**TLR9 Polymorphisms Might Contribute to the Ethnicity Bias for EBV-Infected Nasopharyngeal Carcinoma**

**HIGHLIGHTS**
- EBV-associated nasopharyngeal carcinoma (NPC) is endemic in Mongoloids
- TLR9 is essential for virus recognition and immunity against cancers
- Polymorphisms are common in TLR9 protein in EBV-infected Mongoloids with NPC
- NPC susceptibility prediction by TLR9 mutation screening in people for prevention
TLR9 Polymorphisms Might Contribute to the Ethnicity Bias for EBV-Infected Nasopharyngeal Carcinoma

Nabanita Roy Chattopadhyay, Koustav Chatterjee, Nikhil Tiwari, Sudipta Chakrabarti, Sushil Kumar Sahu, Sankar Deb Roy, Arijit Ghosh, Rajendra Reddy, Piyanki Das, Sudip Mal, Basab Bijay Karnar, Ashok Kumar Das, Sam Tsering, Komri Riba, Zoreng puii, Eric Zomawia, Y. Indibar Singh, Amol Ratnakar Suryawanshi, Abhishek Kumar, Dipyaman Ganguly, Chandan Goswami, and Tathagata Choudhuri

SUMMARY

Nasopharyngeal carcinoma (NPC) is a rare malignancy in most parts of the world, but is endemic in some ethnic groups. The association of NPC with the Epstein-Barr virus (EBV) is firmly established; however, the mechanism is still unclear. TLR9 is well known for its essential role in viral pathogen recognition and activation of innate immunity. Here, we report a set of TLR9 polymorphisms in the TIR-2 domain of the TLR9 protein collected from the EBV-infected NPC samples from northeast Indian populations sharing the aforesaid ethnicity. The occurrence of mutations is significantly high in these samples as we found a p value of <0.0001 at a significance level of 0.05. These might play an important role for the lack of function of TLR9 and thus for the higher occurrence of EBV-mediated NPC in such ethnic groups.

INTRODUCTION

Nasopharynx is a box-like chamber near the base of the skull and covers the upper region of the throat behind the nose. Carcinoma of the nasopharynx, or nasopharyngeal carcinoma (NPC), starts at the mucosal epithelium of the nasopharynx and in the minor salivary glands present there (Chan et al., 2005; Wei et al., 2010; Wei and Sham, 2005; American Cancer Society, 2016). It may occur at any age and occurs much more frequently in the Chinese and Southeast Asian populations, also known as mongoloid populations. NPC accounts for about 0.7% of all cancers in a global perspective, but in endemic populations, it occurs in 0.02%–0.03% males and about 0.01%–0.015% females, and the commons t form is (UC) undifferentiated carcinoma. It is endemic among Chinese (Cancer Incidence in Five Continents, 1987), Maghrebian Arabs in North Africa and Algeria (Cancer Occurrence in Developing Countries, 1986), and the Eskimos in the Arctic (Cancer Incidence in Five Continents, 1982). NPC is uncommon in most regions in India except for the Northeast (NE) region of the Indian subcontinent (Bhatia and Singh, 1981). The NE region in India is well known by a majority of Tibeto-Burman languages and is supposed to be populated by people migrated from East Asia, also bringing the nasopharyngeal cancer with them (Kataki et al., 2011). In non-endemic regions like North America and Europe, this incidence is much lower (about 0.001%, including both genders) and other forms of NPC are seen.

The risk factors of NPC include genetic factors, diet and other environmental factors, and infection with human papillomavirus (HPV) or Epstein-Barr virus (EBV) (http://www.cancer.org/cancer/nasopharyngealcancer/detailedguide/nasopharyngeal-cancer-risk-factors). It is well known that the populations with higher incidences of NPC follow a few interesting dietary habits and their lifestyles maintain poor hygiene. Most people from these populations consume some kinds of preserved foods, for example, the Chinese consume salt-preserved foods, the NE Indian Nagas consume smoke-preserved food, etc. With regard to the lifestyles, all these susceptible populations live in poor hygienic conditions, e.g., staying long time in smoky environment, inhaling various toxic fumes like that of formaldehyde, habits of smoking, etc. The association of NPC with EBV was firmly established in as early as 1973 (Xu et al., 2012). Yet, the role for the virus in the pathogenesis of NPC is still unclear. Almost all cases are EBV positive, irrespective of geographical origin. Although controversial, EBV has been classified as a group I carcinogen by the International Agency for Research on Cancer (IARC), among other reasons, because of its association with NPC (IARC Monographs on the Evaluation of...
Carcinogenic Risks to Humans, Vol 70, IARC press, Lyon, France, International Agency for Research on Cancer, 1997). EBV, a human gammaherpesvirus, was first detected in African patients with Burkitt lymphoma and infects B cells and epithelial cells (Epstein et al., 1964). This virus is very well known for causing infectious mononucleosis (glandular fever) and various types of cancers including Burkitt lymphoma, Hodgkin lymphoma, NPC, and gastric carcinoma (Klein et al., 2007; Young et al., 2016). After initial infection, the EBV nuclear antigens (EBNAs) are expressed; but eventually different latency programs of EBV give rise to different sets of expressed viral antigens (Alfieri et al., 1991; Halder et al., 2009; Houldcroft and Kellam, 2015). EBNA-2 is one of the first genes expressed upon infection of B cells and is essential for establishment of latent infection and cell growth transformation. So this antigen is used widely for the detection of EBV infection in patients.

Toll-like receptors (TLRs) are innate immune sensors that are integral for resisting chronic and opportunistic infections; they recognize pathogen-associated molecular patterns (PAMPs) that are expressed on infectious agents like viruses and bacteria. TLRs are transmembrane proteins expressed on the cell surface and the endocytic compartments. TLR9, also called CD289, is expressed in dendritic cells, macrophages, natural killer cells, and other antigen-presenting cells of the immune system in tissues like peripheral blood, spleen, lymph node, and bone marrow. Various infections, cancers, and diseases including some autoimmune diseases, can modulate TLR9 activation and expression. TLR9 is activated by unmethylated cytidine-phosphate-guanosine (CpG) dinucleotides common in microbes and starts the antiviral responses by triggering the production of antiviral cytokines such as type I interferons (IFNs). The pathway includes MyD88 and TRAF6, leading to inflammatory responses via nuclear factor (NF)-κB activation and cytokine secretion (Du et al., 2000; Takeshita et al., 2001; Doyle et al., 2007). A recent article shows that TLR9 can recognize some other nucleotide patterns present in bacterial or viral genomes (Martinez-Campos et al., 2017). However, the TLR9 protein is usually activated by unmethylated CpGs present in microbial DNA and moves to Golgi apparatus and lysosomes from its initial location, the endoplasmic reticulum. Then the molecule is cleaved to prevent autoimmunity and only a part of the TLR9 original protein is actually expressed on the cell surface. Eventually the signaling pathway leads toward the production of cytokines such as IL-6 (interleukin-6), tumor necrosis factor, IFN-α, and IL-12.

Mounting evidences implicate the role of TLR-polymorphisms in susceptibilities to various infectious diseases, including human immunodeficiency virus (HIV)-1. Pine et al. investigated the impact of TLR single-nucleotide polymorphisms (SNPs) on clinical outcome in a sero-incident cohort of HIV-1-infected volunteers (Pine et al., 2009; Rahman et al., 2016; Medvedev, 2013; El-Omar et al., 2008). However, no report shows the role of TLR9 polymorphisms in patients with NPC infected with EBV. In the current study, we tried to search the TLR9 gene polymorphisms in patients of NE Indian populations with NPC who are EBV-positive, wherein this disease is a common problem. We found some deletions, additions, and point mutations in the TLR9 gene of such patients, suggesting an important role of these SNPs in the patients of NE Indian region with NPC. It should be noted that NPC is endemic in this region, and the affected population follows similar characteristics as the populations of other endemic regions; thereby the findings of this study might indicate a role of TLR9 in EBV-positive NPC with a broad spectrum including other endemic populations. We report for the first time that TLR9 plays an important role in EBV-positive NPC development in the NE Indian population. By performing comparative genomic analyses, we found that TLR9 is conserved on the same loci from ray-finned fishes to mammals. We specifically analyzed the hidden features of TLR9 and the endocytic compartments.

RESULTS
The study was carried out in collaboration with a few NE Indian centers. Seventy freshly diagnosed patients with NPC along with 70 age- and sex-matched controls were registered for this study from all those centers. Routine histopathological analysis was done for each patient to confirm the diagnosis of NPC. Informed consent was obtained from each and every subject as per the guidelines of research review committee. Approval was obtained from the institutional medical ethics committees of the participating institutes for the study.

EBV Is Well Associated with NPC
One of the major problems associated with NPC is the detection of the disease properly. Cellular characteristics are confusing and poorly understood, so the disease is commonly misdiagnosed. The physical
EBV and HPV both have been reported to be associated with NPC. After initial infection, the EBNAs are expressed, EBNA-2 being the first of those (Lung et al., 2014). EBNA-2 is essential for establishment of latent infection and cell growth transformation (Cohen et al., 1989; Hammerschmidt and Sugden, 1989), whereas EBNA-1 is known to function for the lytic phase with its ability to bind DNA in a site-specific manner (Frappier and O’Donnell, 1991). EBNA-1 is a stable homodimer that recognizes a DNA-binding site 16 bp in length (Baer et al., 1984). Soon after the EBNA-2 expression, the EBV latent membrane proteins 1 and 2 (LMP1 and LMP2) are expressed. For the determination of association of EBV infection in NE Indian patients with NPC, we have examined the presence of EBNA-1 and EBNA-2 genes in the blood samples of patients with NPC and controls. After PCR amplification, all those samples showed the presence of EBNA-1 or EBNA-2 (Figure 1B, only a few representatives are shown).

