4 (Table A4.1). The 20 most affected genes in time point t_3 (UTt_3 and Tt_3) are listed in Table 4.4. GO classification of transcripts with the lowest $\log_2$ FC suggested these are involved in cell motility (GO term: 0006928), actin filament organisation (GO term: 0007015) and pathogenesis (GO term: 009405) (Table 4.4). Transcripts which had the highest abundance were involved in cyclic nucleotide metabolic processes (GO term: 0009187), protein polymerisation (GO:0051258) and components of the tricarboxylic acid cycle (GO:0006099).

Table 4.4: **Twenty transcripts that showed the greatest differential decreased and increased abundance during 4PEHz treatment in t_3.**

PlasmoDB ID	PlasmoDB description	log_2FC	Adj. P-value	Annotated GO biological process
Decreased abundance				
MAL7P1.176	erythrocyte binding antigen-175	-4.93	2.99E-12	regulation of immune response
MAL13P1.60	erythrocyte binding antigen-140	-4.74	2.89E-11	cell-cell adhesion
PFL2520w	reticulocyte binding protein homologue 3, pseudogene	-4.63	8.53E-10	-
PF08_0035	conserved *Plasmodium* protein, unknown function	-3.96	6.99E-10	-
PF13_0058	conserved *Plasmodium* protein, unknown function	-3.93	6.47E-10	entry into host cell, pathogenesis
PFA0110w	ring-infected erythrocyte surface antigen	-3.77	1.15E-10	-
PF10_0039	membrane skeletal protein IMC1-related	-3.75	1.71E-10	-
PFD0110w	normocyte binding protein 1,reticulocyte binding protein homologue 1	-3.72	2.53E-09	translation, translational initiation
PF10_0003	rifin	-3.70	6.50E-10	antigenic variation
PF10_0138	conserved *Plasmodium* protein, unknown function	-3.69	2.92E-10	cytoskeleton organization
PF07_0061	conserved *Plasmodium* protein, unknown function	-3.56	3.94E-09	-
PFL2460w	coronin	-3.52	1.71E-10	phosphate ion transport
MAL8P1.70	conserved *Plasmodium* protein, unknown function	-3.49	2.93E-09	entry into host cell
PFL1435c	myosin D	-3.46	2.99E-10	-
PF11_0268	kelch protein, putative	-3.44	6.50E-10	actin filament organization, actin filament-based movement
PF10_0262	conserved *Plasmodium* protein, unknown function	-3.35	6.38E-10	cytokinesis, microtubule-based process
PFI0540w	conserved *Plasmodium* protein, unknown function	-3.34	1.37E-09	-
PF10_0346	merozoite surface protein 6	-3.33	6.50E-10	cell-cell adhesion, pathogenesis
MAL13P1.176	reticulocyte binding protein 2 homologue b	-3.33	5.53E-09	-
PFE1285w	membrane skeletal protein IMC1-related	-3.22	2.73E-09	entry into host cell

Increased abundance				
PF11_0256	pyruvate dehydrogenase E1 alpha subunit	1.65	4.28E-08	acetyl-CoA biosynthetic process from pyruvate, fatty acid biosynthetic process, tricarboxylic acid cycle
PFL0795c	male development gene 1	1.65	4.18E-05	-
MAL13P1.129	conserved *Plasmodium* protein, unknown function	1.66	1.59E-06	-
PF10_0164	early transcribed membrane protein 10.3	1.67	4.17E-08	-
PFI0315c	dynein light intermediate chain 2, cytosolic	1.69	0.002833	-
PF14_0672	phosphodiesterase delta, putative	1.72	4.82E-06	cyclic nucleotide metabolic process, signal transduction
MAL7P1.162	dynein heavy chain, putative	1.81	1.29E-07	microtubule-based movement
PF14_0040	secreted ookinete adhesive protein	1.81	4.73E-05	-
PF11_0092	mechanosensitive ion channel protein	1.84	3.09E-05	-
PF14_0272	CPW-WPC family protein	1.84	2.31E-08	-
PF11_0479	conserved *Plasmodium* protein, unknown function	1.89	7.35E-05	-
PFL1910c	conserved *Plasmodium* protein, unknown function	2.05	1.96E-06	translation
PFL2385c	conserved *Plasmodium* protein, unknown function	2.09	1.44E-07	-
PF14_0467	conserved *Plasmodium* protein, unknown function	2.16	3.50E-06	-
PFB0355c	serine repeat antigen 2	2.40	9.86E-07	immunoglobulin production, proteolysis, regulation of immune response
PFE1205c	conserved *Plasmodium* membrane protein, unknown function	2.69	5.40E-09	-
PFD1050w	alpha tubulin 2	2.71	2.68E-09	microtubule-based movement, protein polymerization
PF14_0708	probable protein, unknown function	3.00	2.31E-08	-
PFI1520w	asparagine-rich antigen, putative	3.10	6.47E-10	-
PFI1230c	conserved *Plasmodium* protein, unknown function	3.43	4.15E-08	-

The adjusted P-value (adj. P-value) indicates the result of a moderated t-statistic test employed in LIMMA which normalises the standard errors from spot or probe results, and is interpreted as a normal t-statistic result. The $\log_2$FC value, or also the $\log_2$(Cy5/Cy3), indicates the relative expression of a transcript in Tt_3 in comparison with the UTt_3 stage, and is considered treatment-specific.

4.3.4 Gene ontology (GO) and functional annotation

GO charts of significantly expressed transcripts were constructed to determine processes affected by 4PEHz treatment (Figure 4.6). The predicted GO process term, a useful descriptive attribute assigned to genes according to the associated/predicted cellular function, cellular process or exact molecular function, was derived from PlasmoDB v7.0. Of the 251 transcripts which had increased abundance in t_3 a large proportion of transcripts (63%) represented conserved proteins with unknown function, and these were omitted from GO charts. Of the 535 transcripts with decreased abundance, the majority of transcripts (73%) similarly represented transcripts of which the product has no known function.

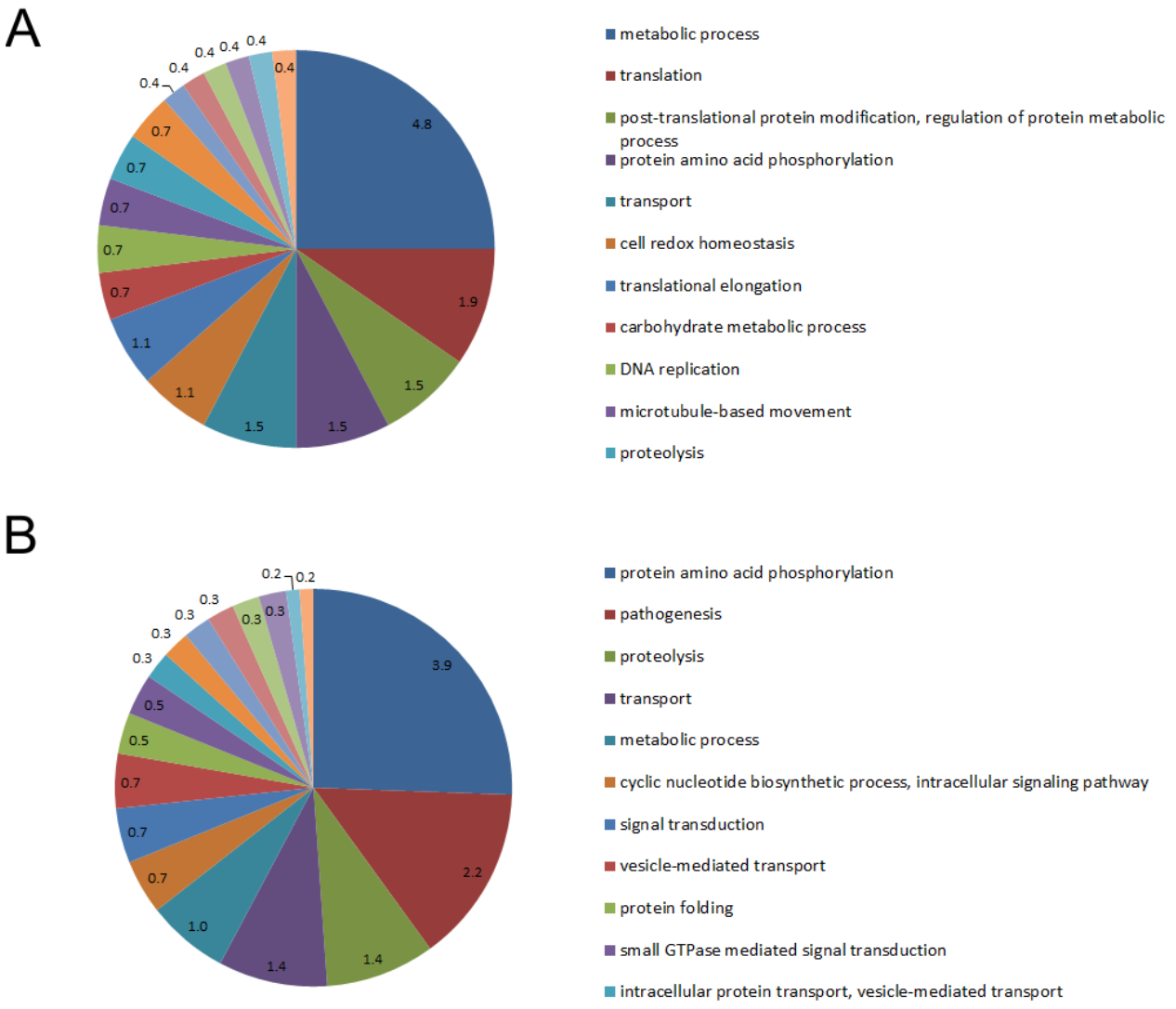

Figure 4.6: Differently expressed transcripts from UTt_3:Tt_3 classified according to PlasmoDB predicted GO processes. A) Transcripts which had increased abundance are associated with metabolic processes, as well as translation, and cell redox homeostasis. **B)** Transcripts which had decreased abundance were shown to be involved in protein amino acid phosphorylation, pathogenesis and proteolysis. Transcripts with unknown or hypothetical function, were not included in both instances. Values represent percentages of the total amount of transcripts.

Functional annotation of differentially expressed transcripts in DAVID was useful to determine additional processes which may have been excluded from the PlasmoDB database [260, 261]. Functional annotation revealed both increased and decreased transcript perturbations in apicoplast- and mitochondrial-related processes (Figure 4.7). Transcripts which had decreased abundance had involvement in phosphorylation, were associated with protein kinase activity, and had contributed in regulation of transcription (Figure 4.7 A). Transcripts with increased abundance were involved in cofactor metabolic processes, as well as acetyl-CoA metabolism (Figure 4.7 B). A few transcripts which had increased expression were also involved in DNA repair such as PF11_0087 (Rad51 homolog), and MAL13P1.346 (DNA repair endonuclease, putative) (Table A4.1).

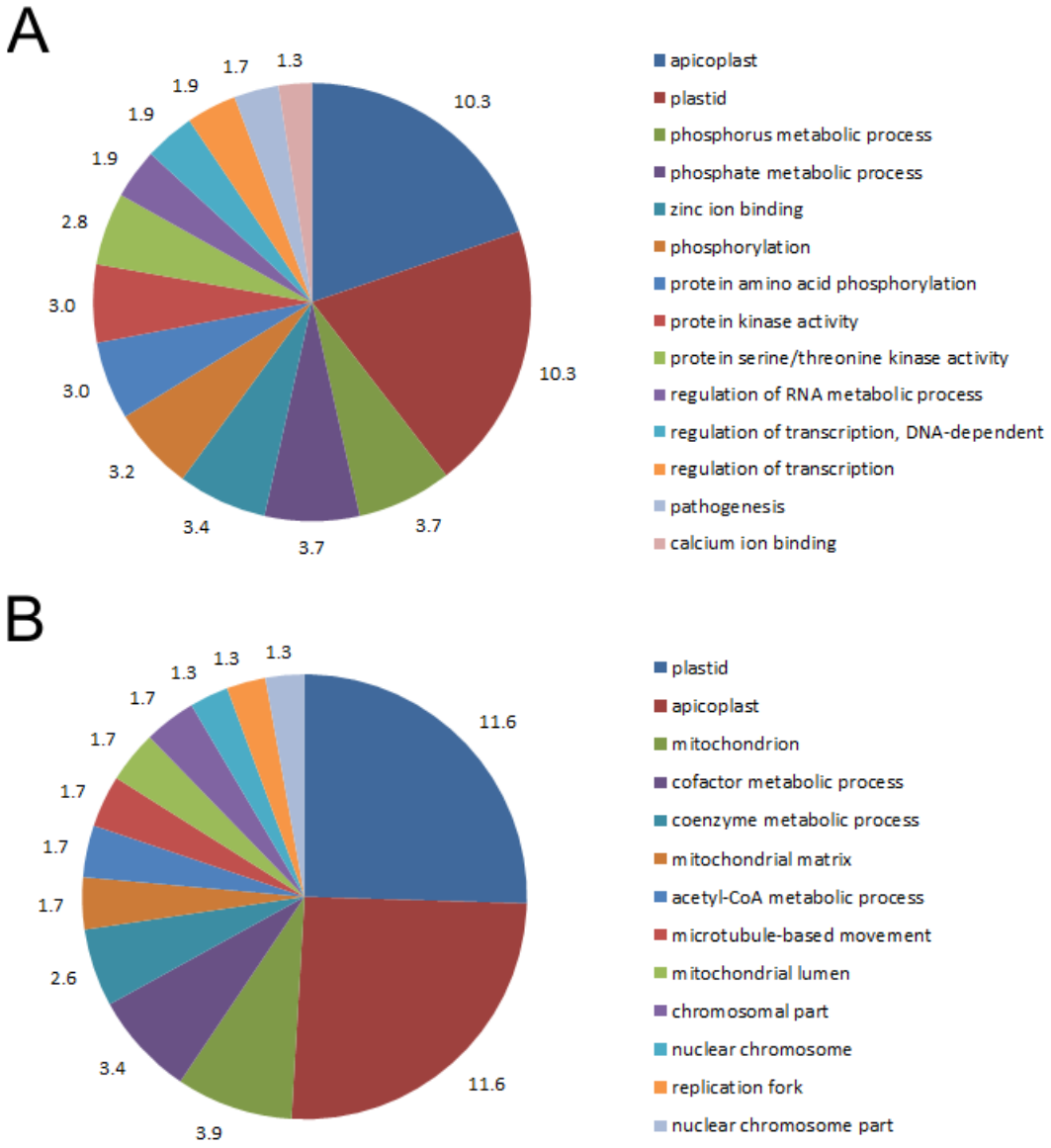

Figure 4.7: Functional annotation of differentially expressed transcripts using DAVID. Functional annotation was performed on **A)** transcripts which had decreased abundance and **B)** transcripts with increased abundance from UTt_3 vs. Tt_3.

