CHAPTER 4

MATERIALS & METHODS
4.1 Study Population

The present hospital based case-control study was undertaken to analyse association of host genetic polymorphisms with rheumatic heart disease. All the patients were recruited from the Department of Cardiology and Department of Cardiovascular and Thoracic Surgery of Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, Uttar Pradesh, India. The controls were randomly selected from a pool of healthy volunteers that visited the general health check-up centre. The required number of patients and controls were calculated using sample-size calculation program QUANTO (Appendix I). The study was approved by SGPGI Ethics Committee and conformed to the norms of World’s Association Declaration of Helsinki.

4.2 Rheumatic Heart Disease (RHD) Patients

The study subjects included echocardiographycally confirmed RHD patients and severity of valve lesions was graded according to echocardiographic findings as mild, moderate or severe. Echocardiography is known to be extremely sensitive for the detection of valve abnormalities. The clinical criteria initially described by Jones in 1944, and revised by other groups (1992; Carapetis et al. 2007) were used for the diagnosis.

Stringent exclusion and inclusion criteria were taken into consideration for recruitment of RHD patients and healthy controls. Subjects who did not fulfil the criteria were excluded from the study and remaining subjects underwent an interview based upon pre-designed proforma (Appendix II). A written informed consent was obtained from each subject.
4.3 Inclusion and Exclusion Criteria

4.3.1 Cases

4.3.1.1 Inclusion Criteria:

Echocardiographycally confirmed RHD patients

Mitral valve Stenosis/ Regurgitation

Mitral and aortal valve Stenosis/ regurgitation or Multivalvular lesion

The patients, free from congenital heart diseases

Ethnicity  (North India)

4.3.1.2 Exclusion Criteria:

Coronary artery disease

Hypertension

Diabetes mellitus

Other cardiac disorder

4.3.2 Controls:

The controls were randomly selected from a pool of healthy volunteers that visited the hospital for general health check-up. Echocardiography was performed to identify silent valve damage.

4.3.2.1 Inclusion Criteria:

Healthy individuals

Mean age & sex matched

Matched females and male ratio with cases

Ethnicity matched

No family history of rheumatic fever
4.3.2.2 Exclusion Criteria:

Any autoimmune disease
Cardiac problems
Family history of RF

4.4 Questionnaire

Complete information about medical history, complications associated with disease was collected from hospital information system (HIS). All the participants of the study underwent a standardized interview using a questionnaire on family history, demographic profile, and their lifestyle.

After obtaining the informed consent, all the individuals were personally interviewed for information on ethnicity, food habits, occupation and living status.

The questionnaire was made up of the following parts (Appendix II):

- Medical history of the participant, of family members, medication intake, and other therapies.
- Demographics, such as gender, age, place of birth, education,
- Anthropometric data, such as weight, height, for calculation of BMI.
- Dietary habits.

4.5 Population Characteristics

The study included 400 RHD cases (218 males and 182 females, mean age $35.71\pm12.68$ years) and 300 healthy controls (178 males and 122 females, mean age $37.36 \pm 13.41$ years). The mean age and gender distributions were not significantly different between cases and controls, suggesting that the frequency matching was adequate. Genomic control method ruled out the possibility of population
stratification in our study. All patients were incident cases and none of the controls had a family history of rheumatic heart disease.

4.6 Study Plan

The overview of the study plan is illustrated in Figure 4.1

![Diagram](Image)

**Figure 4.1 Schematic representations of research plan and statistical analysis**

4.6.1 Collection of Peripheral Blood

After reviewing the exclusion and inclusion criteria, medical charts, and confirmation of RHD, blood samples were collected from both patients and controls.
Five milliliters of peripheral blood was collected in a vial containing 200µl of 0.5 M EDTA (pH-8.0) (Appendix III) as anticoagulant.

4.6.2 DNA Extraction

Frozen blood sample was thawed at room temperature. High molecular weight DNA was extracted by using the salting out method (Miller et al. 1988)(Miller, Dykes et al. 1988 PMID:3344216). The following protocol was used for DNA extraction. 500 µl of EDTA blood was taken in 1.5 ml Eppendorf tube. To each tube 1 ml lysis buffer (Appendix III) was added. The contents were mixed gently by inversion for 1 minute. The tubes were spun in a refrigerated centrifuge at 11,000 rpm for 5 minutes, and the supernatant was discarded. The pellet was again re-suspended in 200 µl lysis buffer, mixed, and then centrifuged. The pellet was dissolved in 200 µl of double distilled water (DDW), mixed thoroughly with a tip and centrifuged at 14,000 rpm for 1 minute. The supernatant was discarded, 80 µl of Proteinase K buffer (Appendix III) and 10 µl of 10% SDS was added to the pellet. Frothing was done in the solution with the help of a micro tip. Then 100 µl of NaCl (5M, cold) was added in the solution. The whole solution was mixed properly by tapping and 200 µl of water was added to the above solution. To the above solution 400 µl of phenol and 100 µl chloroform (4:1) (Appendix III) was added. The tubes were inverted to mix until the contents turned milky. The tubes were centrifuged at 12,000 rpm for 10 minutes. The aqueous layer was taken out in a fresh tube. In the aqueous layer 1 ml chilled absolute alcohol was added. Mixing was done by gently inverting the tubes. The tubes were centrifuged at 14,000 rpm for 4 minutes. The supernatant was discarded and excess fluid was completely drained off. The pellet was washed in 70% ethanol by centrifuging at 14,000 rpm for 1 minute. The excess
liquid was drained off and pellet was dried at 56ºC for at least 3-4 hours. The pellet was finally dissolved in 100 µl of distilled water. The tubes were kept at 56ºC overnight to dissolve the DNA.

