Chapter - 2

Materials and Methods
Chapter - 2
Materials and Methods

2.1. Area

The present study area is Mysore, an administrative district in South Indian state of Karnataka (Figure 2.1). It is bounded by Mandya district in the northeast, Chamrajanagar district in the South-East, Kodagu district in the west, Hassan district in the north and the state of Kerala in the south. The district comprises seven Taluks, of which Mysore Taluk inhabits 40% of the population (Figure 2.2).

According to the Census of India (2001), Mysore was fourth most populous district in Karnataka with a total population of 2,641,027 (Table 2.1), of which 50.9% were male and 49.1% were female. Majority of the people (62.8%) belong to the rural areas.

2.2. Subjects

Unrelated individuals more than 30 years of age with no familial history of type 2 diabetes for at least three generations and resident of Mysore district were included in the study after obtaining informed written consent (Annexure-I). The study, protocols and procedures were approved by Institutional Ethical Committee of the Anthropological Survey of India, Kolkata (Annexure-II).

A total of 420 subjects comprising 214 diabetic cases and 206 glucose tolerant (non-diabetic) controls were studied. Out of the cases, 109 were male and 105 were female and of the controls 91 were males and 115 were females.
2.3. General information

From each individual personal detail like name, gender, date of birth, caste/community, religion, nationality and occupation were collected with the use of a questionnaire (Annexure-III). Questions to obtain information about food and lifestyle habits like tobacco consumption, smoking status, and alcohol consumption status and health history were also included in the questionnaire, which was translated into the local Kannada language by a translator wherever required.

2.4. Anthropometry and physiology

The anthropometric and physiological measurements included were height vertex, body weight, waist circumference, hip circumference, systolic blood pressure and diastolic blood pressure were taken on the participants of the study, following the standard techniques (Singh and Bhasin, 1967; Weiner and Lourie, 1969).

Body mass index (BMI) was calculated with the formula $\text{BMI} = \frac{\text{body weight (kg)}}{\text{height vertex (m)}^2}$. Waist to hip ratio (WHR) was obtained by dividing waist circumference with hip circumference.

2.5. Collection of blood samples

5-10 ml of peripheral blood samples were collected in BD K$_2$EDTA vacutainer®s (BD, NJ, USA) from individuals visiting the outpatient division of JSS hospital, and public health centres in Mysore district. The samples were stored at 4°C to avoid haemolysis and cellular damage. Samples were carried to the DNA laboratory, Anthropological Survey of India, Mysore within 3 h of collection. Blood samples in vacutainers were transferred into labeled sterile polypropylene centrifuge tubes for processing and temporarily stored at 4°C in a refrigerator.
2.6. Diagnosing type 2 diabetes

To ascertain the type 2 diabetic status, WHO (2006) criteria was followed.

2.7. Estimation of plasma glucose levels

Principle: Glucose is oxidized by glucose oxidase (GOD) to produce gluconate and hydrogen peroxide. Hydrogen peroxide is then oxidatively coupled with 4 amino-antipyrene (4-AAP) and phenol in the presence of peroxidase (POD) to yield a red quinoeimine dye that is measured at 505 nm. The absorbance at this wavelength is proportional to concentration of glucose in the sample. Absorbance of the coloured solution is directly proportional to the glucose concentration, when measured at 505 nm (Beach and Turner, 1958).

Method: GOD-POD, End Point.


Reagents/Kits: ERBA LIQUIXX – stable glucose kits (Transasia, Mumbai)

Procedure: Whole blood sample was gently inverted 2-3 times and pulse centrifuged at 500 rpm. 10µl plasma was collected with a micro pipette without disturbing cells. Tests were performed after loading reagents and samples as per manufacturer’s instructions.

Known standard (positive control) was loaded with every batch of samples and calibration of the system was performed as per manufacturer’s instructions.
2.8. *DNA extraction by phenol chloroform method*

For this work, molecular grade DNase free reagents and autoclaved MilliQ water (Millipore Life Sciences, Billerica, MA, USA) were used throughout.

Principle: Phase separation by centrifugation of a mixture of the aqueous sample and organic components like phenol, chloroform and iso-amyl alcohol.