There was some correlation in transcript expression across the different time points (Figure 4.8). In all three time points PFC0120w (cytoadherence linked asexual protein 3.1), PF14_0102 (rhoptry-associated protein 1) and PFI1445w (high molecular weight rhoptry protein 2) had decreased expression (Figure 4.8 A). These transcripts are associated with cytoadherence, and the rhoptries of the parasite. Corresponding transcripts in both t_1 and t_3 time points are shown in Figure 4.8 B. Transcripts MAL7P1.124 (conserved *Plasmodium* protein, unknown function) and PFL1565c (CG2-related protein, putative), had decreased and increased abundance at t_1, respectively, with opposite expression levels at the t_3 time point (Figure 4.8 B). When comparing transcript expression patterns from t_2 and t_3 time points (Figure 4.8 C), the PFD0955w transcript (apical merozoite protein) had a more pronounced suppression at t_3. The gene product of PFD0955w has a predicted role in attachment of GPI anchors to proteins, and plays a role in pathogenesis. Overall, transcript expressions across the different time points were shown to correlate, transcripts which had decreased expression during t_1 and t_2 had decreased expression in the t_3.

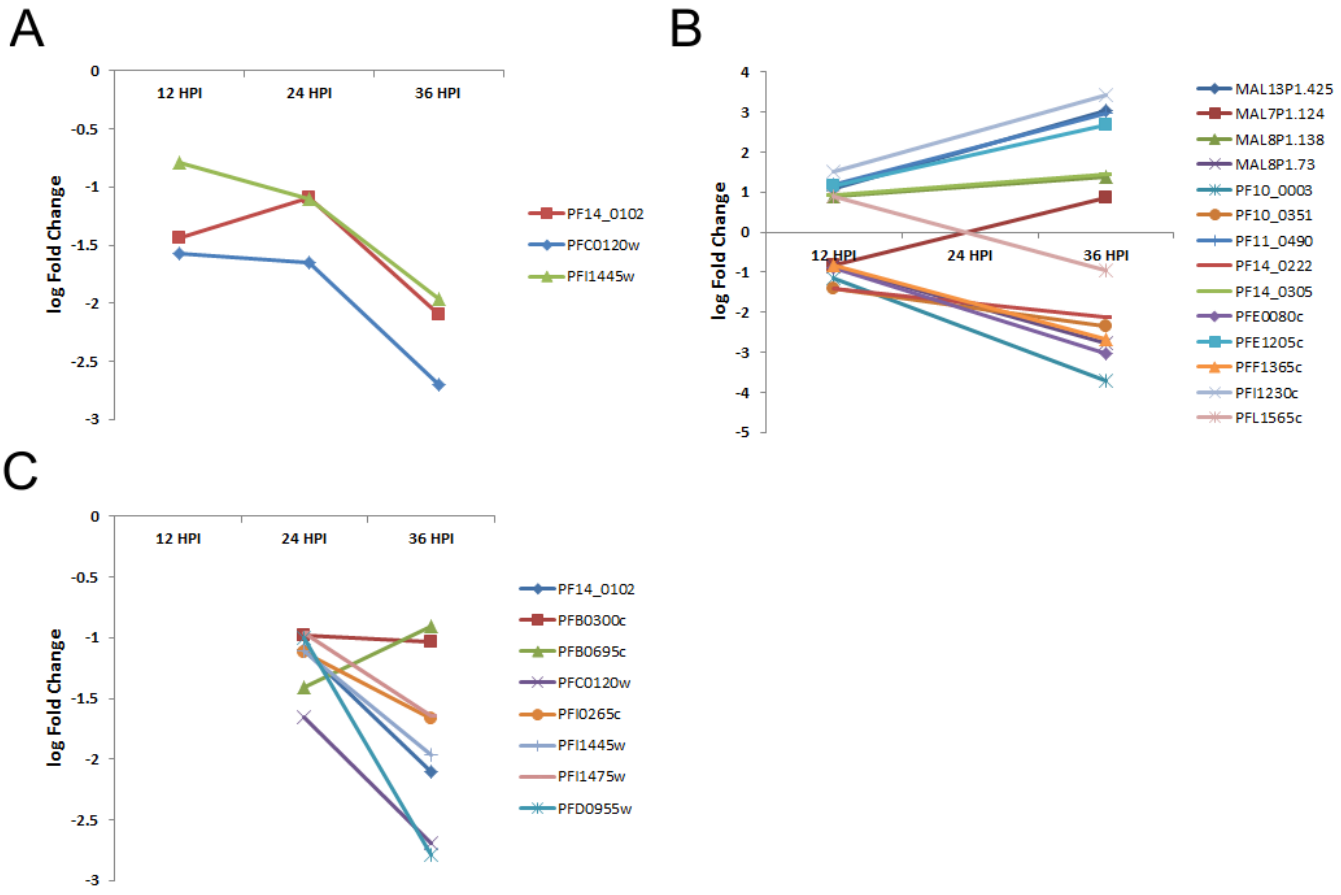

Figure 4.8: Transcripts which had significant expression and correlated between t_1 (12HPI), t_2 (24HPI) and t_3 (36HPI). A) PF14_0102, PFC0120w and PFI1445w had decreased abundance and this became more pronounced during the late t_3 phases. **B)** Transcripts PFL1565c and MAL7P1.124 were shown to have increased and decreased abundance, respectively, which changed to the opposite condition in the late t_3 time point. **C)** PFD0955w and PFC0120w had decreased expression levels during t_2 and greater decreases in expression at t_3.

4.3.5 PLP synthase and PLP-related perturbations

A degree of specificity was shown for 4PEHz in *Pf*Pdx1/*Pf*Pdx2 over-expressing cells, as discussed previously. PLP-related processes as well as processes affiliated with *Pf*Pdx1 and *Pf*Pdx2 substrates were therefore expected to be affected by 4PEHz treatment and it was postulated that these effects were visible in the transcriptome. The *Pf*Pdx2 transcript was significantly affected and had increased expression in t_3 (Table 4.5). Other transcripts of enzymes involved in PLP production such as pyridoxal and pyridoxine kinase (PFF0775w) were not significantly affected by 4PEHz treatment (Table 4.4). Of the PLP dependent enzymes in *P. falciparum*, a putative serine hydroxymethyltransferases (SHMT, PF14_0534) had increased abundance during t_1 and t_2. For these two time points the adjusted *P*-value was 0.049 and 0.319, suggesting that the t_2 expression level could not be considered as statistically significant (Table 4.5). Two other PLP-dependent enzyme transcripts were affected, a putative serine C-palmitoyltransferase (SPT, PF14_0155) and phosphatidylserine decarboxylase (PSDC, PFI1370c) (Table 4.5). The majority of other transcripts of PLP-dependent enzymes showed a general trend of increased expression during the t_3 stages, most of these were however not statistically significant. Of the PLP-dependent enzyme transcripts, AspAT had the most stable expression levels across all three time points.

Of the pathways that involve the substrates of *Pf*Pdx1 and *Pf*Pdx2, some transcripts in the pentose phosphate pathway had increased expression, however the majority were not significantly affected (Table 4.5). Putative deoxyribose-phosphate aldolase (PF10_0210), which reversibly converts G3P to 2-deoxy-D-ribose 5-phosphate had increased abundance (Table 4.5). The transcript of a putative phosphoglucomutase (PF10_0122), which catalyse the interconversion of R5P to D-ribose 1-phosphate had increased expression. A putative glutamate dehydrogenase (PF14_0286) transcript - part of L-glutamine catabolism - had increased abundance with $\log_2$FC of 0.90 in the parasites at t_3 (result not shown). The fact that these transcripts were affected suggested that 4PEHz could have disrupted R5P, G3P as well as L-glutamine homeostasis, thereby resulting in transcriptional perturbations. 4PEHz is a weak inhibitor of spinach R5P isomerase with a K_i of 1.8 ± 0.2 mM [206, 218]. At 20 µM 4PEHz it was anticipated that *P. falciparum* R5P isomerases would be weakly affected. The transcript of a putative D-ribose 5-phosphate epimerase (PFE0730c) was not significantly affected during the course of 4PEHz treatment (Table 4.5), suggesting that 4PEHz did not affect R5P isomerases leading to changes in transcript expression.

Table 4.5: Differential expression of transcript in 4PEHz treated *P. falciparum* parasites. The differential expression levels of transcripts associated with PLP biosynthesis and salvage, as well as PLP-dependent enzyme transcripts during treatment of *P. falciparum* with 4PEHz are shown.

PlasmoDB ID	PlasmoDB description	log_2FC $UTt_1:Tt_1$	log_2FC $UTt_2:Tt_2$	log_2FC $UTt_3:Tt_3$
PLP metabolism				
PF11_0169	pyridoxine biosynthesis protein Pdx2 (Pdx2)	0.05	0.21	0.77*
PF14_0570	pyridoxal 5'-phosphate synthase, putative	0.52	0.29	0.41
MAL13P1.324	aldo-keto reductase, putative	0.44	-0.14	0.32
PF14_0088	aldo-keto reductase, putative	0.29	-0.11	0.26
PFF0775w	pyridoxine kinase	-0.30	0.21	0.23
PFF1025c	pyridoxine biosynthesis protein Pdx1 (Pdx1)	-0.27	0.19	-0.15
PF07_0059	4-nitrophenylphosphatase	0.20	0.15	-0.25
PLP-dependent enzymes				
PF14_0155	serine C-palmitoyltransferase (SPT), putative	0.22	0.17	1.48*
PF14_0557	conserved *Plasmodium* protein, unknown function	0.01	0.69	0.61
PFI0965w	pyridoxal 5'-phosphate dependent enzyme class III, putative	0.01	0.16	0.47
PFL1720w	serine hydroxymethyltransferase (SHMT), cytosolic	-0.07	-0.35	0.47
PFD0670c	lysine decarboxylase-like protein, putative	-0.15	-0.13	0.44
PF07_0068	cysteine desulfurase, putative	0.05	0.21	0.41
PFI1370c	phosphatidylserine decarboxylase (PSDC)	0.15	-0.21	-0.94*
PFL2210w	delta-aminolevulinic acid synthetase (ALAS)	-0.06	-0.08	-0.27
PF14_0534	serine hydroxymethyltransferase (SHMT), mitochondrial	1.72*	0.90	-0.37
PFB0200c	aspartate aminotransferase,aspartate transaminase (AspAT)	-0.15	0.15	0.19
PFF0435w	Ornithine aminotransferase (OAT)	-0.14	-0.14	0.12
Pentose phosphate pathway				
PF10_0122	phosphoglucomutase, putative	-0.08	0.08	0.94*
PF10_0210	deoxyribose-phosphate aldolase, putative	-0.02	-0.07	0.86*
PF14_0520	6-phosphogluconate dehydrogenase, decarboxylating, putative	0.24	-0.23	0.74
PF14_0466	Appr-1-p processing domain protein	0.09	0.06	0.42
PF13_0157	ribose-phosphate pyrophosphokinase, putative	0.20	-0.11	0.38
PF13_0143	phosphoribosylpyrophosphate synthetase	-0.06	0.11	0.33
PFL0960w	D-ribulose-5-phosphate 3-epimerase, putative	-0.18	-0.10	0.18
PFE0730c	D-ribose 5-phosphate epimerase, putative	0.05	0.05	0.02
PFF0530w	transketolase	-0.01	-0.15	-0.15
PF14_0511	glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase	-0.18	-0.04	-0.36

Footnotes: Significant differential expression with a log_2FC above or below 0.75 and -0.75, respectively, with $P \leq 0.05$ is indicated by *.

