Analysis of Anti-malarial drug-resistant genes of Plasmodium falciparum 3D7 to understand its expression: the bioinformatics approach

Gemechis Waktole (gamewa2010@gmail.com)
Cho Donghee
Adama Science and Technology University

Research Article

Keywords: Drug resistance, CpG islands, Plasmodium falciparum 3D7, Promoter region

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1473028/v1

Abstract

Malaria is one of the most important infectious diseases that threaten half of the world's population. Ethiopia is among countries with a high malaria burden. Plasmodium falciparum is the most common (64%) cause of malaria in Ethiopia while P. vivax accounts for the remaining cases (34%). Despite numerous advances in malaria control strategies, the disease still kills countless children worldwide, mainly in sub-Saharan African countries and malaria still remains a major public health problem in Ethiopia. Analyzing parasite genetics has enormous potential to aid both efforts at elimination as well as international efforts toward eradication. Therefore, this study is aimed to insilico analysis of regulatory elements such as CpG islands, transcription factors (TFs) and their corresponding binding sites (TFBSs) involved in the regulation of gene expression of Plasmodium falciparum 3D7 isolates drug resistant genes. In this analysis nine drug resistance affecting gene coding sequences, available in NCBI database with the start codon at the beginning of the sequence and only functional genes (protein coding) were deliberated. Accordingly, the promoter prediction by neural network promoter prediction result showed Plasmodium falciparum 3D7 drug resistance affecting genes have 1–6 TSS and five common candidate motifs (MPfI, MPfII, MPfIII, MPfIV and MPfV) were identified. The study indicates the poor occurrence of CpG islands in both gene body and promoter regions which may affect the access of promoter region of genes to their transcription factors, hence affecting their expression. In general, this in silico analysis of genes encoding for Plasmodium falciparum drug resistance affecting genes could be helpful to add knowledge about the molecular data and supportive to identify gene regulatory elements in the promoter regions.

1. Introduction

Malaria is one of the most important infectious diseases that threatens half of the world's population. The world malaria report 2019 estimated that there were 228 million cases of malaria in 2018 cause deaths [18]. With 272,000 deaths, children less than 5 were the most vulnerable group worldwide. In Africa, 213 million people were affected by malaria, which made it the most vulnerable continent in 2018 [18]. Worldwide, there was significant reduction of malaria cases during 2015– 2017, but there is a major challenge to thoroughly eliminate malaria in many countries by 2030 [17].

Ethiopia is among countries with a high malaria burden. *Plasmodium falciparum* is the most common (64%) cause of malaria in Ethiopia while *P. vivax* accounts for the remaining cases (34%) [19]. *Plasmodium falciparum* causes the most severe form of malaria, however, contrary to popular belief, *P. vivax* can also cause severe malaria and even death.Malaria morbidity and mortality have been significantly decreased in Ethiopia and worldwide in the past decade [18]. Ethiopia's fight against malaria started many years ago and transmission of this infectious disease significantly decreased since 1959 [8].

Despite numerous advances in malaria control strategies, the disease still kills countless children worldwide, mainly in sub-Saharan African countries and malaria still remains a major public health

problem in Ethiopia [18]. The World Health Organization (WHO) reported no significant progress in reducing global malaria burden during the period from 2015 to 2017[18].

Reduced efficacy of chloroquine (CQ) has forced a change in the selection of anti-malaria in the management of *falciparum* malaria. Since 2004, Ethiopia has adopted artemether-lumefantrine (AL) and CQ as first line treatment for infection with *P. falciparum* and *P. vivax*, respectively. In cases of treatment failure of *P. falciparum*, quinine (QN) is the treatment of choice and in cases of severe malaria, artemether, artesunate, or QN can be used [5]. Globally, artemisinin resistance in *P. falciparum* has emerged, especially in Southeast Asia, slowing therapeutic response and increased rates of treatment failures [15]. Similarly, artemisinin resistance has been reported from Africa although there is no evidence that it has taken hold currently [10].

