

Analysis of the genetic diversity in TLR4 3'-untranslated region among Asian populations and the biological effects of altofrequent SNPs

Guodong Chen

Yunnan University

Luhan Yang

Yunnan University

Yun Zhang

Yunnan University

Yang Jiao

Kunming Medical University

Hao Guo

First Affiliated Hospital of Kunming Medical University

Bingrong Zheng (✉ zhengbr@ynu.edu.cn)

Yunnan University <https://orcid.org/0000-0001-9620-286X>

Research article

Keywords: TLR4, SNPs, genetic diversity, site-directed mutagenesis, biological effects, inflammatory diseases

Posted Date: September 12th, 2019

DOI: <https://doi.org/10.21203/rs.2.14417/v1>

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

[Read Full License](#)

Abstract

Background: Numerous case-control studies have demonstrated that single nucleotide polymorphism (SNP) loci in the TLR4 3'-untranslated region (UTR) are associated with multiple inflammatory disorders in various populations, particularly in Asians. However, the distribution frequency of polymorphisms in different Asian races and its effects on TLR4 expression remain unclear. Here, we extracted variants from the 1000 Genomes project database and analyzed the SNPs and haplotypes genetic diversity of TLR4 3'-UTR in 10 Asian populations, and then assessed the biological effects of the altofrequent polymorphic loci on TLR4 expression by luciferase assay. Results: Total 6 SNPs were identified as true polymorphic loci with minor allele frequencies (MAFs) $\geq 1\%$ in Asian populations, including rs41426344, rs7869402, rs11536889, rs7873784, rs11536891, and rs11536896. Considering 4 Tag SNPs, we inferred 5 haplotypes that present a Asiatic frequency higher than 1%. It is noteworthy that rs41426344 is unique to East Asian populations, and the H-5 haplotype frequency was reduced when analyzing pooled data from East and South Asian populations. The MAFs of rs7869402 and rs11536889, and H-2 and H-4 frequencies significantly differed between the populations ($P < 0.001$). We constructed a pGL3-3494-3UTR luciferase plasmid to simulate TLR4 gene structure in vivo, and used PCR-mediated site-directed mutagenesis to construct a series of mutant luciferase constructs corresponding to the six SNPs. In addition, we identified that TLR4 mRNA was selectively expressed in SiHa and THP-1 cell lines, but not in C33A, Hela, and 293T cell lines. The luciferase activity of constructs containing rs7869402 T allele and rs11536889 C allele increased significantly upon LPS or IL-6 stimulation in THP-1 and SiHa cells. Conclusions: The distributions of SNPs and haplotypes in TLR4 3'-UTR were significantly different among Asian populations. The biological effects of rs7869402 and rs11536889 on TLR4 are significant clues that revealing its critical role in harmful TLR4-mediated responses. These results provided a reminder for future investigation on TLR4 related inflammatory diseases susceptibility.

Background

Rapid recognition of Gram-negative bacterial lipopolysaccharides (LPS), and responding to infection are the essential functions of innate immune cells [1, 2]. Toll-like Receptor 4 (TLR4) is a pattern recognition receptor, which are the portal proteins of inflammation signal transduction [3]. Overexpression of TLR4 in the presence of LPS ligands results in the release of a variety of inflammatory cytokines, and effectively activates innate immune responses [4]. Inflammation-induced TLR4 usually plays a specific role in triggering various diseases.

The Human *TLR4* gene (Gene ID: 7099) is located at chromosome 9q33.1, has four exons and three introns, and encompasses a region of 20,332 nucleotides. Polymorphism loci located within the 3'-UTR may influence mRNA translation efficiency via altering the microRNA binding ability, affecting mRNA stability, and modulating disease susceptibility [5]. Analysis of the dbSNP database revealed 669 SNPs in the *TLR4* 3'-UTR (<http://www.ncbi.nlm.nih.gov/SNP/>). Only six of these SNPs (rs41426344, rs7869402, rs11536889, rs7873784, rs11536891, and rs11536896) had been analyzed in specific populations using case-control studies [6-12]. Of these SNPs, rs11536889 is the most extensively studied. The rs11536889

C/C genotype is significantly associated with the risk of periodontitis in Japanese subjects, and the “C” allele frequency is significantly higher than the “G” allele frequency in patients with severe periodontitis [13]. A large study involving ~1400 patients and ~800 controls revealed that the “C” allele obviously increased the risk of prostate cancer in a Swiss population [12]. Sun et al. reported that the “C” allele was significantly associated with ventilator-associated pneumonia in the Chinese population [14]. Another study reported that the rs11536889 “C” allele increases the risk of coronary artery disease in ethnic Zhuang and Han Chinese populations [15, 16]. Chang et al. found that the G/C and C/C genotypes were associated with head and neck cancer in Taiwan's population [17]. This SNP was also associated with increased organ failure risk in patients with sepsis in the Caucasian population [18]. Additionally, a significant association between this SNP and Psoriasis vulgaris was identified in the Southern Chinese population [19].

The 1000 Genomes (1KG) Project is an integrated survey of genetic variation from ~2,500 individual genomes from diverse populations worldwide [20]. This genetic catalogue of human genome information can facilitate improvement of the reference sequence, understanding of population traits and evolutionary histories, and exploration of the genetic basis of many complex diseases [21]. A significant association between genetic variation and population distribution has been previously demonstrated, and population-specific polymorphic loci can be used to further understand the mechanisms of population related disease [22, 23]. In Asia, there are abundant genetic diversities, especially in India, resulting from a long and complex history of mass migration [22, 23].

TLR4 is expressed on myeloid cells, epithelial cells, B cells, muscle cells, and some cancer cells [1, 24-26]. THP-1 is monocyte cell line from the blood of human acute leukemia that expresses endogenous TLR4, and has been used for multiple classical TLR4 studies [1, 8]. Inflammation within the tumor microenvironment is a critical factor in tumor development [27]. Immune system evasion is the main reason of papillomavirus (HPV) infection and cervical cancer development. TLR4 expression plays a role in the oncogenic potential of HPV positive cells [28], and the rs7873784 G allele in the *TLR4* 3'-UTR is significantly associated with increased cervical cancer risk [29]. Therefore, THP-1, HEK 293T, human cervical cancer C33A (HPV-negative), HeLa (HPV-18), and SiHa (HPV-16) cell lines were chosen as alternative hosts.

In this study, we extracted genetic variation data for the full-length *TLR4* 3'-UTR from 993 Asians in 10 Asian populations from the database of 1KG Project (Phase 3 release). We then investigated *TLR4* 3'-UTR variation and haplotype diversity in these populations individually or in combination. We also assessed the effects of SNPs (with MAFs of at least 1%) on *TLR4* expression.

Methods

SNP extraction and haplotype construction

Data from 74 SNPs across the entire *TLR4* 3'-UTR, between positions 120,476,927 and 120,479,769 (GRCh37.p13) of chromosome 9, were downloaded from the 1KG Phase 3 Pipeline using the Ferret

version 2.1.1 JAVA tool [38]. The data used in this Analysis from 10 human populations, included 993 Asian individuals as described in 1KG Browser (<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>) (Additional file 1: Table S1). The two sequences defined by NCBI (DNA: NG_011475.1, and mRNA: NM_003266.3) were used as references in the study. Haploview version 4.2 software was used to filter the SNPs to ensure that the MAF was present at least 1% of the time, to select tag SNPs, calculate Hardy Weinberg equilibrium p value (HWp), and analyze haplotype patterns and linkage disequilibrium [6] among the SNPs [39]. Six SNPs were identified in the *TLR4* 3'-UTR = using this screening approach (Table 1). The Haploview software tagger option was used to predict values r^2 and D' for the remaining SNPs in the dataset. Block-based and block-free approaches were used to increase the accuracy and representativeness of tag SNPs [40]. Haplotypes were evaluated with tag SNPs in 10 populations using Arlequin version 3.5 software [41].

mRNA quantification

Total RNA was isolated from cells using TriPure Isolation Reagent (Roche). Reverse transcription was performed using the PrimeScript™ RT reagent Kit (TAKARA, Japan). Taq Plus DNA Polymerase (TianGen, China) was used for RT-PCR, and TB Green® Premix Ex Taq™ II (TAKARA) was used to perform the qRT-PCR in a CFX96 Real-Time thermocycler platform (Bio-Rad). Relative quantifications were determined by the comparative $2^{-\Delta\Delta Ct}$ method [42]. All primers used are listed in Additional file 2: Table S2.