*Preparation of RBC lysis buffer (pH 8.0)*

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucrose</td>
<td>109.54 g</td>
</tr>
<tr>
<td>1M Magnesium chloride</td>
<td>5 ml</td>
</tr>
<tr>
<td>1M Tris-HCl (pH 8.0)</td>
<td>10 ml</td>
</tr>
</tbody>
</table>

Made up to 900 ml with MilliQ water, autoclaved and added 10 ml Triton X and made up to 1000 ml with MilliQ.

*Preparation of digestion Buffer*

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M Tris-HCl (pH 8.0)</td>
<td>40 ml</td>
</tr>
<tr>
<td>1M Sodium chloride</td>
<td>15 ml</td>
</tr>
<tr>
<td>0.5M EDTA (sodium salt)</td>
<td>10 ml</td>
</tr>
</tbody>
</table>

*Preparation of tris-EDTA buffer (pH 8.0)*

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M Tris-HCl (pH 8.0)</td>
<td>2 ml</td>
</tr>
<tr>
<td>0.5M EDTA</td>
<td>4 ml</td>
</tr>
</tbody>
</table>

Made up to 200 ml with MilliQ water.
Procedure

To the blood sample, 2 volumes of RBC lysis buffer was added and mixed gently by inverting the tube till the solution became transparent.

The tube was centrifuged at 2,500 rpm for 10 min to obtain a pellet. The supernatant containing haemolysed RBCs was discarded carefully.

The pallet was re-suspended in RBC lysis buffer (equal to initial blood volume) and tapped gently to disturb the pallet.

The tube was centrifuged at 2,500 rpm for 10 minutes and the supernatant was discarded to obtain clear white pallet. This procedure was repeated until the pellet was free of RBCs i.e. it became white in colour.

The pallet was disturbed thoroughly, and half the volume of initial volume of blood sample, of digestion buffer was added.

Proteinase K (Sigma Aldrich, India), was added to a final concentration of 10 µg/ml and sodium dodecyl sulphate (SDS) was added to make 1% concentration in the final solution. The whole was mixed thoroughly and gently by inverting the tube for 3-4 minutes till the solution became viscous and thereafter incubated for 3-4 h at 56°C for proper digestion of proteins.

When solution became clear, 1/4th the volume of digestion buffer, 5M sodium perchlorate was added and mixed gently for 3-4 min.

Phenol (tris saturated), chloroform and iso-amyl alcohol in 25:24:1 ratio were added in equal volume to mixture of digestion buffer and 5M NaClO₄. After mixing for 1 minute it was centrifuged at 4,000 rpm for 15 min at 4°C.
Aqueous layer was transferred carefully into another sterile polypropylene centrifuge tube using a broad mouth tip.

Equal amount of chloroform and iso-amyl alcohol in 24:1 ratio was added to the transferred supernatant and mixed gently for 3-4 min and centrifuged at 4,000 rpm for 15 min at 4°C. After centrifugation the aqueous phase was transferred to a fresh labeled tube.

Exactly double the volume of chilled absolute alcohol was added and mixed gently by inverting the tube to precipitate the DNA.

DNA lump was spooled out into a fresh labeled 1.5 ml tube to wash the DNA twice with 70% ethyl alcohol.

The pellet was dried at room temperature to ensure that whole alcohol was evaporated.

DNA pellet was dissolved in 200 µl of tris EDTA (TE) buffer; for optimum dissolution tubes were kept in water bath at 56°C for 20 to 30 min. Dissolved DNA samples were stored at -80°C.

Quantity and quality of extracted DNA was checked by spectrophotometry and gel electrophoresis.

**2.9. Determination of DNA concentration by spectrophotometry**

10 µl DNA sample was diluted to 1000 µl with MilliQ water. Auto zero was set on spectrophotometer with MilliQ water. Optical Density (OD) of the diluted DNA aliquot was measured at 260nm and 280nm using quartz crystal cuvette.
Quality assessment

A ratio of OD values at 260 nm and 280 nm indicates the purity of the extracted DNA sample. If the ratio is within a range of 1.6 to 2.0, the DNA sample is considered as clear and free from contaminants like residual protein and mRNA. An OD ratio less than 1.6 indicates the residual proteins or phenol contamination while a ratio of more than 2.0 indicates residual RNA contamination.

