Insilco Analysis of *Leishmania donovani* Genome to Understand its Gene Expression Regulation

Kefiyalew Jote (✉ kefiyalewjote@gmail.com)

Adama Science and Technology University

Mulugeta Kebede

Adama Science and Technology University

Ashenafi Milkessa

Adama Science and Technology University

Research Article

Keywords: Leishmania donovani, Promoter region, Transcription factor, Motif, CpG

Posted Date: April 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-403172/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Leishmania donovani is an obligate intracellular parasite of human that causes a dreadful visceral infection that may involve liver, spleen, bone marrow, and lymph nodes. The specific feature of the species influences its genetics. Changes in parasite adaptation to different environment, development of silent genes and parasite progression through its lifecycle may be determined by transcriptional organizations affecting its gene expression. Hence, Insilco analysis of *Leishmania donovani* BPK282A1 genome was performed to predict the promoter and regulatory elements involved in gene expression. Forty putative genes of *Leishmania donovani* BPK282A1 genes were obtained from National Center for Biotechnology Information and investigated using bioinformatic tools, such as Neural Network Promoter Prediction, MEME suite, CpG island finder, GOMo and CLC Genomics Workbench. In this analysis, the number of Transcription Start Sites (TSSs) in promoter regions of *Leishmania donovani* BPK282A1 varied greatly between 1 and 4 with majority (60%) of them having two or more TSSs. The study also revealed seven candidate motifs in the promoter regions of the gene. The common promoter motif for all *Leishmania donovani* BPK282A1 genes was MI with an E-value of $4.0e-116$. Six of the identified motifs were related to zinc finger proteins that bind to the GC-rich regions in DNA. The study also revealed that most of both the promoter and the gene body regions contain at least one CpG island. Restriction analysis using MspI enzyme showed that *Leishmania donovani* BPK282A1 genes were CpG rich. In conclusion, the investigation unfolded expression regulation of *Leishmania donovani* BPK282A1 genes contributing to development of a robust drug and vaccine that targets the regulatory regions.

1. Introduction

Leishmania spp. are medically important intracellular parasites that are accountable to cause fatal infections ranging from small cutaneous lesions to more severe and life threatening visceral infections in humans (Holzmuller et al. 2002). Leishmaniasis is a neglected tropical disease that remains to impose a serious public health problems in many countries around the world because of absence of vaccines and limited availability of chemotherapeutic agents in use as of today (Rastrojo et al. 2018). The disease is widespread in tropical and subtropical countries and known to have both zoonotic and human to human transmission patterns with two main clinical forms that include cutaneous and visceral Leishmaniasis (Steverding 2017).

The main etiologic agent of cutaneous leishmaniasis is *Leishmania major*. It is commonly associated with lesions in the area where sandfly bite occurs. Whereas, the cause of visceral leishmaniasis is *Leishmania donovani* which eventually spreads to internal organs culminating in death (Downing et al. 2011). The transmission of *Leishmania donovani*, a parasite that kills millions of people in east Africa in general and thousands in Ethiopia in particular was thought to be anthroponotic. However, recent genetic studies have claimed that there are possibilities of zoonotic transmission (Kassahun et al. 2015).

Leishmania donovani is an obligate intracellular parasite (Mukherjee, Chakraborty, and Chakrabarti 2016; Singh et al. 2016; Verma, Rastogi, and Mukhopadhyay 2017) that is postulated to inhabit

phagolysosomes of the macrophages (Singh et al. 2016). Once the macrophage is inhabited the growth of the parasite will be favored with the help of phagosome maturation regulation which will eventually avoid destruction. As the result, antigen presentation, oxidative damage, apoptosis and immune activation will be depleted and the availability of the nutrients will be favored on the contrary (Podinovskaia and Descoteaux 2015). The parasite almost exclusively multiplies in the phagolysosomes of the infected macrophage and this affects the expression of gene in both the parasite itself and the host cell (Buates and Matlashewski 2001).

Leishmania donovani appears to have two distinct genomic features influencing its genetics; where there are different copy numbers within an individual chromosome or where chromosome dosage varies among clonal population of a cell. On the other hand, once infection occurs it develops a mechanism that weakens microbicidal activity of macrophages interfering with the expression of Major Histocompatibility-II (MHC II) molecules after subsequent stimulation with interferon- γ (IFN- γ) (Buates and Matlashewski 2001). The parasites utilize various proteins and glycoproteins to interact with host cells establishing a long term infection with a clinical presentation of visceral Leishmaniasis (Mukherjee, Chakraborty, and Chakrabarti 2016).

Several complications may result from visceral Leishmaniasis that include hepatic cirrhosis, necrotizing oral infection, pulmonary tuberculosis, ocular complications, disseminated intravascular coagulations, nephritic syndrome, glomerulonephritis, neutropenia (Meyers et al. 2016), immunosuppression and pancytopenia (McGwire and Satoskar 2013) leading to concurrent or inter-current infections finally culminating in death (Meyers et al. 2016). There are various types of therapies to treat leishmaniasis that include first line and second line regimens. The use of the drugs available may rely on regional guidelines and varies among several localities. Despite availability and accessibility of the drugs, ineffectiveness of treatment is increasing due to parasite resistance to the available drugs (McGwire and Satoskar 2013).

Changes in parasite adaptation to different environment, development of silent genes and parasite progression through its lifecycle may be determined by transcriptional organizations where regulatory elements are located; i.e. the Transcriptional Start Sites (TSS) (Duncan et al. 2011). *Leishmania donovani* mainly loses its virulence through transcriptional and translational regulations (Sinha et al. 2018). Therefore, understanding the regulatory elements such as Transcription Factors (TFs), Transcription Factor Binding Sites (TFBSs) and CpG islands in *Leishmania donovani* genome that determines its gene expression is helpful to fully characterize its molecular mechanisms and design effective drugs, vaccines and other control strategies.