4.6.3 Quality of DNA

Integrity of isolated DNA was checked by electrophoresis on 1% agarose gel prepared in 1X TAE buffer containing ethidium bromide (0.5µg/ml). The high molecular weight genomic DNA appeared as a single band near the well when viewed under ultraviolet light (300 nm wavelength). The amount of isolated DNA was quantified using Nanodrop Analyzer (ND-1000) spectrophotometer (Fisher Scientific, USA) by measuring the absorbance at 260 nm and calculating the DNA concentration by using arbitrary value of 1 absorbance unit equals 50 µg cm-3 of DNA:

\[ 50 \times A_{260} = \text{concentration of DNA sample (µg cm}^-3 \text{)} \]

Purity of DNA was determined by taking the optical density of the samples at 260 nm and 280 nm. The ratio of 260/280 nm was calculated and DNA samples for which the ratio was 1.8 or above was considered for further use.

4.6.4 Storage of DNA

The DNA samples were stored at -20ºC for a short time but at -80ºC for longer duration.

4.7 Analysis of Genetic Polymorphisms

A gene or locus is said to be polymorphic if there is more than one allelic form in a population and the rarer allele occur at the frequency of ≥ 1%. The vast majority of human DNA polymorphisms can be split into two groups: those based on
nucleotide substitutions (commonly called SNPs) and those based on insertion or deletion of one or more nucleotides. Other type of polymorphism is present as repetitive sequences, which arises due to variation in the repeat number of the non-coding sequences. In the present study, genotyping of the polymorphisms was performed on DNA samples by polymerase chain reaction (PCR) and restriction fragment length polymorphism analysis (RFLP) or Taqman assay (Applied Biosystems).

### 4.7.1 Genotyping by Polymerase Chain Reaction (PCR)

PCR is an in vitro method for the enzymatic amplification of precise DNA fragment up to million copies from limited starting material. The reaction was carried out using target DNA template, a pair of primers that hybridize to opposite strands of the region of interest in the target DNA, deoxynucleotide triphosphates (dNTPs), PCR buffer containing final concentrations of 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl2 and 1.5 units Taq polymerase in multiple cycles. The primers were custom synthesized and reconstituted in sterile DDW (100 pmol/µl) and stored at -70°C.

### 4.7.2 Genotyping by Restriction Fragment Length Polymorphism (RFLP)

A restriction fragment length polymorphism (or RFLP) is a variation in the DNA sequence of a genome that can be detected by breaking the DNA into pieces with restriction enzymes and analyzing the size of the resulting fragments can be detected by either polyacrylamide gel electrophoresis or agarose gel electrophoresis. If a SNP creates new or abolishes old sites for restriction enzyme to cut the DNA, it is known as RFLP. RFLP makes use of many different restriction endonucleases and
their high affinity to unique and specific restriction sites. By performing digestion on a genomic sample and determining fragment lengths through a gel assay it is possible to ascertain whether or not the enzymes cut the expected restriction sites. Gel analysis was performed under a UV illuminator after staining with ethidium bromide. Gel documentation was done by Bio-rad Gel-doc Tm EZ imager. Table 4.2 shows the analyzed polymorphisms in the present study. As a negative control, PCR mix without DNA sample was used to ensure contamination free PCR product. Ten percent of samples from patients and controls including samples of each genotype were sequenced to evaluate the quality of genotyping which showed 100% concordance. Genotyping was performed without knowledge of the case or control status. Details of PCR reagents and conditions and genotyping method, has been given in following sections (Table 4.1).

4.7.3 Genotyping by Taqman Technique

Taqman probes were used for SNP detection using real time PCR. Taqman probes consist of a fluorophore covalently attached to the 5‘-end of the oligonucleotide probe and a quencher at the 3‘-end. The quencher molecule quenches the fluorescence emitted by the fluorophore when excited by the cycler's light source via FRET (Fluorescence Resonance Energy Transfer). As long as the fluorophore and the quencher are in proximity, quenching inhibits any fluorescence signals. Probes anneal within a DNA region amplified by a specific set of primers. As the Taq polymerase extends the primer and synthesizes the nascent strand, the 5‘ to 3‘ exonuclease activity of the polymerase degrades the probe that has annealed to the template. Degradation of the probe releases the fluorophore from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing
fluorescence of the fluorophore. Hence, fluorescence detected in the quantitative
PCR thermal cycler is directly proportional to the fluorophore released and the
amount of DNA template present in the PCR.

Each TaqMan MGB probe contains:

A reporter dye at the 5’ end of each probe

– VIC® dye is linked to the 5’ end of the Allele 1 probe.

– FAM™ dye is linked to the 5’ end of the Allele 2 probe.

• A minor groove binder (MGB) at the 3’ end of each probe.

• A nonfluorescent quencher (NFQ) at the 3’ end of each probe.

The 5’-exonuclease (TaqMan®) method was used for genotyping.

4.8 Polymorphism Analysis in Candidate Genes

4.8.1 Interleukin-6 (IL-6) -174 G/C

IL-6 -174 G/C genetic variation is localized in promoter region. Polymorphism was
determined by ARMS-PCR described by Ye et al 2001 (Ye et al. 2001).

Primer Sequence:

Outer Forward 5’- GACTTCAGCTTTACTCTTTGTCAAGACA - 3’

Outer Reverse 5’- GAATGAGCCTCAGACATCTCCAGTCCTA - 3’

Inner Forward (G allele) 5’- GCACCTTTCCCCCTAGTTGTGCTTCCG - 3’

Inner Reverse (C allele) 5’- ATTGTGCAATGTGACGTCCTTTAGCTTG - 3’

Reaction Mix (Volume 25 µl)

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin
12.5 pmol each primer
0.2 mM each dNTPs
2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)
DDW water (Made upto 25 µl)
Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

PCR conditions
Initial Denaturation: 94°C for 5 min
Denaturation: 94°C for 20 sec
Annealing: 60°C for 20 sec
Extension: 72°C for 20 sec
Final Extension: 72°C for 5 min

Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. A 325bp common product was obtained. The G allele was characterized by the presence of 205bp while the T allele by 176bp fragment. Figure 4.2 shows the representative gel picture showing all the genotypes of the IL-6 G/C polymorphism.
Materials & Methods

Figure 4.2 IL-6 -174 G/C polymorphism: Lane 1 showing GC genotype, lanes 2 showing CC genotype, and lanes 3, 4 & 5 showing GG genotype and Lane 6 showing 100bp ladder

4.8.2 Interleukin-10 (IL-10) -1082 G/A

IL-10 -1082 G/A polymorphism is a Xag I restriction fragment length polymorphism (RFLP). Flanking 377bp region was amplified using primers described previously by Settin et al; 2007 (Settin et al. 2007).

