Observational Study

Identification of functional tumor necrosis factor-alpha promoter variants associated with *Helicobacter pylori* infection in the Sudanese population: Computational approach

Abeer Babiker Idris, Alaa B Idris, Manal A Gumaa, Mohammed Babiker Idris, Amanda Elgoraish, Mohamed Mansour, Dalia Allam, Bashir MO Arbab, Nazar Beirag, El-Amin M Ibrahim, Mohamed A Hassan

**ORCID number:** Abeer Babiker Idris 0000-0001-6436-5963; Alaa B Idris 0000-0002-2111-8070; Manal A Gumaa 0000-0002-3580-3327; Mohammed Babiker Idris 0000-0002-7316-8504; Amanda Elgoraish 0000-0003-2420-3120; Mohamed Mansour 0000-0002-1679-3071; Dalia Allam 0000-0002-7828-6091; Bashir MO Arbab 0000-0003-4555-9521; Nazar Beirag 0000-0003-2435-9969; El-Amin M Ibrahim 0000-0002-5175-0763; Mohamed A Hassan 0000-0003-3938-6898.

**Author contributions:** Idris AB and Hassan MA conceptualized and designed the study; Idris AB, Idris A, Gumaa MA, Idris MB, Mansour M, Allam D and Arbab MOB participated in sample collection and analysis of the data; Idris AB and Elgoraish A statistically analyzed the data; Idris AB performed the bioinformatics analysis, interpretation of the data and drafting the article; Idris AB, Beirag N, Mansour M, Hassan MA and Ibrahim EM revised the manuscript’s language and critically assessed the intellectual content; all authors approved the final version of the article to be published.

**Institutional review board**

**Abeer Babiker Idris**, Department of Agricultural Science and Technology, Institute of Natural and Applied Sciences, Erciyes University, Kayseri 38039, Turkey

**Abeer Babiker Idris**, Department of Medical Microbiology, Faculty of Medical Laboratory Sciences, University of Khartoum, Khartoum 11111, Sudan. abeer.babiker89@gmail.com

**Alaa B Idris**, Department of Neurosurgery, Ribat University Hospital, Khartoum 11111, Sudan

**Manal A Gumaa, El-Amin M Ibrahim**, Department of Medical Microbiology, Faculty of Medical Laboratory Sciences, University of Khartoum, Khartoum 11111, Sudan

**Mohammed Babiker Idris**, BioMétrie Clinical and Application Advisor, Al-Jeel Medical Co., Riyadh 11422, Saudi Arabia

**Amanda Elgoraish**, Department of Epidemiology, Tropical Medicine Research Institute, Khartoum 11111, Sudan

**Mohamed Mansour, Dalia Allam**, Department of Gastroenterology, Ibn Sina Specialized Hospital, Khartoum 11111, Sudan

**Bashir MO Arbab**, Department of Gastroenterology, Modern Medical Centre, Khartoum 11111, Sudan

**Nazar Beirag**, Biosciences, College of Health, Medicine and Life Sciences, Brunel University, London UB8 3PH, Uxbridge, United Kingdom

**Mohamed A Hassan**, Department of Bioinformatics, Africa city of technology, Khartoum 11111, Sudan

**Mohamed A Hassan**, Department of Bioinformatics, DETAGEN Genetic Diagnostics Center, Kayseri 38350, Turkey

**Mohamed A Hassan**, Department of Translation Bioinformatics, Detuvax Biotech, Kayseri 38350, Turkey

**Corresponding author:** Abeer Babiker Idris, DSc, Academic Research, Department of Agricultural Science and Technology, Institute of Natural and Applied Sciences, Erciyes
Abstract

BACKGROUND
Helicobacter pylori (H. pylori) is a ubiquitous bacterium that affects nearly half of the world’s population with a high morbidity and mortality rate. Polymorphisms within the tumor necrosis factor-alpha (TNF-A) promoter region are considered a possible genetic basis for this disease.

AIM
To functionally characterize the genetic variations in the TNF-A 5’-region (-584 to +107) of Sudanese patients infected with H. pylori using in silico tools.

METHODS
An observational study was carried out in major public and private hospitals in Khartoum state. A total of 122 gastric biopsies were taken from patients who had been referred for endoscopy. Genomic DNA was extracted. Genotyping of the TNF-A-1030 polymorphism was performed using PCR with confronting two-pair primer to investigate its association with the susceptibility to H. pylori infection in the Sudanese population. Furthermore, Sanger sequencing was applied to detect single nucleotide polymorphisms in the 5’-region (-584 to +107) of TNF-A in H. pylori-infected patients. Bioinformatics analyses were used to predict whether these mutations would alter transcription factor binding sites or composite regulatory elements in this region. A comparative profiling analysis was conducted in 11 species using the ECR browser and multiple-sequence local alignment and visualization search engine to investigate the possible conservation. Also, a multivariate logistic regression model was constructed to estimate odds ratios and their 95% confidence intervals for the association between TNF-A -1030, sociodemographic characteristics and H. pylori infection. Differences were statistically significant if \( P < 0.05 \). Statistical analyses were performed using Stata version 11 software.

RESULTS
A total of seven single nucleotide polymorphisms were observed in the TNF-A 5’-region of Sudanese patients infected with H. pylori. Only one of them (T > A, -76) was located at the in silico-predicted promoter region (-146 to +10), and it was predicted to alter transcription factor binding sites and composite regulatory elements. A novel mutation (A > T, +27) was detected in the 5’ untranslated region, and it could affect the post-transcriptional regulatory pathways. Genotyping of TNF-A-1030 showed a lack of significant association between -1030T and susceptibility to H. pylori and gastric cancer in the studied population (\( P = 0.1756 \)) and (\( P = 0.8116 \)), respectively. However, a significant association was detected between T/C genotype and H. pylori infection (39.34% vs 19.67%, odds ratio \( = 2.69 \), 95% confidence interval: 1.17-6.17, \( P = 0.020 \)). Mammalian conservation was observed for the (-146 to +10) region in chimpanzee (99.4%), rhesus monkey (95.6%), cow (91.8%), domesticated dog (89.3%), mouse (84.3%), rat (82.4%) and opossum (78%).

CONCLUSION
Computational analysis was a valuable method for understanding TNF-A gene expression patterns and guiding further \textit{in vitro} and \textit{in vivo} experimental validation.

Key Words: 5’-region; Promoter; TNF-A; \textit{Helicobacter pylori}; \textit{In silico} analysis; Sudan

Core Tip: According to the literature, a considerable number of polymorphisms have
been discovered in the tumor necrosis factor-alpha (TNF-A) promoter. A crucial question is whether these polymorphisms have any significant functional impact on disease incidence or severity. Seven single nucleotide polymorphisms were detected in the TNF-A 5′-region; only one of them, TNF-A-76, was located at the in silico-predicted promoter region (-146 to +10). This single nucleotide polymorphism may lead to the modification of the transcriptional regulation of TNF-A in Helicobacter pylori infection. Nevertheless, this conclusion cannot substitute for the experimental proofs, but it can provide direction or insight for them.

**Citation:** Idris AB, Idris AB, Gumaa MA, Idris MB, Elgoraish A, Mansour M, Allam D, Arbab BM, Beirag N, Ibrahim EAM, Hassan MA. Identification of functional tumor necrosis factor-alpha promoter variants associated with Helicobacter pylori infection in the Sudanese population: Computational approach. *World J Gastroenterol* 2022; 28(2): 242-262

**URL:** https://www.wjgnet.com/1007-9327/full/v28/i2/242.htm

**DOI:** https://dx.doi.org/10.3748/wjg.v28.i2.242

**INTRODUCTION**

*Helicobacter pylori* (*H. pylori*) is a widespread bacterium that affects nearly half of the world’s population with a high morbidity and mortality rate[1]. Generally, the decreasing prevalence of *H. pylori* infection is continuing in many parts of North America and Europe. Still, no such decline has been noted in most developing countries, including Sudan, in which the *H. pylori* infection represents a critical public health challenge[2-4]. Although there is a common consensus that the risk of acquisition and transmission of *H. pylori* can be minimized and prevented to a large extent, this drop in those countries, whether it is a local phenomenon or a national trend, has to be investigated[2].

Infection with *H. pylori* is usually acquired during childhood. However, it may persist if not treated in the host’s stomach throughout life due to the synchronized balance between the bacteria and its host[5,6]. The long duration of *H. pylori* in the gastric environment will develop chronic gastritis in *H. pylori*-infected persons. At the same time, a few may be complicated to more severe diseases such as peptic ulcer, atrophic gastritis, mucosa-associated lymphoid tissue lymphoma or gastric adenocarcinoma[7,8]. The variability in *H. pylori* infection outcomes is determined by bacterial factors, genetic characteristics of the host, especially those regarding polymorphisms in specific cytokines, and the environmental factors[7,9,10]. However, the crucial role of cytokine-related polymorphisms in the susceptibility or severity of *H. pylori* infection has been attracting considerable attention after El-Omar et al[11] proposed that functional polymorphisms in proinflammatory cytokine genes in *H. pylori*-infected patients are associated with a two- to three-fold increased gastric cancer risk. Because tumor necrosis factor-alpha (TNF-a) production has been implicated as an essential factor in immune regulations and inflammatory responses against *H. pylori* infection[12], polymorphisms within the TNF-A promoter are considered a possible genetic basis for this disease[13].

