Chapter 2
Materials and Methods
CHAPTER 2: MATERIALS AND METHODS

STUDY SUBJECTS AND METHODS

2.1 Patient Recruitment and Clinical Evaluation:
This part of the work was done by our clinical collaborators Dr. Madhulika Kabra, Pediatrics Clinic at All India Institute of Medical Sciences (AIIMS), New Delhi, and Dr. Meena Gupta, Department of Neurology at GB Pant hospital, New Delhi. Ethical committee clearance was obtained from the participating hospitals. Informed consent from all participating adult members and parents/guardians of both affected and unaffected minor children included in the study. Families with more than two affected sibs were recruited for the study. Since this project was a continuation of an earlier project on MR in the lab, samples that were already available in the DNA bank in the lab were also utilized for mutation screening. Further, our lab being one of the main centres for Fragile-X testing in the country, samples from various hospitals were available in the lab. Samples that were negative for FMR1 expansion coupled were also used for mutation screening. Consent was not required for stored samples, based on open-ended consent obtained earlier.

All the families recruited for the study were evaluated by our clinical collaborators. Detailed family and clinical history was obtained, followed by physical and neurological examination of the affected members. Further, affected individuals were critically evaluated for dysmorphism, malformations and anomalies to rule out any overlap with known syndromes. Detailed psychological assessments were done in possible affected individuals to assess their social skills and overall development using Wechsler Adult Intelligence Scale (WAIS III) or Vineland social maturity scale (VSMS) as applicable. Detailed checklists were filled and documented. The affected ID individuals were assessed according to Diagnostic and Statistical Manual of Mental Disorders (DSMIV) criteria.
Figure 1: The overall methodology followed is represented in the flowchart below and the steps are detailed in subsequent pages.
2.2. Genetic analysis:

2.2.1. Exclusion/Inclusion of families recruited above: The probands from the above were screened for gross chromosomal abnormalities by karyotyping in the collaborators’ lab at AIIMS followed by Fragile X mutation screening (the most common cause of inherited MR) by Southern blot by dedicated staff in the lab (UDSC). Amino acid profiling of blood/urine samples from probands were carried out and screened for metabolic syndromes [“metabolic syndrome” has a different meaning in heart disease field] where deemed necessary. Families having numerical abnormalities and Fragile X repeat expansion were excluded from the study. Informative families with multiple affected subjects were utilized for linkage analysis to identify putative disease linked locus/loci.

2.2.1.1. DNA Isolation

Approximately 10ml blood (8ml with EDTA blood for DNA isolation+ 2ml whole blood) for clinical, genetic and serological evaluations were drawn from each participating individual. From the 8ml EDTA blood, two aliquots of 1.5ml each were stored at -70°C in 7% DMSO solution (Sigma chemicals) for future use. Remaining 5ml blood was used for DNA isolation using the conventional phenol chloroform method (Sambrook et al 1989), briefly described below:

- 5ml of EDTA blood was transferred to a 50ml sterile polypropylene (pp) tube (Greiner, Germany).
- Added 3 volumes of lysis buffer (1.5mM NH₄Cl; 10mM KHCO₃; 0.1mM EDTA, pH 8.0) to the blood sample and mixed thoroughly. The tubes were kept on ice for 5-10 minutes or till lysis was complete as indicated by clear red solution.
- Centrifuged the sample at 3500 rpm at room temperature (RT) for 15 minutes and discarded the supernatant taking care that WBC-nuclear pellet was not disturbed and retained.
- To this 5ml suspension buffer was added and nuclear pellet resuspended by applying gentle strokes to the tube.
Proteinase K (10mg/ml) to a final concentration of 50μg/ml and 1ml of 20% SDS were added to the resuspended pellet.

Incubated the samples at 37°C for 4-6 hrs with shaking or O/N without shaking.

Added equal volumes of phenol (Tris- equilibrated pH 8.0) to the digested sample, mixed gently for 10 minutes.

Centrifuged at 3500 rpm at room temperature (RT) for 15 minutes leading to separation of two distinct phases. Transferred the upper aqueous phase to a fresh sterilized 50ml pp tube.

Added equal volumes of phenol: CIA (chloroform: isoamyl alcohol, 24:1), mixed gently for 10 minutes.

Centrifuged at 3500 rpm at RT for 15 minutes, transferred the upper aqueous phase to a fresh, sterilized 15ml pp tube.

Added equal volumes of CIA and mixed gently for 10 minutes. Centrifuged as above & transferred aqueous phase to a fresh, sterilized 15ml tube or autoclaved 100 ml conical flask.