1.1 OBJECTIVE

1.1.1 General objective

The general objective of this study was to predict promoter and regulatory elements involved in *Leishmania donovani* BPK282A1 genes expression

1.1.2 Specific objectives

- To detect transcription start sites and determine promoter regions of *Leishmania donovani* *BPK282A1* genes
- To recognize common motifs and transcription factors among *Leishmania donovani* *BPK282A1* genes
- To quest for gene ontology for identified regulatory elements
- To quest for CpG islands in *Leishmania donovani* *BPK282A1* gene body and promoter regions

2. Method

2.1 TSSs and promotor regions determination for *Leishmania donovani* *BPK282A1* genes

Forty functionally putative genes of *Leishmania donovani* *BPK282A1* genome were obtained from National Center for Biotechnology Information (NCBI) genome database (National Center for Biotechnology Information n.d.) on Nov 3, 2020 and analyzed using online regulatory elements determination tools. The corresponding TSS of each gene was analyzed using an online Neural Network Promoter Prediction (NNPP version 2.2) (BDGP: Neural Network Promoter Prediction n.d.), a tool that helps to find promoters in a Deoxyribonucleic Acid (DNA) sequences of eukaryotic and prokaryotic cells. Herein, at least 1kb upstream of the start codon for each gene was excised as it was done by Dinku and colleague (Dinka and Milkesa 2020). The excised segments were fitted to NNPP version 2.2 by setting the minimum standard predictive score (between 0 and 1) with a cut off value of 0.8 (Kanhare and Bansal 2005). In conditions where there were two or more TSSs, the one having the highest prediction value was considered as statistically significant (Michaloski et al. 2011).

2.2 Motifs and transcription factors identification

The identified promoter sequences were transferred to another online resource, Multiple Expression motifs for Motif Elicitation (MEME) suit version 5.3.0 (Introduction - MEME Suite n.d.); and analyzed for common candidate motifs (Release Notes - MEME Suite n.d.). Motif is an approximate sequence pattern that occurs repeatedly in a group of related sequences. MEME discovers novel, un-gapped motifs (recurring, fixed-length patterns) in the identified sequences being submitted. It is a motif-based sequence analysis tool that splits variable-length patterns into two or more separate motifs. It represents motifs as position dependent letter probability matrices that describe the probability of each possible letter at each position in the pattern. Individual MEME motifs do not contain gaps. Patterns with variable-length gaps are split by MEME into two or more separate motifs. This tool also 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 (Bailey and Elkan 1994). The identified motifs were exported to another web-based tool, TOMTOM (Tomtom - Submission form n.d.); which is used to compare one or more query motifs against one or more databases of target motifs (and their reverse complements when applicable), and reports for each query a list of target motifs,

ranked by p -value. The E -value and the q -value of each match is also reported. The q -value is the minimal false discovery rate at which the observed similarity would be deemed significant (Gupta et al. 2007).

2.3 Gene ontology

The identified motifs were also fitted to another motif-based sequence analysis tool, Gene Ontology for Motif (GOMo); under MEME suite version 5.3.0 to scan all promoters using nucleotide motifs provided to determine if any motif is significantly associated with genes linked to one or more Genome Ontology (GO) terms. The significant GO terms can suggest the biological roles of the motifs. The program searches in a set of ranked genes for enriched GO terms associated with high-ranking genes. The genes can be ranked, for example, by applying a motif scoring algorithms on their upstream sequence (Buske et al. 2010a).

2.4 Identification of CpG islands

A server based online CpG islands prediction analysis for genome sequence exploration was used to search for CpG islands (Database of CpG Islands n.d.). Genomic islands play an important role in medical, methylation and biological studies. CpG islands are springs of characteristically unmethylated DNA with a high content of the Cytosine (C) and Guanine (G); i.e., high GC content compared to the bulk DNA. If genes are expressed, the CpG islands of promoters are unmethylated whereas when methylation of these sites occurs it is essential in expression of genes. Majority (40%), of the CpG islands are located in the promoter region. Herein, the CpGPAP platform employs the use of Takai and Jones algorithm. In this algorithm, a DNA sequence with a length of >500bp, GC content >55% in that region and Observed/Expected (O/E) ratio of >0.65 was considered. The algorithm can successfully exclude false positives from short repetitive sequences (Chuang et al. 2012). Additionally, CLC Genomics Workbench ver. 5.5.2 (CLC Genomics Workbench n.d.) was used to search for the restriction enzyme *MspI* cutting sites with a fragment selection of 40 to 220pbs having in mind that the region enriches CpG islands.

3. Results

3.1 Number of TSSs and predictive values of *L. donovani* BPK282A1 genome

The TSS of the parasite was predicted among forty putative genes of *L. donovani* BPK282A1 genome and the important variables such as gene identification number, corresponding promoter region, predictive value and location of the TSS were summarized in table 1. This *in silico* analysis of *L. donovani* BPK282A1 genome showed that a single sequence can have as low as one TSS and as high as four TSSs. Where there were two or more TSSs the one with highest predictive score was considered the best TSS location of the gene. In all cases, the TSSs have a predictive score that ranged from 0.80 to 1.00. Majority (82.3%), of the TSSs identified were located below -500bp away from the start codon of the gene being expressed. Surprisingly, the TSS of one gene was located at -1338 which is very apart from the start codon. The number of TSSs were significantly varied between 1 and 4 among the promoter region sequences. The analysis revealed that about forty percent (40%) of the gene had only one TSS

while only one gene had four TSSs. On the other hand, the location of the TSS of one sequence (*Pro-LDBPK_060-590*) was remarkably close to the start codon which was only 2bps away upstream of the respective gene.

Table 1 Number, predictive score and location of TSSs in *L. donovani* BPK282A1genes

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 the start codon
LDBPK_141460	Pro-LDBPK_141-460	2	0.84, 0.99	-385
LDBPK_230300	Pro-LDBPK_230-300	1	0.88	-835
LDBPK_212210	Pro-LDBPK_212-210	3	0.85, 0.89, 0.99	-679
LDBPK_251640	Pro-LDBPK_251-640	1	0.96	-692
LDBPK_151500	Pro-LDBPK_151-500	3	0.81, 0.86, 0.99	-554
LDBPK_340900	Pro-LDBPK_340-900	1	1.00	-126
LDBPK_262720	Pro-LDBPK_262-720	2	0.97, 0.99	-513
LDBPK_040440	Pro-LDBPK_040-440	1	0.94	-159
LDBPK_171020	Pro-LDBPK_171-020	1	0.88	-1338
LDBPK_151100	Pro-LDBPK_151-100	2	, 0.81, 1.00	-177
LDBPK_151540	Pro-LDBPK_151-540	3	1.00, 0.88, 1.00	-106, -697
LDBPK_060590	Pro-LDBPK_060-590	2	0.99, 0.87	-2
LDBPK_030760	Pro-LDBPK_030-760	2	0.97, 1.00	-132
LDBPK_111140	Pro-LDBPK_111-140	1	0.99	-366
LDBPK_160580	Pro-LDBPK_160-580	3	0.86, 0.89, 0.92	-898
LDBPK_250780	Pro-LDBPK_250-780	2	0.97, 0.92	-85
LDBPK_362640	Pro-LDBPK_362-640	4	0.88, 0.94, 0.99, 0.82	-703
LDBPK_280030	Pro-LDBPK_280-030	3	0.82, 0.96, 0.92	-452
LDBPK_131090	Pro-LDBPK_131-090	1	1.00	-222
LDBPK_040570	Pro-LDBPK_040-570	1	0.89	-431
LDBPK_303160	Pro-LDBPK_303-160	1	0.94	-925
LDBPK_180690	Pro-LDBPK_180-690	2	0.98, 0.96	-466
LDBPK_180620	Pro-LDBPK_180-620	2	0.97, .98	-928
LDBPK_180590	Pro-LDBPK_180-590	3	0.84, 0.98, 0.89	-504
LDBPK_180480	Pro-LDBPK_180-	1	0.92	-195