In conjunction with SPT and phospholipid metabolism, other transcripts such as putative patatin-like phospholipase (PFI1180w) had decreased expression with by 1.69-fold change (Table 4.6).

Similarly, within the phosphatidylserine pathway putative diacylglycerol kinase (PFI1485c) were also found to be less abundant during the t_3 time point (Table 4.6). Transcripts involved in fatty acid (FA) synthesis had increased abundance, favouring the malonyl- and palmitoyl-CoA route of FA synthesis. Transcripts of this pathway include a putative biotin carboxylase subunit of acetyl CoA carboxylase (PF14_0664) and acyl-CoA synthetase, PfACS9 (PFB0685c), with another putative long chain polyunsaturated fatty acid elongation enzyme (PFF0290w) (Table 4.6). Transcripts of the malonyl CoA-acyl carrier protein transacylase precursor (PF13_0066) also had increased expression forming part of the conversion of malonyl-CoA to malonyl-acyl carrier protein (ACP).

Some of the transcripts were also characterised by the GO term - acetyl-CoA biosynthetic process from pyruvate (GO:0006086) – having involvement in acetyl-CoA biosynthesis. The pyruvate dehydrogenase complex (PDC) converts pyruvate to acetyl-CoA, and three of the four components that constitute the PDC in *P. falciparum* had increased expression at the t_3 time point (Table 4.6). These include pyruvate dehydrogenase E1 alpha subunit (PF11_0256); lipoamide dehydrogenase (aLipDH, PF08_0066) and pyruvate dehydrogenase E1 beta subunit (PF14_0441). The other component of the PDC was a putative dihydrolipoamide acyltransferase (PFC0170c which had 0.40 log_2FC greater expression compared to the UTt_3 stage (results not shown). The log_2FC of the latter was not considered as significant, however, for a descriptive purpose was interpreted together within the PDC. Overall there was an increased enrichment of transcripts associated with the PDC.

Table 4.6: Transcripts associated with apicoplast fatty acid biosynthesis in *P.falciparum* affected by 4PEHz. Apicoplast-related transcripts with differential expression during t_3 are listed, with the associated GO terms.

PlasmoDB ID	PlasmoDB description	log_2FC $Tt_3:UTt_3$	adj.P.Val	PlasmoDB annotated processes	GO term
PF11_0256	pyruvate dehydrogenase E1 alpha subunit	1.65	4.28E-08	acetyl-CoA biosynthetic process from pyruvate	GO:0006084~acetyl-CoA metabolic process
PF08_0066	lipoamide dehydrogenase (aLipDH)	1.28	8.55E-05	acetyl-CoA biosynthetic process from pyruvate	GO:0006084~acetyl-CoA metabolic process
PF13_0066	malonyl CoA-acyl carrier protein transacylase precursor	1.19	1.76E-02	fatty acid biosynthetic process	GO:0004314~[acyl-carrier-protein] S-malonyltransferase activity
PFB0685c	acyl-CoA synthetase, PfACS9	1.07	3.47E-04	fatty acid metabolic process	
PF14_0441	pyruvate dehydrogenase E1 beta subunit	0.91	1.07E-04	acetyl-CoA biosynthetic process from pyruvate	GO:0009536~plastid
PFD1035w	steroid dehydrogenase, putative	0.82	2.52E-04	fatty acid biosynthetic process	GO:0006631~fatty acid metabolic process
PF14_0664	biotin carboxylase subunit of acetyl CoA carboxylase, putative	0.80	1.34E-03	fatty acid biosynthetic process	
PFI1180w	patatin-like phospholipase, putative	-0.76	7.06E-05	lipid metabolic process	GO:0016042~lipid catabolic process
PFC0260w	P-loop containing nucleoside triphospahte hydrolase, putative	-0.88	7.55E-04	metabolic process	
PFB0695c	acyl-CoA synthetase, PfACS8	-0.91	2.00E-04	fatty acid metabolic process	
PFI1485c	diacylglycerol kinase, putative	-1.39	8.62E-06	activation of protein kinase C activity by G-protein coupled receptor protein signaling pathway	GO:0007166~cell surface receptor linked signal transduction
PFL0035c	acyl-CoA synthetase, PfACS7	-1.89	2.32E-07	metabolic process	

Footnote: The adjusted P-value (adj. P-value) indicates the result of a moderated t-statistic test employed in LIMMA which normalises the standard errors from spot or probe results, and is interpreted as a normal t-statistic result.

Gene set enrichment analysis (GSEA) was performed to find functionally classified gene sets which were either positively or negatively correlated. The GSEA uses normalised M-values to rank gene sets, and find patterns of enrichment. Individual transcripts in these sets may not have statistically significant differential expression, however as cluster may reveal significant enrichment. Gene set

enriched by 4PEHz treatment were associated with cofactor and coenzyme metabolic processes (Table 4.7, Figure 4.9). Additional information regarding the identities of these transcripts is given in Appendix Table A4.2. Within this group there were several transcripts involved in folate metabolism and these folate-related transcripts were found to be enriched and positively correlated with 4PEHz treatment (Table A4.2). These included dihydrofolate synthase/folylpolyglutamate synthase (DHFS-FPGS), dihydropteroate synthetase (DHPS) and lipoamide dehydrogenase (aLipDH), with $\log_2$FC of 0.49, 0.43 and 1.28 expression in t_3. The transcript of bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS, PFD0830w) had increased expression at t_3 (Table A4.2). DHFR-TS is not annotated as part of folate biosynthesis, however involved in recycling of folate, and was included in this cluster for interpretation (Table A4.2). Transcripts with the Kyoto encyclopaedia of genes and genomes (KEGG) term: Citric acid cycle (TCA cycle) gene set were enriched (Figure 4.9). These include the components of the PDC as well as an ATP-specific succinyl-CoA synthetase beta subunit (Figure 4.9). GSEA also revealed transcripts linked to the pentose phosphate pathway were enriched in the t_3 time point (Table 4.7). The core enriched components of this cluster included a putative deoxyribose-phosphate aldolase (PF10_0210), a putative decarboxylating 6-phosphogluconate dehydrogenase (PF14_0520) and TIM (PF14_0378). These transcripts had a $\log_2$FC expression of 0.86, 0.74 and 0.5, respectively.

Table 4.7: Gene set enrichment analysis (GSEA) of Tt_3 compared to UTt_3.

Gene set name	Cluster size	Enrichment score	Normalised enrichment score	Nominal *P*-value	FDR q-value
GO.0006412. PROTEIN.BIOSYNTHESIS	97	0.50	1.93	0.000	0.012
GO.0051186. COFACTOR.METABOLIC.PROCESS	43	0.54	1.81	0.000	0.025
KEGG.PYRIMIDINE METABOLISM	28	0.58	1.78	0.000	0.028
GO.0006118. ELECTRON.TRANSPORT	34	0.55	1.77	0.002	0.031
GO.0008757. SADENOSYLMETHIONINEDEPENDENT. METHYLTRANSFERASE.ACTIVITY	16	0.61	1.62	0.018	0.083
GO.0016491. OXIDOREDUCTASE.ACTIVITY	19	0.56	1.59	0.015	0.088
KEGG.CITRATE CYCLE (TCA CYCLE)	13	0.64	1.59	0.015	0.091
GO.0006633. FATTY.ACID.BIOSYNTHESIS	4	0.91	1.59	0.002	0.092
GO.0006098. PENTOSEPHOSPHATE.SHUNT	6	0.75	1.54	0.017	0.126
GO.0051188. COFACTOR.BIOSYNTHETIC.PROCESS	24	0.52	1.52	0.032	0.143
GO.0051187. COFACTOR.CATABOLIC.PROCESS	16	0.56	1.48	0.047	0.176
KEGG.GLUTATHIONE METABOLISM	8	0.67	1.47	0.050	0.182
MPM.FOLATE BIOSYNTHESIS	5	0.78	1.47	0.031	0.185

Footnote: Gene sets enrichment scores were calculated with $P \leq 0.05$ and had a false discovery rate (FDR) $\leq$ 0.25 (25%)

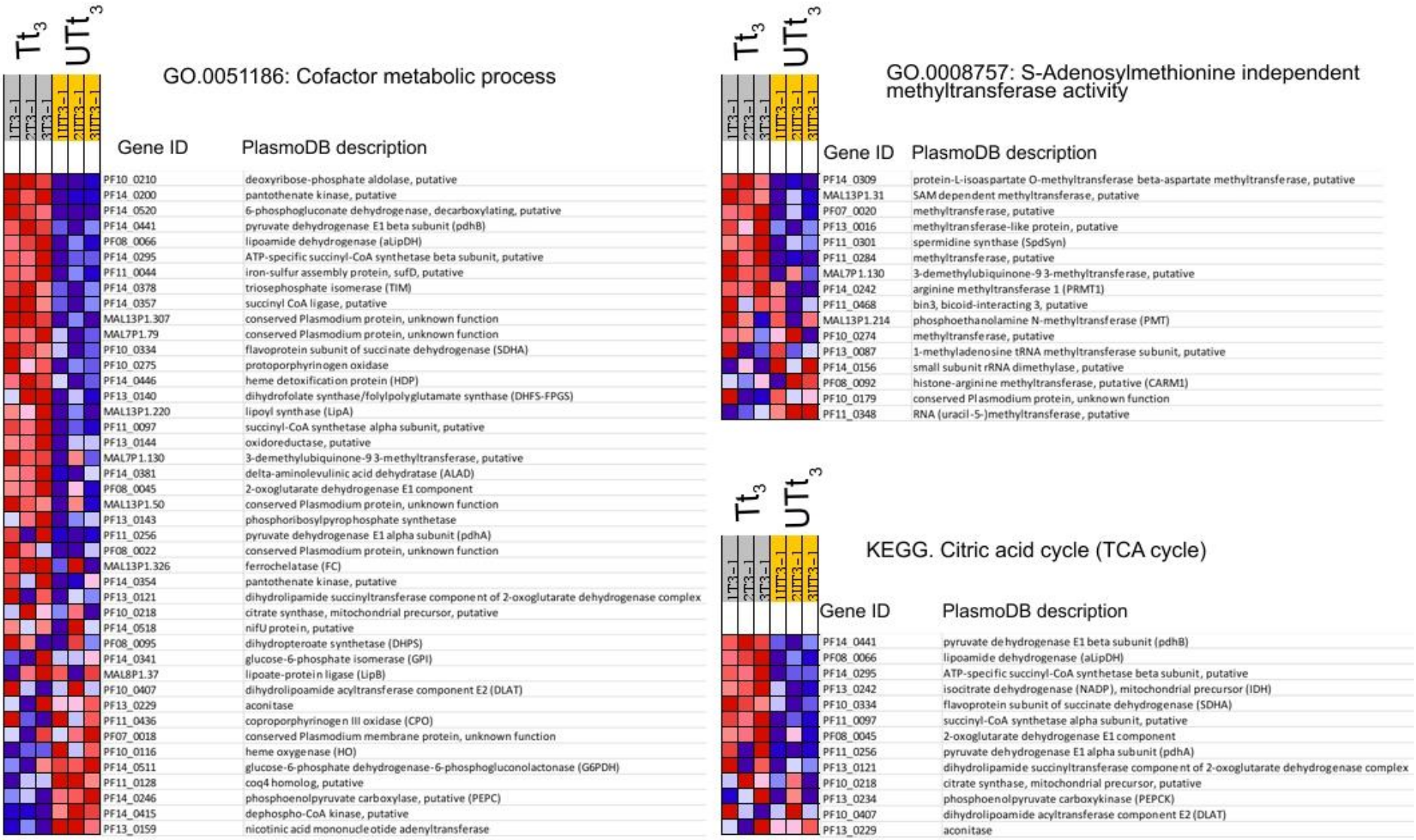

Figure 4.9: GSEA analysis of Tt_3 vs. UTt_3. Heat maps of gene sets found **to be enriched in GSEA. GSEA was calculated with $P \leq 0.05$ and had a false discovery rate (FDR) ≤ 0.25 (25%).** Red indicates increased transcript expression levels in the respective arrays, whereas blue indicates decreased transcript expression.

Transcripts within the purine metabolic pathway were also affected. Both MAL13P1.118 (3',5'-cyclic nucleotide phosphodiesterase, putative) and MAL8P1.150 (adenylyl cyclase beta, putative) had decreased expression at t_3 (Table 4A.1). The former transcript protein product is involved in the recycling of guanosine purine nucleosides, and converts 3', 5'-cyclic GMP (cGMP), a cell signalling metabolite, to GMP. In contrast, adenylyl cyclases convert ATP to 3', 5'-cyclic AMP (cAMP), a pivotal second messenger involved in signal transduction. AMP deaminase (MAL13P1.146), which convert AMP to IMP, had increased expression. Of the 14 genes which are responsive to oxidative stress (GO term: 0006979) only thioredoxin peroxidase 1 (Trx-Px1, PF14_0368) had increased abundance with a $\log_2$FC of 0.94 (Table 4A.1). A large proportion of transcripts which had decreased expression were kinases either involved in protein amino acid phosphorylation or in a phosphorous metabolic process (Table A4.1). A putative calmodulin transcript (PF10_0301) which had decreased expression was also correlated with phosphorylation. The predicted calmodulin transcript encodes for a small calcium-dependent protein critical for the activation of kinases.