Analyzing parasite genetics has enormous potential to aid both efforts at elimination as well as international efforts toward eradication [1]. For example, the use of molecular markers of anti-malarial drug resistance has for some time been used to guide the efficacy studies that define treatment policy; more recently, the detection of parasites mutant for histidine-rich protein 2/3 (pfhrp2/pfhrp3) gene highlighted the need to develop rapid diagnostic tests (RDTs) with alternative targets of detection, and to provide countries with guidance on the implications for case management [17].Therefore, It is the aim of this Insilco analysis to identify regulatory elements such as CpG islands, transcription factors (TFs) and their corresponding binding sites (TFBSs) involved in the regulation of its gene expression so that it provide baseline information to revise existing eradication and elimination strategies and for designing of target specific drugs.

2. Methodology

Determination of transcription starts sites and promoter regions. *Plasmodium falciparum* drug resistance affecting genes (*Pfcrt (*CQ monotherapy, AS-PY and Amodiaquine*), Pfmdr1 (*CQ *monotherapy, Melfoquine, Lumefantrine, Amodiaquine* and limited efficacy of ACT*), Pfdhfr (*ATQ-PGL, Sulfadoxine-Phyrimethamine*), and Pfdhps (*Sulfadoxine-Phyrimethamine*),Pfarps10, Pferredoxin(Ferroquine), Pfexonuclease, PfKpk-13(*AM-LF, AS monotherapy, AS-SP, AS-MQ, AS-PY*) and Pfmdr2* (DHA-PPQ)genes.) were explored from NCBI genome data bank. In this analysis all drug resistance affecting gene coding sequences, available in NCBI database with the start codon at the beginning of the sequence and only functional genes (protein coding) were deliberated. To determine their respective transcription start sites (TSSs), 1 kb sequences upstream of the start codon was excised from each gene [9]. All the TSSs of each of selected *Plasmodium falciparum* 3D7 drug resistance affecting genes was searched within this region by using the Neural Network Promoter Prediction (NNPP version 2.2) tool set with the minimum standard predictive score (between 0 and 1) cutoff value of 0.8 [11]. This tool helps to locate the possible TSSs within the sequences upstream of the start codon where the RNA polymerases start their activity, transcription process. NNPP tool has ability to recognize precisely the position of a TSS for a given gene. For those regions containing more than one TSS, the one with the highest value of prediction score was considered to have trustable and accurate prediction. Therefore, as previously done for SARS-CoV-2 gene

promoter regions determination, *P.falcipurum* drug resistance affecting genes promoter sequences were defined as 1 kb region upstream of each TSS [7].

2.1. Identification of common candidate motifs and transcription factors

Promoter sequences which were identified based on the criteria considered above from *Plasmodium falciparum* drug resistance affecting genes were analyzed using the MEME (Multiple Em for Motif Elicitation) version 5.3.3 searches, via the web server hosted by the National Biomedical Computation Resource [2] to look for common candidate motifs that serves for binding sites of transcription factors that regulate the expression of *Plasmodium falciparum* drug resistance affecting genes. MEME searches for statistically significant candidate motifs in the input sequence set. The MEME output are in the form of XML, text, MAST HTML, MAST XML, MAST text, HTML and shows the candidate motifs as local multiple alignments of the input promoter sequences. Briefly, MEME discovers novel, un-gapped motifs (recurring, fixed-length patterns) in sequences submitted in it. A motif is an approximate sequence pattern that occurs repeatedly in a group of related sequences. MEME represents motifs as position-dependent letter-probability matrices that describe the probability of each possible letter at each position in the pattern. MEME takes as input a group of sequences and outputs as many motifs as requested. MEME uses statistical modeling techniques to automatically choose the best width, number of occurrences, and description for each motif [2]. Buttons on the MEME HTML output allow one or all of the candidate motifs to be forwarded for further analysis, to better characterize the identified candidate motifs, by other web-based programs. In this case TOMTOM [6] web server was used to search for sequences matching the identified motif for its respective TF. The output of TOMTOM includes LOGOS representing the alignment of the candidate motif and TF with the p-value and q-value (a measure of false discovery rate) of the match, and links back to the parent transcription database for more detailed information about it [3]; [4].