Quantity assessment

DNA quantity was estimated from the OD value at 260 nm of extracted sample. If it was 1.00, the concentration of the DNA in the sample was 50µg/ml. Therefore, DNA concentration = OD at 260nm x 50 x dilution factor.

2.10. DNA quality check by electrophoresis

Electrophoretic analysis of DNA using agarose gels can confirm DNA integrity. Typically, intact genomic DNA will be up to 40 kb in size, depending upon the species.

CHEMICALS AND BUFFERS USED FOR GEL ELECTROPHORESIS

Tris-acetate-EDTA (TAE) buffer 20X

Tris 48.4 g
Acetic Acid 11.402 ml
0.5M EDTA 20 ml

Dissolved in 1 l MilliQ water, pH 8.0.

To obtain 1X TAE buffer, 50ml of 20X buffer was made up to 1000ml with MilliQ water.
Chapter – 2

Materials and Methods

Loading dye (stock solution)

Bromo phenol blue 25 mg
Xylene cyanol 25 mg

Dissolved in 10 ml MilliQ water.

5 ml of 40% sucrose solution was added to
1 ml of loading dye stock to make working solution.

Ethidium bromide solution (6%)

0.6 g ethidium bromide was dissolved in

2 ml MilliQ water and made up to 10 ml.

1% agarose gel was prepared by adding required quantity of agarose to
1X Tris-Acetate-EDTA (TAE) buffer and the mixture was heated in a
microwave oven until it became clear. Care was taken to avoid over
boiling and evaporation. The mixture was cooled to ~50°C and 2 µl 6%
ethidium bromide was added. The mixture was poured into a tray in
which a comb was fixed.

After gel formation the tray was placed in buffer tank containing 1X TAE
buffer for submerged gel electrophoresis and comb was removed with
care to avoid rupture of wells. 1 µl of each DNA sample was mixed with 1
µl of loading dye and the mixture was loaded into the wells. Gel was
subjected to electrophoresis at 90 V for 30 min and DNA was visualized
using gel documentation system (Syngene, UK).

2.11. Preparation of DNA working dilutions

100 µl of DNA working dilutions were prepared at a concentration of 50
µg/µl by dissolving required amount of stock DNA sample in MilliQ
water. After this the uniformity of the samples was checked by
performing electrophoresis on a 1% agarose gel. Samples were stored at 4°C. For mitochondrial DNA quantification 5 ng/µl dilutions were prepared.

2.12. Selection of conserved regions in ADIPOQ gene

Despite tremendous progress in vertebrate genomics, it is still not clear how much of the human and other vertebrate genomes are directly functional, in the sense of encoding proteins or RNAs helping to regulate transcription and translation, enabling replication, altering chromatin structure, or performing other important cellular tasks.

Comparatively more is known about the functional roles of sequences in the genomes of model eukaryotes such as *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*, but much remains to be learned in these genomes as well.

One of the best strategies known for finding functional sequences is to look for sequences that are conserved across species. While orthologous sequences from related species might appear “conserved” (i.e. unusually similar) because of reduced mutation rates, the primary reason for cross-species sequence conservation is believed to be negative (purifying) selection.

Thus, orthologous sequences that are significantly more similar than would be expected if they were evolving under some reasonable model of neutral evolution are likely to have critical functional roles (Margulies et al., 2003). To find the conserved regions in the *ADIPOQ* gene, mostConserved28way track of UCSC Genome browser was used (Kent et al., 2002).
Description

This track shows predictions of conserved elements produced by the phast Cons program based on a whole-genome alignment of vertebrates, and for the placental mammal subset of species in the alignment. They are based on a phylogenetic hidden Markov model (phylo-HMM), a type of probabilistic model that describes both the process of DNA substitution at each site in a genome and the way this process changes from one site to the next.

Method

Best-in-genome pairwise alignments were generated for each species using blastz program (Schwartz et al., 2003), followed by chaining and netting (Kent et al., 2003). A multiple alignment was then constructed from these pairwise alignments using multiz program (Blanchette et al., 2004). Predictions of conserved elements were then obtained by running phastCons on the multiple alignments.