Primer Sequence:
Forward 5'- CCAAGACAACACTAAGGCTCCTT'TT - 3'
Reverse 5'- GCTTCTTATATGCTAGTCAGGTA - 3'

Reaction Mix (Volume 25 µl)
10 mM Tris-Cl (pH 8.3)
50 mM KCl
1.5 mM MgCl2
Materials & Methods

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)

DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

PCR conditions

Initial Denaturation:            95°C for 3 min

Denaturation:                       94°C for 45 sec

Annealing:                     61.6°C for 45 sec  \[30 \text{ Cycles}\]

Extension:                           72°C for 45 sec

Final Extension:                  72°C for 5 min

Genotyping

Genomic DNA was amplified in Eppendorf™ Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. The obtained 377 bp PCR product (20 µl) was digested with 10 units XagI (Fermantas INC, USA) restriction enzyme using the restriction enzyme buffer. Digestion was carried out for 6 hrs at 37°C. The digested samples were separated on 2% agarose gel (Appendix III) containing ethidium bromide (0.5 µg/ml) (Appendix III). Electrophoresis was done at 150 V (constant) for 2 -4 hrs. Gel was visualized on UV transilluminator and photographed by Bio-rad Geldoc™ Imager. The corresponding IL-10 A allele (homozygous wild)
introduced a restriction site for the \textit{Xag} I enzyme resulting in fragments of 280 bp and 97 bp whereas the G allele (homozygous variant) produces 253bp and 27bp fragments after digestion by restriction enzyme. Figure 4.3 shows the representative gel picture showing all the genotypes of the IL-10 G/A polymorphism.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4_3.png}
\caption{IL-10 -1082 G/A polymorphism: Lane 1& 2 showing AG genotype, lanes 3 showing AA genotype, lane 4 showing 100bp ladder and lanes 5 showing GG genotype}
\end{figure}

4.8.3 Interleukin-1β (IL 1β) -511 C/T rs2853550

This genetic variation is localized in promoter region of \textit{IL} 1β. A single nucleotide change from T to C produces restriction site for \textit{Ava} I. Flanking 304 bp region was amplified using primers described previously by Chou et al; 2005 (Chou \textit{et al.} 2004a).

\textbf{Primer Sequence:}

Forward 5'- TGGCATTGATCTGTTTCATC - 3'

Reverse 5'- GTTTAGGAATCTTCCACTT - 3'
**Materials & Methods**

**Reaction Mix (Volume 25 µl)**

- 10 mM Tris-Cl (pH 8.3)
- 50 mM KCl
- 1.5 mM MgCl2
- 0.01% Gelatin
- 12.5 pmol each primer
- 0.2 mM each dNTPs
- 2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)
- DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

**PCR conditions**

- **Initial Denaturation:** 95°C for 5 min
- **Denaturation:** 95°C for 30 sec
- **Annealing:** 55°C for 30 sec
- **Extension:** 72°C for 30 sec
- **Final Extension:** 72°C for 5 min

**Genotyping**

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. The obtained 304 bp PCR product (20 µl) was digested with 10 units Ava1 (Fermantas INC, USA) restriction enzyme using the restriction enzyme buffer. Digestion was carried out for overnight at 37°C.
digested samples were run on 2% agarose gel (Appendix II) containing ethidium bromide (0.5 µg/ml) (Appendix III). Electrophoresis was done at 150 V (constant) for 2-4 hrs. Gel was visualized on UV transilluminator and photographed by Bio-rad Geldoc™ Imager.

The corresponding IL 1β T allele (homozygous wild) remained undigested i.e 304bp whereas the C (homozygous variant) introduced a restriction site for the Avo1 enzyme resulting in fragments of 190bp and 114 bp (Figure 4.4).

![Image of gel electrophoresis](attachment:image.jpg)

**Figure 4.4 IL 1β -511 C/T polymorphism:** Lane 1, 2, 5 & 6 showing CT genotype, lanes 3&8 showing TT genotype, lane 4 showing 100bp ladder and lanes 7 showing CC genotype

4.8.4 Interleukin-1RN (IL-1RN) VNTR rs2234663

The IL-1RaVNTR was analyzed using PCR primers flanking the 86-bp repeat region. Allele 1 (4 repeats) will be 410 bp, allele 2 (2 repeats) 240bp, allele 3 (3 repeats) 500 bp, allele 4 (5 repeats) 325 bp, and allele 5 (6 repeats) 595 bp. Flanking
regions were amplified using primers described previously by Chou et. al; 2005 (Chou et al. 2004a).

**Primer Sequence:**

Forward 5'- CTCAGCAACACTCCTAT - 3'

Reverse 5' - TCCTGGTCTGCAGGTAA - 3'

**Reaction Mix (Volume 25 µl)**

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)

DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

**PCR conditions**

Initial Denaturation: 95°C for 5 min

Denaturation: 95°C for 30 sec

Annealing: 55°C for 30 sec \[35\text{ Cycles}\]

Extension: 72°C for 30 sec

Final Extension: 72°C for 10 min
Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. Allele A1, A2, A3, A4 and A5 were characterized by the presence of 410bp, 240bp, 500bp, 325bp, and 595bp fragment respectively. Figure 4.5 shows the representative gel picture showing all the genotypes of the IL-1RN VNTR polymorphism.