Western blot assay

Total protein was extracted separately for five cell cultures using RIPA buffer (Beyotime, China) with added phenylmethylsulfonyl fluoride and phosphatase inhibitor (Beyotime). Equal amounts of proteins were separated by SDS-PAGE on 6% or 12% polyacrylamide separating gels, and then transferred onto polyvinylidene-difluoride-fluoride membranes (PVDF; Roche). The membranes were blocked with 5% Non-Fat powdered milk in TBS-T 1 h at room temperature (15-25°C), then incubated with the primary antibody at 4°C for 12 hours. The antibodies used were: mouse anti-TLR4 (1:500; Santa Cruz, USA), rabbit anti-MyD88 (1:1000, Beyotime, China), rabbit anti- NF-κB p65 (1:1000, Abcam, UK), rabbit anti- NF-κB p-p65 (phospho S536; 1:10000, Abcam, UK), and mouse anti-beta-Tubulin (1:5000, Invitrogen, USA). After washing, the PVDF membranes were incubated with goat anti-mouse IgG or goat anti-rabbit IgG conjugated to horseradish peroxidase (1:10000, Abcam, UK) secondary antibodies at room temperature for 1 hour. Subsequently, an enhanced chemiluminescence kit (CwBio, China,) was used to capture protein signals. A chemiluminescent Western Blot imaging system (Amersham Imager 600, GE, USA) was used to visualize the protein signals with image.

Plasmid construction

The human *TLR4* promoter (-3494 to +235) was amplified from normal human genomic DNA using Cobuddy Super Fidelity DNA Polymerase (CwBio, China) with primer pair TPprimer1/TPprimer2 (Additional file 2: Table S2). The thermal cycling conditions were 2 min at 98°C, 35 cycles of 10 s at 98°C, 30 s at 67°C, 2 min at 72°C, and 5 min at 72°C. The full length *TLR4* 3'-UTR sequence was amplified by

Cobuddy Super Fidelity DNA Polymerase with the primer set TPprimer3/TPprimer4 (Additional file 2: Table S2). The thermal cycling conditions were 2 min at 98°C, 35 cycles of 10 s at 98°C, 30 s at 62°C, 1.5 min at 72°C, and 5 min at 72°C. Amplicons were electrophoresed and the products were gel purified using a Gel Extraction Kit (Omega Bio-Tek, USA). The pGL3-Basic luciferase vector (Promega, USA) was digested with *Xba* I and *Nco* I (TAKARA, Japan), and the linearized vector fragments (1656 bp and 3162 bp) were purified from the electrophoresis gel. Gibson assembly mix was prepared following the one-step isothermal DNA assembly protocol [43]. Approximately 10 ng of the 1656 bp linearized vector fragment, ~15 ng *TLR4* 3'UTR fragment, ~20 ng 3162 bp linearized vector fragment, and the *TLR4* promoter (-3493 to +234) were mixed with 10 µL of Gibson assembly mix, and the final reaction volume adjusted to 20 µL. The mixture was incubated at 50°C for 1 hour, and the entire volume was transformed into 200 µL of DH5α competent cells. Plasmid DNA extracted using an Endo-free Plasmid Mini Kit (Omega Bio-Tek, USA) and validated by sequencing.

Site-directed mutagenesis

To simulate the physiological state of SNPs, PCR-mediated site-directed mutagenesis for circular macromolecules was performed using pGL3-3493-3UTR. For rs41426344, a mutagenic reverse primer (mut344_G/C_R) was designed to introduce the necessary base (Additional file 2: Table S2, see the underlined base) substitutions to convert guanine (G) to cytosine (C). The forward primer (mut344_G/C_F) was designed in the opposite direction from mut344_G/C_R. The annealing positions of the two primers were back-to-back closed to each other. The primers underwent 5' phosphorylation using T4 Polynucleotide Kinase (NEB, USA) to allow for subsequent ligation. PCR was performed with the pGL3-3493-3UTR template to introduce mutation using the Cobuddy Super Fidelity DNA Polymerase system (CwBio, China). The amplification conditions were 2 min at 98°C, followed by 35 cycles of 10 sec at 98°C, 30 sec at 55°C, and 5 min at 72°C, and a 5 min final extension at 72°C. The PCR products were recovered, purified using the Omega Gel Extraction Kit, and self-ligated following the T4 DNA Ligase Protocol (NEB). The constructs were sequenced by the Beijing Genomics Institute for confirmation. After digestion with BglII, the constructs were religated to generate pGL3-684, pGL3-684-3UTR, and seven mutated constructs (-684 to +235).

Cell culture and transfection

The THP-1 human leukemia monocytic and HEK embryonic kidney 293T cell lines were purchased from the Conservation Genetics CAS Kunming Cell Bank (Kunming, China). THP-1 was maintained in L-glutamine-containing RPMI 1640 (GIBCO, Carlsbad, CA, USA) supplemented with 15% fetal bovine serum (GIBCO), 1% non-essential amino acids, 1% sodium pyruvate, and penicillin/streptomycin as previously described [44]. The SiHa human cervical carcinoma cell line, C33A, and HeLa cells were obtained from the American Type Culture Collection (ATCC; Manassas, VA, USA) and cultured at 37°C in Dulbecco's modified Eagle's medium (DMEM; GIBCO) containing 10% fetal bovine serum in an incubator with 5% CO₂. THP-1 and SiHa cells were transiently transfected with luciferase constructs. Exponentially growing THP-1 and SiHa cells were plated onto 24-well plates at a density of 5×10⁵ and 0.5×10⁵ cells/well,

respectively. Twenty-four hours after plating, THP-1 cells were transfected using DNAfectin™ Plus Transfection Reagent (ABM, Canada) and SiHa cells were transfected using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions.

Luciferase reporter assay

THP-1 cells were transiently co-transfected with 350 ng of constructs and 35 ng renilla luciferase plasmid pRL-CMV (Promega, USA), as an internal control for normalizing luciferase activity. After 36 h, the cells were treated with serum starvation for 12 h, and 1 ng/ml recombinant human Interleukin-6 (IL-6; Sangon Biotech, China) or 1 µg/ml LPS (Sigma-Aldrich, USA) was added 6 h before harvesting. SiHa cells were co-transfected with 350 ng of constructs and 17.5 ng of pRL-CMV, and stimulated with 50 µg/ml LPS and 50 ng/ml IL-6. Firefly and renilla luciferase activities were measured following the Dual-Luciferase® Reporter Assay System (Promega, USA) protocol using a Molecular Devices SpectraMax iD5 multi-mode microplate reader. Experiments were performed at least in triplicate.

Statistical calculations

SPSS version 22 software (IBM Corp, USA) was used for statistical analyses. The unpaired Student's t test was used to compare relative luciferase activities between two groups. Frequency differences among haplotypes were assessed using the Chi-square statistical method. Statistical analyses were expressed as mean ± SD of three or more independent experiments. $P < 0.05$ was considered statistically significant.

Results

Variants and haplotypes in the *TLR4* 3'-UTR

The Asian people in the 1KG were classified as being either East or South Asian, and consisted of 994 singletons from 816 families were included (Additional file 1: Table S1). In the *TLR4* 3'-UTR of, 74 SNPs were reported in the 1KG (Phase 3), six of which had $MAF \geq 1\%$ (Table 1). Four polymorphic loci were selected as tag SNPs (bold SNPs), and one LD block was identified in the 3'-UTR (Fig. 1).

The frequency of each SNP in the different Asian populations analyzed is listed in Table 2. The reference MAFs of rs41426344, rs7869402, and rs11536889 significantly differed between the ten populations ($P < 0.001$). rs41426344 is a unique SNP in East Asian populations and has a lower MAF in the Japanese in Tokyo (JPT) population than in the Chinese Dai in Xishuangbanna (CDX), Han Chinese in Beijing (CHB), Han Chinese South (CHS), and Kinh in Ho Chi Minh City (KHV) populations. rs11536889 frequency was significantly higher in the CDX population than in the other nine populations. Moreover, rs7869402 was very rare in the KHV population (Fig. 2, Table 2). Five haplotypes (H-1 to H-5) emerged with an Asiatic frequency higher than 1% (Table 3). The distribution of haplotypes differed between East and South Asian populations. H-4 is more prevalent in South Asian populations ($P < 0.001$) and H-5 is absent from South Asian populations, but has a frequency higher than 1% in East Asian populations (Fig. 3, Table 4). LD analysis revealed modest linkage between rs7873784, rs11536891, and rs11536896 polymorphisms

(rs7873784/rs11536891: $D' = 0.995$, $LOD = 230.51$, $r^2 = 0.984$, and rs11536891/rs11536896: $D' = 1.0$, $LOD = 239.44$, $r^2 = 1.0$). rs11536891 and rs11536896 are in complete linkage disequilibrium (Fig. 1).

