Genomics Data 2 (2014) 393–395

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

An *in vivo* transcriptome data set of natural antisense transcripts from *Plasmodium falciparum* clinical isolates

Amit Kumar Subudhi [a,1], P.A. Boopathi [a,1], Shilpi Garg [a], Sheetal Middha [b], Jyoti Acharya [b], Deepak Pakalapati [a,2], Vishal Saxena [a], Mohammed Aiyaz [c], Harsha B. Orekondy [c], Raja C. Mugasimangalam [c], Paramendra Sirohi [b], Sanjay K. Kochar [b], Dhanpat K. Kochar [d], Ashis Das [a,*]

[a] *Department of Biological Sciences, Birla Institute of Technology and Science (BITS), Pilani, Rajasthan, India*
[b] *Department of Medicine, S.P. Medical College, Bikaner, Rajasthan, India*
[c] *Genotypic Technology Pvt. Ltd., Bangalore, India*
[d] *Rajasthan University of Health Sciences, Jaipur, Rajasthan, India*

ARTICLE INFO

Article history:
Received 20 September 2014
Accepted 1 October 2014
Available online 29 October 2014

Keywords:
Plasmodium falciparum
Uncomplicated malaria
Complicated malaria
Natural antisense transcripts
Strand-specific microarray

ABSTRACT

Antisense transcription is pervasive among biological systems and one of the products of antisense transcription is natural antisense transcripts (NATs). Emerging evidences suggest that they are key regulators of gene expression. With the discovery of NATs in *Plasmodium falciparum*, it has been suggested that these might also be playing regulatory roles in this parasite. However, all the reports describing the diversity of NATs have come from parasites in culture condition except for a recent study published by us. In order to explore the *in vivo* diversity of NATs in *P. falciparum* clinical isolates, we performed a whole genome expression profiling using a strand-specific 244 K microarray that contains probes for both sense and antisense transcripts. In this report, we describe the experimental procedure and analysis thereof of the microarray data published recently in Gene Expression Omnibus (GEO) under accession number GSE44921. This published data provide a wealth of information about the prevalence of NATs in *P. falciparum* clinical isolates from patients with diverse malaria related disease conditions. Supplementary information about the description and interpretation of the data can be found in a recent publication by Subudhi et al. in *Experimental Parasitology* (2014).

Specifications	
Organism	*Plasmodium falciparum*
Sex	Not applicable
Sequencer or array type	Agilent 244 K *P. falciparum* custom designed array (AMADID: 024956)
Data format	Raw data: TXT, Analyzed data: SOFT, MINIML, TXT
Experimental factors	Uncomplicated and Complicated *P. falciparum* malaria
Experimental features	Detection of sense and antisense transcripts from *P. falciparum* clinical isolates from patients with diverse disease conditions.
Consent	Written informed consent were obtained from patients before study entry
Sample source location	Bikaner, Rajasthan, India, geographical coordinates, Latitude- 28.0167° N and Longitude- 73.3119° E

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44921

Experimental design, materials and methods

The 244 K custom P. falciparum array designing

The 244 K *P. falciparum* custom array was designed on an Agilent platform using the RightDesign (Genotypic Technology Pvt. Ltd., Bangalore, India) probe design workflow. This workflow chooses the best probe(s) for a transcript by balancing several criteria: GC content, sequence complexity, cross-hybridization potential and secondary

* Corresponding author. Tel.: +91 1596 245073 Ext.; fax: +91 1596 244183.
E-mail addresses: amit4help@gmail.com (A.K. Subudhi), boopathiarunachalam@gmail.com (P.A. Boopathi), shilpishailley@gmail.com (S. Garg), sheetumig21@gmail.com (S. Middha), jyotiacharya2@gmail.com (J. Acharya), deepakpchowdary@gmail.com (D. Pakalapati), vishalsaxena12@gmail.com (V. Saxena), aiyaz@genotypic.co.in (M. Aiyaz), harsha.ob@genotypic.co.in (H.B. Orekondy), raja@genotypic.co.in (R.C. Mugasimangalam), drpsirohi@gmail.com (P. Sirohi), drskkochar@rediffmail.com (S.K. Kochar), drdkkochar@yahoo.com (D.K. Kochar), ashisd28@gmail.com, adas@pilani.bits-pilani.ac.in (A. Das).
[1] These authors contributed equally to this work.
[2] Current affiliation: Center for Cellular and Molecular Biology, Hyderabad.

http://dx.doi.org/10.1016/j.gdata.2014.10.010

structure. The array contains 241399 probes of 60 nucleotides long. These probes designed in both sense and antisense orientations (sense and antisense probes) representing *P. falciparum* 3D7 transcript sequences from PlasmoDBv5.3 [1,2], *P. falciparum* EST sequences available in the NCBI EST database (2007), apicoplast sequences of *P. falciparum* [10] and apicoplast sequences of *Plasmodium vivax* [7,8]. Annotations of all the probes were updated again according to PlasmoDBv8.2 and NCBI database (2012). Probes were designed with an average of 8 per sequence in both sense and antisense direction. Detailed description about the array can be found in [9]. Array details including the array design, array information such as feature number, Probe ID, probes sequence, target gene identifier ID and target gene description have been submitted to Gene Expression Omnibus (GEO) and available under GEO accession number (GPL16484). Additionally, here we also give the information about each probe orientation in the array (Supplementary Table S1).