![Gel Electrophoresis Image]

**Figure 4.5** *IL-1RN* VNTR polymorphism: Lane 1 showing A1/A1 genotype, lane 2 showing A2/A4 genotype, lane 3 showing A1/A3 genotype, lane 4 showing 100bp ladder, lane 5 showing A1/A4 genotype, lane 6 showing A3/A3, lane 7 showing A1/A2 genotype and lane 8 showing A2/A3 genotype

4.8.5 Transforming Growth Factor-β1 (TGF-β1) -509 C/T rs1800469

The genetic variant of TGF-β1 loci was determined using PCR-RFLP. The TGF-β1 C/T polymorphic site containing fragment was amplified by PCR. Primers
used were as described by Peng et al; 2011(Peng et al. 2011) which amplified the 419 bp fragment. The C>T transversion at the polymorphic site of TGF-β1 gene creates a Eco81I restriction site.

**Primer sequences**

Forward 5`- CAGACTCTAGAGACTGTCAG-3'

Reverse 5` - GTCACCAGAGAAAGAGGAC-3'

**Reaction Mix (Volume 25 µl)**

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 U Taq Polymerase (3U/µl) (Fermantas INC, USA)

DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

**PCR conditions**

Initial Denaturation: 95°C for 3 min

Denaturation: 94°C for 45 sec

Annealing: 58.6°C for 45 sec

Extension: 72°C for 45 sec

Final Extension: 72°C for 5 min
Genotyping

5 µl of amplified 419 bp PCR product was checked by electrophoresis on 2% agarose gel containing ethidium bromide. 20 µl of amplified DNA was digested with 10 units Eco8I restriction enzyme (Fermantas INC, USA) restriction enzyme using the appropriate restriction enzyme buffer. Digestion was carried out overnight at 37°C in an incubator. The digested samples were separated on 15% polyacrylamide gel (Appendix III). Electrophoresis was done at 15 mA constant for 3-5 hrs in mini vertical gel electrophoresis system (Bangalore Genei, India). For visualization under UV light, gel was soaked in ethidium bromide (0.5 µg/ml) (Appendix III) and photographed by Bio-rad Gel-doc™ EZ imager. The C allele was characterized by the presence of two digested fragments of 228bp and 191 bp while the T allele remained undigested. Figure 4.6 shows the representative gel picture showing all the genotypes of the TGF-beta1 C/T polymorphism.

Figure 4.6 TGF-beta1 -509 C/T polymorphism: Lane 1 showing TT genotype, lanes 2& 3 showing CT genotype, and 4 showing CC genotype and Lane 5 showing 50bp ladder
4.8.6 Transforming Growth Factor-β1 (TGF-β1) +869 T/C rs1982073

The sequence variation in exon 1 of TGF-β1 gene i.e. +869 T/C polymorphism results in creation of restriction site, which can be detected using PstI restriction enzyme. Flanking region of polymorphism was amplified using specific primer sequences previously described by Peng et. al; 2011 (Peng et al. 2011).

Primer Sequence:
Forward 5’- ACCACACCAGCCCTGTTCGC - 3’
Reverse 5’- AGTAGCCACAGCAGCGGTAGCAGCTGC - 3’

Reaction Mix (Volume 25 µl)
10 mM Tris-Cl (pH 8.3)
50 mM KCl
1.5 mM MgCl2
0.01% Gelatin
12.5 pmol each primer
0.2 mM each dNTPs
2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)
DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquotted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

PCR conditions
Initial Denaturation: 95°C for 3 min
Denaturation: 94°C for 45 sec
Annealing: 55°C for 45 sec
30 Cycles
Extension: 72°C for 45 sec
Final Extension: 72°C for 5 min
Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. The obtained 123 bp PCR product (20 µl) was digested with 10 units Pst I (Fermantas INC, USA) restriction enzyme using the restriction enzyme buffer. Digestion was carried out for overnight at 37°C in an incubator. The digested samples were run on 15% polyacrylamide gel (Appendix III) stained in ethidium bromide (0.5 µg/ml) (Appendix III). Electrophoresis was done at 150 V (constant) for 2-4 hrs. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. The corresponding TGF-beta1 Exon 1 C allele (homozygous wild) was remained undigested whereas the T (homozygous variant) allele was characterized by the presence of two digested fragments of 89bp and 34bp. Figure 4.7 shows the representative gel picture showing all the genotypes of the TGF-beta1 T/C polymorphism.

![Figure 4.7 TGF-β1 +869 T/C polymorphism](image)

Figure 4.7 TGF-β1 +869 T/C polymorphism: Lane 1 showing CT genotype, lanes 2 showing CC genotype, lanes 3 showing 50bp ladder and Lane 4 showing TT genotype
4.8.7 Tumor Necrosis Factor (TNF-A) -308 G/A rs1800629

TNF-α -308 G/A genetic variation is localized in another promoter region polymorphism. Polymorphism was determined by ARMS-PCR described by Ye et al 2001 (Ye et al. 2001).

Primer Sequence:

Outer Forward 5'- ACCCAAAACACAGGCCTCAGGACTCAACA - 3'
Outer Reverse 5'- AGTTGGGGACACGCAAGCATGAAGGATA - 3'
Inner Forward (A allele) 5'- TGGAGGCAATAGGTTTTGAGGGGAGGA - 3'
Inner Reverse (G allele) 5'- TAGGACCCTGGAGGCTGACCCGTACC - 3'

Reaction Mix (Volume 25 µl)

10 mM Tris-Cl (pH 8.3)
50 mM KCl
1.5 mM MgCl2
0.01% Gelatin
12.5 pmol each primer
0.2 mM each dNTPs
2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)
DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

PCR conditions

Initial Denaturation: 95°C for 2 min
Denaturation: 95°C for 1 min
Annealing: 58°C for 1 min \{ 35 Cycles \}
Extension: 72°C for 1 min
Final Extension: 72°C for 2 min
Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. A 323bp common product was obtained. The G allele was characterized by the presence of 224bp while the A allele by 154bp fragment. Figure 4.8 shows the representative gel picture showing all the genotypes of the TNF-α G/A polymorphism.