**Different Types of Mutations Are Identified in TLR9**

Various signaling pathways have been reported for their contributions in development and pathogenesis of NPC (Chan, 1980; Tulalamba and Janvilisri, 2012). As there is cumulative evidence showing the possible role of TLR9 in the pathogenesis of NPC, we hypothesized that some SNPs/mutations of the TLR9 gene might be associated with susceptibility toward and/or the severity of NPC in the NE Indian population representing a distinct ethnic group. PCR amplifications were performed with the DNA samples isolated from blood samples of patients and controls (all primers are listed in Table S3). For the control group, blood samples were collected from healthy age- and sex-matched volunteers from same ethnic background. The PCR-amplified DNA sequences were purified and sequenced. These sequences were compared with a standard genomic sequence of TLR9 obtained from the published database of National Center for Biotechnology Information (NCBI) database (Database: >gi|74054321|gb|AAZ95520.1| TLR9 [Homo sapiens]). Mutation screening in those sequences was performed by Finch TV software (https://digitalworldbiology.com/FinchTV). In addition, we have checked the mutations by direct sequencing and using the ClustalW software (http://www.ebi.ac.uk/Tools/msa/clustalo/), and all the sequences were authenticated by checking in human BLAST with the help of NCBI (www.ncbi.nlm.nih.gov). We have found some mutations in exon 2 of TLR9 gene.
of the collected NPC samples; of those some mutations are non-synonymous and some are synonymous. It was found that all patients with UC carry at least one mutation, five patients with NKDC had at least one mutation, and 29 patients with KSCC were detected with mutated TLR9. Along with base substitutions, we have also found addition and deletion in the exon 2 of TLR9 gene of some samples (Figures 1C and D, Table 1). A 3-dimensional (3D) model of the wild-type reference TLR9 and the domain where the mutations are found, as predicted by Phyre2 software (E). See also Figure S1 and Table S1.

As stated earlier, TLRs recognize conserved molecular patterns on specific classes of pathogens and initiate a series of signaling events that leads to the expression of pro-inflammatory genes. Recent studies show genetic polymorphisms in TLRs in various pathogenic diseases improving our understanding of the relationship between TLRs and pathogenesis. These in future may provide a rational basis for developing novel therapies to treat these important diseases. A role for TLR9 on viral, fungal, mycobacterial, and Helicobacter pylori infections each has been described by some workers earlier (Berrington and Hawn, 2007; Carvalho et al., 2008, 2009). In humans, the TLR9 gene is located on chromosomes 3 and is polymorphic (Georgel et al., 2009). Among the TLRs that bind nucleic acids, TLR9 recognizes unmethylated CpG DNA motifs (present at a much higher frequency in the genomes of prokaryotes and virus than that of eukaryotes) as a “danger signal” that activates the innate immune system (Carvalho et al., 2012). In humans, this receptor is expressed in plasmacytoid dendritic cells and B lymphocytes, cells that are known to have a
diverse TLR repertoire with the TLR9 being predominant. EBV lives latently in B lymphocytes and can eventually enter the lytic cycle. During an infection, TLR9 is stimulated by unmethylated CpG sites of EBV and triggers the activation of macrophages, B lymphocytes, and dendritic cells, leading to the production of various cytokines, chemokines, and immunoglobulins. Later on, some of these molecules, like IL-2, help the T cells to differentiate into activated T helper 1 and cytotoxic T cells (Krieg, 2007). Interestingly, these initial immune responses toward EBV start to subside as LMP genes start to be expressed. It has been reported recently that EBV persistence may be favored by downregulation of TLR9 via LMP1-mediated NF-kB activation (Fathallah et al., 2010). LMP1 activates NF-kB, which prevents TLR9 promoter activity, and thus prevents it from its antimicrobial activity. Another protein, which is an EBV lytic phase protein called BGLF5, triggers the degradation of TLR9 mRNA (van Gent et al., 2011). Persistence of EBV in humans and generation of various cancers including NPC can be justified by such inhibitory methods adopted by the virus to subside the action of TLR9 (Jordi et al., 2017). The most important finding of these literatures is that loss of function of TLR9 might result in the initiation and progression of EBV-mediated cancers including NPC. In our study, we found various mutations in the coding region of the gene (in exon 2, actually), of which, most result in the loss of function of TLR9. So, such mutations in the NE Indian patients indicate a genetic bias that might be attributed to the higher incidence of NPC in that ethnic group.

### Highly Conserved TLR9 Depicts Mutations in a Specific Domain

The presence of TLRs is restricted to higher eukaryotes only except in most birds and reptiles. The lowest organism where any TLR channel is documented so far is the zebrafish (Yajuan et al., 2017). Comparative genomic analyses of TLR9 depicted that the TLR9 gene is conserved on the same locus from fishes to human over a period of 450 million years with sets of conserved flanking genes (Figure 2A). Syntenic analyses also demonstrated that some fishes have multiple copies of TLR9 like in Atlantic fish; there are five paralogs of TLR9 known as TLR9a–e (Figure 2A).

In most of these organisms including zebrafish, TLR9, if present, recognizes CpG-oligodeoxynucleotides and includes signaling pathways related to those of humans (Mutwiri, 2012; Kumagai et al., 2008). TLR9

| Sl. No. | SNP | Triplet Code Change | Amino acid Change | Mutation | Addition | Deletion |
|--------|-----|---------------------|-------------------|----------|----------|----------|
| 1      | A>52195766T | TGG > AGG | Trp > Lys | + | – | – |
| 2      | G > 52195702Del | AGT > A-T | Ser>__ | – | – | + |
| 3      | 52188102A > Del | GAA > GA- | Glu>__ | – | – | + |
| 4      | 52188114CA > Del | CAG>__R | Gln>__ | – | – | + |
| 5      | 52188159C > G | CCC > GCC | Pro > Ala | + | – | – |
| 6      | 52195757G > C | CCC > GCC | Pro > Ala | + | – | – |
| 7      | A>52188093Del | GAA > GA- | Glu > __ | – | – | + |
| 8      | 52188160A > Del | AAA > AA- | Lys > Lys | – | – | + |
| 9      | 52188093A > Del | AGA> GA | Arg > __ | – | – | + |
| 10     | 52188152C > A | GCC > GCA | Ala > Ala | + | – | – |
| 11     | 52188170A > Del | AAG > Del | Lys>__ | – | – | + |
| 12     | 52188186 A > -Del | AGT> GT | Ser > – | – | – | + |
| 13     | 52188093A > Del | AAG > A_G | Lys>__ | – | – | + |
| 14     | 52188151C > Del | CCA > -AA | Gln > – | – | – | + |
| 15     | 521880107C > Del | TTC > TT_ | Phe > – | – | – | + |
| 16     | 52188151C > Del | CCA > -AA | Pro > – | – | – | + |
| 17     | 52188151C > Del | CCA > -CA | Pro > – | – | – | + |

Table 1. Toll-like Receptor 9 (Third Chromosomal Position) Mutations, Deletions, and Additions in Northeast Indian NPC Samples
gene codes for an 1132-amino acid (aa)-long protein, which contains some conserved domains, most of which are the leucine-rich repeats (LRRs) (Database: https://www.ncbi.nlm.nih.gov/cdd) (Figure 2B). LRRs are generally 20- to 30-aa-long and contain a significantly higher amount of the hydrophobic amino acid leucine; contributing to the β-sheet structures in the corresponding protein. These regions function as the major areas of molecular interactions. In TLR9, LRRs can interact with microbial lipopolysaccharides and/or other pathogenic molecules, thereby facilitating the recognition of, and innate immunity to, foreign invaders. Both the regions (aa 132 and aa 208) that can bind pathogenic unmethylated CpG are located in LRRs. These two amino acids interact with four other peptide stretches along the TLR9 protein, three of which are located in LRRs (upper panel). None of the SNPs in our study has been found in these regions, suggesting that either the IL-mediated signaling is affected or any structural alteration of TLR9 makes it nonfunctional (middle and lower panels). None of the SNPs in our study is found to be located in LRRs; all are present in an N-terminal region belonging to TIR_2 superfamily (Toll/interleukin-2 receptor). Each sequence of exon 2 of TLR9 from each patient with NPV who was EBV positive was aligned with the standard TLR9 sequence, and we have found various sequence alterations in the full TLR9 gene (D). A total of 17 different types of mutations are found to be present in the TIR-2 domain of exon 2 of the TLR9 gene of mutant samples. These mutations are found in a small region of the gene at its 3′ end, a region that can be considered as the mutation hotspot (E). Calculations include 5% or fewer errors. In humans, TLR9 protein can interact with more than 20 other proteins and initiate downstream signaling thereof (F). See also Table S2.
Each sequence of exon 2 of TLR9 from each patient with NPC who is EBV-positive was aligned with the standard TLR9 sequence as mentioned earlier. We have found various sequence alterations in the full TLR9 gene (Figure 2D), of which at least 1 of 17 types of mutations was found to be present in the TIR-2 domain of exon 2 of the TLR9 gene of 74% patient samples (Table S1). This value is quite high, and we found it statistically significant considering the p value of < 0.05 (Figure 1C, right lower part). Depending on these data, it has also been found that these mutations are found in a small region of the gene at its 3' end, a region that can be considered as the mutation hotspot (Figure 2E and Table S2). About half of the patients with mutated TLR9 show less than 80% sequence similarity with the reference TLR9 sequence, and most of those are due to SNPs (Figure 2E). It is of note that the N-terminal region of the final TLR9 protein plays a vital role in the interaction of the protein with the unmethylated CpG of the pathogens, although the exact molecular mechanism is not known (Ohto et al., 2015). The C-terminal domain is responsible for downstream signaling and interactions with various cellular proteins (Figure 2F). Therefore it may be inferred that mutation(s) in the 3' end of the gene results in lack of function of TLR9 for immunity against EBV, leading to higher susceptibility toward NPC (Figure 2C, lower right panel). However, elucidation of the role of any individual SNP needs a detailed investigation of TLR9 structure-function relationship.

As living habitats and environments encountered by different species are tightly linked with detection of immunity and further recruitment of downstream signaling, TLR9 provides an example of a unique molecule that can be studied in the context of molecular evolution. In humans, this protein can interact with more than 20 other proteins and can initiate downstream signaling thereof (Figure 2F). Most of these proteins are found in other higher eukaryotes having TLR9 signaling cascade. Therefore the subtle changes and retention of domains and motifs that are essential for the function can also be explored by comparing the multiple sequences from different organisms. We have retrieved the TLR9 sequences of different species available in public databases and have analyzed the molecular evolution of TLR9 by fragmenting the TLR9 amino acid sequence in different domains and motifs (Figure 3A). We specifically analyzed the regions that are conserved across those species and also analyzed the respective selection pressure imposed on these regions from their degree of conservation. We used MUSCLE alignment program to align the amino acid sequences of TLR9 for the purpose of phylogenetic analysis. We implemented a Bayesian phylogenetic tree constructed by the Bayesian approach (5 runs, 7,500,000 generations, 25% burn-in period, WAG matrix-based model in the MrBayes 3.2 program; Figure 3B). The aligned data were subsequently imported into R statistical tool for statistical analysis. As the complete TLR9 sequences from certain species are not available (mostly due to sequencing errors at certain regions), the analysis was conducted with the available sequences only. As expected, domains and motifs show conservation among varied species in due course of evolution, reflecting the important role of TLR9 in survival. Polymorphisms are present, but they do not affect the structure-function relationship along the evolutionary lineage.