TNF-a is a potent pleiotropic proinflammatory cytokine that mediates diverse biological and pathological processes, in addition to its responsibility for the regulation of two opposite processes: proliferation and apoptosis (reviewed in[14,15]). TNF gene transcription is regulated by nucleoprotein complexes known as enhanceosomes that involve distinct sets of transcription factors and coactivators at the proximal TNF promoter and in the pattern of chromosomal organization of the TNF/lymphotoxin locus that function in synergy to drive transcription[16]. TNF enhanceosome assembly is cell type- and stimulus-specific, partially because of differential occupancy of overlapping DNA motifs in the TNF promoter. For example, after lipopolysaccharide stimulation of monocytes where no inducible translocation of NFATp can be detected, Sp1 and Ets/Elk proteins can successfully bind to sites that NFATp occupies in T cells[13,14]. Furthermore, many TNF activator binding motifs can be recognized by more than one class of transcription factor depending on the ambient concentration of different elements in the nucleus and the stimulus and cell type[16]. These features provide the TNF promoter with a remarkable degree of flexibility to respond to multiple stimuli in a specific manner through a short cis-
regulatory region[13].

The TNF locus lies within the major histocompatibility complex region on the short arm of human chromosome 6p21.3, a highly polymorphic region[17]. Many bi-allelic polymorphisms and microsatellites are found at or around the TNF locus[14,18]. Several have been suggested to be implicated in diseases susceptibility/resistance or severity by modifying the transcriptional regulation of TNF[13,18]. There are different transition variants in the TNF-A promoter region at the positions of -237, -307, -375, -856, -862 and -1030 in which the positions -307 and -237 have been most frequently evaluated for the association with H. pylori infection (susceptibility or progression) [13]. It has been reported that the presence of the adenine allele at -237 or -307 positions causes higher constitutive and inducible transcriptional levels than the guanine allele, which results in enhanced production of TNF-a up to five-fold in vitro [19,20]. In addition, the single nucleotide polymorphism (SNP) at the -1030 position, which is frequently observed in African populations[21,22], has been suggested to be related to high TNF-a production[23]. However, the influence of these SNPs on TNF-a production is not fully known and still a contradictory topic of debate, which is explained, in part, by ethnic differences and genetic variations between populations [19,20].

To our knowledge, there are no previous studies in Sudan that have addressed the association between TNF-A and H. pylori infection or H. pylori-associated diseases. Therefore, in this study, we applied DNA Sanger sequencing to detect genetic variations in the TNF-A 5'-region (-584 to +107) of Sudanese patients infected with H. pylori and predicted if these SNPs could alter transcription factor binding sites (TFBSes) or composite regulatory elements (CEs) using a computational approach. In addition, a comparative profiling analysis was performed to investigate the conservation of these SNPs in 11 species. Regulation mechanisms of TNF-a production may be linked to the difference in responses to H. pylori infection. This emphasizes the importance of studying the regulation of TNF-A gene expression.

**MATERIALS AND METHODS**

**Research methodology**

In the present study, in silico analysis was employed for the TNF-A 5'-region (-584 to +107) of the Sudanese patients infected with H. pylori in two steps: (1) In silico prediction of the promoter region; and (2) In silico analysis of the predicted promoter region (-146 to +10). To obtain accurate results, we applied various software/servers to predict the promoter region, CpG island and regulatory motifs because promoters have complex and specific architecture and contain multiple transcription factors involved in particular transcription regulation[24]. Therefore, the unique composition of TFBSs and different features of each promoter may lead to different powers and algorithms for promoter identification[25]. In addition, genotyping of the TNF-A-1030 T>C polymorphism was performed to assess its association with H. pylori infection in the Sudanese population using PCR with confronting two-pair primer method.

However, the initial studies have numbered the positions of SNPs relative to the TNF mRNA cap site incorrectly (e.g., -308 instead of -307 and -238 instead of -237)[13]. In contrast, the corrected numbering should have proceeded 5’ from the adenine at the +1 position[26] as in the Cytokine Gene Polymorphism In The Human Disease Database[27], and we used it throughout this study. The methodology followed in this study is represented in Figure 1.

**Study population and study setting**

An observational study was carried out in major public and private hospitals in Khartoum state in 2019-2020. These hospitals included Soba teaching hospital, Ibibina special hospital, Modern Medical Centre, Al Faisal Specialized Hospital and Police hospital. In addition, the molecular laboratory processes were conducted in the Molecular Biological Research laboratory at the Faculty of Medical Laboratory Sciences at the University of Khartoum.

The matched case-control formula was selected to calculate the sample size using Epi Info software version 7[28,29], assuming 95% confidence interval (CI), 80% power of the study, one ratio of control to the case, 15% of controls exposed, 3.36 odds ratio (OR) and 37.2% of cases exposed. According to the sample size calculation, the present study included 122 patients referred for upper gastrointestinal tract endoscopy, and most of them were because of dyspepsia. Out of that, 61 were H. pylori-negative patients and regarded as controls. The demographic characteristics of the study
population are shown in Table 1.

The diagnosis of gastroduodenal diseases was based on the assessment of an experienced gastroenterologist, and in cancerous cases, histological examinations were also performed to confirm the diagnosis. The gastric pathologies of the study population are presented in Table 2.

The study population’s demographic and clinical data were obtained by personal interviews, a structured questionnaire and a review of case records. The selection criteria included the Sudanese population from both sexes who were not taking antibiotics or nonsteroidal anti-inflammatory drugs. Written informed consent was granted by the participants after they were informed of the objectives and purposes of the study.

**DNA extraction and PCR amplification of specific 16S rRNA gene of H. pylori**

For DNA extraction purposes, gastric biopsies were collected in 400 µL of phosphate buffered saline. For histological investigation, the biopsies were preserved in 10% formalin. DNA extraction was performed by using innuPREP DNA Mini Kit (analytikjena AG, Germany) and followed the protocol that was given by the manufacturer, as previously described in[30].

Extracted DNA was amplified for the specific 16S rRNA gene to diagnose and confirm the infection of *H. pylori* using the following primers (primers: F: 5’-GGGCAATCAGCTAGGTATTG-3’; and R: 5’-GTTAGAGACAGCTATGCTCC-3’) [31] and the previously described PCR condition[4]. The amplified product for the 16S rRNA is 522 bp.

**Conventional PCR amplification and sequencing of the TNF-A promoter region**

To amplify the promoter polymorphisms -308 and -238 of the TNF-A gene, primers: F: 5’-GCTTGTCCCTGCTACCCGC-3’ and R: 5’-GCTAAGAGACAGCTATGCTCC-3’[32] were used. The following thermal cycling conditions were used: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 94 °C for 60 s, 58 °C for 30 s and 72 °C for 60 s, with a final extension at 72 °C for 10 min. The amplified PCR product is 691 bp located between -585 bp upstream and 107 bp downstream of the TNF-A gene.
Table 1 Demographic characteristics of the study population, n (%)  

| Variable         | Helicobacter pylori (+ve) | Helicobacter pylori (-ve) | Total | P value |
|------------------|---------------------------|---------------------------|-------|---------|
| Mean age in yr ± Standard deviation (range) | 44.00 ± 16.99 (15-85) | 44.74 ± 18.10 (17-89) | 44.37 ± 17.48 (15-89) | 0.9184 |
| Gender           | Male                       | Female                    |       |         |
|                  | 36 (50.00)                 | 25 (50.00)                | 61    |         |
|                  | 36 (50.00)                 | 25 (50.00)                | 61    |         |
| Residence        | Urban                      | Rural                     |       |         |
|                  | 24 (44.44)                 | 37 (54.41)                | 61    |         |
|                  | 30 (55.56)                 | 31 (45.99)                | 61    | 0.3622  |
| Hospital         | Public                     | Private                   |       |         |
|                  | 21 (33.87)                 | 41 (58.57)                | 61    |         |
|                  | 31 (49.21)                 | 29 (41.43)                | 61    | 0.0867  |

1 Mann-Whitney U, Gaussian Approximation. +ve: Positive; -ve: Negative.

Table 2 Gastric pathologies of the study population, n (%)  

| Endoscopy results | n (%) | Helicobacter pylori (+ve) | Helicobacter pylori (-ve) | P value1 |
|-------------------|-------|---------------------------|---------------------------|---------|
| Normal gastric findings | 9 (7.40) | 4 (6.45)                  | 5 (7.94)                  | 0.2686  |
| Esophagitis       | 4 (3.30) | 3 (4.80)                  | 1 (1.59)                  |         |
| Esophageal varices| 6 (4.90) | 1 (1.61)                  | 5 (7.94)                  |         |
| Gastritis         | 55 (45.10) | 28 (45.16)               | 27 (42.86)               |         |
| Duodenitis        | 6 (4.90) | 3 (4.84)                  | 3 (4.76)                  |         |
| Peptic ulcer      | 27 (22.13) | 17 (27.42)              | 10 (15.87)               |         |
| Cancer            | 15 (12.30) | 5 (8.20)                  | 10 (16.4)                 |         |
| Total             | 122 (100) | 61 (50.00)               | 61 (50.00)                |         |

1 χ² test, df: 6. +ve: Positive; -ve: Negative.