![Figure 4.8 TNF-α -308 G/A polymorphism: Lane 1 showing 100bp ladder, lanes 2 & 5 showing GG genotype, lanes 3 & 4 showing GA genotype](image)

4.8.8 Cytotoxic T Lymphocyte-Associated Antigen-4 (CTLA-4) -318 C/T rs5742909

CTLA-4-318 C/T genetic variation is localized in promoter region. Polymorphism was determined by ARMS-PCR described by Balbi et. al; 2007 (Balbi et al. 2007).
Primer Sequence:

Outer Forward 5’- CAATGAAATGAATTGGACTGGATG - 3’

Outer Reverse 5’- TGCACACACAGAAGGCTCTTTGAATA - 3’

Inner Forward (C allele) 5’- CTCCACTTAGTTATCCAGATCTTC - 3’

Inner Reverse (T allele) 5’- ACTGAAGCTTCATGTTCACTCTA - 3’

Reaction Mix (Volume 25 µl)

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)

DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

PCR conditions

Initial Denaturation: 95°C for 10 min

Denaturation: 94°C for 30 sec

Annealing: 56.4°C for 30 sec

Extension: 72°C for 30 sec

Final Extension: 72°C for 7 min
Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. A 296bp common product was obtained. The C allele was characterized by the presence of 201bp while the T allele by 141bp fragment. Figure 4.9 shows the representative gel picture showing all the genotypes of the CTLA-4 C/T polymorphism.

![CTLA-4 C/T polymorphism](image)

**Figure 4.9** CTLA-4 -318 C/T polymorphism: Lane 1, 2 & 4 showing CC genotype and lanes 3 showing CT genotype, lanes 5 showing 50bp ladder

4.8.9 Angiotensin Converting Enzyme Gene Polymorphism (ACE I/D) Intersion/Deletion rs 4340

The ACE I/D polymorphism was analyzed by PCR method. The insertion/deletion (I/D) polymorphism of the ACE gene is characterized by the presence (I) or absence (D) of a 287-bp sequence within intron 16 of the ACE gene.
A PCR product of 490 bp showed the presence of the Alu insertion (I allele) while 190-bp fragment represented the absence of insertion (D allele). Because the D allele in heterozygous sample is amplified preferentially, the ID genotype can be mis-genotyped as DD. To overcome this problem, an insertion-specific primer (IS) was paired with the downstream primer to recheck the DD genotype. An additional 300-bp product confirmed the presence of I allele (Joshi et al. 2009). A 25 µl volume was used for each PCR reaction. We added 100 pM of each primer, together with 50 ng genomic DNA into PCR mix (Lucigen Econo Taq® PLUS GREEN, Master mix).

**Primer sequences**

Forward 5`- CTGGAGACCACCTCCCATCATTTCT -3`

Reverse 5`- GATGTGGCCATCACATTTGTCAGA -3`

Insertion-specific primer (IS) 5`- TGGGACCACAGCGCCCGCCACTAC -3`

**Reaction Mix (Volume 25 µl)**

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 U Taq Polymerase (3U/µl) (Fermantas INC, USA)

DDW water (Made upto 25µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

**PCR conditions**

Materials & Methods
Initial Denaturation: 95°C for 3 min

Denaturation: 94°C for 1 min

Annealing: 48°C for 45 sec

Extension: 72°C for 1 min

Final Extension: 72°C for 5 min

30 Cycles

The PCR products were separated by 2% agarose electrophoresis and visualized with UV light (BIO-RAD Gel-Doc TM EZ image, PA, USA). The PCR product was a 190-bp fragment in the absence of the insertion (D allele) and a 490-bp fragment in the presence of the insertion (I allele).

Figure 4.10 ACE I/D polymorphism: Lane 1 showing ID genotype, lane 2 showing DD genotype, lane 3 showing 100bp ladder and lane 4 showing ID genotype and lane 5 showing II genotype

4.8.10 Signal Transducers and Activators of Transcription 5B (STAT5B)

Intronic Region C/T rs6503691

This genetic variation is localized in Intronic region at chromosome 17. STAT 5B C/T polymorphism is a HaeIII restriction fragment length polymorphism
(RFLP). Flanking region was amplified using primers described previously by Peng et al.; 2012 (Peng et al. 2012).

**Primer Sequence:**

Forward 5’-TGATAGTATTCTGGAAGACAATCGG-3’

Reverse 5’-CCTCCCTACCCTTTCCCCT-3’

**Reaction Mix (Volume 25 µl)**

10 mM Tris-Cl (pH 8.3)

50 mM KCl

1.5 mM MgCl2

0.01% Gelatin

12.5 pmol each primer

0.2 mM each dNTPs

2.0 units Taq Polymerase (3U/µl) (Bangalore Genei, India)

DDW water (Made upto 25 µl)

Depending upon the number of reactions in each experiment, master mix was aliquoted in 200 µl PCR tubes and 100 ng genomic DNA of each sample was added.

**PCR conditions**

Initial Denaturation: 95°C for 5 min

Denaturation: 95°C for 30 sec

Annealing: 55°C for 30 sec

Extension: 72°C for 30 sec

Final Extension: 72°C for 10 min
Genotyping

Genomic DNA was amplified in Eppendorf TM Mastercycler ep Gradient S thermal cycler in 96-well format using specific primer sequences. The amplification was checked by electrophoresis of 5 µl amplified PCR products on 2% agarose gel (Appendix III) containing ethidium bromide. Gel was visualized on UV transilluminator and photographed by Bio-rad GeldocTM Imager. Allele A1, A2, A3, A4 and A5 were characterized by the presence of 410bp, 240bp, 500bp, 325bp, and 595bp fragment respectively. Figure 4.11 shows the representative gel picture showing all the genotypes of the STAT5B C/T polymorphism.