4.3.6 Comparison to other transcriptomes

Transcriptomic fluctuations at the most prominent t_3 stage was compared to that of CQ treated *P. falciparum* (HB3), artesunate treatment of *P. falciparum* (FCR3) after 3h [262], febrile temperature changes (41°C, strain 3D7) [235], and a study which used WR99210 a lethal antifolate compound (Figure 4.10) [241]. The gene expression profiles were also compared to a large study in which 20 different compounds were tested on various late schizont stages of the parasites (3D7) [242]. These comparisons showed that 196 transcripts were uniquely expressed in 4PEHz-treated parasites (Figure 4.10). The majority of these unique transcripts, 54% (90/196), were hypotheticals with no definitive functions ascribed to the transcript products. Some transcripts (6.6%, 13/196) were associated with protein amino acid phosphorylation as well as cell cycle regulation and signal transduction pathways (Table A4.3 in the Appendix). Two unique transcripts are involved in cofactor biosynthetic processes; a putative hydroxyethylthiazole kinase (PFL1920c) and nicotinic acid mononucleotide adenyltransferase (PF13_0159). The former transcript is involved in intermediate steps leading to the production of thiamine pyrophosphate (TPP). The latter is involved in forming nicotinamide adenine dinucleotide (NAD^+).

Of the transcripts differentially expressed in t_3 during 4PEHz treatment, 14 corresponded to transcripts expressed during febrile temperature fluctuations. These were involved in protein phosphorylation, proteolysis and DNA replication. One hundred and twenty three transcripts

corresponded to CQ-treatment expression patterns, and these were involved in pathogenesis and proteolysis. Lipoamide dehydrogenase (aLipDH, PF08_0066) was also differently expressed in CQ-treated parasites, similar to 4PEHz treatment (Figure 4.10). Five hundred and twenty five transcripts also corresponded to a large 20-drug transcriptome study involving parasite-growth inhibitors such as atremisinin, quinine and trichostatin A (Figure 4.10) [242]. According to functional annotation clustering, these transcripts were mainly involved in protein amino acid phosphorylation, as well as cell division and chromosome partitioning (results not shown, annotation in DAVID [260, 261]).

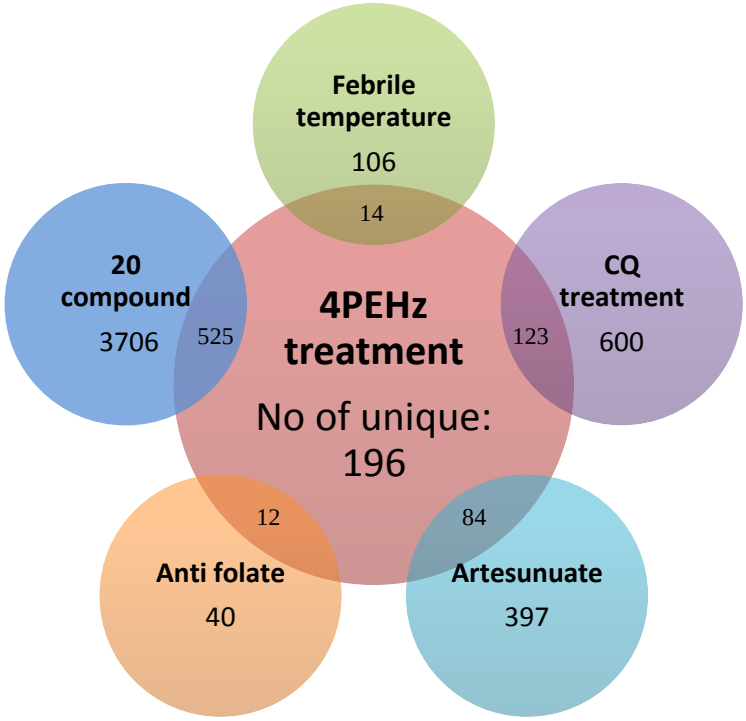

Figure 4.10: Venn diagram of differentially expressed transcripts during t_3 in 4PEHz-treated parasites. The differentially expressed were compared to transcriptome datasets of CQ, artesunate, febrile temperature treatment of *P. falciparum*.

The 4PEHz-treated transcriptome (525 transcripts) of the third time point was clustered, using hierarchical average linkage clustering, to corresponding transcripts from a 20-molecule growth inhibitor transcriptome dataset (Figure 4.11). Clustering revealed that 4PEHz had similar responses to staurosporine, PMSF and apicidin (Figure 4.12). A cluster of transcripts with decreased expression (Figure 4.11) were highly correlated with Pearson correlation factor of 0.93. This cluster of around 40 transcripts were ranked according to their GO-associated biological processes and were found to have involvement protein phosphorylation (GO:0006468), mismatch repair (GO:0006298) and response to DNA damage stimulus (GO:0006974). In the t_3 time point of 4PEHz-treated parasites several putative and known transcription factors had decreased expression (Appendix 4, Table 4A.1). The expression levels of these transcription factors in 4PEHz-treated

parasites were clustered with apicidin and staurosporine transcriptomes (Correlation 0.72, Figure 4.12). Transcripts with similar expression patterns in a sub-cluster (correlation 0.94) were associated with protein phosphorylation (*P* value = 0.008, GO term: 0043687), the cell cycle (*P* value = 0.018, GO term: 0007049) and cytokinesis (*P* value < 0.001, GO term: 0000910). A transcript encoding for a transcription factor, PF11_0442 (transcription factor with AP2 domains, AP2-O) had a decreased expression with $\log_2$FC of -2 in 4PEHz treated parasite, which corresponded with apicidin and staurosporine expression levels (Figure 4.12). Additional clustering was performed to determine the extent to which transcription factor expression was comparable from the different drug treatments (Figure 4.13). Expression of PFL0465c (zinc finger transcription factor (krox1) and PF11_0477 (CCAAT-box DNA binding protein subunit B) in 4PEHz-treated parasites was similarly decreased compared to apicidin transcript levels (Figure 4.13). In a study by Chaal *et al.* several Apicomplexan AP2 (ApiAP2) putative transcription factor proteins, which are thought to play a role in regulating stage specific gene expression, were shown to be affected by apicidin [263]. These results therefore also link up with transcriptome data from Hu *et al.* (Figure 4.13) suggesting disregulation of ApiAP2 transcription factors, moreover suggests that 4PEHz affects the same cluster of transcriptional regulators [242]. Through attenuation of PLP biosynthesis these transcription factors may be disregulated.

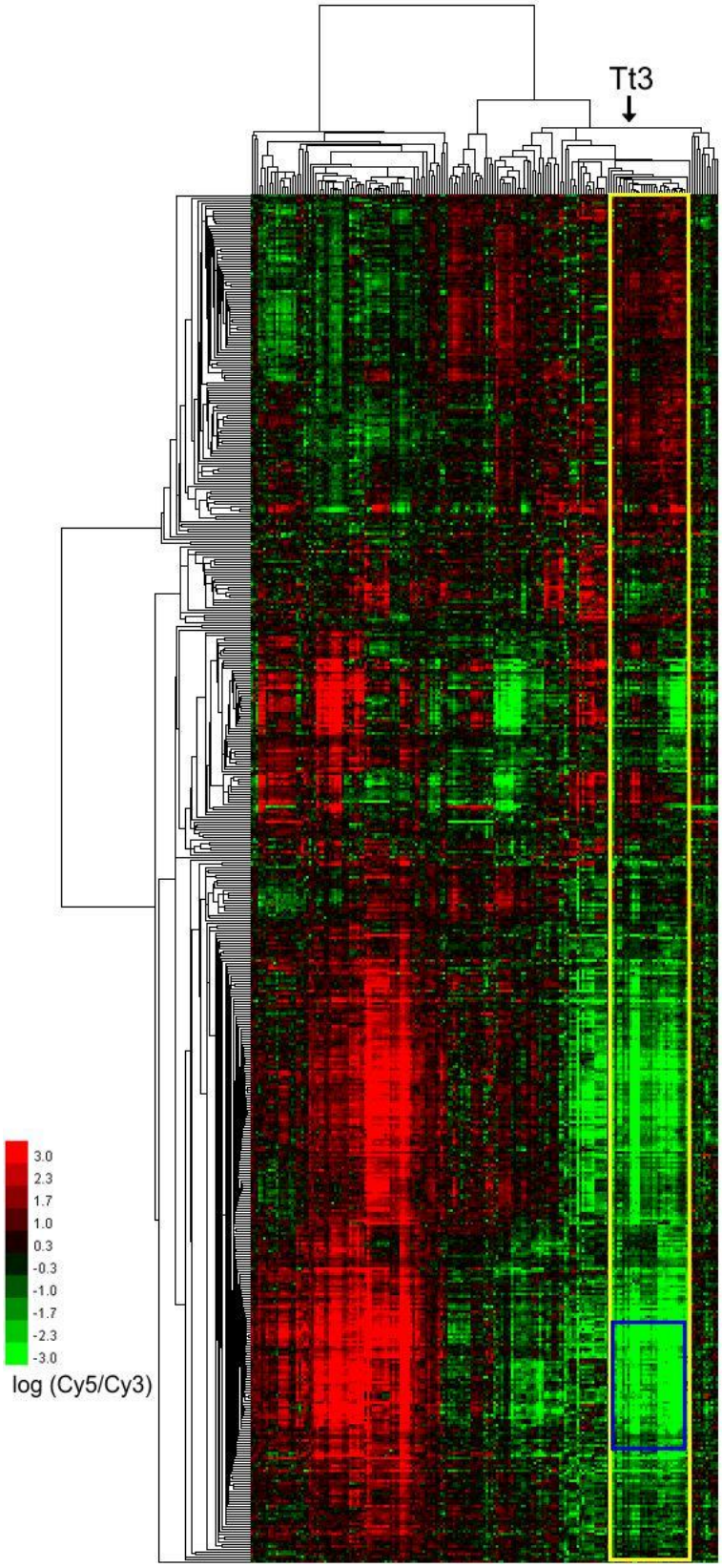

Figure 4.11: Transcriptional responses of *P. falciparum* treated with 4PEHz compared to other parasite growth inhibitors. The third time point t_3 (UT3 in figure) was clustered using hierarchical average linkage clustering with transcripts from a study involving 20 different growth parasite inhibitors. The boxed regions, which shows the correlation of 4PEHz responses, are expanded in the next figure. Clustering was performed in Cluster v 2.11 [255] and cluster trees were generated in TreeView v 1.1.6r2 [264, 265].

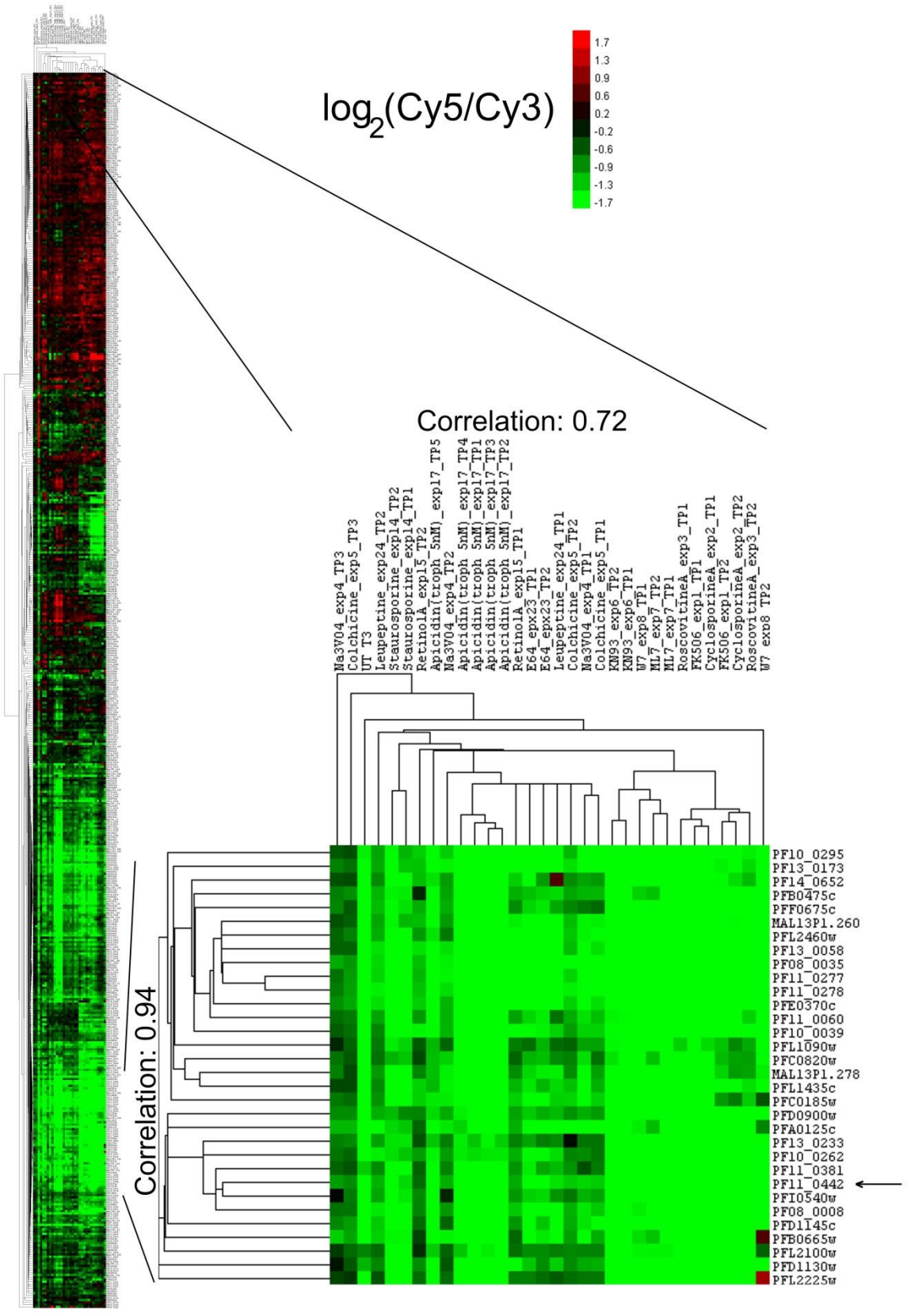

Figure 4.12: 4PEHz treament in t_3 was shown to correlate with apicidin and staurosporine treatment of parasites. Transcripts from 4PEHz-treated parasites in the third time point (UT3 T3) clustered with corresponding transcripts affected by staurosporine, apicidin and cyclosporine [242]. The expression pattern of a cluster (correlation: 0.94) revealed highly similar decreased expression patterns comparable between the treatments.