Absence of TLR9 Protein Is Not Attributed by These Mutations

Previously we have shown that none of the mutations are found in the CpG-interacting region of TLR9; therefore the ability of this protein to interact with and recognize foreign DNA is not compromised. This finding led us to hypothesize two phenomena. First, recognition of EBV as the causative agent of NPC does not play the major role; instead, any loss of function actually contributes to the EBV-mediated carcinogenic process. Like other TLRs, TLR9 can recognize a variety of microbes bearing similar PAMPs, e.g., unmethylated CpG (Martı´nez-Campos et al., 2017). This may initiate similar kinds of downstream signaling, but the success of pathogen clearance depends on activation of specific second messengers and other immune molecules. As mentioned earlier, the proper execution of TLR9 function is dependent on cellular proteins like MyD88 and IRAK4, as well as pathogenic proteins like LMP1 and BGLF5 (Du et al., 2000; Takeshita et al., 2001; Doyle et al., 2007; van Gent et al.). The balance between the direct activation of NF-κB pathway by TLR9 and suppression of TLR9 by LMP1-mediated activation of NF-κB pathway might play a vital role in this process, as the latter mechanism is related to brown adipose tissue (BAT)-generating lipid metabolism seen in the Mongoloids (Fathallah et al., 2010; Jimenez-Preitner et al., 2011, 2011; Talukder et al., 2012; Kostjuk et al., 2012; Nagajyothi et al., 2012; Koberlin et al., 2016).

The second thing we hypothesized is that the presence or absence of TLR9 protein is not critically affected, at least due to these mutations. The functional regulation of TLR9 is outlined in the previous paragraph and is also detailed in Introduction. Therefore we tried to test whether these mutations are responsible for any loss of TLR9 protein or not. We performed immunohistochemical staining of TLR9 in deparaffinized NPC tissue sections by streptavidin-peroxidase method and diaminobenzidine. Hematoxylin was used to stain...
the base tissue, and the sections were examined microscopically. Presence of TLR9 has been seen in all samples tested (Figure 3C, upper panel). We further tested the presence of TLR9 gene, RNA, and protein by performing PCR, real-time RT-PCR, and western blot (C, second, third, and fourth panels, respectively). No significant absence of TLR9 was noted. See also Tables S1–S3.

Figure 3. Functional and Evolutionary Importance of TLR9
Divergence analysis shows that the mutation hotspot is more conserved than the full-length TLR9 throughout the vertebrate evolution, suggesting high selection pressure on this region. However, this region contains conserved sequence patterns within different ethnic groups in humans. Although our study needs further experiments to infer about the structure-function relationship affected by the mutations, presence of polymorphisms in this region suggests a strong such role for the protein (A). Phylogenetic analysis of domains and motifs shows conservation among varied species in due course of evolution, reflecting the important role of TLR9 in survival. Polymorphisms are present in other regions also, but they do not affect the structure-function relationship along the evolutionary lineage. Phylogenetic analysis of domains and motifs shows conservation among varied species in due course of evolution, reflecting the important role of TLR9 in survival. Polymorphisms are present in other regions also, but they do not affect the structure-function relationship along the evolutionary lineage (B). Presence of TLR9 has been seen in all samples tested by streptavidin-peroxidase method and diaminobenzidine (DAB) (C, upper panel). We further tested the presence of TLR9 gene, RNA, and protein by performing PCR, real-time RT-PCR, and western blot (C, second, third, and fourth panels, respectively). No significant absence of TLR9 was noted. See also Tables S1–S3.

The most significant second messenger of TLR9 is MyD88. It interacts directly with TLR9 protein to start immune activation. Therefore we tried to check the effect of presence or absence of TLR9 on MyD88 and some other downstream proteins like TRAF6, IRAK1, and IRAK4. Although these proteins are important in TLR9 signaling, it should be noted that these proteins are expressed from unrelated genes (Du et al., 2000;
Therefore, lack of TLR9 should not imply any compulsion on the presence or absence of these proteins. We have established the cultures of Raji (EBV-positive B cell line of head and neck cancer origin) and Ramos (EBV-negative B cell line of head and neck cancer origin) cells were transfected with siRNA for TLR9. Treated and control sets were examined, but no significant morphological change was observed (A, two upper rows). Each of these cells was then treated with fluorescently tagged antibodies to visualize the presence and interaction of TLR9 and MyD88 (A, third and fourth rows). As seen under the microscope, absence of TLR9 in siRNA-treated cells does not affect the presence of MyD88. However, MyD88 seems to colocalize and interact with TLR9 when both are present (A, lower row). Lack of or reduced TLR9 RNA is seen in RT-PCR in siRNA-treated cells, as expected (B). Presence of MyD88, TRAF6, IRAK1, and IRAK4 were also tested in RT-PCR, and no significant reduction is noted (C). Western blot analyses of TLR9, MyD88, and IRAK4 were done, but no significant alterations were found (D). For experimental control, siRNA to GAPDH was tested. See also Table S3.

Figure 4. Alteration of Downstream Signaling Is Not Attributed to Total Absence of TLR9 Protein

(A–D) Cultures of Raji (EBV-positive B cell line of head and neck cancer origin) and Ramos (EBV-negative B cell line of head and neck cancer origin) cells were transfected with siRNA for TLR9. Treated and control sets were examined, but no significant morphological change was observed (A, two upper rows). Each of these cells was then treated with fluorescently tagged antibodies to visualize the presence and interaction of TLR9 and MyD88 (A, third and fourth rows). As seen under the microscope, absence of TLR9 in siRNA-treated cells does not affect the presence of MyD88. However, MyD88 seems to colocalize and interact with TLR9 when both are present (A, lower row). Lack of or reduced TLR9 RNA is seen in RT-PCR in siRNA-treated cells, as expected (B). Presence of MyD88, TRAF6, IRAK1, and IRAK4 were also tested in RT-PCR, and no significant reduction is noted (C). Western blot analyses of TLR9, MyD88, and IRAK4 were done, but no significant alterations were found (D). For experimental control, siRNA to GAPDH was tested. See also Table S3.
Analyzing these results, we inferred that the mutations do not affect the presence or absence of the protein, but alter the function of TLR9 by altering its structure-function relationship.

Radical Mutations Might Affect the Function of TLR9

The variations in amino acid sequences of TLR9, as revealed in 1,000 human genome sequences collected from public databases (Database: www.ncbi.nlm.nih.gov), have been used in our study for Grantham distance calculation. Align-GVGD is an extension of the original Grantham difference to multiple sequence alignment. It combines the biophysical characteristics of amino acids and protein multiple sequence alignments to predict where missense substitutions in genes of interest might have deleterious or neutral effects. The probability of amino acid substitutions are negatively correlated with the physicochemical distances, which are based on amino acid properties. As higher Grantham score reflects a greater evolutionary distance, higher scores are considered more deleterious and their effects are more damaging. In our study, we have found amino acid substitutions having significant physicochemical distances among them, thus a possible alteration of structure-function relationship leading to inactivation of TLR9. We have also found some deletions in the DNA resulting in frameshift during protein production. This again indicates a loss of function of TLR9 leading to disease susceptibility. A region of the transmembrane (TM) part and two regions in the cytoplasmic (C) part contain the possible mutation hotspot (Figure 5). Hence, these amino acid substitutions or frameshift mutations might affect the signaling cascade initiated by TLR9. The signal peptide shows no mutation; therefore the subcellular localization of TLR9 is not affected in the patients with NPC, at least in our study. Similarly, no mutation is found in the CRAC (cholesterol recognition amino acid consensus) and CARC (the reverse version of CRAC) domains, indicating no alteration for association with cholesterol. Interestingly, none of these domains is present in the TM part of TLR9, but two CRAC domains are present close to the mutation hotspot present in the cytosolic region. Therefore any mutation, which can affect the structure of TLR9, might affect the CRAC-related downstream signaling cascades.

According to the Align-GVGD analysis, it can be noted that amino acid substitutions with high risks of deleterious effects are seen in NE Indian patients. Polymorphisms in various amino acid positions have low through moderate to high risks for NPC (Figure 6A, left panel). In our study, we have found a few polymorphisms that can be considered with a very high risk for disease susceptibility (Figure 6A, middle and right panels, red dots). We have also analyzed the amino acids by predicting their SIFT (scale-invariant feature transform) scores and PolyPhen-2 scores. These two are the two most commonly used algorithms for predicting if an SNP has any effect (generally negative effect) on a protein’s structure. Both scores use the same range, 0.0–1.0, but with opposite meanings; a PolyPhen score of 0.0 is predicted to be benign, whereas a SIFT score of 1.0 is predicted to be benign. After analyzing all the sequences we got a few mutations that can be considered deleterious in accordance with all three scores, i.e., GD score, SIFT score, and PolyPhen-2 score (Figure 6B). Therefore, loss of function of TLR9 is expected in mutants and susceptibility toward NPC may be inferred (Figure 6C). As stated earlier, genetic polymorphisms in TLRs, like other genes, are conserved among ethnic groups. Therefore the findings of our study strongly indicate a role of TLR9 polymorphisms for susceptibility toward NPC in NE Indian populations with higher incidence.