Sample collection and processing

Venous blood samples (~5 ml) were collected from *P. falciparum* infected adult patients ($n = 11$) on in informed consent admitted to S. P. Medical College, Bikaner, India. Infection with *P. falciparum* was confirmed by slide microscopy and RDTs (OptiMal test; Diamed AG, Cressier sur Morat, Switzerland, Falcivax; Zephyr Biomedical System, Goa, India) in the hospital. The patients exhibited symptoms which were categorized as either complicated ($n = 9$) or uncomplicated malaria ($n = 2$) [11]. All other laboratory investigation were performed to rule out possibility of any other possible cause for symptoms exhibited as described elsewhere [3,4]. Infected and uninfected RBCs were separated from peripheral blood mononuclear cells by subjecting the collected blood to a density gradient-based centrifugation and separation according to manufacturer's instructions (Histopaque 1077, Sigma Aldrich, USA). The RBCs and PBMCs fractions were washed twice with phosphate-buffered saline and lysed with four volume of TriReagent (Sigma Aldrich, USA). The lysed fractions were stored immediately in −80 °C deep freezers till further processing. Samples were shipped to BITS, Pilani, Rajasthan, in cold chain where RNA and DNA isolation from samples were carried out. DNA from samples were used to confirm the infection of only *P. falciparum* by performing 18 s rRNA-based multiplex PCR [6] and 28 s rRNA-based nested PCR [5].

RNA quality control checking, labeling, hybridization and scanning

The isolated total RNA from each sample was processed on denaturing agarose gels to assess its integrity. Total RNA integrity was also assessed using RNA 6000 Nano Lab Chip on the 2100 Bioanalyzer (Agilent, Palo Alto, CA) following manufacturer's instructions. RNA purity was assessed using the NanoDrop ND-1000 UV-Vis Spectrophotometer (Nanodrop Technologies, Rockland, USA). Complicated ($n = 9$) and uncomplicated ($n = 2$) malaria parasite RNA samples were pooled in equimolar amount separately. The pooled complicated and uncomplicated malaria parasite RNA sample were amplified and labeled in presence of Cyanine-5 CTP and Cyanine-3 CTP, respectively, using Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies) as described by the manufacturer to use in a two color microarray hybridization experiment. This resulted in production of amplified and labeled cRNA, which was cleaned and quality assessed for yields and specific activity. For hybridization purpose, 1500 ng of each Cy-3 and Cy-5 labeled cRNA samples was mixed, fragmented and hybridized to the 244 K array with Gene Expression Hybridization Kit (Agilent Technologies, Part No. 5188-5242). Hybridization was carried out in Agilent's Surehyb Chambers at 65 °C for 16 h. The hybridized slides were washed using Gene Expression wash buffers (Agilent Technologies, Part No. 5188-5327) and scanned using the Agilent Microarray Scanner (Agilent, Palo Alto, CA, G Model G2565BA) at 5 μm resolution. Data from images were extracted using Feature Extraction software (Agilent Technologies).

Table 1
Types of transcript detected in complicated and uncomplicated malaria isolates.

Type of transcripts	Complicated isolates	Uncomplicated isolates
Only antisense transcripts	34	35
Only sense transcripts	677	359
Both sense and antisense transcripts	736	302
Total number of antisense transcripts detected	797	
Antisense transcripts from Protein coding genes	788	
Antisense transcripts from rRNA coding genes	7	
Antisense transcripts from Non protein coding genes	2	

Data analysis

The data from Red channel (complicated) and Green channel (uncomplicated) were analyzed separately. For each spot, the raw signal intensity and the background signal intensity were obtained from the raw data files. Ratio was calculated by dividing the raw signal intensity with the background signal intensity for each of the probes. Based on the probe orientation, the probes were classified as sense and antisense probes. Probes representing the sense and antisense transcripts were filtered by considering signal intensity, which was ≥twice the background intensity for complicated cases and ≥1.5 times for uncomplicated cases. The probe-based data were converted to gene-based data. For genes represented with multiple probes, median of raw signal intensity and background signal intensity was considered. Genes were further classified based on the type of transcripts they produced: genes expressing only sense transcripts, genes expressing only antisense transcripts and genes expressing both sense and antisense transcripts. Further a stringent filtering criterion was applied where genes detected by at least 3 probes were considered for further analysis. Gene represented by both sense and antisense probes (5330 genes) representing *P. falciparum* transcript sequences from PlasmoDBv8.2 were only analyzed and discussed in the paper [9]. Details about the number and detected transcript type are shown in Table 1. A total of 797 NATs were detected in this study of which 545 were found to be unique to the study reported in [9].