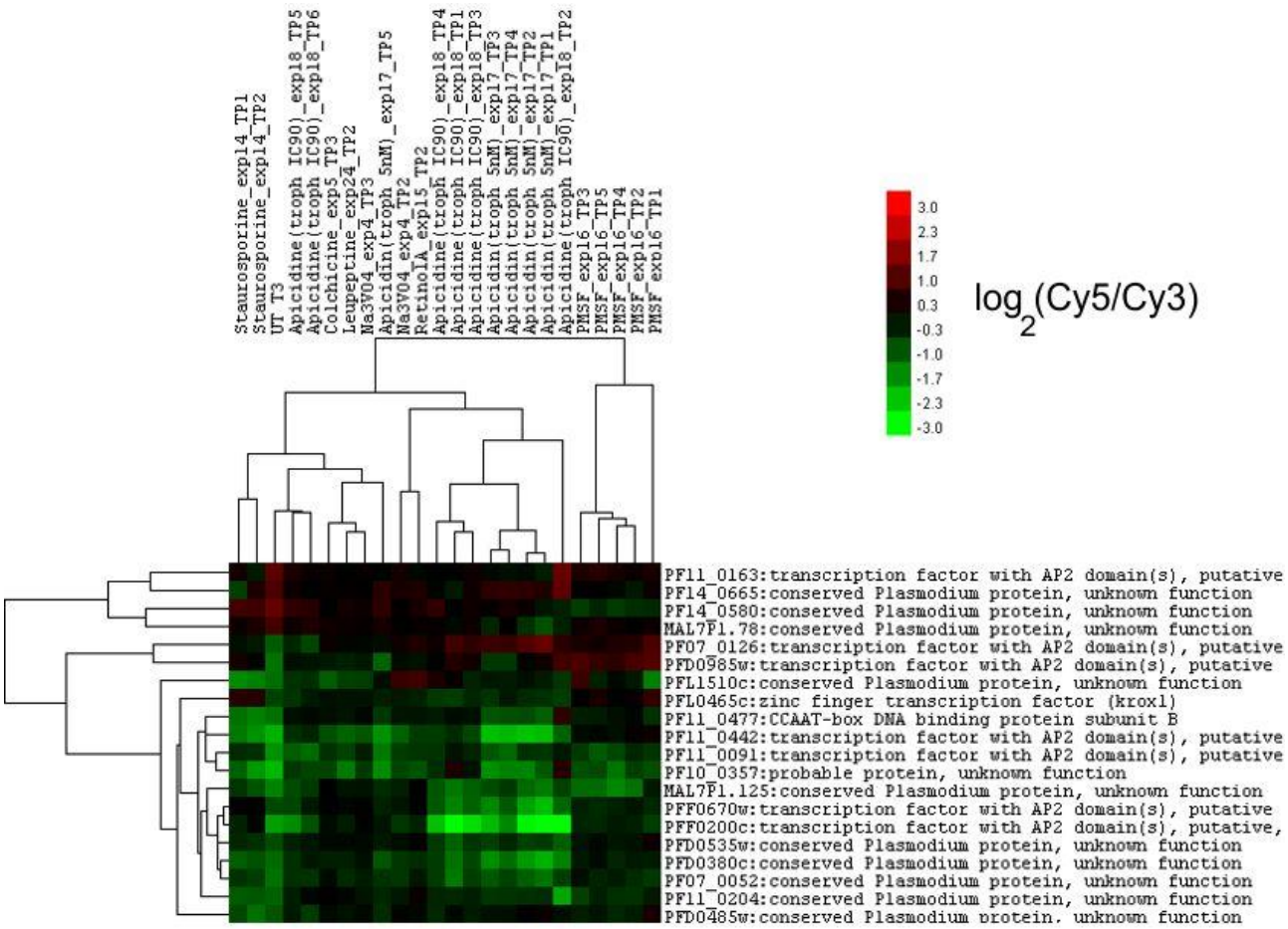

Figure 4.13: Transcription factors affected by 4PEHz-treatement. Several transcription factors with significant differential expression ($log_2FC > 0.75$ or < -0.75) in t_3 during 4PEHz-treatement were clustered with transcript expressions values from apicidin, PMSF and staurosporine-treated parasites [242]. Transcript expression was comparable to staurosporine as well as apicidin-related expression patterns, and the cluster had a correlation of 0.55.

4.3.7 Validation of microarrays using real-time PCR

Microarray data were validated using real-time PCR. The three reference target genes of decreased abundance; (PF14_0224), (PFB0695c), and (PFI1370c) were selected. Three transcripts which had increased expression in t_3 were selected; (PF11_0256), (PF14_0155) and (PF08_0066). Three housekeeping genes; LDH (PF13_0141), cyclophillin (cyc) (PFE0505w) and StSyn (PF07_0073) had geNorm scores values below 0.2 indicating that these reference genes had very stable expression, and had comparable expression in both treated and untreated parasites (Figure 4.14). From the expression of the reference genes expression in both treated and untreated parasites an expression normalisation factor is calculated, which is used to normalise expression of the targets relative to expression of the reference gene.

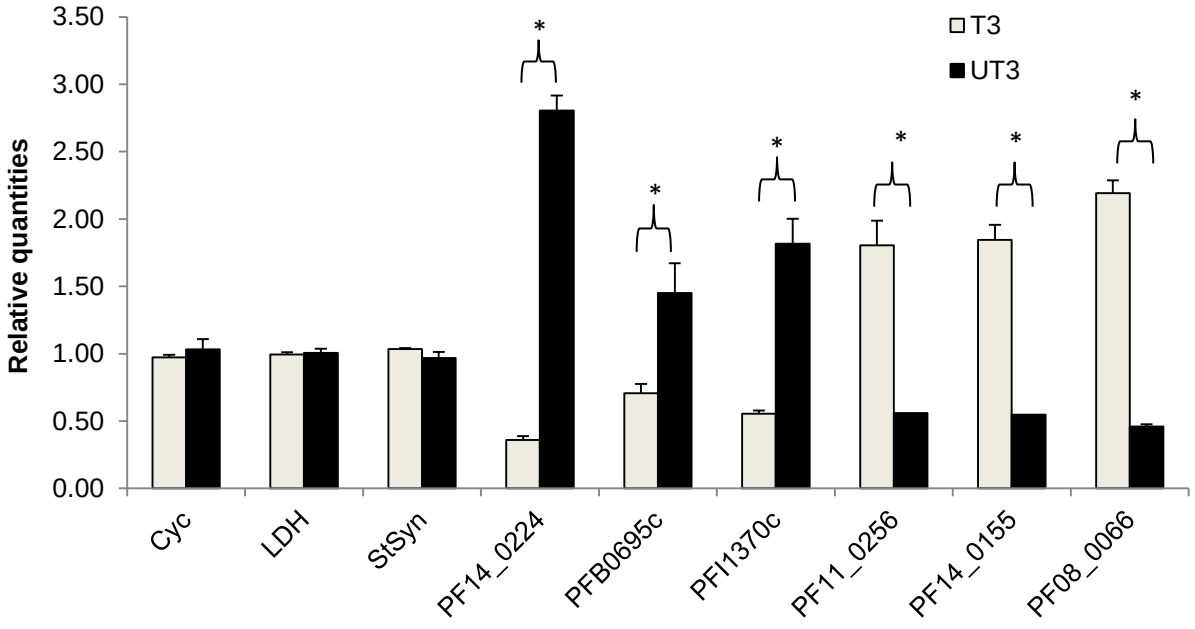

Figure 4.14: Real-time qRT-PCR validation of the t_3 time point transcript expressions from microarrays. Statistical significance is indicated with * which represents $P < 0.05$ in a two-tailed Student's t-test with unequal variance. Error bars indicate the SEM for two independent experiments performed in duplicate.

Six transcripts with differential expression in t_3 time points were selected for real-time PCR validation. Three target transcripts; PF14_0224, PFB0695c, and PFI1370c were had greater abundance in the UTt_3 treated parasite population compared to Tt_3 (Figure 4.14). The relative Tt_3:UTt_3 FC levels of these three transcripts were comparable to FC levels found in microarray comparisons (Table 4.8). Similarly, the three transcripts PF11_0256, PF14_0155 and PF08_0066 were shown to have greater levels in Tt_3 samples (Figure 4.14), and had Tt_3:UTt_3 FC values which correlated with microarray data (Table 4.8).

Table 4.8: Real-time PCR confirmation of microarray data. Six transcripts which had decreased expression and increased expression respectively were validated with real-time PCR. All six transcripts had FC levels comparable to microarray data from t_3.

Target transcript	PlasmoDB description	Real-time PCR Tt_3:UTt_3 FC	Tt_3:UTt_3 FC from microarray	
			FC	Log_2FC
PF14_0224	serine/threonine protein phosphatase	0.13	0.16	-2.65
PFB0695c	acyl-CoA synthetase, PfACS8	0.49	0.53	-0.91
PFI1370c	phosphatidylserine decarboxylase	0.31	0.52	-0.94
PF11_0256	pyruvate dehydrogenase E1 alpha subunit	3.23	3.13	1.65
PF14_0155	serine C-palmitoyltransferase, putative	3.38	2.79	1.48
PF08_0066	lipoamide dehydrogenase (aLipDH)	4.76	2.43	1.28

4.3.8 Proteome analysis

Proteins were extracted from 4PEHz treated and untreated *P. falciparum* parasites to study the effects of the compound on the parasite proteome. Protein extract samples were from the same treatment study used for microarray analysis. The proteins were separated on a 1D SDS-PAGE gel (Figure 4.15) after which the lanes were excised and submitted for MS-based proteomic quantitation and identification. A total of 33 unique proteins were detected in the Tt_3 sample which were not present in UTt_3 parasites. A summarised list is given in Table 4.7. A comprehensive list of proteins, with similarities scores and quantitative information regarding protein identity is given in Table A4.4 and Table A4.5 of Appendix 4.

GO annotation of the proteins present in 4PEHz-treated extracts suggested the involvement of proteins in single-stranded DNA specific endodeoxyribonuclease activity (GO:0000014), RNA-dependent DNA replication (GO term: 0006278), arginine biosynthetic (GO term: 0006526) as well as pyrimidine base biosynthesis (GO term: 0019856) (Appendix 4, Table A4.4). Some GO processes correlated with transcriptome data, and include activation of DNA repair processes with an uncharacterised conserved *Plasmodium* protein (PF13_0155) which is predicted to have involvement in double strand break repair (GO:0006302) and a putative ATP-dependent DNA helicase (PF14_0278). Carbamoyl phosphate synthetases (CPS), which are involved in L-glutamine catabolic processes, were also present in the treated protein fractions. This enzyme utilises ammonia to form carbamoyl phosphate from ATP and CO_2.