2.2. Search for CpG islands

The stringent search criteria, Takai and Jones algorithm: GC content ≥ 55%, ObsCpG/ExpCpG ≥ 0.65, and length ≥ 500 bp will be used [13]. For this purpose, the CpG island searcher program (CpGi130) available at web link http://dbcat.cgm.ntu.edu.tw/ was used to search for particular regions of CpG-rich. Secondly, the CLC Genomics Workbench, CLC bio,Aarhus, Denmark) was used for searching the restriction enzyme *MspI* cutting sites (fragment sizes between 40 and 220 bps). Searching for MspI cutting sites is relevant for detection of CGIs, because studies using whole genome CpG island libraries prepared for different species revealed that, CpG islands are not randomly distributed but are concentrated in particular regions because CpG-rich regions are achieved by isolation of short fragments after MspI digestion that recognizes CCGG sites [14].

3. Results And Discussions

3.1. Identification of transcription start sites (TSSs)

All nine (9) *Plasmodium falciparum 3D7* drug resistance affecting genes transcription start sites (TSS) were summarized below as follows. As a result, the promoter prediction by neural network promoter prediction result showed *Plasmodium falciparum 3D7* drug resistance affecting genes have 1–6 TSS. Accordingly, among identified TSS 44.44% of them (4/9) were less than − 500 compared with ATG start codon. The predictive score of those sequences which have less than − 500 distance from start codon were 0.90, 0.93, 0.97, and 0.98. The rest 55.55% of identified TSS were greater than − 500 with predicted score value 0.84, 0.95, 0.96, 0.97, and 0.98. The relative locations of all TSS with respect to start codon were given in Table 1. The nearest TSS were recorded for *Pfexonuclease* (− 43) followed by *Pfdhfr* (-186) while the outlying TSS was observed for PfArps-10 (− 3381) upstream of the start codons of their respective genes.

Table 1
Number and predictive score value for *Plasmodium falciparum* drug resistance affecting genes TSSs.

Name/Gene ID	Corresponding promoter region name	Number of TSS identified	Predictive score at cutoff value of 0.8	Location of the best TSS from start codon
Pfcrt /2655199	*Pfcrt*	3	0.82,0.80,0.80,0.90	-203
Pfmdr1/813045	*Pfmdr1*	1	0.98	-760
Pfdhfr /9221804	*Pfdhfr*	2	0.84,0.93	-186
Pfdhps/2655294	*Pfdhps*	3	0.81,0.98,0.84	-331
Pfarps10/812163	*Pfarps10*	3	0.86,0.96,0.85	-3381
Pferredoxin /3885862	*Pferredoxin*	6	0.90, 0.95, 0.84, 0.97, 0.86, 0.90	-641
Pfexonuclease /811867	*Pfexonuclease*	5	0.97, 0.89, 0.92, 0.82, 0.92	-43
Pfmdr2/812037	*Pfmdr2*	4	0.92, 0.91, 0.83, 0.95	-517
PfKp-13/814205	*PfKp-13*	2	0.84,0.84	-994

3.2. Common candidate motifs and associated transcription factors in the promoter regions of Plasmodium falciparum drug resistance affecting genes

The insilico analysis tried to identify the best candidate motifs for all (9) promoter sequences of *Plasmodium falciparum 3D7* drug resistance affecting genes. Consequently, five common candidate

motifs (*MPfI, MPfII, MPfIII, MPfIV* and *MPfV*) were identified. Among the identified five candidate motive MPfV shared with only 44.44% of the promoter sequences of *Plasmodium falciparum 3D7* drug resistance affecting genes. Whereas, the rest four candidate motifs are shared by all of *Plasmodium falciparum* drug resistance affecting genes as shown in Table 2. The analysis was performed with minimum and maximum motif width of 6 and 50 residues, respectively, for *Plasmodium falciparum* drug resistance affecting genes whereas, a maximum and minimum number of motifs were 45 and 11, which were used to identify probable promoter regulatory elements (motifs). To determine motifs which are functionally important, motifs which were shared by majority of promoter regions of *Plasmodium falciparum* drug resistance affecting genes was chosen. Accordingly, *MpfIV* was discovered as the common promoter motif for all (100%) genes that serves as binding sites for transcription factors involved in the expression regulation of these genes with lowest E-value 2.2e-002. Sequence logo for *MpfIV* generated by MEME is presented in Fig. 1 below.