***TLR4* luciferase constructs: site-directed mutagenesis to replicate SNPs**

We constructed a luciferase plasmid, pGL3-3494-3UTR, to simulate *TLR4* gene structure *in vivo*. The entire *TLR4* 3'-UTR sequence was inserted in-frame with the *LUC* gene, and approximately 3.7 kb of the *TLR4* promoter (-3494 to +235) was subcloned into the *LUC* NcoI restriction site of (Fig. 4). To detect the effect of SNPs on *TLR4* expression, we used PCR-mediated site-directed mutagenesis to construct pGL3-3493-mutrs344, pGL3-3493-mutrs402, pGL3-3493-mutrs889, pGL3-3493-mutrs784, pGL3-3493-mutrs891, and pGL3-3493-mutrs896, corresponding to rs41426344, rs7869402, rs11536889, rs7873784, rs11536891, and rs11536896, respectively (Fig. 5). To eliminate the effect of the promoter, the deletions (-684 to +235) were generated by truncation of the 5'-proximal promoter region from the constructs.

Effects of Polymorphism loci on *TLR4* Expression

Using RT-PCR, we determined that *TLR4* mRNA was highly expressed in SiHa and THP-1 cells, making them more for the analysis of *TLR4* expression than the C33A, HeLa, and 293T cell lines (Fig. 6a). In SiHa cells, an LPS agonist concentration of 50 $\mu\text{g/ml}$, activated *TLR4* mRNA, and TLR4/MyD88/NF- κ B pathway proteins at the highest level (Fig. 6b, c). In THP-1 cells, 10 $\mu\text{g/ml}$ was an optimal concentration, and increased *TLR4* mRNA expression ~2.6-fold, and activated the TLR4/MyD88/NF- κ B pathway to highest expression level (Fig. 6d, e).

To determine whether the SNPs influence *TLR4* gene expression, we performed transfection analysis. Relative luciferase activities for individual cell lines were compared to those of the pGL3-3494-3UTR construct group and the control group. The relative luciferase activity of the pGL3-3493-3UTR construct decreased ~1.5-fold in SiHa cells and ~1.4-fold in THP-1 cells in the presence of the *TLR4* 3'-UTR (Fig. 7a, c). Both of the promoter regions, -3494 to +235 bp (pGL3-3494) and -684 to +235 bp (pGL3-684), increased luciferase activities following LPS or IL-6 stimulation (Fig. 7). None of the six mutated plasmids showed significant activity changes in the control group. However, the significantly increased activities of pGL3-3494-mutrs402 and pGL3-3494-mutrs889 constructs demonstrate the "T" allele of rs7869402 and "C" allele of rs11536889 may increase the expression levels of *TLR4* gene *in vivo* when stimulated with LPS or IL-6 (Fig. 7). In addition, to investigate whether simultaneous mutation of the moderately linked rs7873784, rs11536891, and rs11536896, which were modest linkage each other, would alter *TLR4* expression, we created the pGL3-3494-TRImut construct, containing all three changes. LPS or IL-6 pretreatment had almost no effect on pGL3-3494-TRImut relative luciferase activity in either cell lines (Fig. 7).

Discussion

We found that SNPs were common in the *TLR4* 3'-UTR. Analysis of 1KG project data from 10 populations from Asian revealed up to 74 SNPs in this region, and that six of these (8.1%) had a MAF > 1%. The

distribution of these SNPs significantly differed among these populations. Moreover, we constructed five haplotypes based on four tag SNPs, rs41426344, rs7869402, rs11536889, and rs11536896. The H-1 and H-3 haplotypes are present in a similar frequency among all Asian populations, possibly indicating a conserved role in these populations. The MAF of rs11536889 is higher in CDX than in the other nine populations, and the highest frequency of haplotype H-2 is also observed in the CDX population. These findings indicate a stronger association between rs11536889 and *TLR4*-related disease in CDX populations than in other populations. The frequencies of rs7869402 and haplotype H-4 both have a higher levels in Bengali in Bangladesh and Sri Lankan Tamil in the UK populations than in other populations, suggesting that rs7869402 and H-4 may have more vital functions in these two populations from South Asia. The H-5 haplotype was not found in the South Asian populations due to the absence of rs41426344. rs41426344 is very rarely observed in JPT populations, where it is present at the lowest frequency of all East Asian populations. These results suggest that rs41426344 and H-5 could play different roles in East and South Asians populations. rs7873784, rs11536891, and rs11536896 were linked and almost always inherited together, making the frequency of these SNPs consistent in each population.

3'-UTR sequence length plays a pivotal role in translational regulation [30]. Moreover, the nucleotides within the in 3'-UTR may post-transcriptionally regulate the observed stimulus-specific induction [31, 32] and repress luciferase translation *in vivo* [33]. Therefore, we cloned the full-length *TLR4* 3'-UTR and inserted it seamlessly into LUC gene construct to imitate the *TLR4* gene structure *in vivo*.

TLR4 is a transmembrane receptor protein expressed on a small number of cell types, including myeloid cells, epithelial cells, B cells, muscle cells, and some cancer cells [1, 24-26]. Upregulated *TLR4* expression, via LPS or inflammatory cytokine stimulation, can cause inflammation [34]. In this study, we found that *TLR4* mRNA is expressed at a higher level in SiHa cells, a cervical carcinoma cell line infected with human papillomavirus 16 (HPV-16), than in HeLa (HPV-18), C33A (HPV negative), and 293T cells. *TLR4* mRNA expression is lower in SiHa cells than in THP-1 cells. The *TLR4* gene can respond to LPS stimulation in these two cells. Taken together, our results show that *TLR4* signal pathway responses to LPS may be controlled by HPV types in cervical cancer cell lines. Indeed, *TLR4*/*MyD88*/*NF-κB* pathway-related proteins were significantly up-regulated following LPS pretreatment. Therefore, we selected SiHa and THP-1 as the host cells for our luciferase assays.

The rs11536889 is located in the center of the *TLR4* 3'-UTR. Sato et al. used truncated fragments and transient transfection of a series of luciferase-reporter plasmids to demonstrate that, after LPS and IL-6 stimulation, the "G" and "C" alleles suppressed and upregulated luciferase activity, respectively, in THP-1 cells [8]. In this study, luciferase reporter experiments were designed to identify the effect of "G" and "C" alleles on *TLR4* expression using the full-length 3'-UTR. Consistent with the previous results, after stimulation with LPS or IL-6, the relative luciferase activities were significantly higher in the "C" allele than in the "G" allele in both of SiHa ($P < 0.01$) and THP-1 cells ($P < 0.01$). To predict SNP effects on *TLR4* mRNA secondary structure, we used RNAfold WebServer (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) and RNAsnp web server (<http://rth.dk/resources/rnasnp/>) [35]. The

transversion had no obvious implication in mRNA folding, and the minimum free energy of optimal secondary structure is -1600.20 and -1600.30 kcal/mol for the “G” and “C” alleles, respectively, and the p-value of maximum structural change is 0.1690 (Additional file 3: Figure S1a, d). The difference in the effect these alleles on gene expression may be due to this location being the target binding site for hsa-miR-1236 [8]. In a large case-control study include 780 controls match with 1383 prostate cancer patients in a Swiss population, Zheng et al. previously demonstrated that the C/C and G/C genotypes of this polymorphism increase prostate cancer risk [12]. In conclusion, these results suggested that rs11536889 may increase the risk of inflammatory disease and some cancers, that HPV-16 positive illnesses such as condyloma acuminatum [36], and cervical and anal carcinoma [37] can be identified as priorities for studies on the diseases caused by rs11536889.

rs7869402 is also located in the center of the 3'-UTR. After LPS or IL-6 stimulation, the luciferase activity of the T/T genotype was significantly up-regulated in both SiHa and THP-1 cells. Only one population-based genetic study showed that genetic variation in rs7869402 is significantly associated with susceptibility to pulmonary tuberculosis in Sudanese population [11]. There is a marginal difference of minimum free energy between the C/C and T/T (-1600.60 kcal/mol) genotypes, no evident change in mRNA folding, and the p-value of maximum structural change is 0.2121 (Additional file 3: Figure S1a, c). Our results indicate that rs7869402 is a pivotal variant that interferes TLR4, and the abnormal expression may be the result by interacted with microRNA.

In addition, limited population-based genetic studies have demonstrated that the polymorphism loci of rs41426344, rs7873784, and rs11536891 are associated with the risk of rheumatoid arthritis and juvenile idiopathic arthritis [9, 10], IgA nephropathy [6], and bipolar disorder [7]. In contrast, the luciferase activities of constructs corresponding to these variations produced almost constant levels of luciferase activity. rs11536891 and rs11536896 both have significant structural changes between the two genotypes ($P < 0.2$) (Additional file 3: Figure S1a, f, g), but no significant differences in minimum free energy. Therefore, further studies need to be performed to identify the possible mechanisms of these SNPs. Additional future research should focus on the associations between polymorphic sites and TLR4-related diseases phenotypes.