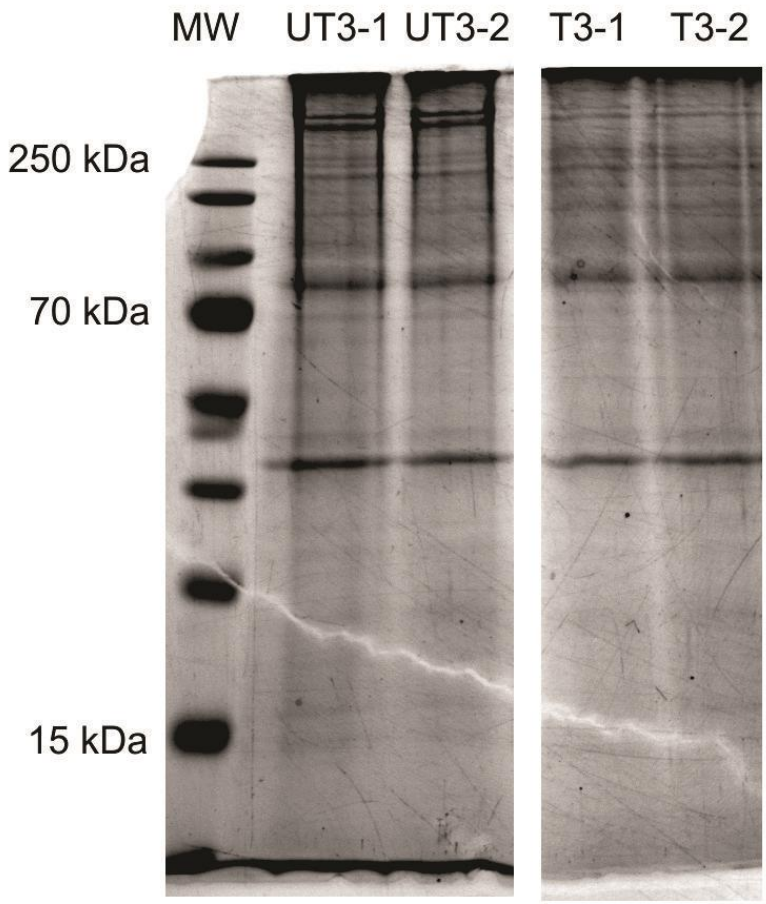

Figure 4.15: SDS-PAGE of protein extracted from untreated and treated 4PEHz parasites at time point t_3. Proteins were extracted from 4PEHz-treated and untreated parasites at time point t_3 of the microarray study. Proteins were quantified using a copper-based colourimetric assay, and samples containing 60 µg of protein were loaded and separated on a SDS-PAGE gel. Both lanes of the each sample were excised and submitted from MS-based protein identification.

Adenylyl cyclase beta (MAL8P1.150) protein was detected in the treated fractions, whereas this particular transcript had decreased expression in transcriptomic profiles of the t_3 stage (Table 4.9). Some of the proteins which correlated with transcriptomic expression patterns included a phosphoinositide-binding protein (MAL7P1.108) The catalytic subunit of DNA polymerase epsilon was also present in the treated protein fraction, corresponding to observed transcript expression data (Table 4.9). These results could suggest that expression of these proteins are regulated at a transcriptional level, rather than post transcriptionally.

Table 4.9: Summarized list of proteins identified in 4PEHz-treated parasites. Some proteins corresponded to differentially expressed transcripts from Tt_3 vs. UTt_3. For a comprehensive list of proteins see Appendix 4, Table A4.4 and Table A4.5.

PlasmoDB ID	Present in Tt_3 vs. UTt_3 transcitpomic data (log_2FC)	PlasmoDB Description
MAL13P1.278	-2.23	serine/threonine protein kinase, putative
MAL7P1.108	-0.94	phosphoinositide-binding protein, putative
MAL8P1.150	-1.50	adenylyl cyclase beta, putative
PF10_0184	-1.89	conserved *Plasmodium* protein, unknown function
PF10_0262	-3.35	conserved *Plasmodium* protein, unknown function
PF11_0417	-1.06	conserved *Plasmodium* protein, unknown function
PF13_0079	-2.56	conserved *Plasmodium* protein, unknown function
PF14_0123	1.12	conserved *Plasmodium* protein, unknown function
PFE0440w	-1.52	conserved *Plasmodium* protein, unknown function
PFF1470c	0.77	DNA polymerase epsilon, catalytic subunit a, putative
PFL0030c	-1.12	erythrocyte membrane protein 1, PfEMP1
PFL2505c	-2.07	conserved *Plasmodium* protein, unknown function
PF13_0080	-	telomerase reverse transcriptase, putative
PF13_0044	-	carbamoyl phosphate synthetase

4.3.9 Interaction networks

None of the proteins detected during 4PEHz treatment formed part of *Pf*Pdx1 or PLP association networks. Therefore, functional linkages were inferred from the transcriptome components of the interactome, and were evaluated based on their transcript expression patterns. Similar correlations from gene expression profiles were used to create a comprehensive interaction network in *P. falciparum* parasites [266]. Using STRING *(*Search Tool for the Retrieval of Interacting Genes/Proteins, v 9.0) [267, 268], most of the transcripts previously listed as PLP-dependent enzymes were flagged as part of the interaction network (Table 4.10). The discovery of Pdx1 and Pdx2 is relatively recent, and for this reason *Pf*Pdx2 has not been classified as an interacting partner of *Pf*Pdx1 in STRING (both *Pf*Pdx1 and *Pf*Pdx2 are not present in PlasmoMAP either). Interacting partners were identified based on previously reported literature or neighbouring transcripts located near to Pdx1, and are included in the interaction network. The predicted and known interacting proteins of *Pf*Pdx1 are shown in Figure 4.16 A, and listed in Table 4.10. There was an overall enrichment of genes associated with *Pf*Pdx1, with the majority of these having increased expression during t_3, however not statistically significant (or above log_2FC of 0.75). Phosphoglycerate kinase (PGK, PFI1105w) were found to be part of the interaction network of *Pf*Pdx1. This enzyme functions in glycolysis to reversibly convert 1,3-bisphosphoglycerate to 3-phosphoglycerate, and follows after the conversion of G3P to 1,3-bisphosphoglycerate by glyceraldehyde-3-phosphate dehydrogenase. This provides evidence that processes indirectly associated with the *Pf*Pdx1 substrate G3P were affected by 4PEHz.

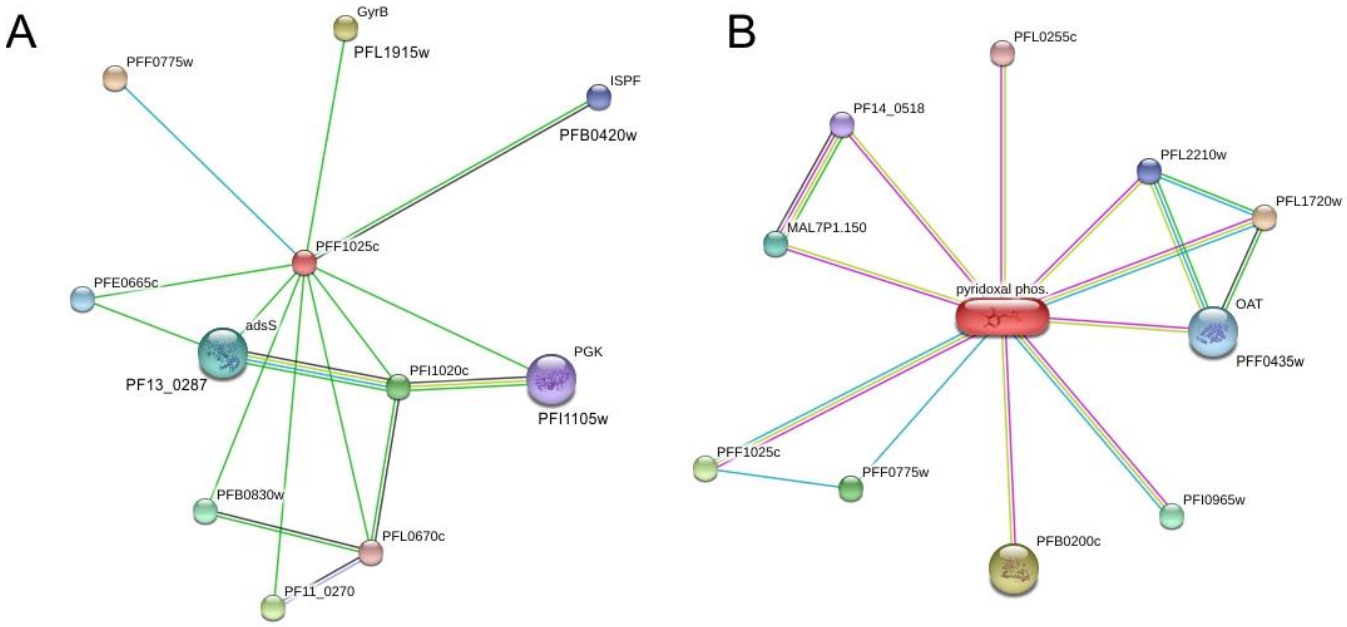

Figure 4.16: Predicted interaction networks of *Pf*Pdx1 and PLP. A) Interacting partners of *Pf*Pdx1 were identified in STRING. These interactions are based neighbouring locations of the transcripts in other organisms, as well as from derived from literature. **B)** The PLP chemical interaction network as predicted in STITCH. These nodes are linked based on neighbouring locations of genes in other genomes, such as SHMT (PFL1720w) and delta-aminolevulinic acid synthetase (PFL2210w). Putative homologues of OAT (PFF0435c) and delta-aminolevulinic acid synthetase (PFL2210w) in *Chlorobium luteolum* are predicted to interact with another [269].

Additional interaction partners which utilise PLP were identified using STITCH, which link chemical interaction groups with proteins and gene information (Table 4.10 and Figure 4.16 B) [269, 270]. PLP-dependent enzyme transcripts have been previously discussed, however STRING identified additional interacting partners based on association in curated databases as well as neighbouring locations of genes in other genomes. Of the predicted transcripts associated with PLP a putative pyridoxal 5'-phosphate dependent enzyme class III (PFI0965w) and SHMT (PFL1720w) were found to be enriched, having low level increased transcript expression. Delta-aminolevulinic acid synthetase (PFL2210w, ALAS) and SHMT (PFL1720w) are situated on the same loci in α-proteobacteria [269], however in *P. falciparum* the expression pattern of these two transcripts suggests that these are not co-regulated (Table 4.10).

Table 4.10: Predicted interacting partners of *Pf*Pdx1 (PFF1025c) and PLP. Transcript and proteins which form part of the *Pf*Pdx1 interaction network were identified using STRINGS. Interaction scores are derived from combined hits of the PFF1025c query based on neighbouring genes, text and database mining results as well as co-expression queries.

PlasmoDB ID	PlasmoDB description	log_2 FC Tt_3:UTt_3	Adj. P-value	Interaction score	Predicted GO process involvement
Interacting partners of *Pf*Pdx1 (PFF1025c) (STRINGS predictions)					
PFF0775w	pyridoxine kinase (PdxK)	0.234	0.117	0.899	salvage of pyridoxine, vitamin B6 biosynthetic
PFL1915w	DNA gyrase subunit B (GyrB)	0.588	0.013	0.717	DNA topological change
PF11_0270	threonyl-tRNA synthetase,Threonine--tRNA ligase (ThrRS)	0.316	0.067	0.596	tRNA aminoacylation,
PFI1020c	inosine-5'-monophosphate dehydrogenase	-0.183	0.264	0.595	oxidation-reduction process
PFB0830w	40S ribosomal protein S26e, putative	0.285	0.121	0.575	translation
PF13_0287	adenylosuccinate synthetase (adsS)	0.313	0.038	0.551	purine nucleotide biosynthetic process
PFE0665c	GTP binding protein, putative	0.574	0.002	0.532	barrier septum formation
PFB0420w	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF)	0.137	0.429	0.478	terpenoid biosynthetic process
PFI1105w	phosphoglycerate kinase (PGK)	0.719	0.003	0.463	glycolysis
PFL0670c	bifunctional aminoacyl-tRNA synthetase, putative	0.304	0.038	0.455	prolyl-tRNA aminoacylation
Interacting partners which utilise PLP (STITCH predictions)					
PFL1720w	serine hydroxymethyltransferase (SHMT)	0.469	0.004	0.994	L-serine and glycine metabolic process
PFB0200c	aspartate aminotransferase,aspartate transaminase (AspAT)	0.188	0.371	0.993	biosynthetic process
PFF1025c	pyridoxine biosynthetic enzyme (Pdx1)	-0.155	0.305	0.977	pyridoxal phosphate biosynthetic process
PFF0775w	pyridoxine kinase (PdxK)	0.234	0.117	0.957	null
PFI0965w	pyridoxal 5'-phosphate dependent enzyme class III, putative	0.471	0.038	0.952	null
MAL7P1.150	cysteine desulfurase, putative (NFS)	-0.348	0.050	0.934	metabolic process
PFF0435w	Ornithine aminotransferase (OAT)	0.120	0.415	0.92	null
PFL2210w	delta-aminolevulinic acid synthetase (ALAS)	-0.267	0.049	0.914	tetrapyrrole biosynthetic process
PF14_0518	nifU protein, putative	0.274	0.063	0.798	iron-sulphur cluster assembly
PFL0255c	UGA suppressor tRNA-associated antigenic protein, putative	0.427	0.080	0.787	null

4.4 Discussion

PLP is a vital co-factor in more than 100 different biochemical reactions, and perturbations in PLP biosynthesis are therefore expected to be severely detrimental to *P. falciparum* parasite growth. PLP plays a role in amino acid metabolism, and acts as a prosthetic group for enzymes predominantly involved amino acid metabolism. In the malaria parasite, PLP is required for the decarboxylation of L-ornithine to putrescine by ODC, where putrescine is subsequently used as scaffold diamine during the formation of polyamines spermidine and spermine [271]. Moreover, parasitic polyamine levels have been correlated to IDC of the parasite during asexual maturation, and depletion of polyamines results in parasite cytostasis [203, 272]. This further underscores the importance of PLP biosynthesis in the parasites. Apart from metabolic processes reliant on PLP, a recent antioxidant role has been ascribed to this molecule and is involved in elimination of ROS [141]. Considering the ubiquitous involvement of PLP in a multitude of different cellular processes, disruptions in PLP biosynthesis could easily result in escalated cellular stress responses. The involvement of PLP in combating oxidative stress would similarly suggest stress-related perturbations would be affected due to diminished PLP levels.