Table 2
Identified common candidate motifs in *Plasmodium falciparum* drug resistance affecting genes promoter regions.

SN	Name	E-value	Sites	N (%) PCEOMs	Width
1	MpfI	9.1e-008	9	9(100)	45
2	MpfII	3.2e-001	9	9(100)	11
3	MpfIII	4.0e-001	9	9(100)	29
4	MpfIV	**2.2e-002**	9	9(100)	26
5	MpfV	2.1e + 001	4	4(44.4)	21
N (%) PCEOMs: Number (%) of promoters containing each one of the motifs					

MEME represents motifs as position-dependent letter-probability matrices that describe the probability of each possible letter at each position in the pattern. MEME takes as input a group of sequences and outputs as many motifs as requested. MEME uses statistical modeling techniques to automatically choose the best width, number of occurrences, and description for each motif [2].

TOMTOM is a motif comparison algorithm that ranks the target motifs in a given data base according to the estimated statistical significance of the match between the query and the target. In similar manner, TOMTOM provides LOGOS that represents the alignment of two motifs and a numeric score for the match between two motifs together with a statistical significance [12].Out of 40 motifs identified from *Plasmodium falciparum 3D7* drug resistance affecting genes promoter regions only 12(30%) were found on the negative strands.

3.3. Analysis for CpG islands (CGIs) in *Plasmodium falciparum* drug resistance affecting genes and promoter regions

In this analysis, nine *Plasmodium falciparum* drug resistance affecting genes promoter and gene body regions were used with two algorithms. First, Takai and Jones' algorithm [13] was used and found no CpG islands in both promoter and gene body regions. Similarly, a second alternative approach to search for the presence of CpG islands is by in-silico digestion using restriction enzyme *MspI* by CLC genomics work bench 3 software revealed No CpG islands in both promoter and gene body regions in *Plasmodium falciparum* drug resistance affecting genes.This result indicates the poor occurrence of CpG islands in both gene body and promoter regions which may affect the access of promoter region of genes to their transcription factors, hence affecting their expression.

Table 3
The list of candidates which could bind to motif *PfMIV*

Candidate transcription factors	Statistical significance			Data Base	Overlap
	p-value	*E*-value	*q*-value		
UP00029/TATA-box-binding protein	9.32e-05	1.68e-01	2.07e-01	Uniprobe mouse	16
FOXB1/FOXB1_DBD_2	8.26e-04	1.49e + 0	7.89e-01	jolma2013	18
UP00150_1 (Irx6_2623.2)	2.92e-03	5.28e + 0	1.00e + 0	uniprobe_mouse	17
UP00094_2 (Zfp128_secondary)	5.40e-03	9.76e + 0	1.00e + 0	uniprobe_mouse	14

4. Conclusions

The result of this analysis could be critically important to understand the nature of promoter regions, the motif discovered in line with the transcription factor binding proteins of *Plasmodium falciparum* drug resistance affecting genes. CpG islands are also regulatory elements in the promoter regions of genome and useful in the detection of promoters. But in this analysis no CpG islands was identified by both algorithms in *Plasmodium falciparum* drug resistance affecting genes. In general, this *in silico* analysis of genes encoding for *Plasmodium falciparum* drug resistance affecting genes could be helpful to add knowledge about the molecular data and supportive to identify gene regulatory elements in the promoter regions. It could also help to predict gene expression profiles in *Plasmodium falciparum* 3D7 which in

turn could be helpful to improve present drug efficacy and to develop new drug with high target specificity. Therefore, knowledge of bioinformatics methods is estimable important to identify gene regulatory regions in the promoter regions and gene body regions could help also to predict gene expression profiles in various pathogens, Since wet molecular laboratory are very expensive and sophisticated.