Discussion

NATs are emerging as a key player in genome regulation. In order to understand the role, it is necessary to explore their prevalence in the transcriptome of different organisms including pathogenic ones. Here we describe the information obtained about the diversity of the NATs population in *P. falciparum* clinical isolates from patients with diverse disease conditions through a 244 K strand-specific microarray experiment. The detailed analysis of which has been reported recently and is the first *in vivo* study to describe about the range of NATs that prevails in *P. falciparum* clinical isolates [9]. The probes representing the transcript sequences from PlasmoDBv8.2 were analyzed in detail in our recent report [9]. However, the array also contains probes representing the *P. falciparum* sequences from NCBI database (2012) detailed here (Supplementary Table S1).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.10.010.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We thank all the patients and technical workers for their participation in and support of this project. A.K.S. acknowledges Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), New Delhi, India, and Project Assistantship from Department

of Biotechnology (DBT), New Delhi, India. P.A.B. acknowledges Basic Scientific Research fellowship from University Grant Commission, New Delhi, India, and Project Assistantship from Department of Biotechnology (DBT), New Delhi, India. A.K.D., S.K.K. and D.K.K. acknowledges Department of Biotechnology (DBT), New Delhi, India for the financial support through the grant BT/PR7520/BRB/10/481/2006 and Birla Institute of Technology and Science, Pilani, India and S.P. Medical College, Bikaner, India for providing the required infrastructural facilities during this study. We thank the PlasmoDB team for making available the whole genome data and the SAGE data for *P. falciparum*. We also thank Genotypic Technology Pvt. Ltd., Bangalore, India, for the microarray hybridization and preliminary data analysis service provided by them.

References

[1] C. Aurrecoechea, J. Brestelli, B.P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O.S. Harb, M. Heiges, F. Innamorato, J. Iodice, J.C. Kissinger, E. Kraemer, W. Li, J.A. Miller, V. Nayak, C. Pennington, D.F. Pinney, D.S. Roos, C. Ross, C.J. Stoeckert Jr., C. Treatman, H. Wang, PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. 37 (2009) D539–543.

[2] M.J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E. Nelson, S. Bowman, I.T. Paulsen, K. James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M.S. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S. Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, B. Barrell, Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419 (2002) 498–511.

[3] D.K. Kochar, V. Saxena, N. Singh, S.K. Kochar, S.V. Kumar, A. Das, *Plasmodium vivax* malaria. Emerg. Infect. Dis. 11 (2005) 132–134.

[4] D.K. Kochar, G.S. Tanwar, P.C. Khatri, S.K. Kochar, G.S. Sengar, A. Gupta, A. Kochar, S. Middha, J. Acharya, V. Saxena, D. Pakalapati, S. Garg, A. Das, Clinical features of children hospitalized with malaria—a study from Bikaner, northwest India. Am. J. Trop. Med. Hyg. 83 (2010) 981–989.

[5] D. Pakalapati, S. Garg, S. Middha, J. Acharya, A.K. Subudhi, A.P. Boopathi, V. Saxena, S.K. Kochar, D.K. Kochar, A. Das, Development and evaluation of a 28S rRNA gene-based nested PCR assay for *P. falciparum* and *P. vivax*. Pathog. Glob. Health 107 (2013) 180–188.

[6] D. Pakalapati, S. Garg, S. Middha, A. Kochar, A.K. Subudhi, B.P. Arunachalam, S.K. Kochar, V. Saxena, R.P. Pareek, J. Acharya, D.K. Kochar, A. Das, Comparative evaluation of microscopy, OptiMAL((R)) and 18S rRNA gene based multiplex PCR for detection of *Plasmodium falciparum* & *Plasmodium vivax* from field isolates of Bikaner, India. Asian Pac. J. Trop. Med. 6 (2013) 346–351.

[7] V. Saxena, S. Garg, S. Ranjan, D. Kochar, A. Ranjan, A. Das, Analysis of elongation factor Tu (tuf A) of apicoplast from Indian *Plasmodium vivax* isolates. Infect. Genet. Evol. 7 (2007) 618–626.

[8] V. Saxena, S. Garg, J. Tripathi, S. Sharma, D. Pakalapati, A.K. Subudhi, P.A. Boopathi, G.S. Saggu, S.K. Kochar, D.K. Kochar, A. Das, *Plasmodium vivax* apicoplast genome: a comparative analysis of major genes from Indian field isolates. Acta Trop. 122 (2012) 138–149.

[9] A.K. Subudhi, P.A. Boopathi, S. Garg, S. Middha, J. Acharya, D. Pakalapati, V. Saxena, M. Aiyaz, H.B. Orekondy, R.C. Mugasimangalam, P. Sirohi, S.K. Kochar, D.K. Kochar, A. Das, Natural antisense transcripts in *Plasmodium falciparum* isolates from patients with complicated malaria. Exp. Parasitol. 141 (2014) 39–54.

[10] R.J. Wilson, P.W. Denny, P.R. Preiser, K. Rangachari, K. Roberts, A. Roy, A. Whyte, M. Strath, D.J. Moore, P.W. Moore, D.H. Williamson, Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 261 (1996) 155–172.

[11] WHO, Severe falciparum malaria. Trans. R. Soc. Trop. Med. Hyg. 94 (2000) 1–90.