Consequences of PLP deficiency in humans include high homocysteine and cystathionine levels during methionine loading [273]. This was attributed to cystathionine β-synthases and cystathionase, both which are PLP-dependent enzymes which function to remove homocysteine. There is currently no evidence for the existence of similar proteins in *P. falciparum* parasites. In yeast (*S. cerevisiae*) depletion of PLP through inhibition of PdxK and knock-out of *de novo* PLP biosynthetic enzymes Pdx1 and Pdx2, resulted in DNA lesions and DNA damage responses [274]. This was thought to be related to reduced formation of N^5,N^{10}-methylene tetrahydrofolate (CH_2-THF) by SHMT, which in turn lead to reduced dTMP levels. These disruptions in nucleotide biosynthesis resulted in accumulation of uracil, and subsequently over-incorporation of this nucleobase into DNA, resulting in DNA lesions and genome instability [274].

Microarrays were utilised to determine the extent of transcriptome level responses caused by 4PEHz treatment. A common reference probe design was used in the microarrays. The reference pool consisted of cDNA from all the T and UT samples, and allows comparison of an individual transcript expression level from each condition in a separate array to a common reference. This indirect method is advantages as transcripts from different arrays, in this case time points as well as treatments, can be accurately compared. Normalisation of the common reference pool, thereby dye intensity, is necessary to account for between-array variations.

A large number of transcripts (787) were affected during the last t_3 time point, corresponding to around 10% of the known transcriptome. Previously classes of small molecule inhibitors that induced similar large changes (between ~5 – 10%) of the transcriptome included EGTA, trichostatin A, staurosporine, and apicidin [242]. These molecules were suggested to arrest the IDC of the parasites, with the exception of apicidin which was thought to deregulate the IDC owed to its HDAC inhibitory action [242]. Moreover, clustering of the transcript expression data from 4PEHz-treated parasites revealed gene expression correlations compared to apicidin and staurosporine treatment of schizonts [242]. The systematic changes in the transcriptome of 4PEHz-treated parasites from a few affected transcripts during the first two time points followed by large scale perturbations during the last time point, comparable to IDC arrest, suggested that 4PEHz had a profound effect on the parasites. These disruptions could have further led to transcriptional deregulation. This underscored the importance of PLP metabolism, more specifically *de novo* biosynthesis within the parasite.

As highlighted in during functional annotation and clustering of differential expressed genes in t_3, more than 10 transcription factors had decreased expression during 4PEHz treatment, and some of these transcript expression levels correlated to apicidin-treated parasites. PLP post-translationally affects the activity of transcription factors [275, 276], and also some transcription factors specific to PLP metabolism which modulate the expression of the Pdx1 and Pdx2 genes, as explained below [277]. PLP had been shown to inhibit the DNA binding capability of the HNF1 transcription factor in *E. coli* [276]. Moreover, PLP is post-translationally conjugated to mouse nuclear receptor interacting protein 140 (RIP140), and increases the co-repressive transcriptional activity, affecting gene expression [275]. The conjugated RIP140 was also shown to interact with HDAC [275]. Decreases in PLP levels of 4PEHz-treated parasites could either result in activation of certain transcription factors, or diminish of potential co-repressive functions (enhance activation), as expected for RIP140. The overall activation of transcription factors was reflected as compensatory responses which aimed to decreased transcription of transcription factors during 4PEHz treatment. This also provides a link between apicidin and 4PEHz treatment of parasites, these molecules could affect the same transcription factor, resulting in similar transcriptomics responses.

Wide-spread transcriptional perturbations during 4PEHz treatment of parasites could be linked to DNA damage repair processes. Topoisomerases regulate the topological states of DNA during transcription and repair [278]. PLP has been postulated to regulate topoisomerases and was shown to reversibly complex with DNA topoisomerases I in *Candida guilliermondii* by binding to the ε-amino group of an active site lysine [278]. Binding of PLP inhibits the DNA cleavage reaction catalysed by the enzyme. Depletion of PLP in 4PEHz-treated parasites could have led to release of

PLP from topoisomerases, activating DNA cleavage. The transcriptional responses were reflected as decreased expression of DNA topoisomerases (MAL13P1.328, DNA topoisomerase VI, b subunit, putative), and supports that DNA damage could have been caused during PLP depletion.

Transcripts of PLP-dependent enzymes were affected by 4PEHz treatment (Figure 4.17). This implies that transcriptomic perturbations could be functionally linked to PLP. This confirms that 4PEHz affected PLP metabolism and lends support to specific targeting of *Pf*Pdx1 by 4PEHz *in vitro*. Of the PLP-dependent enzyme transcripts a putative SHMT (PF14_0534) transcript had increased expression by 3-fold during the first 12 h of the treatment period. SHMT is involved in one carbon transfer reactions, and utilises serine to generate the CH_2-THF cofactor (Figure 4.17). There are two predicted SHMT isoforms in *P. falciparum;* PFL1720w and PF14_0534. The former transcript encodes for the cytosolic form of SHMT or *Pf*SHMTc [279]. The PF14_0534 transcript, which had increased expression during 4PEHz treatment, has 18% identity compared to *Pf*SHMTc, and is the mitochondrial form of SHMT [279]. Both *Pf*SHMTc and *Pf*SHMTm have stage-specific localisation within the parasite, and *Pf*SHMTc has been shown to predominantly have cytosolic localisation throughout the parasite IDC [279]. *Pf*SHMTm localises in the mitochondria of the parasite, and also in the apicoplast during the early trophozoite and schizont stages [279]. Intra-erythrocytic expression of *Pf*SHMTm transcript was previously shown to peak at 28 HPI and the lowest expression levels were at 7 HPI in the 3D7 strain [225]. The deviation of SHMTm transcript from known expression patterns highlighted the importance of this transcript and its dependence on PLP. During the t_1 stage only 25 transcripts were differentially expressed, compared to the t_3 stage with more than 700 affected transcripts. This perturbation in SHMT transcript expression is therefore particularly interesting.

GSEA showed that 4PEHz treatment of *P. falciparum* parasites results in positive enrichment of transcripts involved in THF metabolism. Three other transcripts formed part of this gene set which included dihydrofolate synthase/folylpolyglutamate synthase (PF13_0140, DHFS-FPGS), putative glycine cleavage T protein (GCVT) (PF13_0345), and lipoamide dehydrogenase (aLipDH) (PF08_0066). *P. falciparum* possesses two LipDH isoforms; mitochondrial (mLipDH) and apicoplast (aLipDH). The aLipDH has an apicoplast localisation signal, even though it is annotated to be involved in folate metabolism, aLipDH forms part of the PDC in the apicoplast [280, 281]. Mitochondrial SHMT provides glycine to the glycine cleavage complex (GCV) or glycine decarboxylase complex to generate additional THF [279, 282]. The increased enrichment of the putative glycine cleavage T protein (GCVT) transcript which forms part of the GCV also suggested compensatory THF metabolic responses. CH_2-THF donates a carbon in a reaction catalysed by thymidylate synthases (TS) which converts dUMP to dTMP. Recycling of the oxidised

dihydrofolate from TS is accomplished by dihydrofolate reductase (DHFR), which in *P. falciparum* is a component of a bifunctional complex consisting of DHFR and TS [283]. Surprisingly at t_3 the transcripts of the bifunctional DHFR-TS had increased expression to around 1.7-fold. These results suggests an increased requirement for folates and recycling thereof during 4PEHz treatment, and therefore presumably PLP depletion.

One consequence of PLP depletion is DNA lesions caused by disruptions in dTMP biosynthesis which is further associated with THF biosynthesis [274]. In 4PEHz-treated parasite these disruptions were linked to affected folate biosynthesis by SHMT and down-stream folate utilisation by TS. Disruptions in folate metabolism of parasites treated with 4PEHz could have been due to PLP depletion, resulting in decreased levels of CH_2-THF produced by SHMT, thereby affecting down-stream folate recycling processes. The sequential series in which transcript deviations were observed agrees with this notion. SHMT had increased expression during the t_1 time point, followed by compensatory increased in folate-related transcripts only at the t_3 time point. The onset of PLP depletion could have been rapid, triggering early compensatory responses, followed by a more dramatic rescue response which attempted to balance folate homeostasis.

*Pf*Pdx2 transcripts had increased expression during 4PEHz treatment (Figure 4.17). The *Pf*Pdx1 transcript was not shown to have significant transcript fluctuations throughout the monitored life cycle period. *Pf*Pdx1 and *Pf*Pdx2 are obligate heterodimers which assemble into the PLP synthase complex and the cellular abundance of each counterpart is expected to affect the activity of the other. Increased *Pf*Pdx2 transcript levels demonstrated that a metabolic perturbation can have feedback and results in transcript compensation directly affiliated with the producing enzyme. Presumably increased *Pf*Pdx2 transcript levels aims to increase *Pf*Pdx2 protein abundance, which would have an effect on the activity of the PLP synthase complex that produces PLP. *Pf*Pdx2 monomers are activated by *Pf*Pdx1, which should be inhibited by 4PEHz, potentially resulting in wasteful hydrolysis of L-glutamine and the generation of excess ammonia, and results supporting this speculation are discussed further below. Transcripts of pyridoxine/pyridoxal kinases were not affected by 4PEHz treatment, also corroborating with the notion that salvage of unphosphorylated PN and PL minimally contributes towards intracellular PLP levels (unpublished experiments, C. Wrenger).

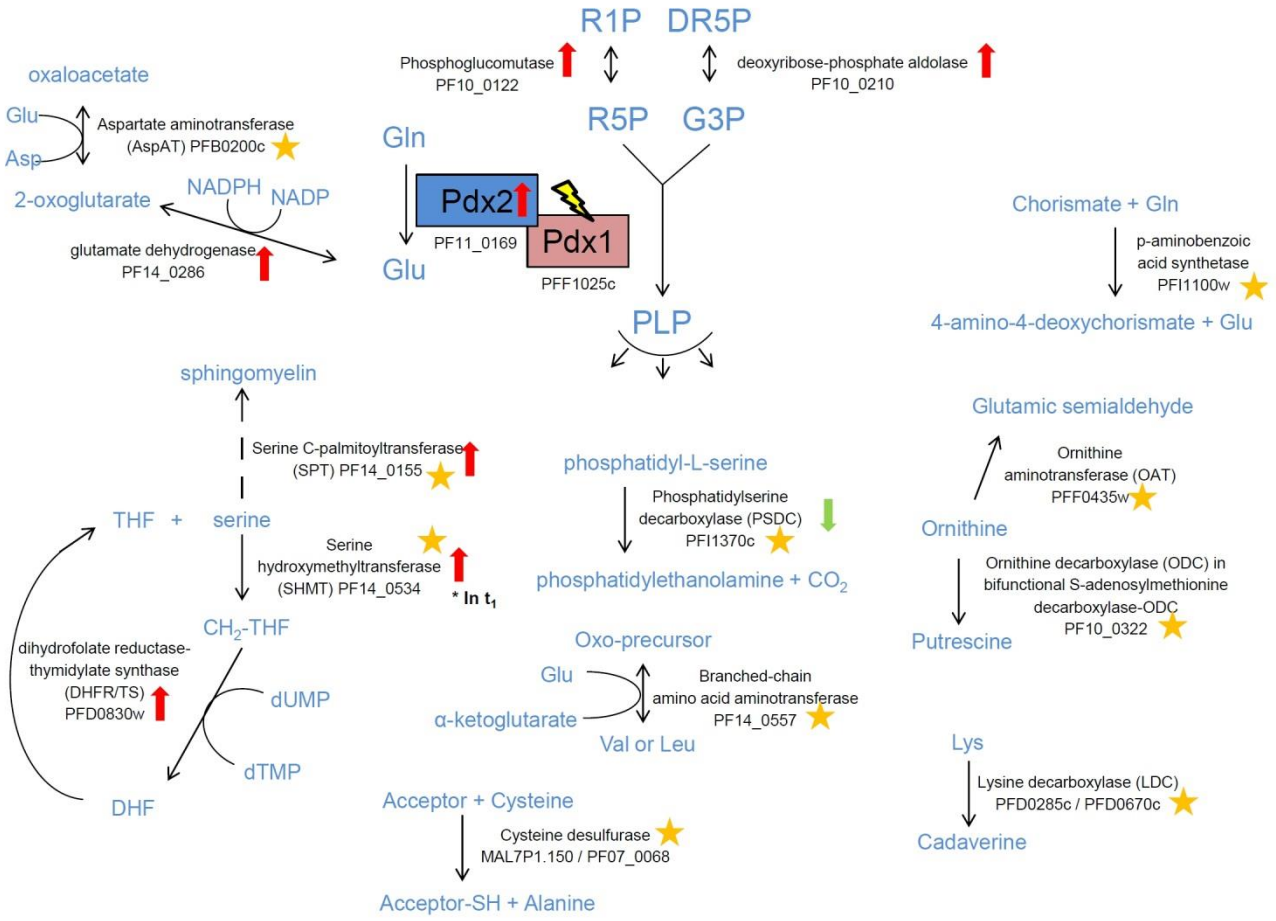

Figure 4.17: Effect of 4PEHz on PLP-dependent enzyme transcripts as well as Pdx1-related transcripts. The protein products of the transcripts indicated by a star are PLP-dependent enzymes. Arrows indicate the transcript expression during time point t_3 (except for SHMT which had increased expression in t_1).