LDBPK_180460	Pro-LDBPK_180-460	2	0.90, 0.86	-601
LDBPK_180370	Pro-LDBPK_180-370	1	0.92	-598
LDBPK_180360	Pro-LDBPK_180-360	2	0.90, 0.80	-368
LDBPK_180270	Pro-LDBPK_180-270	1	0.81	-346
LDBPK_181100	Pro-LDBPK_181-100	2	0.91, 0.97	-927
LDBPK_181080	Pro-LDBPK_181-080	2	0.82, 0.95	-330
LDBPK_180920	Pro-LDBPK_180-920	1	0.90	-857
LDBPK_180810	Pro-LDBPK_180-810	2	0.81, 0.89	-402
LDBPK_180790	Pro-LDBPK_180-790	1	0.97	-504
LDBPK_180650	Pro-LDBPK_180-650	1	0.86	-426
LDBPK_180580	Pro-LDBPK_180-580	3	0.98, 0.91, 0.81	-437
LDBPK_180510	Pro-LDBPK_180-510	1	1.00	-559
LDBPK_170660	Pro-LDBPK_170-660	2	0.99, 1.00	-796
LDBPK_170610	Pro-LDBPK_170-610	3	1.00, 0.97, 1.00	-586, -837
LDBPK_170890	Pro-LDBPK_170-890	2	0.86, 1.00	-422

Results from NNPP tool prediction are considered to be reliable at 0.8 cutoff values (BDGP: Neural Network Promoter Prediction n.d.)

3.2 Promoter regions with potential motifs in *L. donovani* BPK282A1

The spatial distribution and location of the motifs were presented below in figure 1. The motifs were found on both positive and negative strands. The shortest distance of the location of the motif was only 2bps upstream of the TSS while the longest distance was 4941bps upstream of the TSS of the respective gene. Majority, 188 (72.3%) of the motifs were highly distributed on positive strand than the negative strand.

On the other hand, the study also revealed that in about twenty-seven (67.5%) sequences the location of the first motif was held in the first -500bps upstream of the respective gene. Furthermore, the promoter sequence of gene *LDBPK_151500* was exceptionally too long and contains seven motifs among which only one motif was located on the negative strand.

The study has also revealed seven candidate motifs in the promoter regions being analyzed using server-based MEME suite motif identifier. The analysis was performed at a cutoff value of 0.80 and provided MI, MII, MIII, MIV, MV, MVI and MVII motifs. The minimum site per motif was two while the maximum site per motif was forty. The identified motifs were presented below in table 2 with their respective total number of binding sites and the level of statistical significance (*E*-values). It was shown that MI was the common promoter motif for all (100%) *L. donovani* *BPK282A1* genes that regulates transcription and translation of the genes being studied. Statistically, it shares at least 85% of the promoters containing each one of the motifs with an *E*-value of $4.0e-116$. The sequence Logo of the best motif identified was presented in figure 2.

Table 2 Common motifs identified in *L. donovani* *BPK282A1* genome promoter regions with their respective binding sites and e-values

Discovered motif	Number (%) of promoters containing each one of the motifs	E-value	Motif width	Total number of binding sites
MI	34 (85)	4.0e-116	27	34
II	40 (100)	2.6e-095	41	40
III	39 (97.5)	5.4e-051	41	39
IV	40 (100)	8.8e-044	21	40
V	29 (72.5)	9.9e-038	21	29
VI	40 (100)	3.4e-032	21	40
VII	29 (72.5)	5.0e-020	29	29

TOMTOM, a web-based application was used for further analysis of the similarities between the identified common motif and other similar motifs which are publicly available in known registered biological databases of eukaryotes. Accordingly, the web-based tool was used three different databases, namely; jolma2013, JASPAR2018_CORE_vertbrates_non-redundant and uniprobe_mouse to search for similarities and 843, 579 and 386 motifs were used from each of the databases respectively and only seven matches similar to the submitted MI motif were identified. Six of the identified matches were closely related to zinc finger proteins that bind to the GC-rich regions in DNA and regulate various cellular functions. The remaining one of the matches was found to be related to AP2/B3-like transcriptional factor family protein which is found in *Hordeum vulgare* (barley) and the gene functions in vernalization. The possible binding candidate motifs of MI, suggested species, e-value and gene functions were presented in table 3.

Table 3 Possible transcription factors that could bind to common candidate motif MI

Candidate Transcription Factor	Species	E-value	Corresponding gene	Function
zf-C2H2	Mus musculus	1.36e+00	Gm397	Not specified
C2H2 zinc finger factors	Homo sapiens	2.05e+00	RREB1	Increases expression of calcitonin, which may be involved in Ras/Raf-mediated cell differentiation
AP2/B3-like transcriptional factor family protein	Hordeum vulgare	3.97e+00	VRN1	Required for vernalization
EGR2 DBD (Zinc finger, C2H2 type, C2H2-type zinc finger, Zinc finger double domain)	Homo sapiens	6.43e+00	EGR2	Early growth response 2
Three-zinc finger Krüppel-related factors	Homo sapiens	6.43e+00	EGR2	Binds specifically to 9-bp target sequences containing two CpG sites that can potentially be methylated at four cytosine bases
Three-zinc finger Kruppel-related factors	Homo sapiens	6.70e+00	KLF9	Binding of the encoded protein to a single GC box inhibits mRNA expression while binding to tandemly repeated GC box elements activates transcription
EGR2_full (Domain of unknown function (DUF3446), Zinc finger, C2H2 type, C2H2-type zinc finger, Zinc-finger double domain)	Homo sapiens	8.18e+00	EGR2	Early growth response 2

3.3 Gene ontology

Having been compared the common motif using publicly known databases in TOMTOM, another server based investigation using GOMo (Gene Ontology for Motifs) (Buske et al. 2010b; GOMo Results n.d.) was used to look for GO terms for promoter regions of *L. donovani* *BPK282A1* gene and no significant GO term was associated with the common motif (MI) submitted (Figure 3).