To further investigate the effects of these mutations on the structure-function relationship of TLR9, we have performed 3D modeling study with the help of two different software named Phyre² (http://www.sbg.bio.ic.ac.uk/phyre2) and Raswin (http://rasmol.org/) (Figure 7A). Only the reference part of wild-type TLR9 and three other mutated samples are shown for simplicity. The uppermost part of the figure shows the sequences aligned with ClustalW, as mentioned earlier. Base substitutions in the mutated sequences are marked in red. Just below this alignment, 3D-folding prediction of each of the above-mentioned four sequences is shown; against each, its PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) analysis is shown. With the help of this software, each amino acid of a sequence has been studied for its character (aatypes), its possibility to be present in a particular type of protein folding (psipred), and its possibility to interact with the membrane (memsat). The Phyre² results show that the 3D structures are distinctly different from that of the wild-type sequence, indicating the loss of function due to aberrant structure caused by the mutations. The PSIPRED analysis classified as the “psipred“ format shows that none of the mutated sequences carry amino acids required for helical structure formation like the wild-type. This again justifies the possibility of loss of function due to the structure-function relationship. The characters of the substituted amino acids, as shown in the “aatypes“ format, complement this. The “memsat“ format of the sequence analysis indicates a role of the mutation having an even greater importance. The substituted amino acids mostly are unable to interact with the membrane; thereby indicating that mutated TLR9 protein may be produced inside the cell,
but that it is unable to be integrated into the cellular membranes. This lack of integration can destroy the proper functioning of TLR9 and is reported elsewhere (Corne´lie et al., 2004; Mouchess et al., 2011; Majer et al., 2018).

Interestingly, the amino acids found in the transmembrane region and a part of the following carboxy-terminal region in wild-type TLR9 protein matches exactly with the predicted roles of amino acids in

Figure 5. Important Regions of TLR9 and Possible Regulation by Membrane Cholesterol
The signal peptide shows no mutation, therefore the subcellular localization of TLR9 is most likely not affected in patients with NPC. Similarly, no mutation is found in the CRAC (cholesterol recognition amino acid consensus) and CARC (the reverse version of CRAC) domains, possibly indicating no alteration for association with cholesterol. A region of the transmembrane (TM) part and two regions in the cytoplasmic (C) part contain the possible mutation hotspot, and two cytosolic CRAC domains are present close to this hotspot. Hence mutation in this region is expected to affect the structure of TLR9, most likely to be influenced by membrane cholesterol, and thus may affect the CRAC/CARC motif-related downstream signaling cascades. See also Tables S1 and S2.
membrane interaction by the “memsat” format. Thereby we may assume that the prediction is close to real in case of mutated sequences also (Figure 7, these regions are marked by red rectangular boxes). All these findings clearly point toward the loss of function of mutated TLR9 proteins due to aberrant structure-function relationship dependent on protein folding, which in turn is dependent on the amino acid composition. To supplement our data, we have also performed western blots for TLR9, MyD88, and IRAK4 with protein extracted from real NPC tissue samples (Figure 7B). Although some bands for mutated TLR9 are not much clear, and thus may indicate the absence of the protein, it should be noted that the poor condition of the tissue might be responsible for that. Overall, both tissue samples and cultured treated/untreated cells show similar results for the presence of TLR9 and other related proteins.

DISCUSSION

NPC is well reported for its ethnicity-specific higher incidences. Although all higher eukaryotes possess a complex immune system to fight for various diseases, higher incidence of any disease linked with any specific ethnic group is of interest. All these populations, sometimes called “Mongoloids,” share similar characteristics like broad and small nose with smaller nasal passages, broad face, abundant fat tissues in facial region especially surrounding the nose, and mongoloid folds over the eyes. These characteristics have resulted from cold adaptation during the migration of human populations from middle Africa and are related to lipid metabolism. Cold adaptation provides a particular mechanism for thermal homeostasis in humans, called nonshivering thermogenesis (NST), the main site of which is the regions high in BAT (Guthrie, 1996; Montagu, 1951; Cannon and Nedergaard, 2004). In recent literature, it has been reported that biogenesis of BAT requires the transcription factor C/EBPb and the zinc finger protein Prdm16. Both these genes are activated by an evolutionarily conserved protein Plac8, which can be considered as a critical upstream regulator of brown fat differentiation and function. Plac8 expression and function may be
regulated by its interaction with phospholipid scramblase 1 (PLSCR1), which is a calcium-binding protein associated with lipid rafts (Jimenez-Preitner et al., 2011).

Interestingly, this PLSCR1 has been reported to regulate TLR9-mediated type I IFN production. PLSCR1 also contributes to cell proliferation and differentiation and helps avoid apoptosis, indicating that cancer cells might get benefited for their survival (Talukder et al., 2012). Of note, TLR9 itself and its unmethylated
CpG-binding property both are reported to be related to lipid metabolism (Köberlin et al., 2016; Kostjuk et al., 2012; Nagajyothi et al., 2012). The main gene for NST from BAT is the uncoupling protein 1 (UCP1) gene, which uncouples the proton gradient in the inner mitochondrial membrane from the formation of ATP. In a very recent study it has been shown that this gene is also closely related to the expression of TLR9. In colder climate, presence of UCP1 decreases TLR9 function, thereby make the animals susceptible to pathogenic infections (Kazak et al., 2017). After initial infection and EBNA-2 expression, LMP1 starts to be expressed and then activates the NF-κB signaling pathway. This is a broad pathway affecting various cellular processes including lipid metabolism. NF-κB signaling pathway itself can suppress the promoter of TLR9 and make it nonfunctional, as mentioned earlier. Therefore the genes involved in lipid metabolism and TLR9 regulation mentioned earlier, and the NF-κB-mediated signals, might play the contributing role in NPC occurrence. With regard to cell survival, TLR9 helps p53 expression; therefore lack of TLR9 function might result in lack of p53 and thus lack of apoptosis in cancers (Holm et al., 2017). These findings indicate a possible role of TLR9 in controlling susceptibility toward EBV for the mongoloid populations, who show higher occurrence of NPC.

It is very clear that lack of function of TLR9 renders a cell very much susceptible to pathogenic infections like infection with EBV. Therefore the main path of EBV-mediated NPC generation might involve the loss of function of TLR9 (by any pathway as discussed earlier) by deleterious SNPs found in our study (Figure 8). Single-nucleotide variations (SNVs) in TLR9 are documented in various cancers including the head and neck cancers, of which NPC is a part (https://hive.biochemistry.gwu.edu/cgi-bin/prd/biomuta/servlet.cgi?gpageid=11&searchfield1=gene_name&searchvalue1=TLR9). Compared with other cancers, SNVs are less common in head and neck cancers and are much less reported in NPC. Some polymorphisms are also documented in the exon 2 of TLR9 (http://bioinf.umbc.edu/dmdm/gene_prot_page.php?search_type=protein&id=20140872), but no relationship of those with NPC is shown. In the present study, we report for the first time a set of polymorphisms in exon 2 of TLR9 gene, i.e., amino acid modifications in a specific region of the TLR9 protein. We have noted an interesting thing from in silico analyses; all the mutated sequences show modifications in the amino acids responsible for the membrane interaction of TLR9 protein. Membrane-interacting proteins always show functional aberration if their membrane interaction is compromised, and we have already mentioned that with supporting references in a previous part of this article. Although the present study shows no conclusive data for the association of specific TLR9 polymorphisms with NPC, we cannot entirely exclude the role of TLR9 as a candidate gene for this rare but fatal disease. Other TLR9 gene SNPs that were not analyzed in our study may also be associated with NPC, and further studies are needed to elucidate the role of TLR9 in the pathogenesis of NPC in NE Indian populations and other similar ethnic groups. Detailed investigation on the mechanism of UCP1-mediated and PLSCR1-mediated repression of TLR9 will also add to the knowledge for explaining the ethnicity bias of EBV-associated NPC occurrence. Non-mongoloid persons with EBV infection and TLR9 mutation should also be studied in detail to find out the molecular mechanism for such bias. Nevertheless, the importance of our findings lies in the possibility of early detection of the disease and/or susceptibility toward the disease, which can help prevention and cure of this fatal cancer. Most of the patients experienced unexplained chronic headache, and it indeed is an initial symptom of NPC generation (Figure S1B). Complaint of such headache even without any other symptom can be taken care of by screening the polymorphic status of the TLR9 gene/protein of the subject. Presence of any deleterious mutation(s) can then suggest other extensive tests for diagnosis of an early NPC or susceptibility toward it. The treatment plan and/or preventive measures, for example, a change in diet and lifestyle, might therefore save a life. Sequencing of full-length TLR9 or even the “mutation hotspot” at early age can be used as a possible biomarker to access the risk factors associated with fatal disease.

Limitations of the Study
The present report documents a novel possible mechanism for NPC generation and poor prognosis, thereby paving the way for newer preventive and/or curative techniques for this fatal cancer. However, the study has a few limitations that should be considered before planning any diagnostic and/or curative method(s) for NPC in such populations. First, NPC itself is rare; therefore our accessible sample size was small enough to conclude for the model pathway (Figure 8). Second, a complete list of genetic polymorphisms responsible for ethnicity-specific characters is not available and was not created by us due to lack of required data for specific genetic makeup of different human races. Third and the last limitation we are deeply concerned of is whether neither the exact molecular mechanism for cholesterol signaling, diet, and BAT in these ethnic populations nor the details of TLR9 signaling cascade related to BAT is known.
Therefore our model needs a more detailed study for the TLR9 downstream signaling pathway and the structure-function alterations resulting from polymorphisms reported here. Nevertheless, the major importance of our findings is that it provides new insights for studying the ethnicity bias of NPC.

METHODS
All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY
In the Supplemental Information, we have provided all the required data needed for reproducibility of this work. Other data are available in https://data.mendeley.com/datasets/bd6xp274j/draft?a=d7e11c13-04d5-4b65-bd51-1b318c70e245.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100937.

ACKNOWLEDGMENTS
We thank Dr. Luna Goswami, Associate Dean (Research) and Associate Professor (School of Biotechnology), Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, for her kind support in data collection and manuscript preparation. We also acknowledge Mr. Subham Saraswat and Mr. Laltu Hazra, Dept of Biotechnology, Visva Bharati, for their help in file management and laboratory-related matters, respectively. Informed consent was obtained from each subject as per the guidelines of research review.
committee. Approval was obtained from the Institutional Medical Ethical Committees of the participating centers for the study. Post-doctoral fellowship to the first author has been provided by Indian Council of Medical Research, Government of India (Proposal ID. 2019–4919). This work was supported by DBT through its Extramural Grant, Department of Biotechnology, Government of India (Grant No. BT/01/NE/TBP/ 204(MED)/3/2011). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**AUTHOR CONTRIBUTIONS**

Conceptualization, T.C. and C.G.; Methodology, T.C., C.G., and N.R.C.; Sample collections and histology, S.D.R., A.K.D., S.T., K.R., Z.P., E.Z., and Y.I.S.; Investigations, N.R.C. and N.T.; Supportive actions in performing the experiments: K.C., S.C., S.K.S, A.G., R.R.R., P.D.B.B.K., S.M., A.R.S., A.K., and D.G.; Formal Analysis, N.R.C., N.T., C.G., and T.C.; Writing – Original Draft, N.R.C.; Writing – Review & Editing, N.R.C. and T.C.; Resources, C.G., A.K., D.G., and T.C.; Supervision, T.C.