PhastCons constructs a two-state phylo-HMM with a state for conserved regions and a state for non-conserved regions. The two states share a single phylogenetic model, except that the branch lengths of the tree associated with the conserved state are multiplied by a constant scaling factor rho \((0 \leq \rho \geq 1)\). The free parameters of the phylo-HMM, including the scaling factor rho, are estimated from the data by maximum likelihood using an EM algorithm. This procedure is subject to certain constraints on the "coverage" of the genome by conserved elements and the "smoothness" of the conservation scores (Siepel et al., 2005).

The predicted conserved elements are segments of the alignment that are likely to have been "generated" by the conserved state of the phylo-HMM.
Each element was assigned a log-odds score equal to its log probability under the conserved model minus its log probability under the non-conserved model. The "score" field associated with this track contained transformed log-odds scores, taking values between 0 and 1000.

The scores were transformed using a monotonic function of the form $a \cdot \log(x) + b$. The raw log odds scores were retained in the "name" field and can be seen on the details page or in the browser when the track's display mode is set to "pack" or "full". Higher score implies most conserved regions. So a cut off of 200 was used to select most conserved regions in the gene. The regions with score $>200$ were selected for re-sequencing (Table 2.2).

### 2.13. Polymerase chain reaction

The polymerase chain reaction (PCR) is a revolutionary molecular biology method developed by Kary Mullis in 1980s. It is an essential and ubiquitous tool in genetics. With use of this technique DNA can be cloned \textit{in vitro}.

PCR is based on using the ability of DNA polymerase to synthesize new strand of DNA complementary to the offered template strand. Because DNA polymerase can add a nucleotide only onto a pre-existing 3'-OH group, it needs a primer to which it can add the first nucleotide. This requirement makes it possible to amplify a desirable specific region of template sequence.

At the end of the PCR, the specific sequence will be accumulated in billions of copies, referred to as amplicons. The technique of DNA sequencing by dye termination used in the present study requires multiple copies of the region of interest. For this, PCR was performed after designing appropriate primers.
Chapter – 2  Materials and Methods

Primer designing

The main limitation of PCR technique is that one has to provide short pieces of single-stranded DNA (primers) that are complementary to a part of target sequence. With the use of human genome sequence available, it is possible to design primers for any region of interest. The most critical step in PCR experiment is designing oligonucleotide primers. Poor primers may result in little or no PCR product and such primers may also amplify unwanted DNA fragments, thus affecting the downstream analysis.

Primer designing using Primer 3

Primer 3 is a widely used and freely available web based program for designing PCR primers (http://frodo.wi.mit.edu/primer3/input.htm) (Rozen and Skaletsky, 1999). This application analyses target regions and recommend forward and reverse primer sequences. Its analysis can be directed to specific target regions of genes. Many of the factors which affect the primers specificity and sensitivity such as product size, primer size, Tm, GC content, GC clamps and dimer formation, can be adjusted as per the user’s requirement in this application. In the present study, all the primers to amplify and sequence the selected regions of ADIPOQ, HHEX and KCNJ11 genes and 2 SNPs near HHEX region were designed using Primer 3.

Checking primer properties

For the proper optimization of the PCR reaction, knowing the properties of the oligoprimers is an important factor. Presence of the primer secondary structures produced by inter-molecular or intra-molecular interactions can lead to poor or no yield of the product. They greatly
reduce the availability of primers to the reaction and adversely affect primer template annealing and thus the amplification.

*Primer analysis using NetPrimer*

NetPrimer is a free web based tool that analyses primers to optimize PCR, sequencing, and hybridization reactions (http://www.Premierbiosoft.com/netprimer/index.html). Primers are analysed for amplification related properties including melting temperature (Tm) and secondary structures. NetPrimer also analyses the thermodynamically important secondary structures such as hairpins, self and cross dimers, runs and repeats.

*Primer melting temperature (T_m)*

Primer melting temperature (T_m) by definition is the temperature at which one half of the DNA duplex will dissociate to become single stranded and indicates the duplex stability. The GC content of the sequence gives a fair indication of the primer T_m.