3.4 Determination of the CpG islands of *L. donovani* BPK282A1 genome in promotor and gene body regions

CpG islands, stretches of DNA that are 500-1500bp long with CG:GC ratio of more than 0.65% in promoter and gene body regions were determined using two algorithms. First, an online CpG Island finder (Database of CpG Islands n.d.) that depends on Takai and Jones rigorous criteria where parameters were set as ob/ex ratio of the CpG islands ≥ 0.65 , GC content of at least 50% and minimal length of ≥ 500 bps was used to identify possible CpG islands present, the number of CpG islands found and the percent of GC contents in each of the promoter and gene body regions (Takai and Jones 2002). Herein, the search revealed that thirty-one (31) promoter regions of the *L. donovani* BPK282A1 contain at least 1 CpG island. Whereas nine of the sequences did not contain any CpG islands (Table 4).

Due to technical limitations, Table 4 is only available as a download in the supplemental files section.

To examine the presence of CpG islands in the gene body regions using online CpG Island finder (Database of CpG Islands n.d.), genes with less than 500bps were excluded maintaining consistency depending on the predetermined criteria of Takai and Jones as discussed previously. Accordingly, three genes (*LDBPK_151100*, *LDBPK_060590* and *LDBPK_180690*) were excluded from the search. Thus, from the included thirty-seven gene body regions in the search algorithm thirty-two of them had at least one CpG island and two genes (*LDBPK_180360* and *LDBPK_180270*) had 100% CpG concentration. In five of the gene body regions (*LDBPK_212210*, *LDBPK_030760*, *LDBPK_131090*, *LDBPK_303160* and *LDBPK_180790*) no CpG island was found (Table 5).

Table 5 CpG islands and percent GC contents predicted in the gene body sequences of *L. donovani* BPK282A1 genes

Name/gene ID	Possible CpG island (Yes/No)	Number of CpG islands found	%GC content
LDBPK_141460	Yes	1	61
LDBPK_230300	Yes	1	61
LDBPK_212210	No	-	-
LDBPK_251640	Yes	1	59
LDBPK_151500	Yes	1	58
LDBPK_340900	Yes	1	58
LDBPK_262720	Yes	1	57
LDBPK_181120	Yes	1	60
LDBPK_040440	Yes	1	61
LDBPK_171020	Yes	1	64
LDBPK_151540	Yes	1	59
LDBPK_030760	No	-	-
LDBPK_111140	Yes	1	50
LDBPK_160580	Yes	1	66
LDBPK_250780	Yes	1	52
LDBPK_362640	Yes	1	62
LDBPK_280030	Yes	1	53
LDBPK_131090	No	-	-
LDBPK_040570	Yes	1	60
LDBPK_303160	No	-	-
LDBPK_180620	Yes	1	62
LDBPK_180590	Yes	1	63
LDBPK_180480	Yes	1	62
LDBPK_180460	Yes	1	59
LDBPK_180370	Yes	1	59
LDBPK_180360	Yes	1	100
LDBPK_180270	Yes	1	100
LDBPK_181100	Yes	1	56
LDBPK_181080	Yes	1	55
LDBPK_180920	Yes	1	64
LDBPK_180810	Yes	1	60
LDBPK_180790	No	-	-
LDBPK_180650	Yes	1	59
LDBPK_180580	Yes	1	59
LDBPK_180510	Yes	1	60
LDBPK_170660	Yes	1	62
LDBPK_170610	Yes	1	58
LDBPK_170890	Yes	1	60

Another algorithm (CLC Genomics Workbench 5) involving the use of restriction enzyme *MspI* was employed to examine the fragments of CpG islands in both the promoter and gene body regions

regardless of the existence of methyl group. Methylation protects DNA cleavage by endonucleases *HpaII* at internal C position of its recognition site CCGG sequence. *MspI* restriction enzyme is an isoschizomer of *HpaII* that can cleave DNA irrespective of the presence of methyl group at the internal C residue positions (Waalwijk and Flavell 1978).

In four sequences in the promoter region, namely; *Pro-LDBPK_180-690*, *Pro-LDBPK_180-590*, *Pro-LDBPK_181-080* and *Pro-LDBPK_170-890* no restriction site was detected using this *Insilco* digestion. Whereas, at least one and at most fifty-eight restriction sites were detected in all the genes with greater than 500bps in their body sequences. To determine the fragment sizes in both the promoter and gene body regions a size selection of 40 to 220bps were employed to prevent degradation products where the CpG islands are enriched (Gu and Tomazou 2010). On the virtue of this model, fragment sizes of fourteen gene promoter sequences (*Pro-LDBPK_151-100*, *Pro-LDBPK_151-540*, *Pro-LDBPK_111-140*, *Pro-LDBPK_160-580*, *Pro-LDBPK_250-780*, *Pro-LDBPK_280-030*, *Pro-LDBPK_131-090*, *Pro-LDBPK_180-620*, *Pro-LDBPK_180-460*, *Pro-LDBPK_180-270*, *Pro-LDBPK_181-100*, *Pro-LDBPK_180-810* and *Pro-LDBPK_170-660*) and six gene body sequences (*LDBPK_250780*, *LDBPK_280030*, *LDBPK_303160*, *LDBPK_180360*, *LDBPK_181100* and *LDBPK_180650*) did not meet the sizes selection criteria of 40-220bps enrich CpG islands (Table 6).