Conclusions

In summary, our analyses have showed that the distributions of TLR4 variants and haplotypes were significantly different among 10 Asian populations, in 3'-untranslated region. The rs7869402 T allele and rs11536889 C allele that contributed to the up-regulated TLR4 expression upon LPS or IL-6 stimulation might serve as potential functional mutations to apply to the future genetic study. The biological effects of rs7869402 and rs11536889 on TLR4 are significant clues that revealing its critical role in harmful TLR4-mediated responses. These results provided a reminder for future investigation on TLR4 related inflammatory diseases susceptibility.

Abbreviations

TLR4: Toll-like Receptor 4; SNPs: Single nucleotide polymorphisms; UTR: Untranslated region; 1KG: 1000 Genomes project; MAFs: Minor allele frequencies; LPS: Lipopolysaccharides; HPV: Human papillomavirus; IL-6: Interleukin-6; HWp: Hardy Weinberg equilibrium p value.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the results of this article are included within the article (and its additional files). 1000 Genomes project data is publicly available through its website.

Competing interests

The authors declare that they have no competing interests

Funding

This research was funded by the National Natural Science Foundation of China, grant number 81560043.

Authors' contributions

BZ, and HG conceived and designed the study. GC and LY analysed the data, performed the experiments and wrote the original draft, contributed equally to this work. YZ and YJ performed statistical analysis and aided in the editing of the manuscript. All authors contributed to scientific discussions, read and approved the final manuscript.

Acknowledgements

We thank Dr. Xue Zhang for her assistance in the use Arlequin software. Also, we would like to thank Editage (<http://www.editage.com>) for English language editing.

References

1. Rehli M, Poltorak A, Schwarzfischer L, Krause SW, Andreesen R, Beutler B. PU.1 and interferon consensus sequence-binding protein regulate the myeloid expression of the human Toll-like receptor 4 gene. *J Biol Chem.* 2000;275(13):9773-9781.

2. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science*. 1998;282(5396):2085-2088.
3. Mandrekar P, Szabo G. Signalling pathways in alcohol-induced liver inflammation. *J Hepatol*. 2009;50(6):1258-1266.
4. Horng T, Barton GM, Flavell RA, Medzhitov R. The adaptor molecule TIRAP provides signalling specificity for Toll-like receptors. *Nature*. 2002;420(6913):329-333.
5. Day L, Abdelhadi Ep Souki O, Albrecht AA, Steinhöfel K. Accessibility of microRNA binding sites in metastable RNA secondary structures in the presence of SNPs. *Bioinformatics*. 2014;30(3):343-352.
6. Bertinetto FE, Calafell F, Roggero S, Chidichimo R, Garino E, Marcuccio C, et al. Search for genetic association between IgA nephropathy and candidate genes selected by function or by gene mapping at loci IGAN2 and IGAN3. *Nephrol Dial Transplant*. 2012;27(6):2328-2337.
7. Oliveira J, Busson M, Etain B, Jamain S, Hamdani N, Boukouaci W, et al. Polymorphism of Toll-like receptor 4 gene in bipolar disorder. *J Affect Disord*. 2014;152-154:395-402.
8. Sato K, Yoshimura A, Kaneko T, Ukai T, Ozaki Y, Nakamura H, et al. A single nucleotide polymorphism in 3'-untranslated region contributes to the regulation of Toll-like receptor 4 translation. *J Biol Chem*. 2012;287(30):25163-25172.
9. Wang Y, Chen L, Li F, Bao M, Zeng J, Xiang J, et al. TLR4 rs41426344 increases susceptibility of rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA) in a central south Chinese Han population. *Pediatr Rheumatol Online J*. 2017;15(1):12.
10. Yang HJ, Wei CY, Li Q, Shou T, Yang Y, Xiao CJ, et al. Association of TLR4 gene non-missense single nucleotide polymorphisms with rheumatoid arthritis in Chinese Han population. *Rheumatol Int*. 2013;33(5):1283-1288.
11. Zaki HY, Leung KH, Yiu WC, Gasmelseed N, Elwali NE, Yip SP. Common polymorphisms in TLR4 gene associated with susceptibility to pulmonary tuberculosis in the Sudanese. *Int J Tuberc Lung Dis*. 2012;16(7):934-940.
12. Zheng SL, Augustsson-Bälter K, Chang B, Hedelin M, Li L, Adami HO, et al. Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the CAnCER Prostate in Sweden Study. *Cancer Res*. 2004;64(8):2918-2922.
13. Fukusaki T, Ohara N, Hara Y, Yoshimura A, Yoshiura K. Evidence for association between a Toll-like receptor 4 gene polymorphism and moderate/severe periodontitis in the Japanese population. *J Periodontal Res*. 2007;42(6):541-545.
14. Sun D, Wu Y, Wang H, Yan H, Liu W, Yang J. Toll-like receptor 4 rs11536889 is associated with angiographic extent and severity of coronary artery disease in a Chinese population. *Oncotarget*. 2017;8(2):2025-2033.
15. Huang Z, Liu Y, Liang L, Liu W, Sooranna SR, Mo J, et al. Association of Toll-like receptor 4 polymorphisms with the risk of coronary artery disease in the ethnic Zhuang population of the Guangxi Province of China. *Gene*. 2019;708:1-9.

16. Yvan-Charvet L, Welch C, Pagler TA, Ranalletta M, Lamkanfi M, Han S, et al. Increased inflammatory gene expression in ABC transporter-deficient macrophages: free cholesterol accumulation, increased signaling via toll-like receptors, and neutrophil infiltration of atherosclerotic lesions. *Circulation*. 2008;118(18):1837-1847.
17. Chang CC, Lee WT, Hsiao JR, Ou CY, Huang CC, Tsai ST, et al. Oral hygiene and the overall survival of head and neck cancer patients. *Cancer medicine*. 2019;8(4):1854-1864.
18. Mansur A, von Gruben L, Popov AF, Steinau M, Bergmann I, Ross D, et al. The regulatory toll-like receptor 4 genetic polymorphism rs11536889 is associated with renal, coagulation and hepatic organ failure in sepsis patients. *J Transl Med*. 2014;12:177.
19. Shi G, Wang T, Li S, Cheng Y, Sheng P, Fan Y, et al. TLR2 and TLR4 polymorphisms in Southern Chinese Psoriasis Vulgaris patients. *J Dermatol Sci*. 2016;83(2):145-147.
20. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75-81.
21. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073.
22. Das R, Upadhyai P. Application of geographic population structure (GPS) algorithm for biogeographical analyses of populations with complex ancestries: a case study of South Asians from 1000 genomes project. *BMC Genet*. 2017;18:109.
23. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun*. 2014;5:3513.
24. Koff JL, Shao MX, Ueki IF, Nadel JA. Multiple TLRs activate EGFR via a signaling cascade to produce innate immune responses in airway epithelium. *Am J Physiol Lung Cell Mol Physiol*. 2008;294(6):L1068-1075.
25. Reyna SM, Ghosh S, Tantiwong P, Meka CS, Eagan P, Jenkinson CP, et al. Elevated toll-like receptor 4 expression and signaling in muscle from insulin-resistant subjects. *Diabetes*. 2008;57(10):2595-2602.
26. Yu L, Wang L, Li M, Zhong J, Wang Z, Chen S. Expression of toll-like receptor 4 is down-regulated during progression of cervical neoplasia. *Cancer Immunol, Immunother*. 2010;59(7):1021-1028.
27. Balkwill F, Coussens LM. Cancer: an inflammatory link. *Nature*. 2004;431:405-406.
28. Morale MG, da Silva Abjaude W, Silva AM, Villa LL, Boccardo E. HPV-transformed cells exhibit altered HMGB1-TLR4/MyD88-SARM1 signaling axis. *Sci Rep*. 2018;8(1):3476.
29. Jin Y, Qiu S, Shao N, Zheng J. Association of toll-like receptor gene polymorphisms and its interaction with HPV infection in determining the susceptibility of cervical cancer in Chinese Han population. *Mamm Genome*. 2017;28:213-219.
30. Mazumder B, Seshadri V, Fox PL. Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci*. 2003;28(2):91-98.