To the aqueous phase, added 1/10 volume of 3M sodium acetate (pH 5.2) and 0.8 volumes of isopropanol.

Spooled the high molecular weight DNA by gently swirling the tube.

Washed the DNA twice in 70% alcohol to remove salt traces and once in absolute (100%) alcohol.

Vacuum dried the DNA, dissolved in T.E (10mM Tris-.HCl, pH 8.0; 1mM EDTA, pH8.0) and stored at 4°C for genetic analysis.

2.2.1.2. Qualtitative and quantitative assessment of the DNA

2.2.1.2.1. Gel electrophoresis: The quality of isolated genomic DNA was checked by electrophoresis (1-2 μl of DNA) on a 1% agarose gel containing ethidium bromide (0.5 μg/ ml). The gel was prepared and run in 1X TAE buffer at 60V for 30 minutes. DNA was visualized using the UV transilluminator (Fotodyne). The compositions of various stock solutions are given in Appendix.

2.2.1.2.2. NanoDrop analysis: DNA was quantified using Nanodrop ND1000 spectrophotometer (Agilent). The instrument has the sensitivity to measure 1ng/μl and can measure a maximum of 3.7μg/ μl of nucleic acid. One micro litre of the DNA sample was
loaded on to the sample holder which is in contact with the optic fiber and absorbance was measured at wavelengths 260nm and 280 nm. The purity of DNA was checked by taking the ratio at A\textsubscript{260}/A\textsubscript{280}, the ratio being \(\geq 1.8\) for pure DNA. The absorbance of the sample at 260/230 was also assessed to determine the sample impurities.

2.2.1.3. Diagnosis of Fragile-X syndrome by Southern blotting:
Southern blotting has been employed to detect CGG expansion repeats present in the promoter region of \textit{FMR1}. The methylation of the promoter due to excess repeat expansion is the main cause for Fragile-X syndrome, the most common form of inherited MR. The presence of > 200 repeats indicates full mutation and repeats greater than 70 and less than 200 are considered to have premutation. The use of Southern blot for Fragile-X diagnosis was first demonstrated by Oostra et al in 1993. The procedure for Southern blotting for detection of premutation and full mutation used routinely in the lab for diagnostics is as follows.

2.2.1.3.1. Preparation of the samples and Gel: 8µg Genomic DNA was digested with EcoRI (for males), and with EcoRI and a methylation sensitive cutting enzyme EagI for females, at 37°C overnight. The components used for digestion is given in table 1.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Components</th>
<th>Amount(λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>DNA</td>
<td>8µg</td>
</tr>
<tr>
<td>2.</td>
<td>10X Buffer (EcoRI )</td>
<td>5.0</td>
</tr>
<tr>
<td>3.</td>
<td>BSA</td>
<td>0.5</td>
</tr>
<tr>
<td>4.</td>
<td>Spermidine</td>
<td>1.0</td>
</tr>
<tr>
<td>5.</td>
<td>EcoRI</td>
<td>1.5</td>
</tr>
<tr>
<td>6.</td>
<td>EagI</td>
<td>1.5</td>
</tr>
<tr>
<td>7.</td>
<td>H\textsubscript{2}O</td>
<td>-</td>
</tr>
<tr>
<td>8.</td>
<td>TOTAL</td>
<td>50</td>
</tr>
</tbody>
</table>

* DNA 42λ for ECoRI + Eag I Digestion, 43.5λ for EcoRI Digestion
* O/N Digestion at 37°C.

The digested products were run on 0.8% gel overnight at 20V with HindIII marker. The next day, following the visualization of the gel with ethidium bromide, the gel with resolved digested products was processed. The gel was treated for depurination of the DNA using 0.25N HCl for 15 minutes, followed by incubation in denaturation solution for 20minutes twice over (0.5M NaOH and 1.5M NaCl) and treatment in Neutralisation solution (1.5M NaCl, 0.5M Tris, 0.5mM EDTA) for 15 minutes twice. Each of the above steps was followed by rinsing of the gel in double distilled water.
2.2.1.3.2. Transfer of the treated DNA onto the membrane: An apparatus for transfer was set up with 3mm Whatman filterpaper and glass plates. The glasss tray was filled to an inch with transfer buffer ((10XSSC: NaCl 0.075M, Na Citrate 7.5mM, pH=7.0). The wicks and the Nylon membrane were cut according the size of the gel and soaked in the transfer buffer for 15 -20 minutes. The soaked whatman paper wick was placed on the glass plate, followed by inverted gel and Nylon membrane. This gel sandwich was then covered by stacks of similarly cut wicks. A weight was placed on top and the DNA was transferred from the gel to the membrane by capillary transfer mode.