Table 6 Size selection restriction analysis with *MspI* in promoter and gene body regions using CLC genome Workbench

Promoter region name	Number of cut site(s)/Nucleotide cut site(s)	Promoter region	Gene body region		
		Fragment size(s)	Name of gene	Number of cut site(s)/Nucleotide cut site(s)	Fragment size(s)
Pro-LDBPK_141-460	9(257, 457, 837, 882, 960, 1030, 1507, 1799, 1953)	45, 70, 78, 154,200	LDBPK_141460	17(125, 188, 279, 613, 641, 682, 796, 820, 860, 994, 1141, 1216, 1277, 1484, 1642, 1669, 1774)	63, 91, 41, 114, 40, 134, 147, 75, 61, 207, 158, 105
Pro-LDBPK_230-300	3(388, 494, 594)	106, 100	LDBPK_230300	5(155, 194, 625, 673, 1188)	48
Pro-LDBPK_251-640	4(171, 215, 413, 889)	44, 198	LDBPK_212210	10(16, 52, 148, 290, 476, 494, 522, 1177, 1234, 1252)	96, 142, 186, 57
Pro-LDBPK_151-500	17(625, 1444, 1611, 1869, 2581, 2798, 2840, 2890, 2966, 3050, 3299, 3808, 4018, 4049, 4235, 4362, 4574)	31, 42, 50, 76, 84, 127, 167, 186, 212, 210, 217	LDBPK_251640	5(627, 734, 1003, 1224, 1537)	107
Pro-LDBPK_340-900	2(299, 914)	-	LDBPK_151500	3(316, 554, 612)	58
Pro-LDBPK_262-720	7(131, 569, 607, 691, 721, 837, 903)	84, 116, 66	LDBPK_340900	3(558, 613, 764)	55, 151
Pro-LDBPK_040-440	2(380, 425)	45	LDBPK_262720	2(85, 169)	84
Pro-LDBPK_171-020	15(194, 812, 1121, 1255, 1361, 1372, 1390, 1493, 2110, 2205, 2695, 2796, 3426, 3717, 3866)	134, 106, 103, 95, 101, 149	LDBPK_040440	5(34, 229, 595, 707, 1260)	195, 112
Pro-LDBPK_151-100	2(513, 994)	-	LDBPK_171020	37(161, 481, 578, 584, 776, 835, 943, 1379, 1420, 1445, 1652, 1705, 1885, 2021, 2215, 2442, 2491, 2518, 2601, 2713, 2863, 2887, 3344, 3452, 3542, 3816, 3909, 4216, 4861, 4954, 5001, 5150, 5239, 5245, 5449, 5551, 5845)	97, 192, 59, 108, 436, 41, 207, 53, 180, 136, 194, 49, 83, 112, 150, 108, 90, 93, 93,47, 149, 89, 204, 102
Pro-LDBPK_151-540	2(25, 690)	-	LDBPK_151100	2(16, 141)	125
Pro-LDBPK_060-590	2(280, 360)	80	LDBPK_151540	15(47, 190, 627, 801, 963, 1398, 1424, 1442, 1503, 1804, 1864, 1922, 2498, 2628, 2697)	143, 174, 162, 61, 60, 58, 130, 69
Pro-LDBPK_030-760	3(437, 740, 789)	49	LDBPK_060590	2(130, 355)	-
Pro-LDBPK_111-	4(173, 191, 447, 815)	-	LDBPK_030760	58(158, 206, 233, 443, 828, 1228, 1274, 1279, 1336, 1462, 1864, 1907, 1921,	48, 210, 46, 57, 126, 43,

140				1992, 2206, 2282, 2710, 3292, 3473, 3496, 3663, 3911, 3926, 4068, 4268, 4732, 4924, 5014, 5428, 5447, 5473, 5494, 5602, 5621, 5687, 5969, 6089, 6157, 6620, 6665, 6737, 6940, 7159, 7231, 7573, 7790, 7910, 8105, 8138, 8433, 8602, 9012, 9180, 9228, 9233, 9242, 9284, 9380)	71, 214, 76, 181, 167, 142, 200, 192, 90, 108, 66, 120, 68, 45, 72, 203, 219, 72, 217, 120, 195, 169, 168, 48, 42, 96
Pro-LDBPK_160-580	2(300, 618)	-	LDBPK_111140	33(46, 202, 239, 317, 446, 497, 794, 830, 842, 988, 997, 1067, 1273, 1412, 1420, 1688, 2152, 2158, 2179, 2335, 2452, 2522, 2542, 2585, 2641, 2794, 2920, 2939, 2971, 3185, 3306, 3370, 3413)	156, 78, 129, 51, 146, 70, 206, 139, 156, 117, 70, 43, 56, 153, 126, 214, 121, 64, 43
Pro-LDBPK_250-780	1(105)	-	LDBPK_160580	5(313, 374, 763, 800, 907)	61, 107
Pro-LDBPK_362-640	7(22, 86, 98, 158, 397, 606, 735)	64, 60, 209, 129	LDBPK_250780	1(105)	-
Pro-LDBPK_280-030	3(242, 546, 812)	-	LDBPK_362640	11(47, 286, 424, 838, 1026, 1387, 1507, 1973, 2106, 2555, 2662)	138, 188, 120, 133, 107
Pro-LDBPK_131-090	1(83)	-	LDBPK_280030	3(12, 402, 1077)	-
Pro-LDBPK_040-570	9(61, 414, 540, 601, 625, 696, 782, 855, 1105)	126, 61, 71, 86, 73	LDBPK_131090	18(4, 44, 80, 210, 524, 626, 650, 683, 782, 1177, 1331, 1402, 1413, 1600, 1622, 1904, 1912, 2124)	40, 130, 102, 99, 154, 71, 187, 212
Pro-LDBPK_303-160	2(183, 322)	139	LDBPK_040570	7(11, 59, 262, 487, 538, 595, 799)	48, 203, 51, 57, 204
Pro-LDBPK_180-620	3(45, 321, 793)	-	LDBPK_303160	3(421, 425, 448)	-
Pro-LDBPK_180-480	5(673, 830, 887, 980, 1197)	157, 57, 93, 217	LDBPK_180690	2(38, 203)	165
Pro-LDBPK_180-460	3(399, 633, 926)	-	LDBPK_180620	11(307, 416, 452, 568, 973, 1025, 1048, 1145, 1174, 1244, 1352)	109, 116, 52, 97, 70, 108
Pro-LDBPK_180-370	5(98, 128, 495, 545, 616)	50, 71	LDBPK_180590	6(320, 361, 442, 920, 1163, 1198)	41
Pro-LDBPK_180-360	5(318, 485, 715, 925, 970)	167, 210, 45	LDBPK_180480	7(49, 374, 398, 410, 473, 966, 1029)	63
Pro-LDBPK_180-270	1(398)	-	LDBPK_180460	5(52, 124, 419, 424, 491)	72, 67
Pro-LDBPK_181-	2(510, 519)	-	LDBPK_180370	8(23, 33, 147, 634, 640, 767, 1126, 1191)	114, 127, 65