Fourteen PCR products of *H. pylori*-infected patients, which had the clearest bands, were sent for commercial DNA purification and Sanger dideoxy sequencing by Macrogen Inc., Korea.

**Sequence analysis and SNPs detection**

Two chromatograms (forward and reverse) for each sequence were visualized, checked for their quality and analyzed with the Finch TV program version 1.4.0[33]. In addition, the Basic Local Alignment Search Tool nucleotide (https://blast.ncbi.nlm.nih.gov/) was used to look for and assess nucleotide sequence similarities[34].

For SNP detection in the TNF-A promoter region, multiple alignment sequences were performed for the tested sequences with a reference sequence (TNF-A; NG_007462) using Clustal W[35].

**Bioinformatics analysis of the TNF-A promoter region in H. pylori-infected patients**

In silico prediction of the promoter: In this study, we employed five programs for human promoter prediction and recognition of human PolII promoter region and start of transcription (TSS) which include: Berkeley Drosophila Genome Project (http:// www.fruitfly.org/) [36], Promoter 2.0 Prediction Server (http://www.cbs.dtu.dk/) [37], FPROM, TSSG and TSSW (http://softberry.com/) which are based on neural network and linear discriminant approach[25,38,39].

In silico analysis of the predicted promoter region: Assessment for the presence of promoter-associated features: The computationally predicted promoter regions were additionally assessed for the presence of promoter-associated features using the ENCODE data (https://www.encodeproject.org/)[40-42]. These features include promoter-associated histone marks, DNase I hypersensitivity clusters, broad
chromatin state segmentation, transcription factor ChIP-seq and CpG islands.

Prediction of CpG islands: A CpG island is a DNA segment with high GC and CpG dinucleotide content and often regarded as a marker for the initiation of gene expression. In this study, we applied MethPrimer and CpG finder software (http://www.softberry.com/berry.phtml?topic=cpgfinder&group=programs&subgroup promoter) to predict CpG islands using default search parameter values for the software: CpG island length >200 bp, CG% > 50%, and Obs/Exp > 0.6.[43,44]

Prediction of TFBSs: Prediction of the potentially functional TFBSs is a crucial step in the chain of promoter analytical events. Therefore, we applied five prediction software programs to analyze the predicted promoter region for possible TFBSs: (1) AliBaBa2 (http://www.gene-regulation.com/) [45]; (2) Alggen Promo (http://alggen.lsi.upc.es/cgi-bin/promo_v3/)[46,47]; (3) TF-Bind (http://tfbind.hgc.jp/)[48]; (4) Tfsitescan (http://www.ifti.org)[49]; and (5) Gene Promoter Miner (GPMiner) (http://GPMiner.mbc.nctu.edu.tw/), which is an integrated system. For predicting TFBSs, MATCH tool was utilized to scan TFBSs in an input sequence.[50]

Prediction of CEs: CE is the minimal functional unit composed of two closely located DNA binding sites for distinct transcription factors, but its regulatory function is qualitatively different from the regulation effects of either individual DNA binding site. Thus, the MatrixCatch algorithm (http://gnaweb.helmholtz-hzi.de/cgi-bin/MCatch/MatrixCatch.pl) was used to identify the CEs in our region (-146 to +10) [51].

Comparative profiling analysis: ECR Browser (http://ecrbrowser.dcode.org)[52], NCBI Basic Local Alignment Search Tool nucleotide (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Clustal W (https://www.genome.jp/tools-bin/clustalw)[35] were utilized to analyze the possible conservation of the predicted promoter region. The conservation was assessed in 11 species which include: Cow (Bos taurus), chimpanzee (Pan troglodytes), rhesus monkey (Macaca mulatta), dog (Canis lupus familiaris), mouse (Mus musculus), rat (Rattus norvegicus), chicken (Gallus gallus), zebrafish (Danio rerio), frog (Xenopus laevis), opossum (Monodelphis domestica), fugu pufferfish (Takifugu rubripes) and spotted green pufferfish (Tetraodon nigroviridis). Furthermore, multiple-sequence local alignment and visualization search engine (https://mulan.dcode.org/) was used to evaluate the conservation of the promoter’s SNPs and to screen the possible conservation of TFBSs at these SNP locations.[53].

Molecular detection of the TNF-A-1030 C/T polymorphism using PCR with confronting two-pair primer method

Two sets of primers were used in PCR with confronting two-pair primer to detect the TNF-A-1030 C/T polymorphism as follow: for allele T, the sequences of the primers were F: 5’-AAGGCTCTGAAAGCCAGCT-3’ and R: 5’-CCAGACCCT-GACTTTTCTTCA-3’; and for allele C, F: 5’-GAAGCAAGGAGAGCTGAGAAGAC-3’ and R: 5’-CTTCCATAGCCTGGACATTCT-3’ [54]. The PCR mixture components were the same as in [30]. The amplification conditions were initial denaturation at 95 °C for 10 min, followed by 30 cycles of denaturation at 95 °C for 1 min, annealing at 66 °C for 1 min, extension at 72 °C for 1 min and final extension at 72 °C for 5 min. The resulting PCR products were visualized by electrophoresis on a 1% agarose gel containing ethidium bromide. The predicted size was confirmed using a 50 bp DNA molecular weight marker. Genotyping was distinguished as follows; 444 bp and 316 bp for TT genotype, 444 bp, 316 bp and 174 bp for TC genotype and 444 bp and 174 bp for CC genotype.

Statistical analysis

Fisher’s test or χ² test investigated Hardy-Weinberg equilibrium of allele and genotype distributions. Descriptive statistical analysis was applied to calculate percentages, proportions and means ± standard deviation. Regarding the incidence of H. pylori infection, differences in frequency distribution by age were examined by the Mann-Whitney test, while differences in the frequency distribution of the categorical demographic and clinical variables of the study population were assessed by Fisher’s test or χ² test. A multivariate logistic regression model was constructed to estimate OR and their 95% CI for the association between TNF-A-1030, sociodemographic characteristics and H. pylori infection. Differences were considered to be statistically significant if P < 0.05. The statistical analyses were performed using Stata version 11 (StataCorp, College Station, Texas) software.
RESULTS

**Nucleotide variations in the 5′-regulatory region (-585 to +107) of the TNF-A gene**
In the Sudanese *H. pylori*-infected patients, a total of seven nucleotide variations were detected in the 5′-regulatory region (six substitution mutations and one insertion mutation of C at +70). Among which, five were bimodally mutated heterozygous SNPs, and two are homozygous SNPs, as illustrated in Table 3 and Figure 2. The results of the multiple sequence alignment were visualized in Jalview. The nucleotide sequences of the TNF-A 5′-region (-584 to +107) were deposited in the GenBank database under the following accession numbers: from MW251712 to MW251725.

**In silico prediction of the TNF-A promoter regions**
Five types of promoter prediction programs were applied to predict the promoter regions of the TNF 5′-region (-584 to +107), and the results are overviewed in Table 4. Neural Network Promoter Prediction (NNPP version 2.2) predicted one promoter region, located at +1 bp relative to the TNF-α mRNA cap site. In addition, FPROM and TSSG programs predicted two promoter regions -146 bp and +5 bp, respectively. The TSSW program also expected these two regions. In comparison, no promoter region was predicted by Promoter 2.0 Prediction Server.

**In silico analysis of predicted TNF-A promoter region**

**Presence of promoter-associated features:** ENCODE data revealed a high level of transcription factor occupancy patterns, promoter-associated histone modifications and DNase I hypersensitivity at the in silico predicted promoter region, from -146 to +10 nt relative to the TNF-α mRNA cap site. Furthermore, CpfFinder and MethPrimer software were employed to predict the presence of CpG islands. However, both software predicted no CpG islands present in the predicted promoter region, and ENCODE data confirmed no presence of the CpG island in the predicted promoter (see Figure 3).

**Conservancy of the predicted promoter:** As illustrated in Figure 4, the ECR Browser showed mammalian conservation for the (-146 to +10) region in chimpanzee (*Pan troglodytes* - pan-Tro2) (99.4%), rhesus monkey (*Macaca mulatta* - rheMac2) (95.6%), cow (*Bos taurus* - bosTau3) (91.8%), domesticated dog (*Canis lupus familiaris* - canFam3) (89.3%), rat (*Rattus norvegicus* - rn4) (82.4%), mouse (*Mus musculus* - mm9) (84.3%) and opossum (*Monodelphis domestica* - monDom5) (78.0%). But the region was not conserved in chicken (*Gallus gallus*), frog (*Xenopus laevis*), zebrafish (*Danio rerio*), fugu pufferfish (*Takifugu rubripes*) and spotted green pufferfish (*Tetraodon nigroviridis*). An overview of the mammalian conservation of the in silico-predicted promoter of TNF is shown in Figure 5.