31. Gallego X, Cox RJ, Laughlin JR, Stitzel JA, Ehringer MA. Alternative CHRN4 3'-UTRs mediate the allelic effects of SNP rs1948 on gene expression. *PLoS One*. 2013;8(5):e63699.
32. Watanabe S, Takeshige K, Muta T. A cis-element in the 3'-untranslated region of IkappaB-zeta mRNA governs its stimulus-specific expression. *Biochem Biophys Res Commun*. 2007;356(3):785-791.
33. Otero LJ, Devaux A, Standart N. A 250-nucleotide UA-rich element in the 3' untranslated region of *Xenopus laevis* Vg1 mRNA represses translation both in vivo and in vitro. *RNA*. 2001;7(12):1753-1767.
34. Medzhitov R, Janeway CJ. Innate immune recognition: mechanisms and pathways. *Immunol Rev*. 2000;173(1):89-97.
35. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. The RNAsnp web server: predicting SNP effects on local RNA secondary structure. *Nucleic Acids Res*. 2013;41:W475-479.
36. Maniar KP, Ronnett BM, Vang R, Yemelyanova A. Coexisting high-grade vulvar intraepithelial neoplasia (VIN) and condyloma acuminatum: independent lesions due to different HPV types occurring in immunocompromised patients. *Am J Surg Pathol*. 2013;37(1):53-60.
37. Sannigrahi MK, Sharma R, Singh V, Panda NK, Rattan V, Khullar M. Role of Host miRNA Hsa-miR-139-3p in HPV-16-Induced Carcinomas. *Clin Cancer Res*. 2017;23(14):3884-3895.
38. Limou S, Taverner AM, Winkler CA. Ferret: a user-friendly Java tool to extract data from the 1000 Genomes Project. *Bioinformatics*. 2016;32(14):2224-2226.
39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-265.
40. Halperin E, Kimmel G, Shamir R. Tag SNP selection in genotype data for maximizing SNP prediction accuracy. *Bioinformatics*. 2005;21:i195-203.
41. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 2007;1:47-50.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-408.
43. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6(5):343-345.
44. Schnoor M, Buers I, Sietmann A, Brodde MF, Hofnagel O, Robenek H, et al. Efficient non-viral transfection of THP-1 cells. *J Immunol Methods*. 2009;344(2):109-115.

Tables

Table 1 List of all SNPs (MAF \geq 1%) in the TLR4 3'-untranslated region, their genomic positions on chromosome 9, and their allele frequencies in Asians represented in the 1000 Genomes Project (Phase 3)

NO.	Name	Chr9 Position	HWp	%Geno	MAF	Allele 1	Allele 1 frequency	Allele 2	Allele 2 frequency
24	rs41426344	120477933	1	100	0.016	G	0.9844	C	0.0156
27	rs7869402	120478032	1	100	0.055	C	0.9451	T	0.0549
30	rs11536889	120478131	0.5218	100	0.22	G	0.78	C	0.22
47	rs7873784	120478936	0.7411	100	0.108	G	0.8917	C	0.1083
57	rs11536891	120479337	0.7686	100	0.109	T	0.8912	C	0.1088
71	rs11536896	120479734	0.7686	100	0.109	T	0.8912	C	0.1088

Table 2 The frequencies of SNPs (MAF \geq 1%) among the 1000 Genomes Project (Phase 3) in ten Asian populations

NO.	Name	East Asians					South Asians					P value
		CDX	CHB	CHS	JPT	KHV	BEB	GIH	ITU	PJL	STU	
24	rs41426344	0.032	0.044	0.029	0.01	0.04	/	/	/	/	/	2.93E-01
27	rs7869402	0.022	0.058	0.057	0.034	0.015	0.11	0.058	0.074	0.026	0.098	9.10E-05
30	rs11536889	0.339	0.228	0.243	0.26	0.247	0.238	0.121	0.196	0.188	0.152	1.60E-05
47	rs7873784	0.145	0.078	0.081	0.072	0.146	0.11	0.112	0.078	0.172	0.098	8.17E-03
57	rs11536891	0.145	0.078	0.081	0.082	0.146	0.11	0.107	0.078	0.172	0.098	1.38E-02
71	rs11536896	0.145	0.078	0.081	0.082	0.146	0.11	0.107	0.078	0.172	0.098	1.38E-02

Table 3 The most frequent haplotypes in the *TLR4* 3'-untranslated region generated by Tag SNP, with frequencies higher than 1% in Asian populations

NO.	Chr9 Position	Tag SNP	H-1	H-2	H-3	H-4	H-5
24	120477933	rs41426344	G	G	G	G	C
27	120478032	rs7869402	C	C	C	T	C
30	120478131	rs11536889	G	C	G	G	C
71	120479734	rs11536896	T	T	C	T	T
Asians frequency, n = 993			0.604	0.198	0.105	0.055	0.016
S.D.			0.011	0.009	0.007	0.005	0.003

Table 4 The frequencies of haplotypes (MAF \geq 1%) among the 1000 Genomes Project (Phase 3) in ten Asian populations

Name	East Asians				South Asians					<i>P</i> value	
	CDX	CHB	CHS	JPT	KHV	BEB	GIH	ITU	PJL		STU
H-1	0.495	0.636	0.619	0.625	0.591	0.541	0.714	0.652	0.615	0.652	1.30E-03
H-2	0.306	0.184	0.214	0.250	0.207	0.238	0.121	0.196	0.188	0.152	6.39E-04
H-3	0.145	0.078	0.081	0.082	0.146	0.110	0.107	0.078	0.172	0.098	1.35E-02
H-4	0.022	0.058	0.057	0.034	0.015	0.110	0.058	0.074	0.026	0.098	9.10E-05
H-5	0.032	0.044	0.029	0.010	0.040	/	/	/	/	/	2.93E-01

Figures

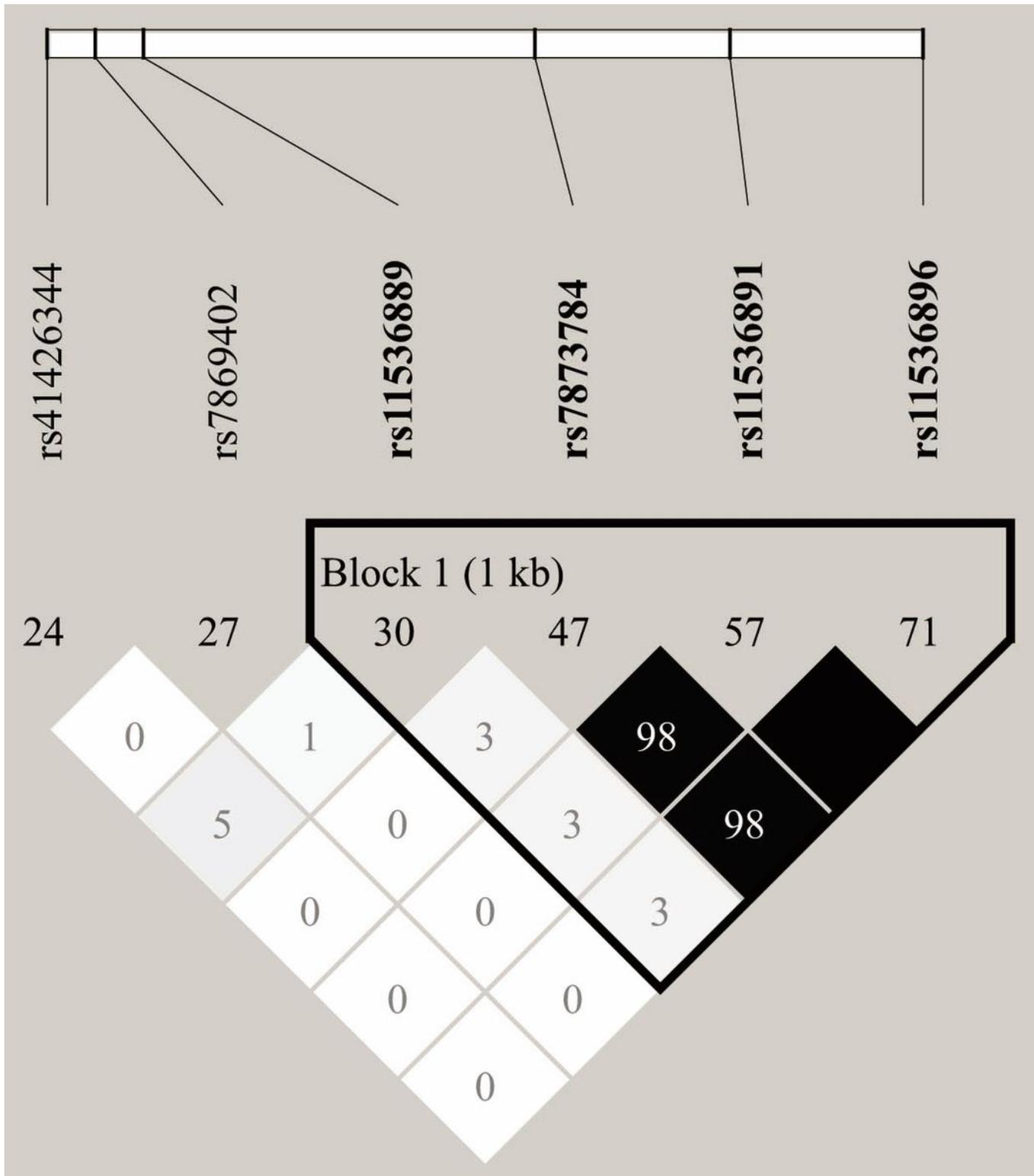


Figure 1

Linkage disequilibrium (LD) analysis for six SNPs (MAF \geq 1%) in the TLR4 3'-UTR in Asian populations. The number in the boxes indicates the R-squared value ($\times 100$) between SNPs, absence of value represents r^2 of 1. The LD color scheme is determined via r^2 : Darker box represents higher pairwise r^2