100					
Pro-LDBPK_180-920	2(817, 999)	182	LDBPK_180360	2(1023, 1030)	-
Pro-LDBPK_180-810	3(230, 827, 1372)	-	LDBPK_180270	5(54, 310, 698, 892, 988)	194, 96
Pro-LDBPK_180-790	3(327, 651, 757)	106	LDBPK_181100	2(290, 811)	-
Pro-LDBPK_180-650	9(173, 502, 546, 660, 723, 785, 872, 1098, 1816)	44, 114, 63, 62, 87	LDBPK_181080	7(112, 452, 550, 887, 1403, 2345, 2400)	98, 55
Pro-LDBPK_180-580	3(141, 510, 626)	116	LDBPK_180920	8(223, 257, 308, 1088, 1246, 1282, 1389, 1790)	51, 158, 107
Pro-LDBPK_180-510	9(226, 422, 902, 918, 1106, 1175, 1473, 1586, 1929)	196, 188, 69, 113	LDBPK_180810	5(32, 202, 701, 730, 759)	170
Pro-LDBPK_170-660	1(149)	-	LDBPK_180790	5(193, 257, 304, 704, 1014)	64, 47
Pro-LDBPK_170-610	3(124, 192, 783)	68	LDBPK_180650	2(53, 403)	-
Pro-LDBPK_212-210	2(263, 427)	164	LDBPK_180580	2(15, 113)	98
			LDBPK_180510	23(305, 554, 560, 631, 722, 1073, 1142, 1243, 1423, 1447, 1484, 1573, 1628, 1718, 1831, 1985, 2059, 2063, 2086, 2258, 2306, 2510, 2540)	71, 91, 69, 101, 180, 89, 55, 90, 113, 154, 74, 172, 48, 204
			LDBPK_170660	19(403, 553, 578, 617, 833, 938, 1069, 1205, 1298, 1416, 1464, 1774, 2104, 2150, 2423, 2446, 2486, 2692, 2840)	150, 216, 105, 131, 136, 93, 118, 48, 46, 40, 206, 148
			LDBPK_170610	3(124, 192, 783)	68
			LDBPK_170890	7(290, 296, 357, 579, 746, 1033, 1685)	61, 167

4. Discussion

Reliable identification of the TSSs is a prerequisite to realize the regulation and expression of genes in any organism in a promoter sequence (Halees, Leyfer, and Weng 2003). As stated by Dinka and colleague, genes may have one or more TSSs contributing to genetic variation of an organism (Dinka and Le 2018) which is in agreement with this study where the analysis revealed about forty (40%) percent of the genes had one TSS and sixty percent of them had two or more TSSs. They also showed that the locations for 80% of the TSS were within -518bp relative to the translation start codon which is also

similar to 82.3% in this study where the location of TSSs identified were located below -500bp away from the start codon of the gene being expressed.

In eukaryotes, the occurrence of transcription regulation within 25bps upstream of the TSS were associated with core promoter while it was associated with promoter proximal elements anywhere within the range of 50 to 200bps upstream of the TSS. On the other hand, transcription regulation in either direction and orientation distant from transcription initiation sites were associated with distal enhancer promoter (Nikolov and Burley 1997). In this study, the analysis showed that the location of the best TSS for one promoter sequences (Pro-LDBPK_060-590) was at -2bps from the start codon where transcription is associated with core promoter. Whereas, the location of best TSSs in seven promoter sequences were associated with promoter proximal elements. The other 32 promoter sequences had the best TSS location of greater than -200pbs away from start codon and therefore may be associated with distal promoter enhancer elements.

Searching for motifs is not an insignificant task and seems bottleneck where efforts are required to large scale projects that attempt to exactly identify the binding positions of TFs and to understand mathematical models employed for TFs (Reid et al. 2010). The most commonly used mathematical model is position matrix weight (PMW) model that can be used to understand the position dependent frequency or probability of each nucleotide in a motif (Boeva 2016). With this regard, the information contents of the most conserved nucleotide for the seven motifs identified were 28.9, 26.4, 23.8, 17.2, 22.3, 18.3 and 19.7 for MVI, MVII, MVIII, MVIV, MVV, MVVI and MVVII respectively as shown by sequence logo where the total height of each bin was considered the information content in bits of the corresponding position. Searching the binding sites in long promoter regions is likely to yield false prediction of the binding sites while it can also help to exceptionally provide true predictions (Boeva 2016). In this study, 20% of the promoter regions included in the analysis had a nucleotide length of greater than 2000bps. Thus, additional analysis using different models is required whether the predicted values for those promoter regions were true or false.

Razavi *et al.* in their study on cloning and expression of leishmanolysin gene from *Leishmania major* in primate cell lines found that *L. guyanensis* and *L. major* possess zinc binding motifs which is essential for attachment of the parasite to the host cell and internalization (Razavi et al. 2001). Accordingly, from the seven matches that were identified in web based TOMTOM search in this study, six of them were associated with zinc finger protein binding TF in *Homo sapiens* and *Mus musculus*. Hence, there is possible functional similarities between the identified motif of *L. donovani* BPK282A1 gene and motifs of the *L. guyanensis* and *L. major* being reviewed. However, cloning and expression studies are required to establish scientifically plausible similarities.

Up to 20% of eukaryotic genes and promoters contain tandem repeats. Tandem repeats are sequences that are highly unstable and subject to increased mutation rates which is usually 10 to 100,000times compared to mutations in other parts of a gene (Liang et al. 2012; Tsang 2014). The occurrence of the tandem repeats in a motif sequence is responsible for evolution of certain traits (Liang et al. 2012). With

this regard, the identified best motif sequence logo of the *L. donovani* *BPK282A1* gene in this study seems to have a tandem repeat sequence which probably contributes to the emergence of drug resistant strain of *L. donovani* as mutations in DNA binding motifs are responsible to changes in gene expression and function. Furthermore, it was also known that *Leishmania donovani* inhabits phagolysosomes of the macrophages (Singh et al. 2016) where it undergoes phagosome mutation regulation which will eventually avoid its destruction. As the result, antigen presentation, oxidative damage, apoptosis and immune activation are depleted and the availability of the nutrients are favored on the contrary (Podinovskaia and Descoteaux 2015). The reason for such escape mechanism may be due to the presence of the tandem repeats in the protein binding regions of the promoter sequences as discussed earlier. However, a thorough study that includes wet laboratory and bioinformatic tool must be conducted to understand in which stage of the parasite mutation typically undergoes.