5. References

1. Auburn, S., & Barry, A. E. (2017). Dissecting malaria biology and epidemiology using population genetics and genomics. *International journal for parasitology*, *47*(2–3), 77–85.
2. Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers.
3. Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic acids research*, *34*(suppl_2), W369-W373.
4. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, *37*(suppl_2), W202-W208.
5. World Health Organization. (2015). Food, Medicine and Healthcare Administration and Control Authority of Ethiopia. *Standard Treatment Guidelines for General Hospital*,.
6. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. *Genome biology*, *8*(2), 1–9.
7. Dinka, H., & Milkesa, A. (2020). Unfolding SARS-CoV-2 viral genome to understand its gene expression regulation. *Infection, Genetics and Evolution*, *84*, 104386.
8. McCann, J. C. (2015). *The historical ecology of malaria in Ethiopia: Deposing the spirits*. Ohio University Press.
9. Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, *13*(4), 233–245.
10. Lu, F., Culleton, R., Zhang, M., Ramaprasad, A., von Seidlein, L., Zhou, H., … Cao, J. (2017). Emergence of indigenous artemisinin-resistant Plasmodium falciparum in Africa. *New England Journal of Medicine*, *376*(10), 991–993.
11. Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Computers & chemistry*, *26*(1), 51–56.
12. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research*, *32*(suppl_1), D91-D94.
13. Takai, D., & Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the national academy of sciences*, *99*(6), 3740–3745.
14. Takamiya, T., Hosobuchi, S., Noguchi, T., Paterson, A. H., Iijima, H., Murakami, Y., & Okuizumi, H. (2009). The Application of Restriction Landmark Genome Scanning Method for Surveillance of Non-

Mendelian Inheritance in Hybrids. *Comparative and functional Genomics*, *2009*.

15. Tun, K. M., Imwong, M., Lwin, K. M., Win, A. A., Hlaing, T. M., Hlaing, T., … Woodrow,C. J. (2015). Spread of artemisinin-resistant Plasmodium falciparum in Myanmar: a cross-sectional survey of the K13 molecular marker. *The Lancet infectious diseases*, *15*(4), 415–421.
16. World Health Organization. (2020). World malaria report 2019. 2019. *Reference Source*. https://www.who. int/malaria/publications/worldmalaria-report-2019/*en*.
17. World Health Organization. (2018). *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*. World Health Organization.
18. World Health Organization. World Malaria Report 2018. Geneva: WHO; 2018. p 24.
19. World Health Organization. (2016). *World malaria report 2015*. World Health Organization.
20. World Health Organization. (2017). *False-negative RDT results and implications of new reports of P. falciparum histidine-rich protein 2/3 gene deletions* (No. WHO/HTM/GMP/2017.18). World Health Organization.

Figures

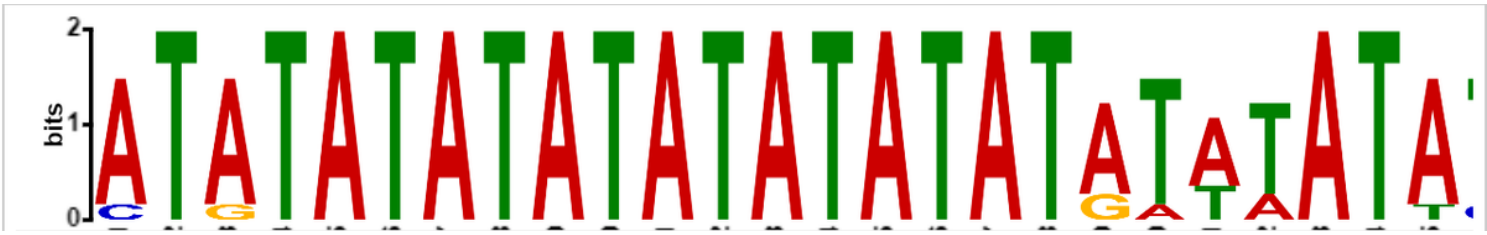

Figure 1

Sequence logo for the identified common promoter motif *MpfIV* gene for *Plasmodium falciparum* drug resistance affecting genes