2.2.1.3.3. Preparation of the probe: 100-200ng of the DNA fragment (pP2, pX6 probes) was diluted with double distilled water to 33μl, denatured at 100°C for 12minutes followed by immediate cooling on ice. The probe was labeled with α-P32 dCTP by random primer labeling method. The reaction consisted of 200μM dATP, dTTP, dGTP each, and 50μCi (α-P32) dCTP, denatured probe, 1X labeling buffer and 2U of Klenow enzyme. The labeling reaction was carried out at 37°C for 90 minutes. The volume was made up to 100μl by addition of 30μl of TE pH 8.0 and 20 μl of 6X bromophenol blue dye. The probe thus made was eluted using a G-50 Sephadex column equilibrated with T.E (pH 8) to remove the unincorporated radioactive nucleotides. The labeled probe was loaded on to the column drop by drop and centrifuged at 3000 rpm for 3 min. A fresh tube was used to collect the probe eluted subsequently using T.E.

2.2.1.3.4. Hybridisation, washing and Autoradiography: The Nylon membrane with DNA was rinsed in 6X SSC, dried and cross linked using UV. The treated Nylon membrane was incubated in pre-warmed pre-hybridisation solution at 65°C for 3 hrs. Following this the pre-hybridisation solution was decanted and replaced with hybridization buffer. Hybridisation buffer was made with approximately 1ml of the eluted probe mixed with 4ml of the pre-hybridisation solution (0.5M Phosphate buffer (0.4M Na2HPO4, 0.1M NaH2PO4,7% sodium dodecyl sulphate; pH 7.4) and boiled to 100°C for 10 minutes. To the membrane hybridization buffer with denatured radioactive probes was added and incubated at 65°C for 18-20hrs in the hybridization oven. The membrane in the hybridization bottle was then washed twice for 20 minutes with 2XSSC with 0.1% SDS, which was followed by washing in 1XSSC with 0.1% SDS solution for 5 minutes. The last step was continued till the radioactive count on the blot was between 10-20mCi. The blot was wrapped in saran wrap and exposed to an X-ray film (Kodak) in an X-ray cassette. The cassette was stored at -70 overnight for optimum exposure and developed to reveal the autoradiogram.
2.2.1.3.5. Analysis of the results: The hybridization band observed for the test samples were compared with the positive control and expected band size. The expected band size for Pp2 is 5.2kb and pX6 1.1kb for the control samples. Individuals with premutation and full mutation had bands higher than that expected indicating increase in size.

Based on the inclusion criteria, four families with multiple affected individuals were prioritised for further analysis.

2.2.2. Array CGH using cytochip to detect copy number variant(s):
The role of copy number variants involving DNA segments that are larger than 1kb has been reported in both familial and sporadic cases of intellectual disability. The proband(s) in the families were screened for known/novel copy number variants using array CGH.

2.2.2.1. Illumina HumanCytoSNP-12: Illumina HumanCytoSNP-12 was used to detect sub-microscopic chromosomal aberrations which cannot be detected by routine karyotyping. The chip contains 300,000 probes across the whole genome targeting SNPs in regions prone to aberrations. The chip can detect imbalances as small as 30kb with an average marker spacing of 6.2kb. The chip has been validated to detect a plethora of common microdeletion/duplication and subtelomeric deletions/duplications. However balanced translocation and inversion cannot be detected. The intensity ratio and allelic profiling allow for both copy number variation and copy neutral LOH. 250ng of high quality genomic DNA was used for hybridization on to the chip containing the probes using the manufacturer’s protocol at a commercial facility.

2.2.2.2. Analysis of the data: Intensity and allele frequencies of SNPs were utilized for calculating the confidence of the aberrations using Karyostudio software (Illumina). The karyostudio software consists of integrated chromosomal browser and cross matching capabilities with multiple cytogenetic databases and linked genomic databases such as UCSC, DGV, ENSEMBL, DECIPHER, CYTO etc. The clinical significance of the aberrations i.e deletions and duplications are interpreted based on the size of gain/loss, gene content, inheritance pattern and frequency in control populations. To determine the frequency of the observed copy number changes in the
general population previously reported copy number variation (CNV) data sets ascertained from apparently normal individuals and consulted CNVs reported in online databases (http://www.genome.ucsc.edu; http://projects.tcag.ca/variation/) were searched for. The chromosomal aberrations obtained were also cross matched with DECIPHER and CHOP database.

**Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER):** The database is maintained by Sanger Institute, where array comparative genomic hybridisation results of patients DNA are compiled and maintained. It not only includes individual deletion and duplications observed in cases, but also catalogues clinical characteristics associated with the aberration and maintains a database of known microdeletion/duplication syndromes.

**The Copy Number Variation project at the Children's Hospital of Philadelphia (CHOP):** The database has catalogued the commonly occurring copy number variants from ~ 2026 control samples by high resolution genome wide scanning. This repository is an excellent source of frequent copy number variants and allows for the effect assessment of role of the CNVs in disease susceptibility and normal variation.

**Database of Genomic Variants:** This database provides the catalog of structural variation identified in control samples to correlate genomic variation with phenotypic data. The database consists of DNA variants with > 1kb and also indels in 100bp-1kb range. Array CGH is now routinely used for detecting recurrent chromosomal aberrations associated with developmental disorders and diagnosis in Clinics. Those families with known/novel sub microscopic chromosomal aberrations with implication for intellectual disability were not analysed further.

2.2.2. **Genome-wide scan for identifying disease linked locus/loci:**

2.2.2.1. **Whole genome scan using Linkage mapping set:** The whole genome scan was performed using Linkage Mapping Sets v. 2.5 ABI MD10 panel. It consists of 400 markers spaced over the entire genome with an average spacing of 10cM. The forward primer of each marker was labeled with one of the three fluorescent dyes (FAM, VIC & NED). The markers are selected from Genethon sex-averaged maps. (reference). The markers were grouped in panels. Each panel consisted of a group of markers that can be loaded onto one capillary in the sequencing machine. The overlapping alleles in every panel were distinguished by labeling with three different
Materials and Methods

Fluorescent dyes; FAM, VIC and NED which were played as blue, green, and yellow peaks respectively. Fragment analysis of the fluorescent labelled microsatellite markers and sizing was done using the LIZ (500-250) as size standard on a ABI 3100 genetic analyser.

True Allele PCR Premix (ABI) was used to amplify the PCR products. The master mix consisted of 2.5mM MgCl2, Amplitaq gold and dNTPs. Forward and reverse primers were combined and supplied in a tube at 10 mM concentration (5 mM of each primer) in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0. A PCR reaction of 7.5µl was set with 15-25ng of DNA, 5pmoles of primer (0.5µl), 4.5µl of the master mix and 1.5µl deionised double distilled water. The cycling conditions followed were: initial denaturation at 95°C for 12 minutes; denaturation at 94°C for 15 seconds, annealing at 55°C for 15 seconds, and extension at 72°C for 30 seconds for the first 10 cycles; denaturation at 95°C for 15 seconds, annealing at 55°C for 15 seconds, and extension at 72°C for 30 seconds for the next 20 cycles; and then final extension at 72°C for 10.

The products thus amplified for individual samples were pooled according to the panels in the ratio of 1:1:2 (FAM: VIC: NED). 1µl of each pooled product was mixed with a cocktail of 8.5µl of Hi-Di formamide (Applied Biosystems) and 0.5µl of LIZ (500-250) internal size standard (Applied Biosystems) and denatured at 95°C for 5 minutes. The products were loaded onto an ABI3100 Genetic Analyzer for sizing and fragment analysis. The data thus obtained were imported into Genemapper (v.3) for allele sizing and calling. The bins provided by the manufacturer were modified according to the allele sizes observed in the samples.

Genome wide scans for identifying linked locus/loci was carried out either by using all 400 microsatellite markers or only 18 of the markers of the X chromosome depending on the suggestive inheritance of the disease phenotype observed in the family. Though the markers of this kit had a mean heterozygosity >75%, a proportion of the MS markers were uninformative in individual families. To overcome this limitation and to refine the genomic region of interest, additional microsatellite markers were designed using universal HEX and NED fluorescent probes.

2.2.2.2. Additional micro-satellite markers: Additional micro-satellites genotyped in this study were obtained from Genethon sex-averaged maps. The markers were amplified by using primers designed using Primer3 software or with those available
from the database (NCBI 37/ UCSC 18). Labeling each of the primers designed with fluorescent markers is expensive. Hence the universal primer labeling method was adopted and standardized (Schuelke, 2000).