Determination and identification of CpG islands is one of the integral part of biological, methylation and medical studies (Chuang et al. 2012). In vertebrates, 50% of the genes contain CpG islands (also called gene markers) in their promoter region. In spite of the fact that CpG islands are unmethylated in promoter regions, 80% of mammalian genome are methylated. As mutation rate in methylated CpG is higher than that of the unmethylated CpG, there is greater chance of loss of CpG islands which subsequently results in change of gene expression and function (Han and Zhao 2008). As previously discovered by Williams et al promoters with high CpG content are largely hypomethylated (Williams, Christensen, and Helin 2012). Das *et al* in 2014 discovered that *L. donovani* genes are highly enriched for CpG islands. In addition, they also revealed that CpG motifs delayed programmed cell death in macrophages which may be helpful to extend the parasite lifespan conferring the advantage of infection progression (Das et al. 2015). In line with this, the current *Insilco* analysis also discovered that *L. donovani* *BPK282A1* genome is rich in CpG in both the promoter and gene body regions. On the other hand, It was also explained that antiapoptotic effect of *L. donovani* was observed where there was DNA methylation (Das et al. 2015).

The GC coverage of a gene region can be affected by its GC concentration. Hence, regions with 50 to 60% GC content do have the highest coverage compared to gene regions with GC content of 70-80% and gene regions with GC content of 30 to 40% (Amr and Funke 2015). With this regard, this *Insilco* analysis showed that majority (57.5%) of the gene promoters and close to half (47.5%) of the gene body regions do have the highest GC coverage. On the other hand, this study also indicated a great GC content variation that ranged from 50% to 100% (in two genes) which agrees with a previous literature indicating GC content in a eukaryotic organisms is highly variable and may be from as low as 30% to as high as 90% (Sharp 2001).

5. Conclusion

This *Insilco* analysis has unraveled great variations in the number of TSSs in promoter region and percent GC concentration in both promoter and gene body sequences. The study also discovered CpG rich sequences and the common motifs shared by *Leishmania donovani* *BPK282A1* genes and the possible transcription factors associated with the sequence. Hence, the investigation unfolded gene expression

regulation of *Leishmania donovani* BPK282A1 genes bringing new insight and contributing to understand mechanisms that underly its adaptation to different environments and development of robust drug and vaccine that targets the regulatory regions.

Declarations

Data availability

All the data were obtained from NCBI and available online.

Competing interest

All authors do not have any conflict of interest and have seen and read the final version of this manuscript.

Authors' contribution

KJ conceived the idea and developed the proposal. He also obtained the data from NCBI and performed all the statistical analysis using different software and wrote the initial and final version of the paper. MK supervised the overall work and gave advise on how to write the manuscript. AM helped in editing and providing technical support on how to use online tools.

References

- Amr, Sami S., and Birgit Funke. 2015. "Targeted Hybrid Capture for Inherited Disease Panels." In *Clinical Genomics*, Elsevier Inc., 251–69.
- Bailey, Timothy and, and Charles Elkan. 1994. "Fitting a Mixture Model by Expectation Maximization." *AAAI Press*: 3–9.
- "BDGP: Neural Network Promoter Prediction." https://fruitfly.org/seq_tools/promoter.html (February 3, 2021).
- Boeva, Valentina. 2016. "Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells." *Frontiers in Genetics* 7(FEB).
- Buates, Sureemas, and Greg Matlashewski. 2001. "General Suppression of Macrophage Gene Expression During Leishmania Donovanii Infection ." *The Journal of Immunology* 166(5): 3416–22.
- Buske, Fabian A., Mikael Bodén, Denis C. Bauer, and Timothy L. Bailey. 2010a. "Assigning Roles to DNA Regulatory Motifs Using Comparative Genomics." *Bioinformatics* 26(7): 860–66.
- Buske, Fabian A, Mikael Bodén, Denis C Bauer, and Timothy L Bailey. 2010b. "Assigning Roles to DNA Regulatory Motifs Using Comparative Genomics." *Bioinformatics* 26(7): 860–66.

<https://doi.org/10.1093/bioinformatics/btq049>.

Chuang, Li Yeh, Cheng Huei Yang, Ming Cheng Lin, and Cheng Hong Yang. 2012. "CpGPAP: CpG Island Predictor Analysis Platform." *BMC Genetics* 13.

"CLC Genomics Workbench." https://download.cnet.com/CLC-Genomics-Workbench-64-bit/3000-2054_4-75445915.html (November 6, 2020).

Das, Sushmita et al. 2015. "Unmethylated CpG Motifs in the L. Donovanii DNA Regulate TLR9-Dependent Delay of Programmed Cell Death in Macrophages ." *Journal of Leukocyte Biology* 97(2): 363–78.

"Database of CpG Islands." <http://dbcat.cgm.ntu.edu.tw/> (November 3, 2020).

Dinka, Hunduma, and Minh Thong Le. 2018. "Analysis of Pig Vomeronasal Receptor Type 1 (V1R) Promoter Region Reveals a Common Promoter Motif but Poor CpG Islands." *Animal Biotechnology* 29(4): 293–300. <https://doi.org/10.1080/10495398.2017.1383915>.

Dinka, Hunduma, and Ashenafi Milkesa. 2020. "Unfolding SARS-CoV-2 Viral Genome to Understand Its Gene Expression Regulation." *Infection, Genetics and Evolution* 84(May): 104386. <https://doi.org/10.1016/j.meegid.2020.104386>.

Downing, Tim et al. 2011. "Whole Genome Sequencing of Multiple Leishmania Donovanii Clinical Isolates Provides Insights into Population Structure and Mechanisms of Drug Resistance." : 2143–56.

Duncan, Robert et al. 2011. "Identification and Characterization of Genes Involved in Leishmania Pathogenesis: The Potential for Drug Target Selection." 2011.