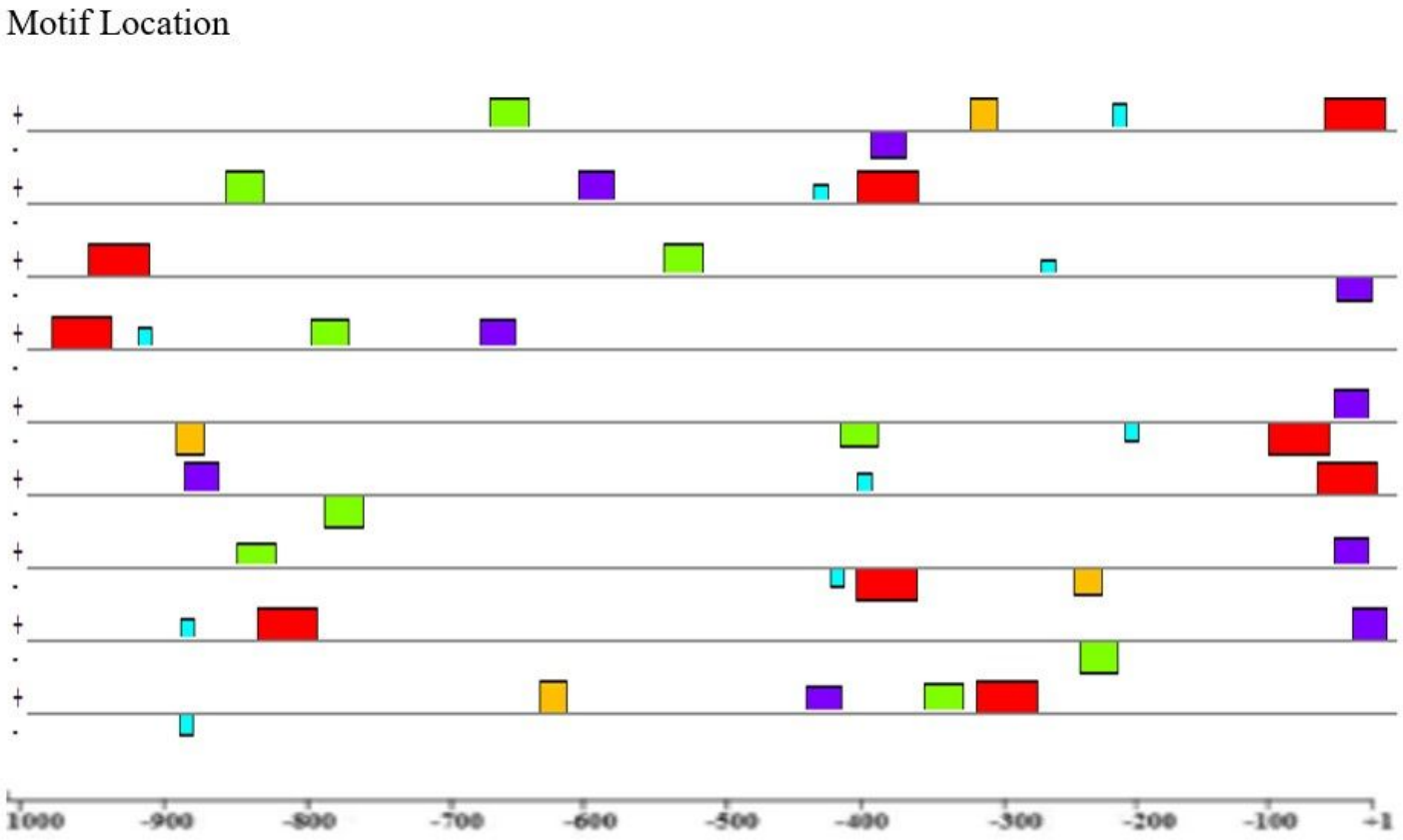

Figure 2

The relative positions of motifs in different *Plasmodium falciparum* drug resistance affecting gene sequences relative to TSSs.