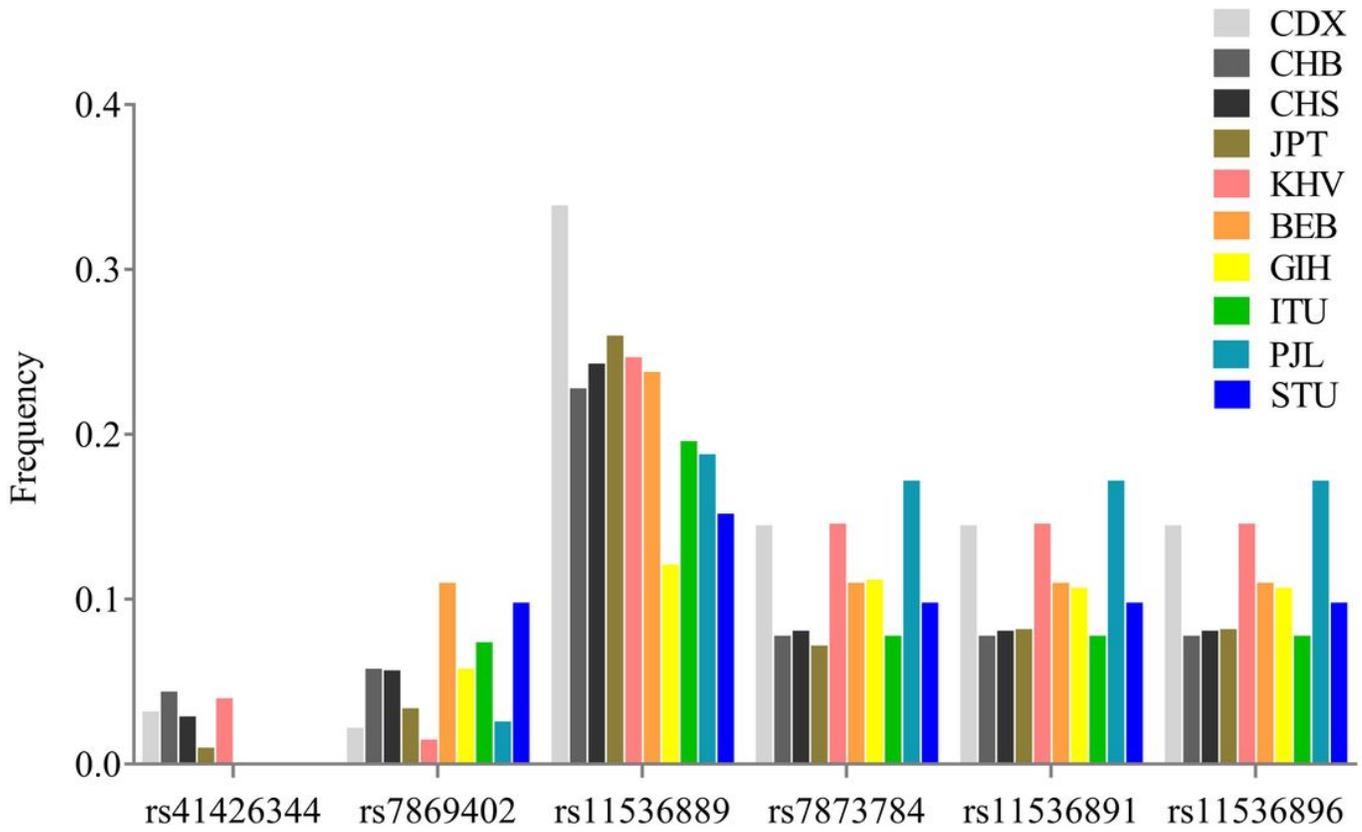


Figure 2

Distribution of the allele frequency of six SNPs (MAF \geq 1%) in 10 Asian populations

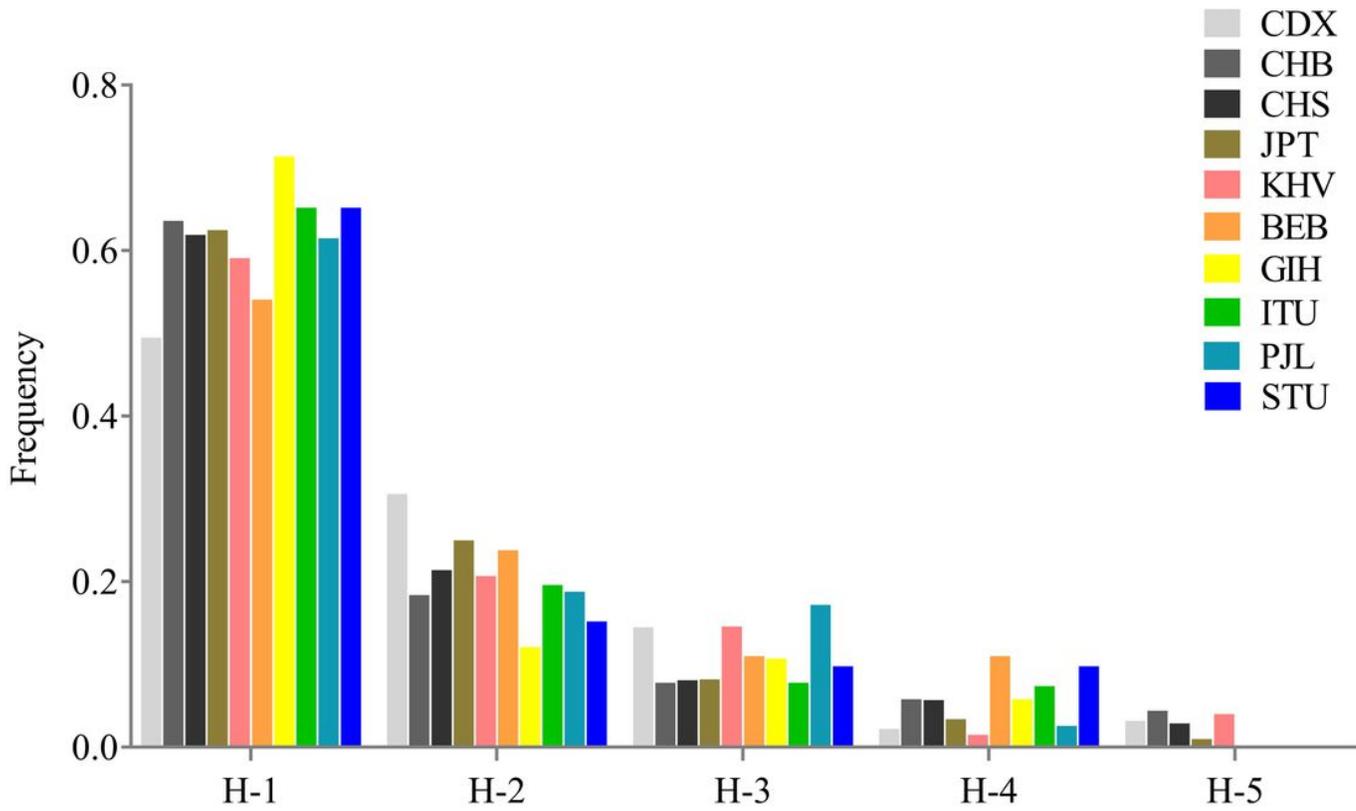


Figure 3

Frequency distribution of the most frequent haplotypes in the TLR4 3'-UTR region in different populations