In this study, 4PEHz was shown to affect *Pf*Pdx2 transcript levels, with no statistical significant changes in *Pf*Pdx1 transcript abundance. Several studies have shown that *P. falciparum* *Pf*Pdx1 and *Pf*Pdx2 are regulated transcriptionally, and often co-expressed [141, 242, 284]. Both *Pf*Pdx1 and *Pf*Pdx2 transcripts had increased expression in staurosporine and apicidin transcriptomes, and in the majority of time points sampled showed that these transcripts were co-regulated [242]. In response to cercosporin, parasites also exhibited increased expression of both *Pf*Pdx1 and *Pf*Pdx2 transcripts [141]. Isoleucine starvation also resulted in patterns of co-regulation of both *Pf*Pdx1 and *Pf*Pdx2 transcripts throughout the IDC and these transcripts had decreased expression compared to normal culture conditions [284]. The *Pf*Pdx1 and *Pf*Pdx2 transcripts were not significantly affected when parasites were exposed to increased temperature [235] or during hyperoxia [234], which suggests that these are not involved during general stress responses. Additional investigations into the co-regulation of *Pf*Pdx1 and *Pf*Pdx2 and the associated triggers involved in transcription expression could therefore be of interest. Interestingly, inhibition of both enzyme activities in bifunctional

*Pf*AdoMetDC/ODC resulted in increased *Pf*Pdx1 protein levels, however *Pf*Pdx2 was not detected [259]. This also supported the view that *Pf*Pdx1 could be regulated translationally.

PLP is capable of transcriptionally regulating expression of Pdx1 and Pdx2 transcripts, as well as other PLP-dependent enzymes, and this is thought to be mediated by specific transcription factors. A transcription factor termed PdxR was discovered in *Corynebacterium glutamicum* ATCC 13032 (NCBI Sequence ID: NP_600015.1) and is adjacent to Pdx1 and Pdx2 genes, *PdxS* and *PdxT*, respectively [277, 285]. Inactivation of PdxR resulted in decreased expression of these PLP synthase genes [277, 285]. DNA mobility shift assays confirmed that PdxR was capable of binding in the PdxR-Pdx1/Pdx2 intergenic region, however how PdxR senses cellular levels of PLP to affect biding could not be established [277]. In *Streptococcus pneumonia* the transcript expression of Pdx1 was decreased by exogenously supplied PLP and the presence of a PdxR-like transcription factor was shown to modulated these expression levels [286]. Another example of a transcription factor which is activated by PLP is the GabR transcription factor in *B. subtilis* [287]. GabR elegantly communicates PLP levels and affects the transcription of *GabT*, which encodes GABA-aminotransferases - a PLP-dependent enzyme [287, 288]. In the presence of PLP and GABA the GabR mediated expression of the *GabT* promoter [288]. These transcription factors, PdxR and GabR , share protein architecture and contain a short helix-turn-helix (HTH) domain, which is believed to bind DNA, with a C-terminal aminotransferase-like domain, which could mediate sensing of PLP [288]. The common HTH-fold linked to a PLP-recognising domain suggests that these transcription factors have similar functions to relay the cellular status of PLP concentrations to genes involved in PLP metabolism. BLAST analysis using PlasmoDB did not reveal any positive hits for PdxR-like or GabR transcription factors with similar sequences in *P. falciparum,* however does not exclude the possibility that a similar transcription factor might be present. *Pf*Pdx1 and *Pf*Pdx2 are situated on distant loci in *P. falciparum* (*Pf*Pdx1 on chromosome 6, *Pf*Pdx2 on chromosome 11). Further characterisation of upstream binding elements of *Pf*Pdx1 and *Pf*Pdx2 might reveal a common binding element, such as PdxR in *C. glutamicum,* which could regulate *Pf*Pdx1 and *Pf*Pdx2 expression. Additionally, uncharacterised sequences in *P. falciparum* containing putative aminotransferase or HTH domains could form part of a transcription factors which are able to recognise and relay cellular PLP levels.

The correlation between 4PEHz-treated and apicidin transcriptomes may be linked through a common transcription factor which is affected, such as an ApiAP2 family protein. ApiAP2 proteins are involved in stage-specific gene expression [263], however it is not known whether these are affected by PLP. Two independent microarray studies have also shown that apicidin growth inhibition of *P. falciparum* parasites results in increased transcription of either *Pf*Pdx1 or *Pf*Pdx2 or

both [242, 263]. The disregulation of transcription through inhibition of HDAC could affect a common transcription factor, which activates or derepresses PLP biosynthesis.

PLP biosynthesis is part of antioxidant responses in the parasite; more specifically is essential in response to ROS [141]. Interestingly, DHA resistance in *P. falciparum* parasites is associated with increased transcript expression of *Pf*Pdx2 compared to non-resistant cell lines [238]. Part of the mechanism of action of DHA is the release of singlet oxygen, which leads to formation of peroxides [289, 290]. Increased PLP levels could therefore play an essential role in resistance to DHA. This could also motivate the use of *Pf*Pdx1 or PLP synthase inhibitors in combination with DHA to prevent possible resistance development. Glutamate dehydrogenases involved in the synthesis of glutamate from 2-oxoglutarate and L-glutamine in a reaction requiring NADPH had increased expression due to 4PEHz treatment. This suggested that the transcripts associated with L-glutamine catabolism were more active, possibly to compensate for glutamine accumulation generated through lowered utilisation of the substrate by *Pf*Pdx2.

In this study transcripts of the pentose phosphate pathway (PPP) were also affected by 4PEHz. The deoxyribose-phosphate aldolase transcript which had increased expression encodes for an enzyme that reversibly converts G3P into 2-deoxy-D-ribose 5 phosphate (Figure 4.17). This could have been the cellular response towards increased G3P levels due to decreased utilisation of the substrate by *Pf*Pdx1. Another transcript of the PPP affected was phosphoglucomutase, which catalyses the interconversion of α-D-ribose 1-phosphate to R5P (Figure 4.17). This enzyme also catalyses the interconversion of α-D-glucose 1-phosphate to D-glucose 6-phosphate. Presumably disrupted R5P homeostasis could have led to increases in this transcript. During GSEA other transcripts with lower differential increased expression of the PPP were identified. A putative 6-phosphogluconate dehydrogenase (PF14_0520) and TIM are predicted to have involvement in the conversion of 6-phosphogluconate to Ru5P and isomerisation of G3P into DHAP. These perturbations in the PPP are directly linked to PLP metabolism through the substrate dependence of *Pf*Pdx1.

Other PLP-dependent enzymes affected by 4PEHz were putative serine C-palmitoyltransferases (SPT, PF14_0155) which had increased expression around 2.8-fold. The enzyme condenses serine with palmitoyl CoA to form 3-ketodihydrosphingosine, which is the precursor required to produce sphingomyelin. Sphingomyelin is believed to be one of the most prominently produced sphingolipids in *P. falciparum* [291]. SPT is a key enzyme that regulates not only sphingomyelin but also other sphingolipid levels within cells, and these are involved in signal cascades that can stimulate apoptosis or cell growth [291-293]. Moreover, previous evidence has suggested SPT are transcriptionally up-regulated in response to cellular stress including endotoxins [292]. SPT could

therefore be directly stimulated by PLP depletion, resulting in diminished levels of sphingomyelin precursors, which leads to activation or deactivation of signalling pathways. Of all the transcript with decreased expression during t_3 0.7% were involved in signalling pathways, however these were not sphingolipid-dependent, and mostly involved small GTPase mediated signal transduction factors.

Phosphatidylserine decarboxylase (PSDC, PFI1370c) had 1.9-fold decreased expression during t_3. PSDC utilise PLP to decarboxylate phosphatidyl-L-serine (PS) forming phosphatidylethanolamine (PE) and CO_2 in the process. Decreased levels of PSDC transcripts appeared to be linked to the down-regulation of a putative patatin-like phospholipase, the latter which uses PS for the production of L-2-lysophosphatidylethanolamine. Other transcripts involved in PS and PE metabolism included putative diacylglycerol kinases (PFI1485c) and putative patatin-like phospholipase (PFI1180w), and also had decreased abundance. Diacylglycerol kinases down-regulation suggested lowered levels of phosphatidate, a precursor of PS, which in turn would suggest lowered PS production. The down-regulation of a putative patatin-like phospholipase (PFI1180w) would suggest that triacylglycerol metabolism is favoured, with decreased 1,2-diacylglycerol production.

Expression patterns of mRNA transcripts are generally linked to delayed expression of the corresponding protein [294]. However, very few proteins were detected in 4PEHz treated parasites, and no definitive correlations could be made regarding linked expression patterns. Some proteins were involved in annotated processes such as DNA damage repair, which was similarly found to be active from transcriptome data. From transcriptome data glutamate dehydrogenases were shown to have increased expression which suggested L-glutamine catabolism was promoted. In the proteome the CPS protein was found in the treated parasite protein extractions and was not present in the untreated parasites. The transcriptomic expression of the corresponding transcript was not significant. CPS utilise L-glutamine together with CO_2 and ATP to generate carbamoyl phosphate [295]. Similar to *Pf*Pdx2 this enzyme is an amidotransferase, capable of providing ammonia through the hydrolysis of L-glutamine [296]. Two classes of CPS enzymes exist; CPS I and CPS II, the former is restricted to mitochondria and involved in arginine biosynthesis as well as the urea cycle. The latter CPS II has cytosolic expression and is involved in pyrimidine biosynthesis [295]. CPS I uses free ammonia instead as nitrogen source, whereas CPS II utilise L-glutamine. The CPS enzyme performs the first committed step in the urea cycle, and also provides the carboymyl phosphate precursor for synthesis of the pyrimidine UMP and arginine [297, 298]. The CPS protein identified during 4PEHz treatment represented a class II CPS which is associated with L-glutamine hydrolysis. This suggested that L-glutamine catabolic processes were active in the proteome, linking

with observations that other transcripts involved in L-glutamine catabolism were active in the transcriptome.

Drug efficacy and specificity cannot easily be assessed based on transcriptome data alone. Relatively few genes of low amplitude were expressed during CQ and the antifolate drug WR99210 (DHFR inhibitor) treatment of *P. falciparum* [240, 241]. These minor transcriptional changes were also not functionally linked to proposed drug targets [240]. In these cases the antagonists could affect very specific biologically-essential processes, thereby arresting transcriptomic response altogether, or simply there is no transcriptional response due to the drugs [240]. In stark contrast, other potent antimalarials such as artesunate resulted in 398 differentially expressed genes [262], and doxycycline induced responses were delayed however resulted in around 100 differentially expressed genes which were predominantly apicoplast-associated [207]. Similarly, artemisinin affected 270 transcripts with at least threefold change [242]. As in the case of co-inhibition of the bifunctional *S*-adenosylmethionine decarboxylase/ODC, 538 transcripts had differential expression, however transcripts directly linked to polyamine metabolism were affected, revealing a highly specific transcriptomic response [259]. This was also confirmed using proteomics and metabolomics [259]. Overall, dramatic transcriptional changes which appear unrelated may be of consequence due to initial target attenuation – possibly reflecting the cellular importance of the targeted enzyme *Pf*Pdx1. As demonstrated for 4PEHz, transcripts associated with PLP metabolism were affected, suggesting specificity of 4PEHz on *Pf*Pdx1. As a consequence additional indirectly-associated processes were affected and were suggested to be cellular compensatory responses due to disruptions in PLP biosynthesis. This supports the view that 4PEHz is specific for *Pf*Pdx1, and also highlights the dependence of the parasite on the PLP cofactor.

4.5 Conclusions

Inhibition of *Pf*Pdx1 using 4PEHz resulted in genome-wide perturbations within *P. falciparum* parasites. Several PLP-dependent enzyme transcripts had differential expression compared to untreated parasites, lending support to the proposed targeting of *Pf*Pdx1 by 4PEHz. Processes involved in folate biosynthesis and recycling were shown to be affected by 4PEHz. These results support previous observations that *Pf*Pdx1 is inhibited by 4PEHz resulting in PLP depletion, which in turn affected PLP-related processes.