"GOMo Results." https://meme-suite.org/meme//opal-jobs/appGOMO_5.3.11612355265197394517341/gomo.html (February 3, 2021).

Gu, Hongcang, and Eleni Tomazou. 2010. "NIH Public Access." (February).

Gupta, Shobhit, John A. Stamatoyannopoulos, Timothy L. Bailey, and William Stafford Noble. 2007. "Quantifying Similarity between Motifs." *Genome Biology* 8(2).

Halees, Anason S., Dmitriy Leyfer, and Zhiping Weng. 2003. "PromoSer: A Large-Scale Mammalian Promoter and Transcription Start Site Identification Service." *Nucleic Acids Research* 31(13): 3554–59.

Han, Leng, and Zhongming Zhao. 2008. "Comparative Analysis of CpG Islands in Four Fish Genomes." 2008.

Holzmuller, Philippe et al. 2002. "Nitric Oxide-Mediated Proteasome-Dependent Oligonucleosomal DNA Fragmentation in Leishmania Amazonensis Amastigotes." 70(7): 3727–35.

"Introduction - MEME Suite." <https://meme-suite.org/meme/> (February 3, 2021).

- Kanhere, Aditi, and Manju Bansal. 2005. "A Novel Method for Prokaryotic Promoter Prediction Based on DNA Stability." *BMC Bioinformatics* 6(February).
- Kassahun, Aysheshm et al. 2015. "Detection of Leishmania Donovanii and L. Tropica in Ethiopian Wild Rodents." *Acta Tropica* 145: 39–44. <http://dx.doi.org/10.1016/j.actatropica.2015.02.006>.
- Liang, Tong, Xiaodan Fan, Qiwei Li, and Shuo Yen R. Li. 2012. "Detection of Dispersed Short Tandem Repeats Using Reversible Jump Markov Chain Monte Carlo." *Nucleic Acids Research* 40(19).
- McGwire, Bradford S, and Abhay Satoskar. 2013. "Sir William Osler Medicine Master Class Leishmaniasis: Clinical Syndromes and Treatment." (June).
- Meyers, Wayne, Mary Klassen-fischer, Walter Reed, and National Military. 2016. "Visceral Leishmaniasis." (January 2011).
- Michaloski, Jussara S. et al. 2011. "Common Promoter Elements in Odorant and Vomeronasal Receptor Genes." *PLoS ONE* 6(12).
- Mukherjee, Ishita, Abhijit Chakraborty, and Saikat Chakrabarti. 2016. "Identification of Internalin-A-like Virulent Proteins in Leishmania Donovanii." *Parasites and Vectors* 9(1). <http://dx.doi.org/10.1186/s13071-016-1842-5>.
- "National Center for Biotechnology Information." <https://www.ncbi.nlm.nih.gov/> (November 2, 2020).
- Nikolov, D. B., and S. K. Burley. 1997. "RNA Polymerase II Transcription Initiation: A Structural View." *Proceedings of the National Academy of Sciences of the United States of America* 94(1): 15–22.
- Podinovskaia, Maria, and Albert Descoteaux. 2015. "Leishmania and the Macrophage: A Multifaceted Interaction." *Future Microbiology* 10(1): 111–29.
- Rastrojo, Alberto et al. 2018. "Genomic and Transcriptomic Alterations in Leishmania Donovanii Lines Experimentally Resistant to Antileishmanial Drugs." *International Journal for Parasitology: Drugs and Drug Resistance* 8(2): 246–64. <https://doi.org/10.1016/j.ijpddr.2018.04.002>.
- Razavi, Mohamad Reza, Mohammad Azizi, Amir Amanzadeh, and Mehdi Assmar. 2001. "Cloning and Expression of Leishmanolysin Gene from Leishmania Major in Primate Cell Lines." (January).
- Reid, John E et al. 2010. "Variable Structure Motifs for Transcription Factor Binding Sites."
- "Release Notes - MEME Suite." (MEME - Submission form (meme-suite.org)).
- Sharp, P.M. 2001. "Codon Usage Bias." In *Encyclopedia of Genetics*, Elsevier, 402–6.
- Singh, Alok Kumar, Rajeev Kumar Pandey, Chandrima Shaha, and Rentala Madhubala. 2016. "MicroRNA Expression Profiling of Leishmania Donovanii-Infected Host Cells Uncovers the Regulatory Role of

MIR30A-3p in Host Autophagy." *Autophagy* 12(10): 1817–31.

<http://dx.doi.org/10.1080/15548627.2016.1203500>.

Sinha, Roma, Mathu Malar C, Subhadeep Das, and Sonali Das. 2018. "Genome Plasticity in Cultured Leishmania Donovanii: Comparison of Early and Late Passages." 9(July): 1–20.

Steverding, Dietmar. 2017. "The History of Leishmaniasis." *Parasites and Vectors* 10(1): 1–10.

Takai, Daiya, and Peter A. Jones. 2002. "Comprehensive Analysis of CpG Islands in Human Chromosomes 21 and 22." *Proceedings of the National Academy of Sciences of the United States of America* 99(6): 3740–45.

"Tomtom - Submission Form." <https://meme-suite.org/meme/tools/tomtom> (February 3, 2021).

Tsang, Adrian. 2014. "Fungal Genomics." *Briefings in Functional Genomics* 13(6): 421–23.

Verma, Jitender Kumar, Ruchir Rastogi, and Amitabha Mukhopadhyay. 2017. "Leishmania Donovanii Resides in Modified Early Endosomes by Upregulating Rab5a Expression via the Downregulation of MiR-494." *PLoS Pathogens* 13(6): 1–25.

Waalwijk, C., and R. A. Flavell. 1978. "MspI, an Isoschizomer of HpaII Which Cleaves Both Unmethylated and Methylated HpaII Sites." *Nucleic Acids Research* 5(9): 3231–36.

Williams, Kristine, Jesper Christensen, and Kristian Helin. 2012. "DNA Methylation: TET Proteins-Guardians of CpG Islands?" *EMBO Reports* 13(1): 28–35. <http://dx.doi.org/10.1038/embor.2011.233>.

Figures

Figure 1

Block diagram showing number and relative positions of motifs in promoter regions of *L. donovani* BPK282A1 genes

Figure 2

Common motif sequence logo of Leishmania donovani BPK282A1 genes

INVESTIGATED MOTIFS

Figure 3

GO-term showing no significant association with the submitted motif

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table4.docx](#)