Figure 4.11 STAT5B C/T polymorphism: Lane 1 showing 100bp ladder, lane 2 showing TT genotype, lane 3 showing CT and lane 4 showing CC genotype

4.8.11 Signal Transducers and Activators of Transcription 3 (STAT3) 5' - Upstream Region C/G rs4796793

Samples were genotyped for the STAT3 C/G polymorphism by a custom designed Taqman based allelic discrimination assay (Applied Bio systems, Foster City, CA, U.S.A.). Assay conditions were in accordance with manufacturer’s
protocols. Fluorescence outputs were quantified in real time by using a 7500HT Fast Real Time PCR System (Applied Bio systems) (Fig.4.12)

The 10 μl PCR reaction mix contained 100 ng genomic DNA, 1X TaqMan® SNP genotyping assay custom designed (two primers and two TaqMan® MGB probes), 1X Universal PCR master mix SNP genotyping master mix and DDW. Genotyping was performed on an ABI 7500 Real Time PCR system using 96-well plates. All plates included negative controls (wells containing no DNA). The software used for amplification detection was 7500 sequence detection software (version 1.3).

The thermal cycler conditions were:

- **Hold:** 95°C for 11 min
- **Denaturation:** 95°C for 20 sec
- **Annealing:** 95°C for 3 sec
- **Extension:** 60°C for 30 sec

40 cycles

The genotype was determined by the fluorescence signal generated by PCR amplification as follows

The resulting plots are shown in the figure 4.12.
Materials & Methods

4.8.12 Protein Tyrosine Phosphatase Nonreceptor 22 (PTPN22) G/A rs2476601

Samples were genotyped for the PTPN22 G/A polymorphism by a custom designed Taqman based allelic discrimination assay (Applied Bio systems, Foster City, CA, U.S.A.). Assay conditions were in accordance with manufacturer’s protocols. Fluorescence outputs were quantified in real time by using a 7500HT Fast Real Time PCR System (Applied Bio systems). Details of assay as mentioned above in 4.8.11.

Figure 4.12 Genotyping of polymorphisms by Taqman assay

- a – Homozygosity for Allele 1
- b- Homozygosity for Allele 2
- c- Allele 1 and Allele 2 Heterozygosity
4.8.13 Protein Tyrosine Phosphatase Nonreceptor 22 (PTPN22) C/A rs1217406

Samples were genotyped for the PTPN22 G/A polymorphism by a custom designed Taqman based allelic discrimination assay (Applied Bio systems, Foster City, CA, U.S.A.). Assay conditions were in accordance with manufacturer’s protocols. Fluorescence outputs were quantified in real time by using a 7500HT Fast Real Time (Applied Bio systems) Details of assay as mentioned above in 4.8.11.

4.8.14 Protein Tyrosine Phosphatase Nonreceptor 22 (PTPN22) C/T rs3789609

Samples were genotyped for the PTPN22 G/A polymorphism by a custom designed Taqman based allelic discrimination assay (Applied Biosystems, Foster City, CA, U.S.A.). Assay conditions were in accordance with manufacturer’s protocols. Fluorescence outputs were quantified in real time by using a 7500HT Fast Real Time PCR System (Applied Bio systems). Details of assay as mentioned above in 4.8.11.
Table 4.1 Primer sequences of the polymorphisms and technique used for study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Sequence</th>
<th>Technique used</th>
<th>Restriction Enzyme</th>
<th>AP*</th>
<th>Fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inflammatory pathway genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL-6 (G/C)</td>
<td>(F) Inner primer (G allele) 5'-GCACCCCCCCCTAGTTGCTCTTCCG-3' (R) Inner primer (C allele) 5'-ATTGTGCAAATGTGACGTCTTTAGCTTG-3' (F) Outer primer 5'-GACTTCAGCTTTACTCTTTGCAAGACA-3' (R) Outer primer 5'-GAATGAGCCTCAGACATCTCCAGTCT-3'</td>
<td>ARMS PCR</td>
<td>-</td>
<td>K*= 325bp</td>
<td>K*= 325bp G=205bp C=176bp</td>
</tr>
<tr>
<td>IL-10 (G/A)</td>
<td>(F) 5'-CCAAGACAACACTAAGGTCCTCCTT-3' (R) 5'-GCTTCTTTATATGCTAGTCTCAGGTA-3'</td>
<td>PCR</td>
<td>XagI</td>
<td>377bp</td>
<td>A=280bp+97bp G=253bp</td>
</tr>
<tr>
<td>IL 1β</td>
<td>(F) 5'-TGG CAT TGA TCT GGT TCA TC- 3' (R) 5'-GTT TAG GAA TCT TCC CAC TT -3'</td>
<td>PCR</td>
<td>Ava I</td>
<td>304bp</td>
<td>T=304bp C=190bp+14bp</td>
</tr>
<tr>
<td>IL 1 VNTR</td>
<td>(F) 5'-CTC AGC AAC ACT CCT AT- 3' (R) 5'-TCC TGG TCT GCA GGT AA- 3'</td>
<td>PCR</td>
<td>-</td>
<td>-</td>
<td>86 bp repeats</td>
</tr>
<tr>
<td>TGF-beta1 (C/T)</td>
<td>(F) 5'-CAGACTCTAGAGACTGTCAG-3' (R) 5'-GTCAAGAGGAGGAC-3'</td>
<td>PCR/RFLP</td>
<td>Eco81I</td>
<td>419bp</td>
<td>C=228bp+191bp T=419bp</td>
</tr>
<tr>
<td>TGF-beta1 (T/C)</td>
<td>(F) 5'-ACCACACAGCCCCTTCCGC-3' (R) 5'-AGTAGCCACAGCAGGGTACGCAG-3'</td>
<td>PCR/RFLP</td>
<td>Pst I</td>
<td>123bp</td>
<td>C=123bp T=89bp</td>
</tr>
<tr>
<td>TNF-α (G/A)</td>
<td>(F) Inner primer (A allele) 5'-TGGAAAGCAATAGGTTTGGAGGGCAGGA-3' (R) Inner primer (G allele) 5'-TAGGACCTGGAGGCTGAACCCCGTACC-3' (F) Outer primer</td>
<td>ARMS PCR</td>
<td>-</td>
<td>K*=323bp G=224bp A=154bp</td>
<td>K*=323bp G=224bp A=154bp</td>
</tr>
<tr>
<td><strong>Materials &amp; Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CTLA-4 (C/T)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-ACCCAAACACAGGCCTCAGGACTCAACA-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)Outer primer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-AGTTGGGGACACGCAAGCATGAAGGATA-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F)Inner primer (C allele)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-CTCCACTTAGTTATCCAGATCTTC-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)Inner primer (T allele)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-ACTGAAGCTTCATGTTCACTCTA-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F)Outer primer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-CAATGAAATGAATTGGGACTGGATG-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)Inner primer (T allele)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5’-TGCAACACACAGAAGGCTCTTGGAATA-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F)Outer primer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ARMS PCR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em><em>K</em>=296bp</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C=201bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T=141bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Angiotensin converting enzyme gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACE (I/D)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F) <strong>5’CTGGAGACCACTCCCATCATTTCT-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R) <strong>5’-GATGTGGGCATCACATGTGTCAG-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IS) <strong>5’-TGGGGACACACAGGCGCGCGCCACTAC-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PCD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>-</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>287bp &amp; 190bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>I=287bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D=190bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JAK-STAT pathway</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STAT3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq man Probe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq-Man PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STAT5B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F) <strong>5’-TGATAGTATTCTGGAAGACATCGG-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R) <strong>5’-CCTCCCTTACCTTCCCTCCCT-3’</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PCR/RFLP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HaeIII</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>111bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T=111bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C=85bp+26bp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protein Tyrosine Phosphatase non receptor22</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTPN22 (G/A)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq man Probe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq-Man PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTPN22 (C/A)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq man Probe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq-Man PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTPN22 (C/T)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq man Probe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq-Man PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **AP*= Amplified Product, K*= size of the control amplicon**
4.9 Data Analysis