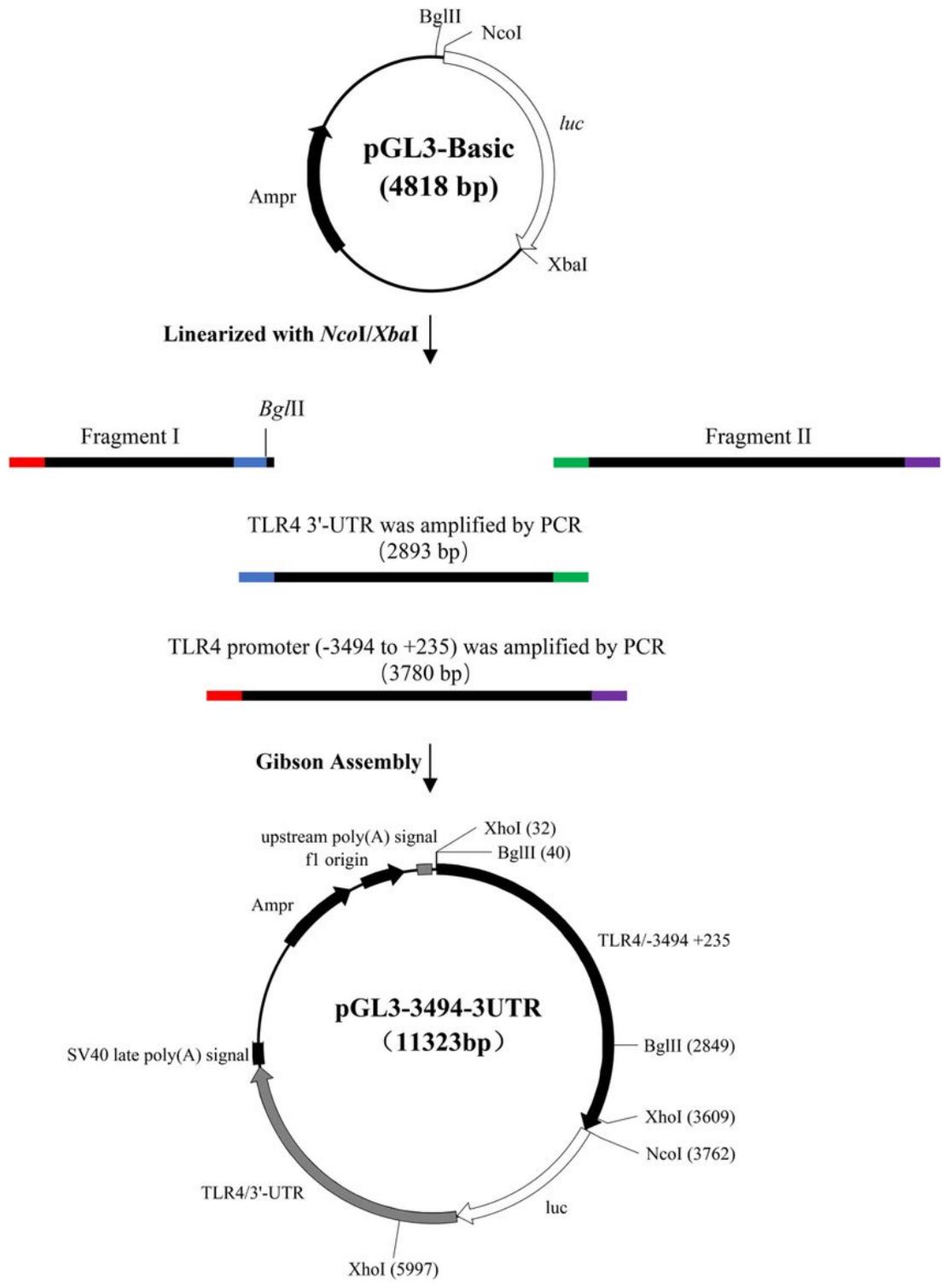


Figure 4

Schematic diagram showing the construction procedure and gene structure of the TLR4 luciferase reporter plasmid. Sequence homology is depicted with red, blue, green, and purple. The four homologous sequences are introduced into both ends of the promoter and 3'-UTR sequence using PCR primers. The four DNA fragments are assembled with Gibson assembly to yield plasmid pGL3-3493-3UTR

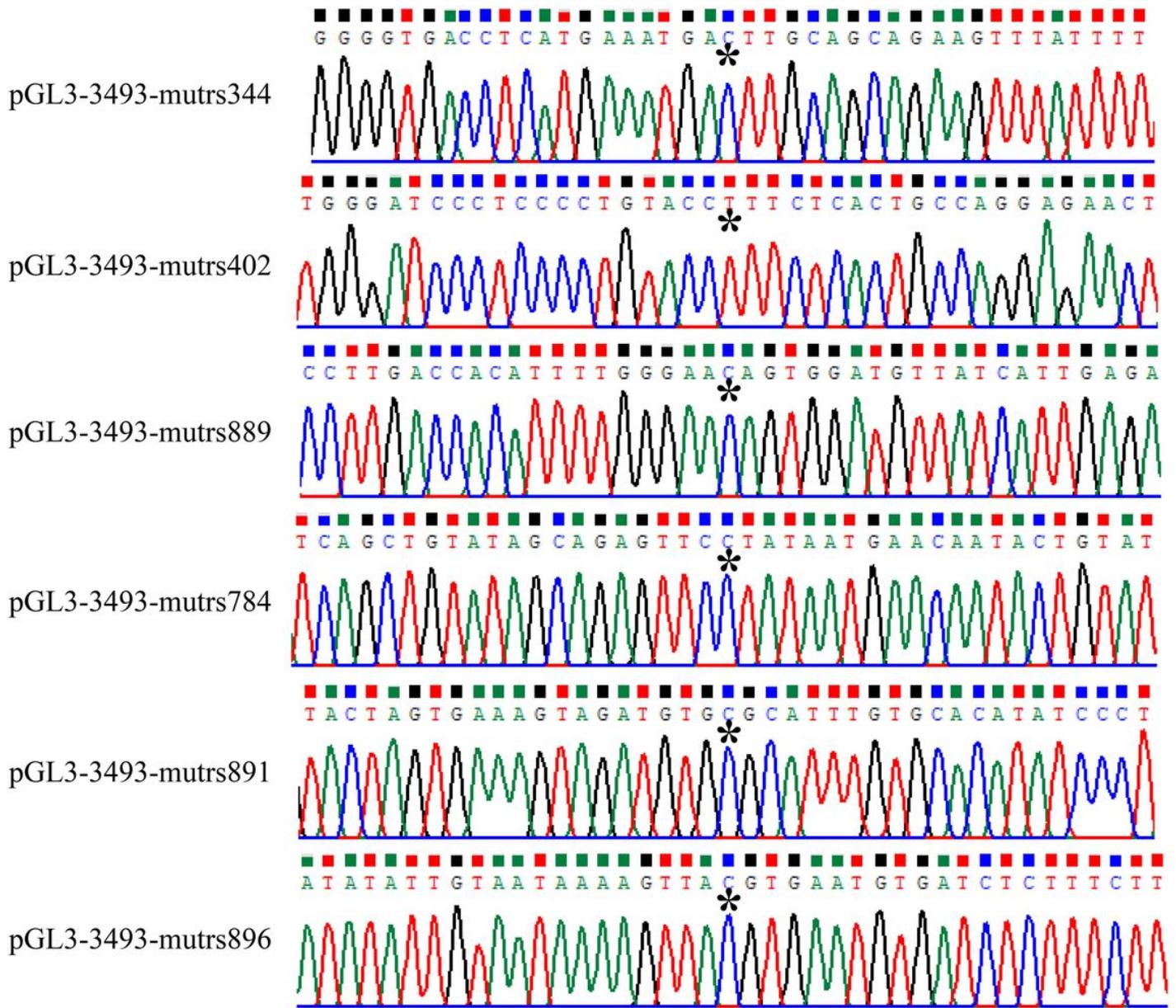


Figure 5

Verification of site-directed mutagenesis by DNA sequencing. Mutant loci highlighted with the asterisk (*) correspond to rs41426344, rs7869402, rs11536889, rs7873784, rs11536891, and rs11536896