**Prediction of TFBSs:** In the present study, we used five software programs to predict TFBSs. To ensure accurate analysis, only factors predicted by three out of the five programs or predicted by two programs but verified in the literature were selected. The five prediction programs reported multiple putative TFBSs within the (-146 to +10) region, as illustrated in Table 5. Moreover, screening of this region using the NCBI SNP databases showed four SNPs, as shown in Table 5. The ECR Browser and NCBI Basic Local Alignment Search Tool nucleotide revealed the conservation of these SNPs in chimpanzees, rhesus monkey, cow and dog. Multiple-sequence local alignment and visualization presented multiple TFBSs to be located at rs765073823 and rs537401710. The overview from the multiple-sequence local alignment and visualization software for conserved TFBSs predicted to be conserved (100%) between humans, chimpanzees, rhesus monkeys, cows, and domesticated dogs is summarized in Table 6.

**Prediction of CEs:** In the present study, CEs were predicted using MatrixCatch, which can find experimentally known regulatory elements (both single sites and pairs) and novel regulatory elements by computational comparison to the known ones in a library of CE models. An overview of predicted CEs by MatrixCatch is presented in Table 7.

**Allele and genotype frequencies of TNF-A-1030 and susceptibility to *H. pylori* infection**
Sixty-one participants from each group (*H. pylori*-infected and uninfected) were successfully genotyped for TNF-A-1030 polymorphism; a representative gel is shown
Table 3 Nucleotide variations in the 5'-regulatory (-584 to +107) region of tumor necrosis alpha in Sudanese Helicobacter pylori-infected patients

| SNP                        | Event          | SNP locus1 | Chromosomeposition | Serialnumber (rs) |
|----------------------------|----------------|------------|--------------------|-------------------|
| C>G (homozygous)           | Transversion   | -567       | NC_007462.1:g.4422 C>G | rs2857711         |
| G>A (heterozygous)         | Transition     | -375       | NC_007462.1:g.4614 G>A | rs1800750         |
| G>A (heterozygous and homozygous) | Transition | -307       | NC_007462.1:g.4682 G>A | rs1800629         |
| G>A (heterozygous)         | Transition     | -237       | NC_007462.1:g.4752 G>A | rs361525          |
| T>A (heterozygous)         | Transversion   | -76        | NC_007462.1:g.4913 T>A | rs41297589        |
| A>T (heterozygous)         | Transversion   | +27        | NC_007462.1:g.5016 A>T2 | -                |
| Insertion of C             | -              | +70        | NC_007462.1:g.5058-5059 InsC | -                |

1Position of the single nucleotide polymorphism is relative to the tumor necrosis factor-alpha mRNA cap site[26].
2Novel mutation at the 5' untranslated region detected in this study. SNP: Single nucleotide polymorphism.

Table 4 Overview of the in silico predicted tumor necrosis factor-alpha promoter regions for the respective prediction programs. The most predicted region is indicated in bold

| Prediction program | Predicted promoter regions1 |
|--------------------|----------------------------|
| BDGP (NNPP version 2.2)2 | -40 to +10 |
| Promoter 2.0 Prediction Server | no promoter |
| FPROM | +5 |
| TSSG3 | -146 |
| TSSW4 | -140 |
| TSSW+6 | +6 |

1All positions are relative to the tumor necrosis factor-alpha mRNA cap site[26].
2NNPP: Neural network promoter prediction; Threshold 0.80.
3Threshold for LDF- 4.00.
4Thresholds for TATA+ promoters - 0.45, for TATA-/enhancers - 3.70.

in Figure 6. Allele and genotype frequencies of TNF-A-1030 polymorphism in patients infected with H. pylori and uninfected controls are presented in Table 8. Although the TNF-A-1030-C allele was more prevalent in patients infected with H. pylori compared to uninfected controls, no significant association was found between the allele frequencies of TNF-A-1030 and H. pylori infection (27.87% vs 19.67%, OR = 0.6339, 95%CI: 0.3490-1.151, P = 0.176). However, a significant association was detected between T/C genotype and H. pylori infection (39.34% vs 19.67%, OR = 2.69, 95%CI: 1.17-6.17, P = 0.020) (Tables 8 and 9). In contrast, no significant association was observed between allele and genotype frequencies of TNF-A-1030 and gastric cancer (P = 0.8116 and P = 0.3900, respectively) (Table 10). The allelic and genotype distributions at -1030 of TNF-A were in Hardy-Weinberg equilibrium, with non-significant χ2 values (for uninfected subjects P = 0.1520; and subjects with benign disorders P = 0.2089).

DISCUSSION

Many SNPs have been observed in the vicinity of the TNF locus, particularly in the distal TNF promoter. Some of these have been implicated in the disease susceptibility, resistance or severity by modifying the transcriptional regulation of TNF[13,56-58]. In this study, we functionally analyzed SNPs in the TNF-A 5'-region (~584 to +107) of the Sudanese patients infected with H. pylori and had divergent clinical outcomes that ranged from simple asymptomatic gastritis to more serious conditions, i.e. peptic ulcer disease and gastric cancer. The exploratory sequencing of TNF-A 5'-region (~584 to +107) of the Sudanese patients infected with H. pylori revealed six previously reported
Table 5 Summary of the in silico predicted transcription factor binding sites for tumor necrosis factor-alpha (-146 to +10) region

| Transcription Factor | Position | SNP in binding site | Software |
|----------------------|----------|---------------------|----------|
|                      | Start    | End                 | AllBaba2.1 | Alggen Promo | Tfsitescan | TF-Binding | GPMiner |
| Elk-1                | -145     | -130                | X         |
| Sp1                  | -138     | -126                | X         |
| AP-2                 | -129     | -118                | X         |
| Ets-1                | -118     | -111                | X         |
| PEA3/ Ets-1          | -118     | -110                | X         |
| T3R_TRE1             | -115     | -108                | X         |
| TNF-alpha-CRE        | -112     | 92                  | X         |
| C/EBP-beta           | -100     | -74                 | X         |
| NF-kappaB            | -99      | -86                 | X         |
| AP-1                 | -67      | -57                 | X         |
| NF-kappaB            | -55      | -46                 | X         |
| Sp1                  | -54      | -41                 | X         |
| Sp1                  | -48      | -39                 | X         |
| GATA                 | -40      | -30                 | X         |
| AP-2                 | -38      | -27                 | X         |
| TBP                  | -29      | -20                 | X         |

1) Position is relative to the tumor necrosis factor-alpha transcription start site.

2) X means the transcription factor is predicted by the software that is marked with X.

3) Observed in this study. SNP: Single nucleotide polymorphism.

Table 6 Overview of conserved transcription factor binding sites predicted by multiple-sequence local alignment and visualization

| Transcription Factor | Binding sequence | Position | Conserved SNPs |
|----------------------|------------------|----------|---------------|
|                      |                  | Start    | End           |               |
| VS.CETS168_Q6        | gCTTCCTC         | -115     | -108          | rs765073823   |
| VS.ETS6_Q6           | gCTTCCTc         | -115     | -108          | rs765073823   |
| VSSP1.Q4.01          | ttccC GCCCCttc   | -51      | -39           | rs765073823   |
| VSSP1.Q6.01          | tcCCCCGCCct      | -50      | -41           | rs765073823   |
| VSSP1.Q2.01          | cCCCCGCCct      | -49      | -40           | rs765073823   |

1) All positions are relative to the tumor necrosis factor-alpha mRNA cap site. Multiple-sequence local alignment and visualization (https://mulan.dcode.org/) to be conserved (100%) between human, chimpanzee, rhesus monkey, cow, and domesticated dog at the (-146 to +10) region. The crucial transcription factor for tumor necrosis factor transcription, nuclear factor, was not included.

SNPs located at -567, -375, -307, -237, -76 and +70 and one novel SNP at +27 relative to the TNF-A transcription start site. A C to G homozygous transversion at -567 nt was found in all our studied populations. At the same time, the novel A to T heterozygous transversion at +27 nt was observed in five patients who had normal upper gastroendoscopy or gastritis. This SNP is located in the 5' untranslated region. Therefore, it could affect the post-transcriptional regulatory pathways that control mRNA localization, stability and translation efficiency [39]. Further studies are encouraged to investigate this SNP in terms of the frequency of the minor allele (T) in the Sudanese population and its functional significance using computational and experimental...
Intriguingly, the heterozygous SNP T>A at -76 was located at the in silico-predicted promoter region. However, in this study, the in silico analysis predicted promoter region was located from -146 to +10 nt relative to the TNF-α mRNA cap site using different algorithms [26]. This finding is in agreement with previous studies conducted on 411 TNF promoters from individuals of distinct ethnic backgrounds that concluded that the human TNF promoter sequences up to 200 bp relative to the transcription start site involved in enhanceosome formation and the regulation of the gene are completely conserved in humans [21, 26, 60, 61]. Also, the ENCODE data showed a high level of the hallmarks of cis-regulatory regions at the in silico predicted promoter region (-146_+10), which include nuclease hypersensitivity and histone modifications [40-42].