**Figure 2:** Illustration of universal primer labeling method employed in this study.

The PCR mix contained 15-25ng of DNA, 2.5mM Mgcl₂, 0.5U of Biotools taq polymerase, 5pmol of forward primer with the 5’ tail, 10 pmol of the reverse primer and 10 pmol of the labelled M13 sequence. Conditions of the PCR amplification are as follows: 94°C (5 min), 30 cycles at 94°C (30 s)/ 60°C (30 s)/ 72°C (30 s) and a final extension at 72°C for 7 mins. The PCR products were pooled depending on non overlapping allele sizing and differential labelling by FAM and HEX. After checking on 1% agarose gel, the PCR products were pooled in the ratio of...
1:1 (HEX: FAM). Subsequently 1µl of the pooled PCR product is added to 8.5µl of formamide and 0.5µl of LIZ standard, denatured for 5mins and loaded on the ABI genetic analyser.

For example, D15S146: primers from database:
GGAAGCCTGACTTTATATCCG
ATGTCTGTTCAGATCCTTTGC

**Design:**
TGGTAAAACGACGCCGACGGAAGCCTGACTTTATATCCG (primer with tail, concentration 5 pico moles)
GTGTCTTATGTCTGTTCAGATCCTTTGC (other primer, with pigtail for reduction of stutterpeaks, concentration 10 pico moles)
HEX/FAM-TGGTAAAACGACGCCGAC (tail sequence, concentration 10 picomoles)

Polymorphic microsatellite markers were typed among the most informative family members. Genotypes called were used for haplotyping and calculation of LOD scores. In families with a known syndrome and with genetic heterogeneity, it was more economical and less time consuming to carry out chromosome specific scans of chromosomal regions encompassing known genes than mutation screening by resequencing. Haplotype sharing among affected individuals was assessed before proceeding with these families for whole-genome scans.

**2.2.2.3. Mendelian inconsistencies and genotyping errors:** Haplotyping and linkage analysis requires the genotype data to be accurate and reproducible. Errors while genotyping are inevitable which can affect the results significantly resulting in skewed interpretation. Undetected genotyping errors can reduce the power to detect linkage by introducing spurious recombinants or negate them giving rise to erroneous results. Genotyping error occurs when the observed genotype does not correspond with the actual genotype of the individual. This includes typing errors, misinterpretation of the genotypes and data entry errors. Reading of heterozygotes as homozygotes due to signal intensity or bin size and missing of the allele by single repeat are common causes of mistyping. Genotyping errors may or not result in Mendelian inconsistency. Mendelian inconsistency can also be caused due to misrecording/misspecifying of relationship in the pedigree. It is also likely that marker mutations may mimic genotyping errors. Errors were detected both manually and by MERLIN software. Mendelian inconsistency errors are easier to detect than
mistyping/allele calling errors, which can be achieved by checking if the data is in accordance with inheritance. The software helps in detecting false double recombinants by constructing haplotypes. To reduce the genotyping error, the alleles were called manually and tallied/compared with the software allele calls. This is tedious but it is the only way to reduce genotyping errors. Retyping of the incorrect markers was carried out with appropriate controls. Random allele errors and mutations were more difficult to detect. The genotypes thus sorted for errors were used for linkage analysis and haplotype reconstruction.

2.2.3: Linkage Analysis:
Meiotic recombination occurs due to crossing over and gene conversion resulting in genetic variation. Linkage analysis exploits this natural mechanism in the identification of disease trait locus. It estimates/tests the co-segregation of a marker with the disease phenotype in a family. With more individuals available for genotyping belonging to different generations, there is increase in the number of recombination observed. Thus the likelihood of identifying the disease locus is greater in a large multigenerational pedigree or by using multiple pedigrees having the same disease phenotype. In our study, we have analysed both a large kindred with X-linked inheritance and also small/nuclear families with putative X-linked and autosomal recessive mode of inheritance. Due to both clinical and genetic heterogeneity observed among families with ID, the genotypes were not pooled for linkage analysis in this study. For small/nuclear families, linkage analysis facilitated exclusion mapping rather than identifying significant linked regions due to the limited number of informative individuals available. Further, it also identified multiple suggestive linked loci which were analysed using a combination of additional experimentation and bioinformatics tools.