NetPrimer calculates T_m using the nearest neighbour thermodynamic theory, accepted as a much superior method for estimating it, which is considered the most recent and best available. T_m is calculated as follows.

\[
\text{Melting Temperature } T_m (^oK) = \frac{\Delta H}{\Delta S + R \ln (C)}, \text{ or}
\]

\[
\text{Melting Temperature } T_m (^oC) = \left( \frac{\Delta H}{\Delta S + R \ln (C)} \right) - 273.15
\]

where, H is the enthalpy which is the amount of heat energy possessed by substances. \( \Delta H \) is the change in enthalpy and is obtained by adding up all the di-nucleotide pairs’ enthalpy values of each nearest neighbour base pair.
S is the amount of disorder a system exhibits and is called entropy. \( \Delta S \) is change in entropy and is obtained by adding up all the di-nucleotide pairs’ entropy values of each nearest neighbour base pair. An additional salt correction is added as the Nearest Neighbour parameters were obtained from DNA melting studies conducted in 1M Na+ buffer and this is the default condition used for all calculations.

\[
\Delta S \text{ (salt correction)} = \Delta S \text{ (1M NaCl)} + 0.368 \times N \times \ln([\text{Na}^+])
\]

where, \( N \) is the number of nucleotide pairs in the primer (primer length -1) and \([\text{Na}^+]\) is salt equivalent in mM i.e. monovalent ion concentration +4 x free Mg\(^{2+}\).

Primers with melting temperatures in the range of 52-65\(^\circ\)C were considered for this study.

**Primer annealing temperature**

The primer annealing temperature \( (T_a) \) is the estimate of the DNA-DNA hybrid stability and it is critical in determining the annealing temperature. Too high \( T_a \) will produce insufficient primer-template hybridization resulting in low PCR product yield. Too low \( T_a \) may possibly lead to non-specific products caused by a high number of base pair mismatches. Mismatch tolerance is found to have the strongest influence on PCR specificity.

\[
T_a = 0.3 \times T_m \text{ (primer)} + 0.7 \times T_m \text{ (product)} - 14.9
\]

where, \( T_m \text{ (primer)} \) = Melting temperature of the primer

\( T_m \text{ (product)} \) = Melting temperature of the product
Chapter – 2

Materials and Methods

Secondary structures

Presence of the primer secondary structures produced by inter-molecular or intra-molecular interactions can lead to poor or no yield of the product. They greatly reduces the availability of primers to the reaction thereby adversely affecting primer template annealing and thus the amplification. The strength of the primer secondary structures can be measured by the use of Gibbs free energy (G) which is a measure of the amount of work that can be extracted from a process operating at a constant pressure. It is the measure of the spontaneity of the reaction. The stability of hairpin is commonly represented by its ΔG value, the energy required to break the secondary structure. Larger negative value for ΔG indicates stable, undesirable hairpins. Presence of hairpins at the 3’ end most adversely affects the reaction.

\[ \Delta G = \Delta H - T\Delta S \]

Table 2.3 shows cut off values for the primers used in this study to ensure proper amplification of the PCR product.

Assessing specificity of the primers

The specificity of primer is very important in view of amplifying a specific gene region. If the primer is not specific for the region of interest, problem of obtaining reliable sequence data arises. To avoid this all the primers will be checked for their specificity by performing similarity search with the entire genome of the organism for which the primers are designed.

The general approach involves the use of a set of algorithms such as the BLAST programs to compare a query sequence to all the sequences in a specified database. Comparisons are made in a pair wise fashion. Each comparison is given a score reflecting the degree of similarity between the
query and the sequence being compared. The higher the score, the greater is the degree of similarity. The similarity is measured and shown by aligning two sequences.

Alignments can be global or local (algorithm specific). A global alignment is an optimal alignment that includes all characters from each sequence, whereas a local alignment is an optimal alignment that includes only the most similar local region or regions. Discriminating between real and artifactual matches is done using an estimate of probability that the match might occur by chance. Primers which fit the specified criteria were checked for their specificity using NCBI BLAST.