Functional elements in the mammalian gene promoters can be detected by high conservation between species distantly related on the evolutionary scale [62, 63]. In this study, the computational comparative analysis revealed mammalian conservation for the -146 to +10 region in chimpanzee (Pan troglodytes - pan-Tro2) (99.9%), rhesus monkey (Macaca mulatta - rheMac2) (93.7%), cow (Bos taurus - bosTau3) (78.2%), domesticated dog (Canis lupus familiaris - canFam2) (77.8%), rat (Rattus norvegicus - rn4) (76.3%), mouse (Mus musculus - mm9) (78.7%) and opossum (Monodelphis domestica - monDom5) (72.2%). In an earlier study conducted by Leung et al [64], the region from positions -131 to -63 and -53 and -45 relative to the start site of TNF gene transcription showed high sequence conservation among primates, with complete conservation approaches.

Table 7 Summary of the in silico predicted composite regulatory elements for the -146 to +10 region of tumor necrosis factor-alpha

| Composite element | MatrixName for a 1st element | Score, strand | Distance in between | MatrixName for a 2nd element | Score, strand | Position of CE, orientation | Composite score | P value | Sequence |
|------------------|-----------------------------|--------------|---------------------|-------------------------------|--------------|-----------------------------|----------------|---------|----------|
| CE00266          | V$ETS_Q6                   | 0.991+, 12   |                     | V$AP1_Q2_01                   | 0.831+, -118 | +                           | 0              | 1.08e-03 | GCTTCCTCAGGTGAGCTCATGGG |
| CE00109          | V$AP1_C                    | 0.853-, 8    |                     | V$NFAT_Q6                    | 0.943-, -66  | -                           | 0.007          | 3.42e-03 | TTGAATGATCTCTTCCCCGC  |
| CE00150          | V$AP1_C                    | 0.853-, 8    |                     | V$NFAT_Q6                    | 0.943-, -66  | -                           | -0.237         | 3.42e-03 | TTGAATGATCTCTTCCCCGC  |
| CE00058          | V$HMGIY_Q6                 | 0.955-, 5    |                     | V$NFKB_Q6_01                 | 0.759-, -78  | -                           | -0.09          | 6.89e-03 | GTTTCCCCCTG        |
| CE00058          | V$HMGIY_Q6                 | 0.955-, 4    |                     | V$NFKB_Q6_01                 | 0.662-, -78  | -                           | 0.005          | 9.20e-03 | GTTTCCCCCTG        |
| CE00064          | V$HNF3_Q6                  | 0.780+, 15   |                     | V$NF1_Q6                     | 0.956+, -28  | +                           | -0.254         | 4.83e-02 | ACATATAAAAGGCAGTGTGGCAGCAG |

CE: Composite regulatory elements.

Table 8 Allele and genotype frequencies of tumor necrosis factor-alpha-1030 polymorphism among Helicobacter pylori-infected and uninfected subjects and their contributions to Helicobacter pylori infection, n (%)

| Allele frequency | H. pylori (+ve), n = 61 | H. pylori (-ve), n = 61 | P value | OR (95%CI) |
|------------------|-------------------------|-------------------------|---------|------------|
| **Allele frequency** |                         |                         |         |            |
| TNF-A-1030-T     | 88 (72.13)              | 98 (80.33)              | 0.176   | 0.63 (0.349-1.151) |
| TNF-A-1030-C     | 34 (27.87)              | 24 (19.67)              |         |            |
| **Genotype frequency** |                      |                         |         |            |
| T/T              | 32 (52.46)              | 43 (70.49)              | -       | 1 (reference) |
| T/C              | 24 (39.34)              | 12 (19.67)              | 0.020a  | 2.69 (1.17-6.17) |
| C/C              | 5 (8.20)                | 6 (9.84)                | 0.862   | 1.12 (0.31-3.99) |

*P value < 0.05 statistically significant difference. OR: Odds ratio; 95% CI: 95% confidence interval; H. pylori: Helicobacter pylori; +ve: Positive; -ve: Negative; TNF-A: Tumor necrosis factor-alpha.
Table 9 Multivariate analysis of genotype frequencies of tumor necrosis factor-alpha-1030 polymorphism and sociodemographic characteristics in *Helicobacter pylori* infection

| Variables                  | P value | OR (95%CI) |
|----------------------------|---------|------------|
| Age                        | 0.946   | 0.99 (0.98-1.02) |
| Gender                     |         |            |
| Male                       |         | 1 (reference) |
| Female                     | 0.845   | 0.93 (0.42-2.02) |
| Residence                  |         |            |
| Urban                      |         | 1 (reference) |
| Rural                      | 0.341   | 1.46 (0.67-3.18) |
| Hospital                   |         |            |
| Private                    |         | 1 (reference) |
| Public                     | 0.045<sup>1</sup> | 0.45 (0.21-0.98) |
| Genotype frequency         |         |            |
| T/T                        |         | 1 (reference) |
| T/C                        | 0.016<sup>a</sup> | 2.87 (1.22-6.78) |
| C/C                        | 0.982   | 1.02 (0.27-3.88) |

<sup>1</sup>Weak evidence of significant different.

<sup>a</sup>P value < 0.05 statistically significant different. OR: Odds ratio; 95%CI: 95% confidence interval.

Table 10 Tumor necrosis factor-alpha-1030 polymorphism and gastric cancer association risk in gastric cancerous patients and subjects with benign gastric disorders, n (%)

| Allele frequency          | Benign disorders<sup>1</sup>, n = 108 | Cancer, n = 14 | OR (95%CI)     |
|---------------------------|--------------------------------------|----------------|----------------|
| TNF-A-1030-T              | 176 (90.28)                          | 21 (75.00)     | 1.150 (0.4627 - 2.8600) |
| TNF-A-1030-C              | 51 (9.72)                            | 7 (25.00)      |                |
| P value                   | 0.8116                               |                |                |
| Genotype frequency        |                                      |                |                |
| T/T                       | 68 (62.96)                           | 7 (50.00)      | 1.700 (0.5556 - 5.2020) |
| C carrier                 | 40 (37.04)                           | 7 (50.00)      |                |
| P value                   | 0.3900                               |                |                |

<sup>1</sup>Benign disorders, i.e. gastritis, duodenitis or ulcer.

OR: Odds ratio; 95%CI: 95% confidence interval; TNF-A: Tumor necrosis factor-alpha.

Intriguingly, we detected one mutation (T>A, -76) located at the identified TNF primate phylogenetic footprints<sup>64</sup> in a patient with reinfection with *H. pylori* multiple times. The presence of this SNP in the conserved region that is involved in regulating the TNF gene could affect the expression and create a condition of hypoacidity that favors the survival and colonization of *H. pylori*<sup>76,77</sup>. The -76 SNP within phylogenetic footprints across seven primate species, including Old World monkeys and New World monkeys<sup>64</sup>.

However, the TNF-A gene is located within the major histocompatibility complex region of chromosome 6, the highly polymorphic region of the human genome, which includes the human leukocyte antigen (HLA) genes<sup>65</sup>. In fact, all of the common ten detected SNPs (-1030, -862, -856, -375, -307, -243, -237, and +70 nt relative to the start site of transcription) fall outside of the identified TNF primate phylogenetic footprints<sup>66-70</sup>. This feature is consistent with the accrual of mutations under neutral evolution<sup>64,72</sup>. Furthermore, several of these SNPs (-1030, -862, -856, -307, -243, -237) have been shown to be in non-random association with extended HLA haplotypes<sup>21,26,71-73</sup>, which are HLA regions that exhibit fixity of the genomic interval between HLA-B and HLA-DR, that includes TNF-A and complement (complotype) regions<sup>21,74,75</sup>.

The -76 SNP
Figure 2 Nucleotide variations of the tumor necrosis factor-alpha 5'-region (-584 to +107) in the Sudanese Helicobacter pylori-infected patients. A: PCR amplification results of the tumor necrosis factor-alpha 5'-region (-584 to +107) examined on 1% agarose gel electrophoresis; B: Multiple sequence alignment of the nucleotide sequences of the tumor necrosis factor-alpha 5'-region performed in Clustal W and visualized in Jalview [55]. The levels of identity are visible as histograms at the bottom.

was first discovered with an overall minor allele frequency of 0.031 in the Indian population while studying polymorphisms of TNF-enhancer and the gene for FcγRIIa correlated with the severity of *Falciparum malaria* in the ethnically diverse Indian population [78]. It is present at the low-affinity transcription factor Ets-1 binding site in macrophages and the high affinity NFATp binding site in T and B cells [79]. Also, it is located at an in silico-predicted C/EBP-beta (also known as NF-IL6) binding site (in this study, see Table 5), which was experimentally proven in myelomonocytic cells [80-82].

Sequence analysis of the TNF-A 5'-region (-584 to +107) of the Sudanese patients revealed relative variability of TNF-A SNPs (Figure 2 and Table 3). This finding is in agreement with a study conducted by Baena et al [21] that found that the highest relative diversity of TNF-A SNPs was revealed among the three African studied populations (Malawian, Southern Nigerian, African-American and African-Caribbean). Also, it is consistent with the “out of Africa” hypothesis based upon human mitochondrial DNA analyses [83]. Despite this high level of variation, here in this study, the sequences involved in enhanceosome formation and gene regulation are highly conserved [64]. Remarkably, we detected only a single SNP (-76) in this region. By contrast, the novel +27 variant and other TNF-A SNPs described in this study occur within or immediately adjacent to promoter regions with high diversity in the primate lineage. Unexpectedly, the -243 variant, the first TNF-A SNP linked to an African-derived extended HLA haplotype [70], was not detected in our studied population.
Idris AB et al. In silico analysis of TNF-A promoter variants

Figure 3 MethPrimer software prediction of no CpG islands in the in silico predicted promoter region.