**DECLARATION OF INTERESTS**

The authors have declared that no competing interests exist.

Received: January 17, 2019

Revised: January 28, 2020

Accepted: February 19, 2020

Published: March 27, 2020

**REFERENCES**

International Agency for Research on Cancer (1997). Epstein Barr virus and Kaposi’s sarcoma, herpes virus/human herpes virus. In IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 70 (IARC press), pp. 164–194.

Alfieri, C., Birkenbach, M., and Kieff, E. (1991). Early events in Epstein-Barr virus infection of human B lymphocytes. Virology 181, 595–608.

American Cancer Society (2016). What are the risk factors for nasopharyngeal cancer?. http://www.cancer.org/cancer/nasopharyngealcancer/detailedguide/nasopharyngeal-cancer-risk-factors.

Baer, R., Bankier, A.T., Biggin, M.D., Denninger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., et al. (1984). DNA sequence and expression of the B9s-8 Epstein-Barr virus genome. Nature 310, 207–211.

Berrington, W.R., and Hawn, T.R. (2007). Mycobacterium tuberculosis, macrophages, and the innate immune response: does common variation matter? Immuno. Rev. 219, 167–186.

Bhatia, P.L., and Singh, L.S. (1981). Evaluation of contrast radiography in nasopharyngeal malignancy. Indian J. Cancer 18, 141–146.

Cancer Incidence in Five Continents. (1982). Cancer Incidence in Five Continents, Vol. IV (IARC press).

Cancer Incidence in Five Continents. (1987). Cancer Incidence in Five Continents, Vol. V (IARC press).

Cancer Occurrence in Developing Countries. (1986). Cancer Occurrence in Developing Countries (IARC press).

Cannon, B., and Nedergaard, J. (2004). Brown adipose tissue: function and physiological significance. Physiol. Rev. 84, 277–359.

Cavalho, A., Pasqualotto, A.C., Pitzurra, L., Romani, L., Denning, D.W., and Rodrigues, F. (2008). Polymorphisms in toll-like receptor genes and susceptibility to pulmonary aspergillosis. J. Infect. Dis. 197, 618–621.

Cavalho, A., Cunha, C., Carotti, A., Aloisi, T., Guerrero, O., Di lanni, M., Falzetti, F., Brizto n, F., Aversa, F., Pitzurra, L., et al. (2009). Polymorphisms in Toll-like receptor genes and susceptibility to infections in allogeneic stem cell transplantation. Exp. Hematol. 37, 1022–1029.

Cavalho, A., Cunha, C., Almeida, A.J., Osório, N.S., Saraiva, M., Teixeira-Coelho, M., Pedreiro, S., Tomrado, E., Domingues, N., Gomez-Alves, A.G., et al. (2012). The rs5743836 polymorphism in TLR9 confers a population-based increased risk of non-Hodgkin lymphoma. Genes Immun. 13, 197–201.

Chan, S.H. (1980). Immunogenetics of nasopharyngeal carcinoma – Position in 1980. Ann. Acad. Med. Singapore 9, 296–299.

Chan, J.K.C., Pilch, B.Z., Kuo, T.T., Wenig, B.M., and Lee, A.W.M. (2005). Tumours of the nasopharynx. In WHO Classification of Tumours: Pathology and Genetics of Head and Neck Tumours, L. Barnes, J.W. Eveson, P. Reichart, and D. Sidransky, eds. (IARC press), pp. 83–97.

Cohen, J.J., Wang, F., Mannick, J., and Kieff, E. (1989). Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation. Proc. Natl. Acad. Sci. U S A 86, 9558–9562.

Cornélle, S., Hoebeke, J., Schacht, A.M., Bertin, B., Vicogne, J., Capron, M., and Riveau, G. (2004). Direct evidence that toll-like receptor 9 (TLR9) functionally binds plasmid DNA by specific cytosine-phosphate-guanine motif recognition. J. Biol. Chem. 279, 15124–15129.

Doyle, S.L., Jeffries, C.A., Feighery, C., and O’Neill, L.A. (2007). Signaling by Toll-like receptors 8 and 9 requires Bruton’s tyrosine kinase. J. Biol. Chem. 282, 36953–36960.

Du, X., Poltorak, A., Wei, Y., and Beutler, B. (2000). Three novel mammalian toll-like receptors: gene structure, expression, and evolution. Eur. Cytokine Netw. 11, 362–371.

El-Omar, E.M., Ng, M.T., and Hold, G.L. (2008). Polymorphisms in Toll-like receptor genes and risk of cancer. Oncogene 27, 244–252.

Epstein, M.A., Aichang, B.G., and Barr, Y.M. (1944). Virus particles in cultured lymphoblasts from Burkitt’s lymphoma. Lancet 17, 702–703.

Fathallah, I., Parroche, P., Gruffat, H., Zannetti, C., Johansson, H., Yue, J., Manet, E., Tommasino, M., Sylia, B.S., and Hasan, U.A. (2010). EBV latent membrane protein 1 is a negative regulator of TLR9. J. Immunol. 185, 6439–6447.

Frappier, L., and O’Donnell, M. (1991). Overproduction, purification, and characterization of EBNA1, the origin binding protein of Epstein-Barr virus. J. Biol. Chem. 266, 7819–7826.

van Gent, M., Griffin, B.D., Berkhoff, E.G., van Leeuwen, D., Boer, I.G., Buissen, M., Hartgers, F.C., Burmeister, W.P., Wiertz, E.J., and Ressing, M.E. (2011). EBV lytic-phase protein BGLF5 contributes to TLR9 downregulation during productive infection. J. Immunol. 186, 1694–1702.

Georgeil, P., Macquin, C., and Bahram, S. (2009). The heterogeneous allelic repertoire of human toll-like receptor (TLR) genes. PLoS One 4, e7803.
Guthrie, R.D. (1996). The mammal steppe and the origin of mongoloids and their dispersal. In Prehistoric Mongoloid Dispersals, T. Akazawa and E.J.E. Szathmary, eds. [Oxford University Press], pp. 172–186.

Halder, S., Murakami, M., Verma, S.C., Kumar, P., Yi, F., and Robertson, E.S. (2009). Early events associated with infection of Epstein-barr virus infection of primary B-cells. PLoS One 4, e7214–e7229.

Hammerschmidt, W., and Sugden, B. (1989). Genetic analysis of immortalizing functions of Epstein-Barr virus in human B-lymphocytes. Nature 340, 393–397.

Holm, K.L., Sylljüsen, R.G., Hasvold, G., Alseë, L., Nilsen, H., Ivanaukiené, K., Collins, P., Shaposhnikov, S., Collins, A., Indreivu, R.L., et al. (2017). TL9R stimulation of B-cells induces transcription of p53 and prevents spontaneous and irradiation-induced cell death independent of DNA damage responses: implications for Common variable immunodeficiency. PLoS One 12, e0185708.

Houldcroft, C.J., and Kellam, P. (2015). Host genetics of Epstein-Barr virus infection, latency and disease. Rev. Med. Virol. 25, 71–84.

Jimenez-Peitner, M., Berney, X., Uldry, M., Vitali, A., Cinti, S., Ledford, J.G., and Thorens, B. (2011). Plac8 is an inducer of C/EBPα required for Brown adipocyte differentiation. Nat. Genet. 43, 658–670.

Jordi, M., Marty, J., Mordasini, V., Lüinemann, A., McComb, S., Bernasconi, M., and Nadal, D. (2017). IRAK4 is essential for TL9R-induced suppression of Epstein-Barr virus BZLF1 transcription in Alata Burkitt’s lymphoma cells. PLoS One 12, e0186614.

Kataki, A.C., Simons, M.J., Das, A.K., Sharma, K., and Mehra, N.K. (2011). Nasopharyngeal carcinoma in the Northeastern states of India. Chin. J. Cancer 30, 106–113.

Kazak, L., Chouchani, E.T., Stavrovskaya, I.G., Lu, G.V., Jedrychowski, M.P., Egan, D.F., Kumor, M., Kong, X., Erickson, B.K., Szpkyt, J., et al. (2017). UCPI deficiency causes brown fat respiratory chain depletion and sensitizes mitochondria to calcium overload-induced dysfunction. Proc. Natl. Acad. Sci. U.S.A. 114, 7981–7986.

Klein, E., Ka, L.L., and Klein, O. (2007). Epstein-Barr virus infection in humans: from harmless to life endangering virus-lymphocyte interactions. Oncogene 26, 1297–1305.

Koberlin, M.S., Heinz, L.X., and Superti-Furga, G. (2016). Functional crosstalk between membrane lipids and TL9R biology. Curr. Opin. Cell Biol. 39, 28–36.

Kostjuk, S., Loseva, P., Chvartatskaya, O., Ershova, E., Sminova, T., Malinovskaya, E., Roginko, O., Kuzmin, V., Izhevskaya, V., Baranova, A., et al. (2012). Extracellular GC-rich DNA activates TL9R- and NF-kB-dependent signaling pathways in human adipose-derived mesenchymal stem cells (hMSCs). Expert Opin. Biol. Ther. 12 (Suppl 1), S99–S111.

Krieg, A.M. (2007). Antifective applications of toll-like receptor 9 agonists. Proc. Am. Thorac. Soc. 4, 289–294.

Kumagai, Y., Takeuchi, O., and Akira, S. (2008). TL9R as a key receptor for the recognition of DNA. Adv. Drug Deliv. Rev. 60, 795–804.

Li, Z., and Zong, Y.S. (2014). Review of the histological classification of nasopharyngeal carcinoma. J. Nasopharyng Carcinoma 1, e15.