**BLAST**

The BLAST programs (Basic Local Alignment Search Tools) are a set of sequence that is used to search sequence databases for optimal local alignments to a query (Altschul et al., 1997).

The BLAST programs improve the overall speed of searches while retaining good sensitivity by breaking the query and database sequences into fragments, and initially seeking matches between fragments. The fragment which is having high similarity score will be extended in both the directions in an attempt to generate an alignment with a score exceeding the threshold.

All the designed primers which fit the specified criteria were tested for their similarity by performing BLAST against the *Homo sapiens* genome build 36 version 3 (36.3). Primers with multiple hits, max score less than 40 and E value less than 0.015 were not considered. The designed oligo primers with good score and e value were synthesized (Sigma-Aldrich, Bangalore, India) and standardized to obtain good amplicons. Table 2.4 lists the primers designed and used in this study.
PCR standardization

All the primers were standardized for their optimal performance after using gradient PCR technique to identify the optimal hybridization temperature for each primer. To avoid primer dimers Mg++ ion and primer concentration were kept under 50 mM and 100 nM, respectively. The PCRs were carried out as per the manufacturer’s instructions of Taq polymerase enzyme (New England Biolabs, Ipswich, MA, USA).

Amplicon check by electrophoresis

After the completion of 35 cycles of polymerase chain reaction 1 µl of the amplicon product was electrophoresed on a 2% agarose gel containing ethidium bromide. DNA bands were visualized using gel documentation system (Syngene, UK). The samples which were amplified successfully were used for the sequencing after post PCR clean up.

Post PCR clean up

Typically, the DNA fragment is purified prior to sequencing to remove proteins, salts, left out primers and dNTPs which may have detrimental effects on the sequencing reaction. A combination of exonuclease I and shrimp alkaline phosphatase enzymes (ExoI/SAP) were used to clean the PCR product.

PCR cleanup master mix was prepared by adding 5 U of Exo I and 0.5 U of SAP in final volume of 8.5 µl MilliQ water and this master mix was added to 8 µl of PCR product. Exo I/SAP enzymatic reaction was allowed to proceed by heating the samples up to 37°C for 30 min using a thermal cycler and then denaturing the enzymes by heating at 80°C for 15 min. Thereafter 0.1 volumes of 3 M sodium acetate (1.65 µl) and 2.0 volumes of chilled 100% ethanol (33 µl) were added and mixed well. Samples were centrifuged at 4,000 rpm for 30 minutes at 4°C. Ethanol was decanted
and folded paper towels were used to remove the excess ethanol by blotting the plate. 100 µl of cold 70% ethanol was added and mixture was centrifuged at 4,000 rpm for 10 min at 4°C. After decanting the ethanol PCR plate was blotted on folded paper towels to remove the excess ethanol and plate was centrifuged inversely at 180 rpm for 30 sec to remove remaining ethanol. The pellet was re-suspended in 8µl of water and stored at 4°C.

**2.14. DNA sequencing**

There are various methods available for DNA sequencing like chemical degradation, chain termination, sequencing by ligation and micro fluidic Sanger sequencing, among others. Advances in automation have opened gates to new, fast and reliable automated DNA sequencing technologies. Owing to its greater expediency and speed, the dye-terminator sequencing is now the mainstay in automated DNA sequencing.

The dye-terminator sequencing is a slight modification of the Sanger’s original chain termination method; it utilizes labeled chain terminator dideoxynucleotide (ddNTPs) which permit sequencing in a single reaction. In this method, each of the four ddNTP chain terminators are labeled with different fluorescent dyes with different wavelength of emission. The dye labeled DNA amplicons are capillary electrophoresed and a detection system identifies the labeled bases as they pass through a laser that activates the dyes.