Figure 4 Overview from the ECR Browser shows mammalian conservation of the tumor necrosis factor 5'-upstream region compared to the human sequence in the region (-584 to +107) (hg19 chr6:31542751-31543444). Blue boxes represent the first tumor necrosis factor exon, while yellow indicates the tumor necrosis factor 5' untranslated region. Intragenic positions are highlighted in red or in green when corresponding to transposable elements and simple repeats. UTR: Untranslated region.

Other previously identified SNPs in this region (-584 to +107), which include -567, 375, -307, -237, -76 and +70 nt relative to the start site of transcription, were observed[66-68, 84].

The -1030 (T/C; rs1799964) SNP is frequently observed in African populations (17%-25% Malawian, 7% Nigerian, 29% African-American/Caribbean and 14% Gambia)[21, 22]. In this study, the -1030 SNP was detected in 47 Sudanese (38% of our studied population). Many studies have been published that examine the association between the -1030 SNP and multiple diseases[60,85,86] and diseases related to H. pylori such as gastroesophageal reflux disease, gastritis, peptic ulceration and gastric cancer[87-90]. In this study, we found a lack of association between -1030T and susceptibility to H.
Figure 5 Mammalian conservation of (AT; single nucleotide polymorphism databases dbSNP: rs41297589) at position -76 and the novel mutation (A>T at +27) in the 5’ untranslated region among various species. The nucleotides are enumerated at each line on the right side, and the in silico predicted TATA-, C/EBP-beta and transcription start site have marked inboxes. The chromatogram results of the polymorphisms are visualized using Finch TV software. TSS: Transcription start site.

Figure 6 Molecular detection of the tumor necrosis factor-alpha-1030 C/T polymorphism. A: PCR with confronting two-pair primer products analyzed on a 1% agarose gel stained with ethidium bromide. Three genotypes can be seen. Lanes 2, 3 and 5 showed 444 bp, 316 bp and 174 bp, indicating a heterozygous genotype. Lanes 4 and 6 showed 444 bp and 316 bp, which indicated a homozygous T genotype. Lane 1 showed 444 bp and 174 bp, which indicated a homozygous C genotype; B: Mammalian conservation of (TC; dbSNP: rs1799964) position -1030 among different species. MW: Molecular weight marker; C-ve: Negative control.

Idris AB et al. In silico analysis of TNF-A promoter variants

pylori and gastric cancer in the studied population \((P = 0.1756 \text{ and } P = 0.8116, \text{ respectively})\). This finding is in agreement with previous studies \([87, 88, 91]\). In contrast to a study conducted in Japan that observed a significant association between -1030T and gastric cancer \([91]\). This variation could be attributed mainly to differences in genetic backgrounds of the studied population, the method of genotyping and sample size \([76, 92]\).

The study’s limitations include the relatively small number of the enrolled subjects and depending on the in silico tools for studying the influence of the detected promoter variants on the TNF-A gene expression and \(H. \text{ pylori}\) infection (susceptibility and progression). Further large cohort studies are needed to validate the study findings. However, the computational analysis provides a basis for identifying
promoter regions, recognizing regulatory motifs and understanding gene expression patterns[25].

According to the literature, a considerable number of polymorphisms have been discovered in the TNF-A promoter. However, a crucial question that remains to be answered is whether these polymorphisms have any functional effect that may significantly impact disease incidence or severity. Identifying which of these many variants are functional is of great significance for discovering new preventive, diagnostic and therapeutic strategies against the incidence and/or progression of multifactorial diseases such as H. pylori infection. This study detected seven SNPs in the TNF-A 5'-region; only one of them (T>A, -76) was located at the in silico-predicted promoter region (-146 to +10). This SNP could lead to the modification of the transcriptional regulation of TNF-A in H. pylori infection. However, this conclusion cannot substitute for the experimental proofs (in vitro or in vivo), but it can provide a direction or insight for such experiments to validate the in silico predictions.

**CONCLUSION**

In the Sudanese H. pylori-infected patients, seven SNPs were observed in the TNF-A 5'-region; only one of them (T>A, -76) was located at the in silico-predicted promoter region (-146 to +10). In addition, it was predicted to alter TFBSs and CEs. Furthermore, a novel mutation (A>T, +27) was detected in the 5' untranslated region, and it could affect the post-transcriptional regulatory pathways. In addition, computational analysis was a valuable method for understanding gene expression patterns and providing guidance for further in vitro and in vivo experimental validation.

**ARTICLE HIGHLIGHTS**

**Research background**

*Helicobacter pylori* (H. pylori) infection represents a major public health challenge in Sudan. However, functional polymorphisms within the tumor necrosis factor-alpha (TNF-A) promoter are associated with the incidence and progression of H. pylori infection by increasing TNF-α production.

**Research motivation**

TNF lies within the major histocompatibility complex region of chromosome 6, which is a highly polymorphic region. Therefore, many polymorphisms have been detected in the TNF-A promoter and studying which variants could affect TNF-A gene expression is relevant. However, the influence of these single nucleotide polymorphisms (SNPs) on TNF-α production is not fully known and still a contradictory topic of debate due to the ethnic differences between populations. Furthermore, to our knowledge, there are no previous studies in Sudan that have addressed the association between TNF-A and H. pylori infection or H. pylori-associated diseases.

**Research objectives**

To functionally characterize the genetic variations in the TNF-A 5'-region (-584 to +107) of Sudanese patients infected with H. pylori and predict if these SNPs could alter the regulatory motifs using bioinformatics analyses. Also, to investigate the mammalian conservation of these SNPs using comparative profiling analysis in 11 species.

**Research methods**

An observational study was conducted in the major hospitals in Khartoum state. Genomic DNA was extracted from 122 gastric biopsies of patients who had been referred for endoscopy. Genotyping of the TNF-A-1030 polymorphism was performed using PCR with confronting two-pair primer to investigate its association with H. pylori infection in the Sudanese population. Sanger sequencing was applied to detect SNPs in the 5'-region (-584 to +107) of TNF-A in H. pylori-infected patients; in silico tools were used to predict whether these mutations would alter transcription factor binding sites or composite regulatory elements in this region. In addition, the ECR browser and multiple-sequence local alignment and visualization search engine were
used to study the conservation of the detected SNPs among 11 mammalian species.

Research results
A total of seven SNPs were observed in the TNF-A 5′-region of Sudanese patients infected with *H. pylori*. Among them, the SNP (T>A, -76) was located at the in silico-predicted promoter region (-146 to +10), and it was predicted to alter transcription factor binding sites and composite regulatory elements, while the novel mutation (A>T, +27) was detected in the 5′ untranslated region. It could affect the post-transcriptional regulatory pathways. Mammalian conservation was detected for the (-146 to +10) region in chimpanzee (99.4%), rhesus monkey (95.6%), cow (91.8%), domesticated dog (89.3%), mouse (84.3%), rat (82.4%) and opossum (78.0%). Furthermore, genotyping of TNF-A-1030 revealed a lack of significant association between -1030T and susceptibility to *H. pylori* and gastric cancer in the studied population (*P* = 0.1756 and *P* = 0.8116, respectively).

Research conclusions
Despite the high level of genetic variation in the TNF-A 5′-region (-584 to +107) of the Sudanese patients, the sequences involved in enhanceosome formation and gene regulation are highly conserved. Remarkably, only a single SNP (-76) was detected in this region. In addition, computational analysis was a valuable method for studying gene expression patterns and insights for further *in vitro* and *in vivo* experimental proofs.

Research perspectives
Further large cohort studies are needed to assess the association between (T>A, -76) mutation and *H. pylori* infection (susceptibility and progression). Also, further studies are encouraged to investigate the novel mutation (A>T, +27) in terms of the frequency of the minor allele (T) in the Sudanese population and its functional significance using computational and experimental approaches. Identifying which of these detected variants are functional is of great relevance for discovering new preventive, diagnostic and therapeutic strategies against the incidence and/or progression of multifactorial diseases such as *H. pylori* infection.

ACKNOWLEDGEMENTS
We gratefully acknowledge the patients and the staff of the endoscopic unit in Soba teaching hospital, Ibin Sina specialized hospital, Police hospital, Modern Medical Centre, and Al Faisal Specialized Hospital. Also, we would like to show our gratitude to the Department of Molecular Research Laboratory at the Faculty of Medical Laboratory Sciences, the University of Khartoum, for their cooperation.