Li, J.X., Lu, T.X., Huang, Y., and Han, F. (2012). Clinical characteristics of recurrent nasopharyngeal carcinoma in high-incidence area. Sci. World J. 2012, 8.

Lung, M.L., Cheung, A.K., Ko, J.M., Lung, H.L., Cheng, Y., and Dai, W. (2014). The interplay of host genetic factors and Epstein-Barr virus in the development of nasopharyngeal carcinoma. Chin J. Cancer 33, 556–568.

Mayer, J., Woo, B.J., Liu, B., Van Dis, E., and Barton, G.M. (2018). An essential checkpoint for TL9R signaling is release from Unc93b1 in the development of nasopharyngeal carcinoma. Cytokine Res. 38, 467–484.

Martinez-Campos, C., Burguete-García, A.I., and Madrid-Marina, V. (2017). Role of TL9R in oncogenic virus-produced cancer. Viral Immunol. 30, 98–105.

Medvedev, A.E. (2013). Toll-like receptor polymorphisms, inflammatory and infectious diseases, allergies, and cancer. J. Interferon. Cytokine Res. 33, 467–484.

Montagú, A. (1951). An Introduction to Physical Anthropology: A Revised, Second Edition (Springfield).

Mouchess, M.L., Arpaia, N., Souza, G., Barbalat, R., Ewald, S.E., Lau, L., and Barton, G.M. (2011). Transmembrane mutations in Toll-like Receptor 9 bypass the requirement for ectodomain proteolysis and induce fatal inflammation. Immunity 35, 721–732.

Mutiwirii, G. (2012). TL9R agonists: immune mechanisms and therapeutic potential in domestic animals. Vet. Immunol. Immunopathol. 148, 85–89.

Nagaiyothi, F., Desruisseaux, M.S., Machado, F.S., Upadhya, R., Zhao, D., Schwartz, G.J., Teixeira, M.M., Albanese, C., Lisanti, M.P., Chua, S.C., Jr., et al. (2012). Response of adipose tissue to Toll-like receptor 9 signaling in obesity. J. Immunol. 188, 205, 830–840.

Ohito, U., Shibata, T., Tanji, H., Ishida, H., Krayukhina, E., Uchiyama, S., Miyake, K., and Shimizu, T. (2015). Structural basis of GpG and inhibitory DNA recognition by Toll-like receptor 9. Nature 520, 702–705.

Pine, S.O., McElrath, M.J., and Bochud, P.Y. (2009). Polymorphisms in toll-like receptor 4 and toll-like receptor 9 influence viral load in a seroconversion cohort of HIV-1-infected individuals. AIDS 23, 2387–2395.

Rahman, S., Sheering, M., Ogden, N.H., Lindsay, R., and Badawi, A. (2016). Toll-like receptor cascade and gene polymorphism in host–pathogen interaction in Lyme disease. J. Inflamm. Res. 9, 91–102.

Takeshita, F., Leifer, C.A., Gursel, I., Ishii, K.J., Takeshita, S., Gursel, M., and Klimin, D.M. (2001). Cutting edge: role of Toll-like receptor 9 in CpG DNA-induced activation of human cells. J. Immunol. 167, 3555–3558.

Talukder, A.H., Bao, M., Kim, T.W., Facchinetti, V., Hanabuchi, S., Bover, L., Zel, T., and Liu, Y.J. (2012). Phospholipid Scramblase 1 regulates Toll-like receptor 9-mediated type I interferon production in plasmacytoid dendritic cells. Cell Res. 22, 1129–1139.

Talambu, W., and Janvlisiri, T. (2012). Nasopharyngeal carcinoma signaling pathway: an update on molecular biomarkers. Int. J. Cell Biol. 2012, https://doi.org/10.1155/2012/594681.

Wang, W.H., Lin, Y.C., Chen, W.C., Chen, M.F., Chen, C.C., and Lee, K.F. (2012). Detection of mucosal recurrent nasopharyngeal carcinomas after radiotherapy with narrow-band imaging endoscopy. Int. J. Radiat. Oncol. Biol. Phys. 83, 1213–1219.

Whee, J.T., Ha, T.C., Loong, S.L., and Qian, C.N. (2010). Is nasopharyngeal cancer really a “Cantonese cancer”? Chin J. Cancer 29, 517–526.

Wei, W.L., and Sham, J.S. (2005). Nasopharyngeal carcinoma. Lancet 365, 2041–2054.

Wei, K.R., Xu, Y., Liu, J., Zhang, W.J., and Liang, Z.H. (2011). Histopathological classification of nasopharyngeal carcinoma. Asian Pac. J Cancer Prev. 12, 1141–1147.

Xu, F.H., Xiong, D., Xu, Y.F., Cao, S.M., Yue, W.Q., Qin, H.D., Liu, W.S., Cao, J.Y., Zhang, Y., Feng, Q.S., et al. (2012). An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation. J. Natl. Cancer Inst. 104, 1396–1410.

Yajuan, L., Yuelong, L., Xiaocong, C., Xiangyu, J., and Tengchuan, J. (2017). Pattern recognition receptors in zebrafish provide functional and evolutionary insight into innate immune signaling pathways. Cell. Mol. Immunol. 14, 80–89.

Young, L.S., Yap, L.F., and Murray, P.G. (2016). Epstein–Barr virus: more than 50 years old and still providing surprises. Nat. Rev. Cancer 16, 789–802.
Supplemental Information

**TLR9 Polymorphisms Might Contribute to the Ethnicity Bias for EBV-Infected Nasopharyngeal Carcinoma**

Nabanita Roy Chattopadhyay, Koustav Chatterjee, Nikhil Tiwari, Sudipta Chakrabarti, Sushil Kumar Sahu, Sankar Deb Roy, Arijit Ghosh, R. Rajendra Reddy, Piyanki Das, Sudipa Mal, Basab Bijay Karnar, Ashok Kumar Das, Sam Tsering, Komri Riba, Zoreng puii, Eric Zomawia, Y. Indibar Singh, Amol Ratnakar Suryawanshi, Abhishek Kumar, Dipyaman Ganguly, Chandan Goswami, and Tathagata Choudhuri
**Supplementary Information**

Table S1: Alignment of reference TLR9 gene done with TLR9 sequences retrieved from EBV-infected NPC patients. Related to Figures 1c, 3a, 5, 6, and 7.

| Sl. No. | Sequence |
|--------|----------|
| 1      | ACGACCAMATAGAGGCGGCTTGAGMCKCCTCCGCCGGGGGGAGAGYTAAAGCMCCCACCCCAGMGC |
| 2      | CSGTMSCCTTCCTGKMGGGGGCGGGAANMTGGGCAGATGKAAGTAGGCTCCTGCCCTACGATGCTTCTGGTGCTCWTARAACAAAAAMGCAAGGCGCAKTGGGCAGACTGGGTGTGATCAAACGAACTTAAAAAAAAARG |
| 3      | TCSKKGCTMMTGGGCGGCGGAAGTGAGGAGGAGTGCCTGCCCTACGATGCTTCGTTCTCGAMAAACGCACAGGCAGCAGTGGCCAGACTGGGTGTTTACCAACAGAGCTTTYYAGGGT |
| 4      | GMSSKCGCTYMTKGGGCGGCGGCGGCGAGAGTTGGGGGAGATGAGGATGCACTGCAGTGGGAGACTGGGTGTACAACGCTTTTAARGCCTTGCTCCTACGATGTGCTCCTTTCGAGCCAGACGAGACTGGGTGTTTACCAACAGAGCTTTYYAGG |
| 5      | CMYKKRSSTTTYWGGCTGAGACMGCAAGGCAATGGGCATGAGCTGAGGCGCAGCAGTCTCCATCTGCCCAACTTGGGCCGCCCCGCGARGGGAAGCGGCAAGGCAAGGCAAGGTGGGAGACTGGGTGTTTACCAACAGAGCTTTYYAGG |
| 6      | GCTGYAWCCTCCCKRRWRRGGGGCGGCAAGTGGGCGAGATGAGGATGCCCTGCCCTACGATGGCTTCGTTCTTTCGACAAAACGCAGAGCGGAGTGGCAGACTGGGTGTACAACGAGCTTYAAAAGTGCTT |
| 7      | GWTTCGYTTMTTETTWTMTMMCAKAMCACGAAGGCATCGTAGGGCAGGGCATCCTCATCTCGCCCACTTTGCCGCCCCCGGAAAGGGAGGCCAGGCCAGGCACAGGTGGAAGCAGTACCAGAGGTTTCCAAARG |
| 8      | TCCCKGGCCTMCTGGCGGCGGCGGCAATGGGGAGATGAGGATGCCCTGCCCTACGATGCTGTTGCTTTRAMAAAACGCAGACGAGCTGAGACTGGGTGTTTACCAACAGAGCTTTYYAGG |
| 9      | GTWAAGGGGGCARGGGCATCCTCATCTCGCCCACTTTGCCGCCCCCGCAGGGAAGMSAGGCCAGGCCAGGCACAGGTGGAAAGCAGACGAGCTTCCCGCTTCCTG |
| 10     | TCGKCTCTGGCCGGGCGCCTGGCCGAGATGAGGATGCCCTGCCCTACGATGCTGTTGACAAAAACGCAGACGAGCTGAGACGAGCTTCCCGCTTCCTG |
| 11     | ASGTATAGCCCTATWTCGAGCTACGKATRRKGAKKAKAGKTTTCCCCCCGCGGTGCTGAGGTAAGAWATMGCAACAGCSWSGTYACAAGACGKACTGCTGRGGCRTCCTGATCTCCGCCCCACTCTTGCGCCGCCCCGCAAGGAAGMSAGGCCAGGCACAGGTGGAAGCAGACGAGCTACAGAGGYCTT |
| 12     | TASSAAATGGCGGCAATACARAAGGTAAGTGACGCCTCCGCTCTGCTGCTCTTAAAA MGCCGAGATGCTACGAGTCGAGGTTGAGAAGAACATCTTCCCTGCTCGCCCTTCCTTACAGATG TCTTCTKMAGAGCCACCCTTTMACTAGGTGTTAAAYATTATAGGTGCTACTACATAAAGGTTAGCAGAGGCTT |
| 13     | TTTCATTTTTCTCGCMMCBCCTTTGCCGGGGCSCWTKRAARCCAGGCAAGGCAAGGCAAGGCTTGGAAAGCA GTACCAGAGGTTTTYYWWTG |
| 14     | }
|    | Sequence             |
|----|---------------------|
| 35 | TTTGCGCCCCCGAARGGAAGMSRGGCCAGGCACAGGTGGAAAGCAGTACCAGAGGTCCCTAAKRGGGC |
| 36 | GYCKKGCCCTTMMCTGGCGGGGGCGGCAAGGTGGGCGAGATGAGGATGCCCTGCCCCTACGATGCCTTCGTGGTCTTGAMAAAAAMGSAGAGCGCAGTGGCAGACTGGGTGTAACAACAGAGCTTYTAGGGGT |
| 37 | TMMWYKKSSGTYMKYRAAGACCAGAAGGCATCGTAGGCGAGGGCATCTYATCTGCCTACCTTTGCCGCCCCCGAAGGGAAGCMAGGCCAGCACAGTGGAAGCAGTACCAGAGGTCCYATAGGGG |
|    | TWSSRWMGGTCYTKRYYGAAGACCAGAAGGCATCGTAGGCAGGGGATTCCTCATCTGCCCACGGGCAGGGAAGCCGGCAGGACAGTTGGAAAAAGCAGTACCAGAGGTCCYATAAAGGG |
Table S2: The probable mutation hotspot. Related to Figures 2c, 2e, 3a, 5, 6, and 7.