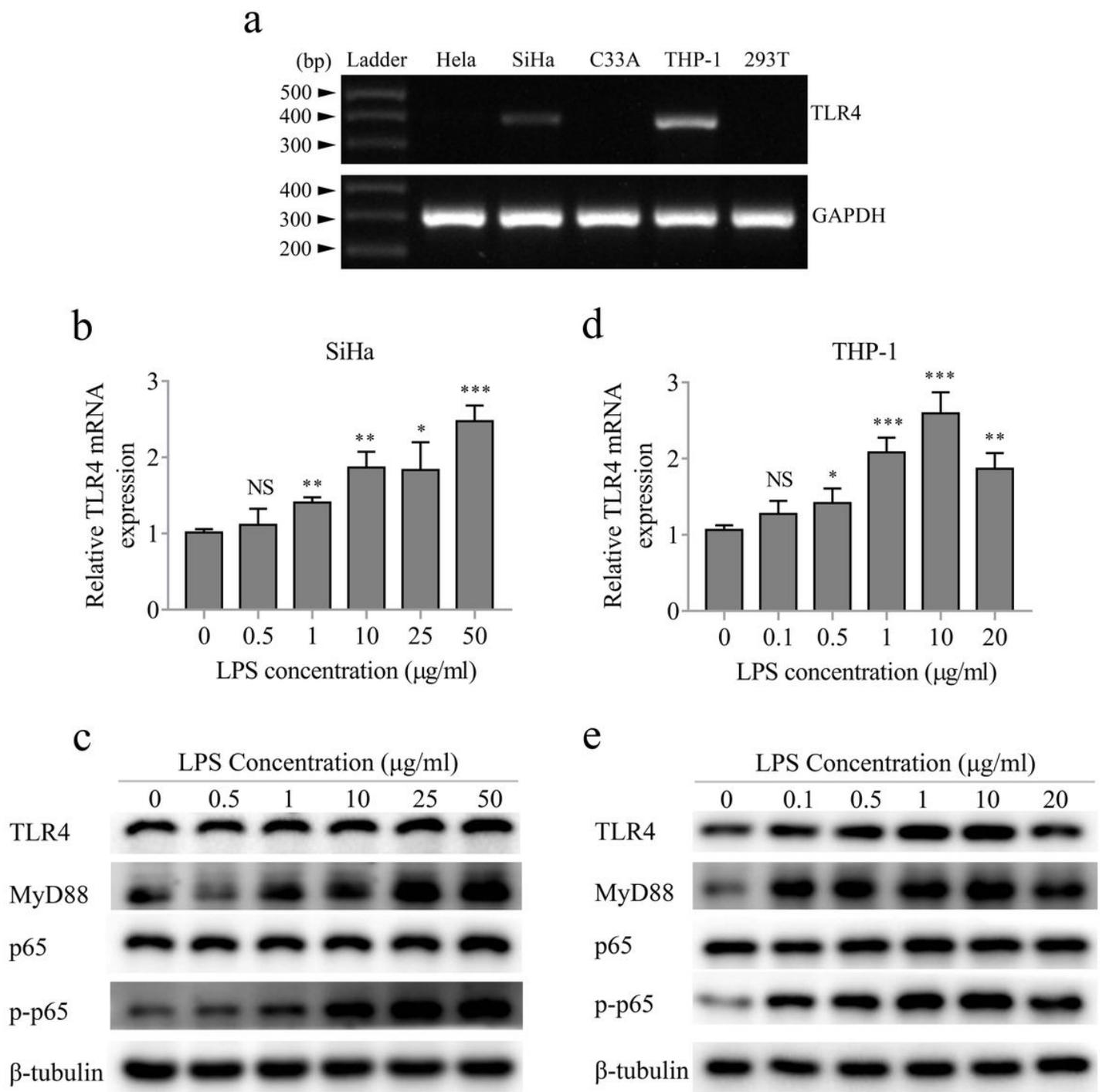


Figure 6

TLR4 expression in different cell lines. (a) RT-PCR analysis of TLR4 mRNA expression in C33A, HeLa, SiHa, 293T, and THP-1 cells respectively, GAPDH is used as a loading control. qRT-PCR analysis TLR4 mRNA expression in SiHa cells (b) and in THP-1 cells (d) treated with different concentrations of LPS. Western blot analysis of TLR4/MyD88/NF-κB pathway-related protein expression in SiHa (c) and THP-1 cells (e) treated with different concentrations of LPS. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, compared with control group

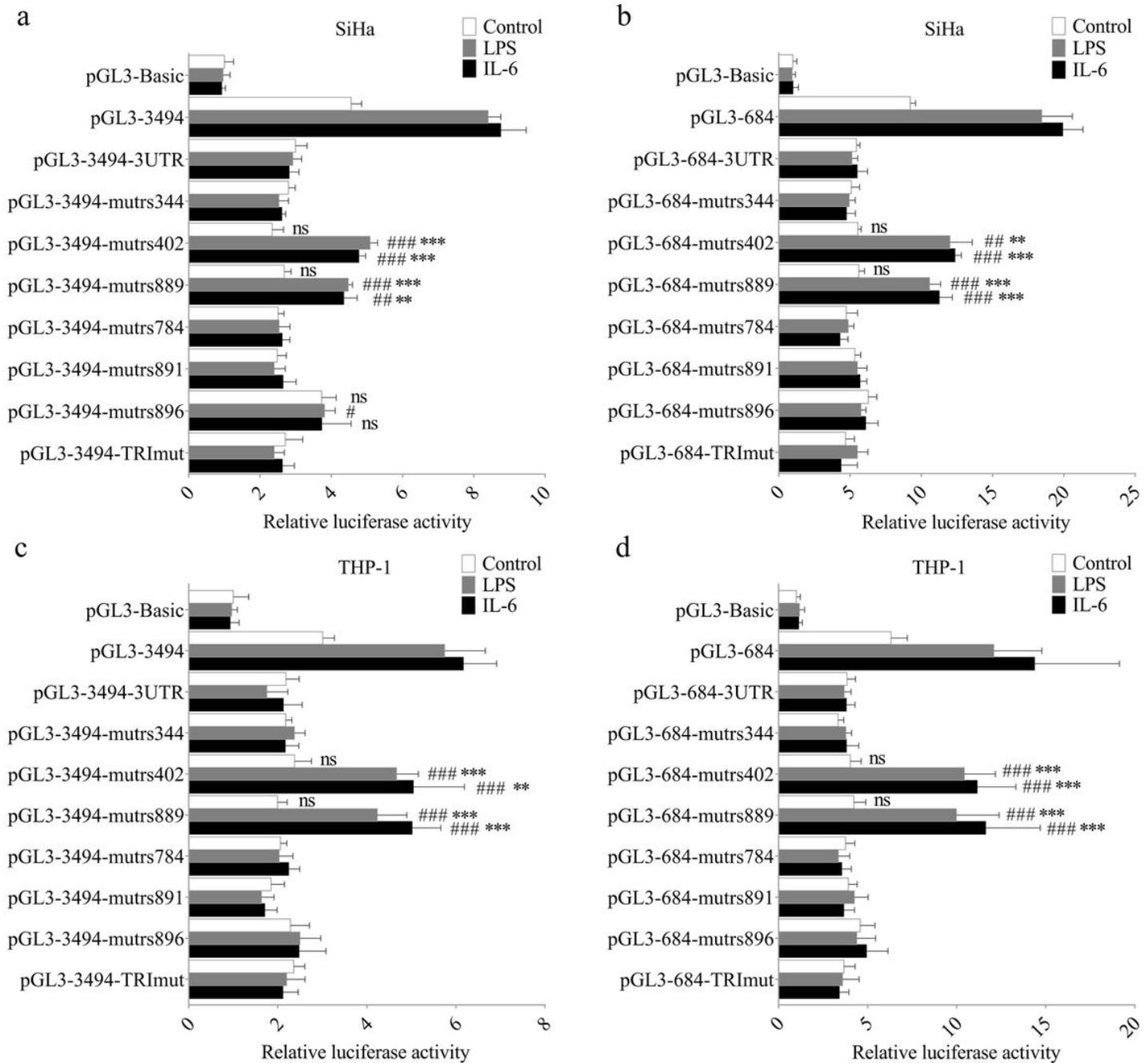


Figure 7

Effect of six SNPs on TLR4 promoter-driven luciferase activity in SiHa and THP-1 cells. SiHa cells were transfected with a series of chimeric luciferase constructs containing the TLR4 promoter of -3494 to +235 (a) or -684 to +235 (b). THP-1 cells were transfected with luciferase constructs containing the TLR4 promoter of -3494 to +235 (c) or -684 to +235 (d). After the transfectants were unstimulated, or stimulated with LPS or IL-6, luciferase activity was measured and normalized to Renilla luciferase activity. The values were obtained from three separate transfections in SiHa, and five independent transfections in THP-1 cells, and are presented as the means + S.D. Compared with control group **: $P < 0.01$; #: $P < 0.05$; ns: not significant; ###: $P < 0.001$; ####: $P < 0.0001$; ***: $P < 0.001$.

0.01, ***: $P < 0.001$; and compared with wild-type group of pGL3-3493-3UTR or pGL3-684-3UTR ns: $P > 0.05$, #: $P < 0.05$, ##: $P < 0.01$, ###: $P < 0.001$

Supplementary Files

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

- [supplement1.docx](#)
- [supplement2.docx](#)
- [supplement3.docx](#)