REFERENCES
1. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, Ng SC. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* 2017; 153: 420-429
2. Salih BA. Helicobacter pylori infection in developing countries: the burden for how long? *Saudi J Gastroenterol* 2009; 15: 201-207
3. Sjomina O, Pavlova J, Niv Y, Leja M. Epidemiology of Helicobacter pylori infection. *Helicobacter* 2018; 23 Suppl 1: e12514
4. Idris AB, Hassan HG, Ali MAS, Eltaher SM, Idris LB, Altayb HN, Abass AM, Ibrahim MMA, Ibrahim EM, Hassan MA. Molecular Phylogenetic Analysis of 16S rRNA Sequences Identified Two Lineages of Helicobacter pylori Strains Detected from Different Regions in Sudan Suggestive of Differential Evolution. *Int J Microbiol* 2020; 2020: 8825718
5. Go MF. Review article: natural history and epidemiology of Helicobacter pylori infection. *Aliment Pharmacol Ther* 2002; 16 Suppl 1: 1-15
6. Brown LM. Helicobacter pylori: epidemiology and routes of transmission. *Epidemiol Rev* 2000; 22: 283-297
7. Cadamuro AC, Rossi AF, Maniezzo NM, Silva AE. Helicobacter pylori infection: host immune response, implications on gene expression and microRNAs. *World J Gastroenterol* 2014; 20: 1424-1437
8. Sharma SA, Miller GG, Peek RA Jr, Pérez-Pérez G, Blaser MJ. T-cell, antibody, and cytokine
responses to homologs of the 60-kilodalton heat shock protein in Helicobacter pylori infection. Clin Diag Lab Immunol 1997; 4: 440-446
9 Chang WL, Yeh YC, Sheu BS. The impacts of H. pylori virulence factors on the development of gastroduodenal diseases. J Biomed Sci 2018; 25: 68
10 de Sabeti P, Piazuelo MB, Shaffer CL, Schneider BG, Asim M, Chaturvedi R, Bravo LE, Sicinschi LA, Delgado AG, Mera RM, Israel DA, Romero-Gallo J, Peek RM Jr, Covey TL, Correa P, Wilson KT. Phylogeographic origin of Helicobacter pylori is a determinant of gastric cancer risk. Gut 2011; 59: 1189-1195
11 El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA, Herrera J, Lissowska J, Yuan CC, Rothman N, Lanyon G, Martin M, Fraumeni JF, Jr, Rabkin CS. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature 2000; 404: 398-402
12 Mayot M, Velin D. Immune responses to Helicobacter pylori infection. World J Gastroenterol 2014; 20: 5583-5593
13 Falvo JV, Tsytyskova AV, Goldfeld AE. Transcriptional control of the TNF gene. Curr Dir Autoimmun 2010; 11: 27-60
14 Wang H, Czura CJ, Tracey KJ. Chapter 35 - tumor necrosis factor. In: Thomson AW, Lotze MT. The cytokine handbook. 4th ed. London: Academic Press, 2003: 837-860
15 Atzeni F, Sarzi-Puttini P. Tumor necrosis factor. In: Maloy S, Hughes K. Brenner’s encyclopedia of genetics. 2nd ed. San Diego: Academic Press, 2013: 229-231
16 Tsytyskova AV, Goldfeld AE. Inducer-specific enhanceosome formation controls tumor necrosis factor alpha gene expression in T lymphocytes. Mol Cell Biol 2002; 22: 2620-2631
17 Tumor necrosis factor alfa. In: Aronson JK Meyler’s side effects of drugs. 16th ed. Oxford: Elsevier, 2016: 230-232
18 Dahlb B, Mistry KN. In silico analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. Meta Gene 2014; 2: 586-595
19 Galbraith GM, Hendley TM, Sanders JJ, Palesch Y, Pandey JP. Polymorphic cytokine genotypes as markers of disease severity in adult periodontitis. J Clin Periodontol 1999; 26: 705-709
20 Kroeger KM, Steer JH, Joyce DA, Abraham LJ. Effects of stimulus and cell type on the expression of the -308 tumour necrosis factor promoter polymorphism. Cytokine 2000; 12: 110-119
21 Jauca A, Leung JY, Sullivan AD, Landires I, Vasquez-Luna N, Quiliones-Berroc J, Fraser PA, Uko GP, Delgado JC, Clavijo OP, Thim S, Molyneux M, Dunai V, Nyirenda T, Yunis EJ, Goldfeld AE. TNF-alpha promoter single nucleotide polymorphisms are markers of human ancestry. Genes Immun 2002; 3: 482-487
22 Richardson A, Sisay-Joof F, Ackerman H, Usen S, Katundu P, Taylor T, Molyneux M, Pinder M, Kwiatkowski D. Nucleotide diversity of the TNF gene region in an African village. Genes Immun 2001; 2: 343-348
23 Majumder P, Thom K, Bhattacharya M, Nair V, Ghosh S, Dey SK. Association of tumor necrosis factor-α (TNF-α) gene promoter polymorphisms with aggressive and chronic periodontitis in the early and chronic Indian population. Biosci Rep 2018; 38
24 Zhang MQ. Computational analyses of eukaryotic promoters. BMC Bioinformatics 2007; 8 Suppl 6: S3
25 Solovyev VV, Shahmuradov IA, Salamov AA. Identification of promoter regions and regulatory sites. Methods Mol Biol 2010; 674: 57-83
26 Ugliaroli AM, Turbay D, Pesavento PA, Delgado JC, McKenzie FE, Gribben JG, Hartl D, Yunis EJ, Goldfeld AE. Identification of three new single nucleotide polymorphisms in the human tumor necrosis factor-alpha gene promoter. Tissue Antigens 1998; 52: 359-367
27 Hollegraard MV, Bidwell JL. Cytokine gene polymorphism in human disease: on-line databases, Supplement 3. Genes Immun 2006; 7: 269-276
28 Harbage B, Dean AG. Distribution of Epi Info software: an evaluation using the Internet. Am J Prev Med 1999; 16: 314-317
29 Su Y, Yoon SS. Epi info - present and future. AMLA Ann Symp Proc 2003: 1023
30 Idris AB, Idris EB, Atsamanan AE, Mohamed AE, Osman Arbab BM, Ibrahim EM, Hassan MA. First insights into the molecular basis association between promoter polymorphisms of the IL1B gene and Helicobacter pylori infection in the Sudanese population: computational approach. BMC Microbiol 2021; 21: 16
31 Carbone M, Maugeri TL, Gugliandolo C, La Camera E, Biondo C, Fera MT. Occurrence of Helicobacter pylori in the coastal environment of southern Italy (Strait of Messina). J Appl Microbiol 2005; 98: 768-774
32 Zhang G, Zhou B, Li S, Yue J, Yang H, Wen Y, Zhan S, Wang W, Liao M, Zhang M, Zeng G, Feng CG, Sassetti CM, Chen X. Allele-specific induction of IL-1β expression by CEBPβ and PU.1 contributes to increased tuberculosis susceptibility. PLoS Pathog 2014; 10: e1004426
33 Treves DS. Review of Three DNA Analysis Applications for Use in the Microbiology or Genetics Classroom. J Microbiol Biol Educ 2010; 11: 186-187
34 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-410
35 Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics 2002; Chapter 2: Unit 2.3
melanogaster genome. *Comput Chem* 2001; 26: 51-56

37. *Knudsen S*. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* 1999; 15: 356-361

38. *Solovyev V*, *Kosarev P*, *Seledsov I*, *Vorobyev D*. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 2006; 7 Suppl 1: S10.1-S1012

39. *Hernandez-Garcia CM*, *Finer JJ*. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci* 2014; 217-218: 109-119

40. *ENCOD Resource Consortium*. A user’s guide to the encyclopedia of DNA elements (ENCOD). *PLoS Biol* 2011; 9: e1001046

41. *Qu H*, *Fang X*. A brief review on the Human Encyclopedia of DNA Elements (ENCOD) project. *Genomics Proteomics Bioinformatics* 2013; 11: 135-141

42. *ENCOD Resource Consortium*. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57-74

43. *Li LC*, *Dahiya R*. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002; 18: 1427-1431

44. *Xin S*, *Wang X*, *Dai G*, *Zhang J*, *An T*, *Zou W*, *Zhang G*, *Xie K*, *Wang J*. Bioinformatics Analysis of SNPs in IL-6 Gene Promoter of Jinghai Yellow Chickens. *Genes (Basel)* 2018; 9: 446

45. *Grabe N*. AliBab2: context specific identification of transcription factor binding sites. *In Silico Biol* 2002; 2: S1-15

46. *Farré D*, *Roset R*, *Huerta M*, *Adsuara JE*, *Roselló L*, *Albá MM*, *Messeguer X*. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res* 2003; 31: 3651-3653

47. *Lin S*, *Zhang W*, *Shi Z*, *Tan L*, *Zhu Y*, *Li H*, *Peng X*. β-Catenin/LEF-1 transcription complex is responsible for the transcriptional activation of LINC01278. *Cancer Cell Int* 2021; 21: 380

48. *Tsunoda T*, *Takagi T*. Estimating transcription factor bindability on DNA. *Bioinformatics* 1999; 15: 622-630

49. *Ghosh D*. Object-oriented transcription factors database (ooTFD). *Nucleic Acids Res* 2000; 28: 308-310