TLR9 reference

5.084KB
>gi|224589815:c52260179-52255096 Homo sapiens chromosome 3, GRCh37.p5 Primary Assembly
1. GGAGGCTCTGTGTTTCCGGAAGATGTTGCAAGGCTGTGGTGAAGGCAGGTGCAGCCTAGCCTCCTGCAAG
2. CTACACCTTGGCCCTCCACCGCATAGGCCCTGAGAATTCTGAGATGTTCCTACAAAGGCGAAAGAAAAAGG
3. ACAATGGCAGACCGAGCTGCTGGTGAGGGCCATTGGTGAGAGGCAAGACCTTACGGGAGGTGAAAAGG
4. CCTTTTGAATGGGAGATGCCAGGCTCTGATCTGGGCTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
5. TTGCTCGTCTGAGCTGTGGGGCTTAGGAGCTGCTGCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTG
6. GGGATGTAGGCTGTCTGAGAGGGGAGTGGAAAGAGGAAGGGGTGAAGGAGCTGTCTGCCATTTGACTATG
7. CAAATGGCCTTTGACTCATGGGACCCTGTCCTCCTCACTGGGGGCAGGGTGGAGTGGAGGGGGAGCTACT
8. AGGCTGGTATAAAAATCTTACTTCCTCTATTCTCTGAGCCGCTGCTGCCCCCTGTGGGAAGGGAGCTCGAG
9. TGTGAAGCATCCTTCCCTGTAGCTGCTGTCCAGTCTGCCCGCCAGACCCTCTGGAGAAGCCCCTGCCCCC
10. CAGCAGTATGACAAGCTCTCAGTCCCCTGGGGAAGGGGATATCCTGGGAAGGGGTGGGGAAAGAAG
11. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
12. TGAAGACTTCAGAGCCGAAGCCCTCTTCTTTTTTCTTTTTTTTTTTTTTTGAGACGGAGTTTCGCTCTTG
13. TTGCCCAAGCTGGTCAAGCTGGTTCTACATGCTAATCTCCCTCACTCTCCCTCCTCGGCTGCTGCTGCTG
14. TCTTGCTCAGCCTCGCTCCCAAGTAGCTGGGATTATGGCTTGCGCCACCATGCCTGGCTAATTTTGTATTTT
15. TAGTAAAGCCAAGGTTTNTTTTCCTCATGTGTCAGCTGTCTGAGCTGACATCGCTGTGGCTGTTCTGAGC
16. TTCGGCTTCCCAAACTGCTGGGATTACAGGTGTGAGCCACCGTGCCCAGCCACTGAAGCCCTCTGCTGCT
17. CAATGAGACCCAAACCCCCAGAGGGGACGAGGAGACTGAGGCACTGATGGAGCTGGGGCTGGAGCCTGGG
18. TTTGGGCTCTACTGGGCGCTGTAGCTTGGGGCTGGCTCTGTCTGGCACTTTGCAGGCCACCCCTCTCTCCA
19. TCTGTTCTGCACCAATAGCGAAAGGCTCTCTCTACCACTGTTGGGGCTGGCTCTCTCTCTCTCTCTCTCT
20. TGGGAAGGGTTGGAAGATGCTAGAAGATGCCCATGAAGTGGAAGTGGGTGGAGGTAGAGCTGGGGGCC
21. TGACCATCGACAGCCCGCTCCCAACCCACAGGGTTTCTGCCGCAGCGCCCTGCACCCGCTGTCTCTCCTG
22. ATCTTGCGCCAAGGAAAAAGGAAGGAGGAGAAGGAGGAGAAGGAGGAGAAGGAGGAGAAGGAGGAG
23. GGGGGGAGCCACCTCTGCCGAACCTTCTGATAGAGGGATTTGGAAGTGGTAAGTCGGAGCAACCTTCTGAT
24. TGAATTTGCTAGTGCTAGTGCCTCTCTCTGTTCTGAGCAGCTCTTCTAGGAGCTGAGCTG(Adapter)
25. AGGGCCACAAACGGCTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
26. TGGGAAGGGTTGGAAGATGCTAGAAGATGCCCATGAAGTGGAAGTGGGTGGAGGTAGAGCTGGGGGCC
27. GGGCCGGAAGGAGCTGGACATGCACGGCATCTTCTTCCGCTCACTCGATGAGACCACGCTCCGGCCACTGG
28. GCCGAGATCAATGTGGCTCGCTCAGCTCAACTGGAGGTGCTGGACATGAGAGGAGGAGGAGGAGGAGG
29. GGCCTGCGCTACGTGGACCTGTCGGACAACCGCATCAGCGGAGCTTCGGAGCTGACAAG
30. GCCACCATGGGGGAGGCAGATGGAGGGGAGAAGGTCTGGCTGCAGCCTGGGGACCTTGCTCCGGCCCCAG
31. TGGTCATGGTGTGGGCAAGGATGGGAATGGGGCTTGGCAGCCAGGAAGAACTTCTGCAGGTAGGGCT
32. TGGAGAGAGGGGTTGGAAGATGCTAGAAGATGCCCATGAAGTGGAAGTGGGTGGAGGTAGAGCTGGGGGCC
33. GTGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
34. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
35. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
36. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
37. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
38. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
39. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
40. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
41. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
42. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
43. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
44. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
45. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
46. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
47. TGGGCTTGGCAGCTGCAAGCAACAGTGACGGGTTGTCACCCACATGGCCCTGGGGACACTCACTGAATCC
The probable mutation hotspot.

"GGCCCAGCCGCGCCAGCTCCGGCAAGCTCGGCGCTGGGC"
Table S3: List of primers used for PCR. Related to Figures 3c, 4b, 4c, and Table 1.

| For Mutation screening | T.C_TLR9_Exon2.1_ LEFT PRIMER | ATGAAGTGGAGTGGGTGGAG | GAGTGCACAGGGTGGTGAAGGT |
|------------------------|--------------------------------|-----------------------|-------------------------|
|                        | T.C_TLR9_Exon2.1_ RIGHT PRIMER |                       |                         |
|                        | T.C_TLR9_Exon2.2_ LEFT PRIMER  | TTCATGGACGGCACTGTATA  | GGGCCTGGTTGATGAAAGTG   |
|                        | T.C_TLR9_Exon2.2_ RIGHT PRIMER |                       |                         |
|                        | T.C_TLR9_Exon2.3 LEFT PRIMER   | TCTTTCTCGCTCACTCGAT   | CTCAGGCCTTGGGAAGAAGTG  |
|                        | T.C_TLR9_Exon2.3 RIGHT PRIMER  |                       |                         |
|                        | T.C_TLR9_Exon2.4 LEFT PRIMER   | CGAGGGGAGACCTCTATCTGC | CACAGGTGGAAGCAGTACCA   |
|                        | T.C_TLR9_Exon2.4 RIGHT PRIMER  |                       |                         |
|                        | T.C_TLR9_Exon2.5 LEFT PRIMER   | CCCTCTCCTGGGACTGTTC   | ATTCTCCCTCCTCCCTCCT    |
|                        | T.C_TLR9_Exon2.5 RIGHT PRIMER  |                       |                         |
| For Real-time RT-PCR   | MyD88 Forward:                 | GGTTGGTGTAGTCGAGACA   | GTCCTCCCTACACTCCTCCT   |
|                        | Reverse:                       |                       |                         |
|                        | IRAK4 Forward:                 | CAACATATGTCGCTGCTTC  | GACTTGAGGAGTGAGTGAG    |
|                        | Reverse:                       |                       |                         |
|                        | IRAK1 Forward:                 | CTCTCCCGAGCTTTTCCAGG  | ACACCCTGTTCCTCACTACC   |
|                        | Reverse:                       |                       |                         |
|                        | TLR9 Forward:                  | GTGTACAACGAGCTTGGGG   | GAGGCCCACAGGTCTCAAA    |
|                        | Reverse:                       |                       |                         |
Figure S1

(a) Risk factors for NPC

(b) Symptoms and Combined symptoms

(c) Types of EBV infection

(d) Total=70

Legend:
- Stage IV
- Stage III
- Stage II
- Stage I
- Disputed
- Type-1
- Type-2
- Both types
Figure S1. Risk factors, symptoms, and stages of NPC in NE Indian populations. Related to Figure 1. The risk factors of NPC includes genetic factors; diet and other environmental factors; and infection with HPV (Human Papilloma virus) or EBV (Epstein-Barr virus). A few dietary habits and lifestyles with poor hygiene are known to be associated with NPC in endemic populations like those in specific NE Indian populations. Most people from these populations, colloquially called ‘Mongoloids’, consume some kinds of preserved foods, live in poor hygienic conditions like a long time in smoky environments, inhaling various toxic fumes like that of mosquito repellents, habits of smoking etc. (a). Symptoms include headache, nasal obstruction, nose bleeding, deafness, loss of vision etc. (b); of which, headache is found in most patients and in earlier stages (c). Type 1 EBV is present in most of the cases, type 2 is much less common, and only a few have both types of infection (d).
Transparent Methods

Patients’ history

The study was carried out in collaboration with Civil Hospital Dimapur, Nagaland, RIIMS, Imphal, Manipur and other NE hospitals of India. Seventy freshly diagnosed patients with NPC and seventy age and sex matched controls were taken for this study. Routine histopathological analysis was done to confirm the diagnosis. Detailed questionnaire was completed, in which patients were interviewed particularly for dietary habits, tobacco consumption and type of dwelling unit. Detailed evaluation of clinico-pathological parameters and identifiable clinical variables such as disease severity, age of onset and initial clinical manifestation was obtained from patient’s clinical records, operative notes and pathologic reports. Informed consent was obtained from all subjects as per the guidelines of research review committee. Approval was obtained from the Institutional Medical Ethical Committees of the participating centers for the study.