4.9.1 Demography

Chi-square analysis or two-sided Fisher’s exact test was used to compare the differences in demographic variables and genotype distributions of the polymorphisms between cases and controls. To calculate the chi-square for the data, a 2 x 2 table was first constructed, actual number of occurrences was placed and the expected frequencies in each cell were calculated. The expected frequency in a cell is the product of the relevant row and columns total divided by the sample size \[X = \frac{(a+b)(a+c)}{N}\]. The difference between observed (O) and expected (E) for each cell was then calculated. The chi square test was then calculated as the sum of \[(O-E)^2/E\] for all four cells:

\[\chi^2 = \Sigma \frac{(O-E)^2}{E}\].

4.9.2 Hardy-Weinberg Equilibrium (HWE)

Observed genotype frequencies for the polymorphisms in controls were examined for deviation from Hardy Weinberg equilibrium (HWE) using a goodness-of-fit \[\chi^2\] test with one degree of freedom. G.H. Hardy (English Mathematician) and W. Weinberg (German Physician) independently worked out the mathematical basis of population genetics in 1908. It describes and predicts genotype and allele frequencies in a non-evolving population. The model has five basic assumptions: 1) the population is large (i.e., there is no genetic drift); 2) there is no gene flow between populations, from migration or transfer of gametes; 3) mutations are negligible; 4) individuals are mating randomly; and 5) natural selection is not operating on the population. Under these assumptions, a population’s genotype and
allele frequencies will remain unchanged over successive generations, and the population is said to be in HWE. The Hardy-Weinberg model can also be applied to the genotype frequency of a single gene. Hardy and Weinberg showed that if the population is very large and random mating is taking place, allelic frequencies remain unchanged (or in equilibrium) over time unless some other factors intervene. Their formula predicts the expected genotype frequencies using the allele frequencies in a diploid Mendelian population.

\[ p^2 + 2pq + q^2 = 1 \]

It describes the binomial distribution of genotypes that result in a population when there are no external pressures that cause deviation from \( p^2 + 2pq + q^2 \). In this formula, \( p^2 \) corresponds to the frequency of homozygous genotype AA, \( 2pq \) to Aa and \( q^2 \) to aa. The AA, Aa, aa are the three possible genotypes for a biallelic locus. If the difference between observed and expected frequency was >0.05 using chi-square goodness of test statistics, the population for that genotype was considered in HWE.

Alleles that are causative of disease would be expected to be disproportionately over represented among cases than controls. Deviation from HWE might provide clue to the role of genes in etiology of the disease. Hardy-Weinberg equilibrium (HWE) in control population was assessed using the online test tool offered by the Institute for Human Genetics, Technical University Munich (http://ihg2.helmholtz-muenchen.de/cgibin/hw/hwa1.pl) and \( p \)-value was considered significant at <0.05 level.

### 4.9.3 Regression Analysis

The association of alleles, genotypes and haplotypes of genes were explored with RHD using case-control study. Binary logistic regression, a statistical modeling
which can be used to predict a dependent outcome variable (GS) on the basis of continuous (age) and/or categorical independent variables (allele or genotype) and to determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents and to understand the impact of covariate control variables. Confounding could arise from non-comparability between cases and controls which may bias the estimation of exposure on disease. Confounders may be age, gender which may be risk factors for disease and will always show a relationship with disease risk. Hence, statistical adjustment methods are used to reduce the bias caused by measured confounders. In the present study, unconditional univariate and multivariate logistic regression analysis was used to fit statistical models to explain predictors of RHD. Risk was expressed as odds ratios (OR) with 95% confidence intervals (CI) and a two-tailed $P$-value of $<0.05$ was considered significant. Statistical analysis was performed using SPSS software version 16.0 (SPSS, Chicago, IL, USA).