50. *Kel AE*, *Gössling E*, *Reuter I*, *Cheremushkin E*, *Kel-Margoulis OV*, *Wingender E*. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 2003; 31: 3576-3579

51. *Deyneko IV*, *Kel AE*, *Kel-Margoulis OV*, *Deineko EV*, *Wingender E*, *Weiss S*. MatrixCatch—a novel tool for the recognition of composite regulatory elements in promoters. *BMC Bioinformatics* 2013; 14: 241

52. *Zhang Q*, *Li T*, *Zhang Y*, *Lin J*, *Chen X*, *Gai Y*, *Li Q*. Comparative sequence analysis and functional validation identified a retina-specific enhancer around zic5 and zic2a. *Gene Expr Patterns* 2021; 39: 119162

53. *Ovcharenko I*, *Loots GG*, *Giardine BM*, *Hou M*, *Ma J*, *Hardison RC*, *Stubbs L*, *Miller W*, *Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res* 2005; 15: 184-194

54. *Atsuta Y*, *Kawase H*, *Hamajima N*, *Nishio K*, *Niwa Y*, *Tanaka D*, *Yamamoto K*, *Tanaka D*, *Yamamoto K*, *Tanaka D*, *Yamamoto K*. Use of duplex PCR-CTPP methods for CYP2E1RsaI/IL-2 T-330G and IL-1B C-31T/TNF-A T-1031C polymorphisms are associated with the inability to control HIV-1 viremia. *Proc Natl Acad Sci U S A* 2000; 97: 6614-6618
65 Erickson RP. Natural history of the major histocompatibility complex. *Am J Hum Genet* 1987; 40: 468-469
66 Wilson AG, di Giovine FS, Blakemore AI, Duff GW. Single base polymorphism in the human tumour necrosis factor alpha (TNF alpha) gene detectable by Ncol restriction of PCR product. *Hum Mol Genet* 1992; 1: 353
67 D’Alfonso S, Richiardi PM. A polymorphic variation in a putative regulation box of the TNFA promoter region. *Immunogenetics* 1994; 39: 150-154
68 Brinkman BM, Kaiteil EL, Huizinga TW, Giphart MJ, Breedveld FC, Verweij CL. Detection of a C-insertion polymorphism within the human tumor necrosis factor alpha (TNF) gene. *Hum Genet* 1995; 96: 493
69 Hamann A, Mantzoros C, Vidal-Puig A, Flier JS. Genetic variability in the TNF-alpha promoter is not associated with type II diabetes mellitus (NIDDM). *Biochem Biophys Res Commun* 1995; 211: 833-839
70 Zimmerman PA, Gudierian RH, Nutman TB. A new TNF-a promoter allele identified in South American Blacks. *Immunogenetics* 1996; 44: 485-486
71 Higuchi T, Seki N, Kamizono S, Yamada A, Kinura A, Kato H, Itoh K. Polymorphism of the 5'-flanking region of the human tumor necrosis factor (TNF)-alpha gene in Japanese. *Tissue Antigens* 1998; 51: 605-612
72 Goldfeld AE, Leung JY, Sawyer SA, Hartl DL. Post-genomics and the neutral theory: variation and conservation in the tumor necrosis factor-alpha promoter. *Gene* 2000; 261: 19-25
73 Wilson AG, de Vries N, Pociot F, di Giovine FS, van der Putte LB, Duff GW. An allelic polymorphism within the human tumor necrosis factor alpha gene promoter region is strongly associated with HLA A1, B8, and DR3 alleles. *J Exp Med* 1993; 177: 557-560
74 Alper CA, Awdeh ZL, Raum DD, Yunis EJ. Extended major histocompatibility complex haplotypes in man: role of alleles analogous to murine t mutants. *Clin Immunol Immunopathol* 1982; 24: 276-285
75 Yunis EJ, Larsen CE, Fernandez-Villa M, Awdeh ZL, Romero T, Hansen JA, Alper CA. Inheritable variable sizes of DNA stretches in the human MHC conserved extended haplotypes and their fragments or blocks. *Tissue Antigens* 2003; 62: 1-20
76 Ma J, Wu D, Hu X, Li J, Cao M, Dong W. Associations between cytokine gene polymorphisms and susceptibility to *Helicobacter pylori* infection and *Helicobacter pylori* related gastric cancer, peptic ulcer disease: A meta-analysis. *PLoS One* 2017; 12: e0176463
77 El-Omar EM. The importance of interleukin 1beta in *Helicobacter pylori* associated disease. *Gut* 2001; 48: 743-747
78 Sinha S, Mishra SK, Sharma S, Patibanda PK, Mallick PK, Sharma SK, Mohanty S, Pati SS, Ramteke BK, Bhatt R, Joshi H, Dash AP, Ahuja RC, Awasthi S, Indian Genome Variation Consortium, Venkatesh V, Habib S. Polymorphisms of TNF-enhancer and gene for Fc gamma RIa correlate with the severity of falciparum malaria in the ethnically diverse Indian population. *Malar J* 2008; 7: 13
79 Tsai EY, Falvo JV, Tsytyskova AV, Barczak AK, Reimold AM, Glimcher LH, Fenton MJ, Gordon DC, Dunn IF, Goldfeld AE. A lipopolysaccharide-specific enhancer complex involving Ets, Elk-1, Sp1, and CREB binding protein and p300 is recruited to the tumor necrosis factor alpha promoter in vivo. *Mol Cell Biol* 2000; 20: 6084-6094
80 Pope RM, Leutz A, Ness SA. C/EBP beta regulation of the tumor necrosis factor alpha gene. *J Clin Invest* 1994; 94: 1449-1455
81 Liu Y, Nommacher MR, Wigdahl B. CCAAT/enhancer-binding proteins and the pathogenesis of retrovirus infection. *Future Microbiol* 2009; 4: 299-321
82 Zagariya A, Mungre S, Lovis R, Birrer M, Ness S, Thimmapaya B, Pope R. Tumor necrosis factor alpha gene regulation: enhancement of C/EBP beta-induced activation by c-Hun. *Mol Cell Biol* 1998; 18: 2815-2824
83 Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. African populations and the evolution of the mitochondrial DNA. *Science* 1991; 253: 1503-1507
84 Qidwai T, Khan F. Tumor necrosis factor gene polymorphism and disease prevalence. *Scand J Immunol* 2011; 74: 522-547
85 Cay HF, Sezer I, Dogan S, Felek R, Aslan M. Polymorphism in the TNF-alpha gene promoter at position -1031 is associated with increased circulating levels of TNF-alpha, myeloperoxidase and nitrotyrosine in primary Sjogren’s syndrome. *Clin Exp Rheumatol* 2012; 30: 843-849
86 Darvishi E, Aziziaram Z, Yari K, Bagheri Dehbaghi M, Kahrizi D, Karim H, Vaziri S, Zargooshi J, Ghadiri K, Muhammedi S, Kazemi E, Moradi MT, Shokrinia M, Mohammadi N. Lack of association between the TNF-α-1031 genotypes and generalized aggressive periodontitis disease. *Cell Mol Biol (Noisy-le-grand)* 2016; 62: 63-73
87 Lu CC, Sheu BS, Chen TW, Yang HB, Hung KH, Kao AW, Chuang CH, Wu JJ. Host TNF-alpha-1031 and -863 promoter single nucleotide polymorphisms determine the risk of benign ulceration after *H. pylori* infection. *Am J Gastroenterol* 2005; 100: 1274-1282
88 Zheng W, Zhang S, Min L, Wang Y, Xie J, Hou Y, Tian X, Cheng J, Liu K, Xu D, Yu X, Liu Z, Lv Y, Liang N, Zhang J, Liu F, Tian Y. The relationship between tumor necrosis factor-α polymorphisms and gastric cancer risk: An updated meta-analysis. *Biomed Rep* 2017; 7: 133-142
89 Hou L, El-Omar EM, Chen J, Grillo P, Babkin CS, Baccarelli A, Yeager M, Chanock SJ, Zatonski
Idris AB et al. In silico analysis of TNF-A promoter variants

W, Sobin LH, Lissowska J, Fraumeni JF Jr, Chow WH. Polymorphisms in Th1-type cell-mediated response genes and risk of gastric cancer. *Carcinogenesis* 2007; 28: 118-123

90 Sugimoto M, Furuta T, Shirai N, Nakamura A, Xiao F, Kajimura M, Sugimura H, Hishida A. Different effects of polymorphisms of tumor necrosis factor-alpha and interleukin-1 beta on development of peptic ulcer and gastric cancer. *J Gastroenterol Hepatol* 2007; 22: 51-59

91 Yang JJ, Ko KP, Cho LY, Shin A, Gwack J, Chang SH, Shin HR, Yoo KY, Kang D, Park SK. The role of TNF genetic variants and the interaction with cigarette smoking for gastric cancer risk: a nested case-control study. *BMC Cancer* 2009; 9: 238

92 Xue H, Lin B, Ni P, Xu H, Huang G. Interleukin-1B and interleukin-1 RN polymorphisms and gastric carcinoma risk: a meta-analysis. *J Gastroenterol Hepatol* 2010; 25: 1604-1617