*Cycle sequencing*

To label each DNA fragments of PCR amplicon the final product was subjected to cycle sequencing reaction with one primer using ABI Prism® BigDye™ terminator v3.1 cycle sequencing ready reaction kits (Applied Biosystems, USA) following the manufacturer’s guidelines.
Chapter – 2  Materials and Methods

Sequencing clean up

To remove the remnants of the above reaction, to each sample 2 µl of 3 M sodium acetate and 50 µl of 100% ethyl alcohol were added and incubated at room temperature for 15 min to precipitate the DNA. The samples were centrifuged at 4,000 rpm for 25 min at 4°C. The supernatant was discarded and the reaction plate was centrifuged inversely at 300 rpm for 20 sec. 100 µl of 75% alcohol was added to each sample and centrifuged at 4,000 rpm for 15 min at 4°C. The supernatant was discarded and plate was inversely centrifuged at 300 rpm for 20 sec to remove alcohol completely. The plate was dried at room temperature until left out alcohol to evaporate.

Sequencing run

10 µl of Hi-Di™ formamide was added to each well of the sample plate. The samples were heated to 96°C and immediately cooled to 4°C to denature the DNA. Sample information sheets which contain analysis protocol along with the sample details were prepared and fed to the data collection software. Prepared samples were analysed on Applied Biosystems 3730 genetic analyzer (Applied Biosystems, USA) to generate DNA sequences.

Sequence quality check

After completion of sequencing reaction, the quality of generated sequences was checked by using Sequencing Analysis 5.2 software (Applied Biosystems, USA), designed to analyse, display, edit, save, and print sample files generated from Applied Biosystems DNA analyzers and other ABI PRISM® genetic analyzers. The program has a novel basecaller algorithm that performs basecalling for pure and mixed base calls also. It provides quality values (QV) for every single base and sample scores for
the assessment of Average quality value of the bases in the clear range sequence for the sample. The QV is a per-base estimate of the basecaller accuracy. The QVs are calibrated on a scale corresponding to:

\[ QV = -10 \log_{10}(Pe) \]

where, \( Pe \) is the probability of error

For this study typical high-quality pure base QVs were set at 20 to 50 and typical high-quality mixed bases QVs were set at 10 to 50. The samples which did not followed the above conditions were re-sequenced after fresh PCR amplification.

### 2.15. Sequence alignment

The generated sequences were aligned to their respective reference sequences with the use of SeqScape v2.5 software (Applied Biosystems, USA), designed for automated sequence data analysis. It performs sequence comparisons for variant identifications, SNP discovery and validation. It allows analysis of the re-sequenced data, comparing the consensus sequences to a known reference sequence.

The reference sequences for the genes studied were obtained from NCBI Genbank data base. For \textit{ADIPOQ}, NC_000003.11 was used as reference sequence. For \textit{HHEX} and \textit{KCNJ11}, NC_000010.10 and NM_000525.3, respectively, were used as reference sequences. To set clear range of the sequence, a method that considers quality values of the bases was used, which removes bases from the ends of sequences until fewer than 4 bases out of 20 have QVs < 20. Filter setting values to filter the inappropriate sequences were set as maximum mixed bases to 20 and minimum sample score to 25. Depending on the sequence quality and the criteria specified for filtering, the samples with low quality were not
assembled by the program. These unassembled samples were re-sequenced until it satisfied the quality.

**2.16. Sequence editing and mutation scoring**

By checking electropherograms of aligned sequences, manual editing was done to remove inappropriate mixed baseballs. Heterozygotes were confirmed by performing re-sequencing of the samples. 5% of the samples where the proportion of heterozygotes was high were re-sequenced for confirmation. All the deviations from the reference sequences were manually checked by examining electropherograms thoroughly. Genotypes were exported from the software for further analysis.

**2.17. Statistical analysis**

Descriptive statistics were calculated for the anthropometric and physiological/clinical variables with the use of SPSS v13 (SPSS, USA). Allele frequencies were calculated for all the SNPs and were tested for Hardy-Weinberg equilibrium using PLINK (Purcell et al., 2007). Fisher’s exact test was used to assess the allelic association with type 2 diabetes. Linkage disequilibrium (LD) between all the SNPs was estimated using Haploview (Barrett et al., 2005). Genotype association with the phenotypes was tested under different genetic models viz., additive, dominant and recessive for both quantitative and qualitative traits by regression analysis. All tests were performed after adjusting for confounding factors to know the effect of genetic risk factors. Correction for multiple testing was applied wherever appropriate. Haplotype analysis was performed using Haploview software.