**4.9.4 Odds Ratio (Measure of Risk)**

Risk or protection for disease due to presence of particular genotype or allele was calculated using odds ratio. The odds ratio is the ratio of odds of exposure in diseased subjects to the odds of exposure in the non-diseased group. For each polymorphism, the contingency table was constructed as shown below:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Allele 1 Present</th>
<th>Allele 1 Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>A</td>
<td>B</td>
<td>A+B</td>
</tr>
<tr>
<td>Control</td>
<td>C</td>
<td>D</td>
<td>C+D</td>
</tr>
<tr>
<td>Total</td>
<td>A+C</td>
<td>B+D</td>
<td>A+B+C+D</td>
</tr>
</tbody>
</table>

The odds in the cases and controls were calculated as shown below:

Odds of Exposure in Cases = $A (A + B) / B (A + B)$
Odds of Exposure in Controls = \(\frac{C (C + D)}{D (C + D)}\)

\[
\text{Odds Ratio} = \frac{A / B}{C / D} = \frac{(AD)}{(BC)}
\]

The odds ratio takes values between zero (‘0’) and infinity (‘\(\infty\)’). One (‘1’) is the neutral value and means that there is no difference between the groups compared; close to zero or infinity means a large difference. An odds ratio larger than 1 means that group one has a larger proportion than group two; if the opposite is true the odds ratio will be smaller than one.

4.9.5 Confidence Interval for Odds Ratio

95% Confidence Interval (CI) for odds ratio was calculated as provided by (Woolf 1955)

\[
95\% \; \text{CI of } \ln (OR) = \ln (OR) \pm 1.96\left(\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}\right)^{0.5}
\]

A confidence interval (CI) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is an observed interval (i.e. it is calculated from the observations), that frequently includes the parameter of interest if the experiment is repeated.

Risk estimates were calculated for co-dominant and dominant genetic models using the most common homozygous genotype as reference. All tests were two-sided by using SPSS software version 16.0 (SPSS, Chicago, IL, USA).

4.9.6 Haplotype Construction (by using Software SNPstatwww.snpstats.in)

A haplotype (Greek- haploos=simple) is a combination of alleles at multiple linked loci that are transmitted together. A multilocus haplotype refers to the set of alleles, one from each marker, which are inherited from a single parent, from either
the mother or the father. Haplotype analysis is the study of the pattern of descent of a combination of alleles at different sites on a single chromosome (haplotype). It is used for the identification of recombination events between markers and traits during linkage studies, thereby establishing the boundaries of the location of a phenotype associated locus. Such information is very valuable for investigating the genetics behind common diseases.

4.9.7 Linkage Disequilibrium (LD)

The term linkage disequilibrium (LD) refers to a nonrandom relationship between 2 alleles that typically arises because those alleles are closely linked on a chromosome and therefore infrequently separated from one another by recombination. In this case, the frequency of each allele in the population does not allow one to predict the frequency at which they occur together. Typically one of the loci is an observed marker locus, and the other is the disease locus. One of the common measures of LD is D’ a standardized measure of LD (Risch & Merikangas 1996) which ranges between 0 and 1. When 2 loci are completely unlinked (for example, if they are on different chromosomes), D’=0. If the two loci are identical (for example, a marker locus is the disease locus), then D’=1. Values of D’ closer to 1 suggests that the marker locus is closer to the disease locus. Linkage disequilibrium between pairs of loci was calculated and P-value <0.02 was considered significant for LD after Yates correction.

4.9.8 Higher Order Gene Gene Interaction

A variety of parametric and non-parametric methods have been proposed for modeling and detecting gene-gene interaction, e.g. support-vector machines, random
forests, multi-factor dimensionality reduction (MDR), combinatorial partitioning methods, focused interaction testing framework, Classification and regression tree (CART), logic regression, and lasso regression. Despite the wealth of these approaches, none of the proposed methods is optimal for all two-locus disease models. Consequently, there is no established method for analyzing gene-gene interaction so far. Since parametric methods have problems to detect interaction in the absence of main effects and non-parametric approaches are ineffective when main effects are present, it might well be that there is no single approach appropriate for all types of biological interactions. Therefore, multi-analytical approaches are applied to elucidate the combination pattern of genetic variants in genetic predisposition to disease risk. In the present study, single locus analysis was followed by multi-analysis with the help of non-parametric methods as CART and MDR.

4.9.8.1 Multifactor Dimensionality Reduction (MDR)

Multifactor dimensionality reduction (MDR) method is non-parametric, genetic model-free method for overcoming some of the limitations of logistic regression (i.e. sample size limitations) for the detection and characterization of gene–gene interactions (Hahn et al. 2003). In MDR, multilocus genotypes are pooled into high risk and low risk groups, effectively reducing the genotype predictors from n dimensions to one dimension (i.e. constructive induction). The new one-dimensional multilocus genotype variable is evaluated for its ability to classify and predict disease status through cross-validation and permutation testing. The MDR software (version 2.0 beta8) was applied to identify high-order gene-gene interactions associated with RHD risk. In our study, the best candidate interaction
model was selected across all multilocus models that maximized testing accuracy and the cross-validation consistency (CVC). Furthermore, validation of models as effective predictors of disease status was derived empirically from 1000 permutations, which accounted for multiple comparison testing as long as the entire model fitting procedure was repeated for each randomized dataset to provide an opportunity to identify false positives. The MDR permutation results were considered to be statistically significant at the 0.05 level. All the variables identified in the best model were combined and dichotomized according to the MDR software and their ORs and 95% CIs in relation to GBC risk were calculated. Finally, combined effect of the variables in the best model by the number of risk genotypes was evaluated using logistic regression analysis.

4.9.8.2 Classification and Regression Tree (CART)

Classification and regression tree (CART) analysis was performed using the SPSS ver. 16 software to build a decision tree via recursive partitioning. For the analysis, decision tree was created by splitting a node into two child nodes repeatedly, beginning with the root node that contains the total sample. Before growing a tree, we choose measure for goodness of split using Gini criteria, by which splits were found that maximize the homogeneity of child nodes with respect to the value of the target variable. After the tree was grown to its full depth, a pruning procedure was performed to avoid over fitting the model. Finally the risk of various genotypes was evaluated by using the logistic regression analysis. The ORs and 95% CIs were adjusted for age and sex, with treating the least percentage of cases as the reference.