Sample Collection

Peripheral blood sample was taken from each patient prior to treatment for genomic DNA extraction. Tumor tissue sections were obtained from paraffin block and were used for histopathological confirmation of diagnosis. For control group, peripheral blood samples were collected from healthy age and sex matched volunteers from same ethnic background.

Histology

Paraffin sections (5 μm) from samples of nasopharyngeal carcinoma specimens were deparaffinized in 100% xylene and re-hydrated in descending ethanol series(100%, 90%,80%,
70% ethanol) and water according to standard protocols. Then the sections were stained by hematoxylin and eosin. The slides were examined under the microscope.

**Processing of Samples**

Genomic DNA was extracted and purified from each blood sample using commercially available kits following manufacturer’s protocol (The GenElute™ Blood Genomic DNA Kit, Sigma-Aldrich). DNA samples were used immediately after isolation or stored at -80°C for further analysis.

**PCR based Detection of EBV**

PCR amplifications of EBNA1 and EBNA2 genes were done for each sample. Real-time RT PCR method was applied in both cases and control samples to standardize the optimum PCR conditions for detection of EBV. Primers used in the study are as follows: EBNA-1F: 5’-TGAATACCACCAAGAAGGTG-3’ and EBNA-1R: 5’-AGTTCCCTTCGTCGGTAGTC-3’. EBNA-2F: TGGAAAACCGTCACTCTC, EBNA-2R: TAATGGCATAGGTGGAATG. The primers were synthesized from IDT.

**PCR amplification and mutation screening for TLR9 gene**

PCRs were performed in the DNA samples isolated from blood of each subject. For each sample, the PCR amplified DNA was purified and placed for sequencing. Then mutation screening was performed by Finch TV software. The standard genomic sequence of TLR9 was obtained from the published database of National Center for Biotechnology Information (NCBI) database. Primer pairs were designed for the target region using the web-based software Primer 3.0. Primers used for TLR9 amplification are 5’TCTGGAGTGACGTGGTGT3’(F)
3’CTTCCCAGGATATCCCTTC5’(R) for exon 1, and for exon 2_1
5’ATGAAGTGGAGTGGGTGGAG3’(F) 5’GAGTGACAGGTGGTGAGG3’(R), exon2_2
5’TTCATGGACGGCAACTGTTA3’(F), 5’ GGGCCTGGTATGAGTT3’(R) and exon2_3
5’TCTTCTTCCGCTCACTCGAT3’(F), 5’ CTAGGCCTTGAAGATG3’(R). The PCR
products were sequenced using Big Dye Terminator Sequencing Kits (ABI) and by loading the
samples onto an ABI Automatic Sequencer (Applied Biosystems, Foster City, CA).

Sequence retrieval and alignment

The TLR9 sequences were retrieved from Ensemble and National Centre for Biotechnology
Information (NCBI) database. The TLR9 sequence from Zebrafish was retrieved from NCBI. The
sequence alignment was done by using MUSCLE alignment software with its default values. As
a highly conserved protein, sequences for histone H4 from different species were downloaded
from the Ensembl site (http://www.ensembl.org/index.html).

Phylogenetic tree formation

We used MUSCLE alignment program to align the amino acid sequences of TLR9 for the
purpose of phylogenetic analysis. We implemented a Bayesian phylogenetic tree constructed by
the Bayesian approach (5 runs, 7500,000 generations, 25% burnin-period, WAG matrix-based
model in the MrBayes 3.2 program. Fragmentation of TLR9 in different domains and motifs was
done. To analyze the conservation of different small regions of the TLR9 that are important
structurally and/or functionally, different domains and motifs were characterized before we
analyzed those separately. MUSCLE software was used to align and find out the respective
regions present in other species. The aligned data were subsequently imported into R statistical
tool for statistical analysis. As the complete TLR9 sequences from certain species are not
available (mostly due to sequencing errors at certain regions), the analysis aimed to understand the conservation of different domains and motifs of \textit{TLR9} were conducted with the available sequences only (Supplementary material, Table 1).

\textbf{Box-plot using distance Matrix generation and Statistical tests:}

Using the saved alignment files in MEGA6, distance matrices were generated for different aligned data sets. Using this method, pair-wise distances of any two different amino acid sequences within a group can be measured. To estimate the variance, bootstrap method was used. In substitution method, amino acid p-distance was used. In case of data gaps/data missing pair-wise deletion method was used. For each data set there will be one matrix which informs about the pair-wise distances of all sequences in a group. In the matrix window distances between each sequence with another is calculated along with overall mean distance of all sequences. Then the pair-wise distance values (generated in the distance matrix) were imported in “R” software for statistical analysis and graphical representation.

Using R, box-plots were generated to represent the evolutionary relationship of different protein sequences. The Kruskal-Wallis analysis of variance test was done for each set of data to check the reliability and significance of the data points. As we have measured the pair-wise evolutionary distances of protein sequences, the graphical representation reflect values in the Y-axis which is inversely proportional with the conservation. Therefore, the conserved sequences show lower values and divergent sequences show higher values in the Y-axis. Along with this calculation, the median values of each data set were calculated and also represented along with conservation figures. The distances generated were imported to Graphpad Prism and boxplot were generated for different domains.
**Immunohistochemistry**

Thin paraffin sections (5 μm) from samples of nasopharyngeal carcinoma specimens were deparaffinized in 100% xylene and re-hydrated in descending ethanol series (100%, 90%, 80%, 70% ethanol) and water according to standard protocols. For increased specificity and sensitivity, tissues have been warmed in microwave at 95 °C for 15 min to retrieve the antigen. After cooling and rinsing in distilled water, endogenous peroxide activity has been blocked in each with 3% H₂O₂ for 15 min, after which samples were rinsed in 0.01 mol/L phosphate-buffered saline (PBS), pH 7.4 for 10 mins. After blocking of any unwanted proteins, primary antibodies (mouse monoclonal anti-human TLR9) has been applied in proper dilution as per manufacturer’s protocol. After secondary anti-mouse biotinylated antibody treatment, antigen-antibody complexes have been detected using the streptavidin-peroxidase method with diaminobenzidine (DAB) as the chromogenic substrate. The sections were stained lightly with hematoxylin, and PBS has been used as negative control. Then the sections were examined under the microscope.

**siRNA treatment**

Both Raji and Ramos cells were transfected with siRNA against TLR9 as per manufacturer’s protocol. siRNA against GAPDH was used as control and both the siRNAs were gifted kindly by Dr. Dipyaman Ganguly, Head, Dendritic Cell Laboratory, IICB, Kolkata and were manufactured by Santa Cruz Biotechnology, Inc.

**Immunofluorescence and colocalization**

siRNA treated cells and control cells were washed with PBS and pelleted by centrifugation. Cells were then spreaded on clean glass slides and air dried. After acetone-methanol treatment and blocking, each slide was incubated with appropriate antibody. After incubation with appropriate
secondary antibody and addition of antifed, ceels were visualized under microscope. Merging was done by in-built software. Alexa Fluor 488 was used against TLR9 and PE was used against MyD88. All antibodies were procured from Santa Cruz Biotechnology, Inc.

**PCR based Detection of various gens**

PCR amplifications of TLR9, MyD88, IRAK1, IRAK4, and TRAF3 genes were done for 20 patient samples and 20 controls, chosen randomly. Real-time RT PCR method was applied in both cases and control samples to standardize the optimum PCR conditions for detection of each gene. Primers used in the study are listed in Table 3 of Supplementary material. The primers were synthesized from Eurofins Genomics India Pvt. Ltd.

**Western blot analyses**

Presence of various proteins checked in the current study is done by standard western blotting techniques. Total protein has been extracted from each cell culture set or tissue sample provided. 8% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) electrophoresis, followed by transfer to PVDF membrane, has been done and the membrane has been incubated with different monoclonal primary antibodies overnight at 4°C. After washing, membranes have been incubated with appropriate polyclonal secondary antibody at room temperature for 4 h. After washing with PBS, color development has been performed with ECL. GAPDH has been used as loading control. Antibodies were purchased from Santa Cruz Biotechnology, Inc.

**Grantham Distance calculation**

The variations in amino acid sequence of *TLR9* as revealed in 1K humans genome sequence database was used for Grantham Distance calculation. All these changes in a single amino acid
coordinate were analysed and fed to Align GVGD (http://agvgd.hci.utah.edu/index.php), an online tool by IARC, WHO, which combines biophysical characteristics of amino acids and protein multiple sequence alignments to predict where amino acid substitutions in protein of interest fall in a spectrum from enriched deleterious to enriched neutral. Align-GVGD is an extension of the original Grantham differences to multiple sequence alignments and true simultaneous multiple comparisons. Thus, we obtained the GD (Grantham Deviation) score for each amino acid substitution in TLR9. The GD scores were plotted against the respective amino acid coordinates and different colour codes were given for different classifiers and were plotted.

**Structure-function analysis studies in silico**

3D-modelling study was done by the help of two different software named Phyre² (http://www.sbg.bio.ic.ac.uk/phyre2) and Raswin (http://rasmol.org/). PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) software analysis of each mutated sequence and reference TLR9 sequence has been done to check the effect of mutations on protein structures. By the help of this software, each amino of a sequence has been studied for its character (aatypes), its possibility to be present in a particular type of protein folding (psipred), and its possibility to interact with the membrane (memsat).