Linkage analysis was performed using genotyped microsatellite markers for the available family members. Markers were checked for Mendelian inconsistencies using pedigree information (Pedcheck) for two-point/multi point LOD (Logarithm of odds) score calculations using MLINK and MERLIN respectively. The LOD score is a statistical test of linkage. The probability of two loci present adjacent to each other on the same chromosome to be inherited together is higher than two loci present on different chromosomes. Parametric LOD scores tests the likelihood of observing
marker(s) with the disease phenotype under a predetermined model of inheritance. Depending on the inheritance pattern documented in the family, either an autosomal recessive or X linked inheritance was assumed while calculating the LOD scores. Multiple marker haplotypes were reconstructed and compared between affected and unaffected individuals for narrowing down the linkage interval, wherever possible. Haplotypes were not only used for positive reinforcement of the region in disease aetiology, but also for exclusion mapping.

Parametric multipoint linkage analysis was carried out with autosomal recessive mode of inheritance with parameters of complete penetrance and 0.0001 disease allele frequency. The values for penetrance in unaffected and carrier males/females were set as 0.0001 and 1.0 for affected individuals (males/females). Complete penetrance was also assumed in case of X linked inheritance, thus the values of affected and unaffected males were set to 1.0 and 0.0, whereas in females both unaffected and carriers were set to 0.0. The markers were assumed to have equal allele frequencies and no difference in recombination rates between males and females. Two-point linkage analysis was done using MLINK and multipoint was calculated using MERLIN. A LOD score value > 3 was considered to be significant, where the odds of it being false positive are 1 in 1000. LOD score values between 1.8-2.5 was considered be suggestive in nature.

2.2.4: Reconstruction of haplotypes:

The ordered genotypes in the offspring were determined based on the inheritance of the alleles from their parents. Since the markers used were polymorphic, the alleles were assigned with ease in the offspring. Thus the phase was determined in the offspring. Recombination positions in the offspring were inferred by analyzing identity by descent (IBD) status of the alleles between the siblings. The change in inheritance pattern was observed as the change of IBD status from one locus to another. The main limitation of this method is that while using sparingly spaced microsatellite markers, we might miss double recombinants. However the possibility of recombination in a small genomic region is negligible. Based on inferred recombination positions, the chromosome was portioned into segments, each of which was recombination free. The predominant order of markers among the children with minimum recombination was assumed to be the parental haplotype. All possible
configurations with Mendelian inheritance and recombinant free segment were obtained. Thus genotype information in the children was used in reconstructing the parental haplotypes.

The total number was recombinations and precise loci involved are uncertain. However minimum numbers of recombinants were used to find the possible assignment of the genotypes. Though the markers used are polymorphic with significant heterozygosity in the population, uninformative markers constitute 15-20% of the total markers genotyped. The parental genotypes are uninformative when they are (a) homozygous for a marker or (b) when both the parents have the same heterozygous genotype. In the above situations, inferring the haplotype in the offspring becomes difficult or impossible depending on the combination of genotypes inherited by the offspring. When both the parents are heterozygous for the same genotype, it is possible to deduce the exact phase in 50% of the times in the offspring. The ambiguous intervals consisting of uninformative markers were placed according to recombination free segments. When these uninformative markers are present in between the fixed phase of the recombination free segment, it does not affect the original ordering of genotype/haplotype configuration. However, if the uninformative marker is present at the juncture of recombination, the precise recombination cannot be assumed, thus it was placed without assuming Phase. The haplotypes were constructed manually assuming minimum recombination and by using MERLIN software.

2.2.5: Refining the critical region of Interest:
Depending on the pattern of linkage obtained, and the information content of the markers, a denser map of the region was made by using additional 5cM markers or multiple SNPs; haplotypes were constructed and compared between affected and unaffected family members using standard protocols described above. Genotyping of additional markers facilitated the exclusion of false positive scores due to uninformative markers and reiterated the role of the genomic region in disease aetiology.
2.2.6. Prioritization of putative candidate genes:

Bioinformatic tools were employed to look for candidate genes in the linked locus (orthologous genes and their function, genes expression pattern, interacting partners of known disease genes). Due to genetic heterogeneity underlying the disease condition, the families were not pooled for analyses. Thus the families did not have enough power to obtain significant LOD scores. Thus regions showing suggestive LOD scores obtained in small families were processed for prioritisation of the list of candidate genes based on bioinformatic tools. The interacting partners and pathway genes of known disease genes were obtained from BIOGRID (http://thebiogrid.org/), STRING (Szklarczyk et al., 2011; Jensen et al., 2009), KEGG (http://www.genome.jp/kegg/) and REACTOME (http://www.reactome.org/ReactomeGWT/).

2.2.7. Resequencing of putative candidate genes and candidate disease genes:

Polymerase chain reaction was carried out to resequence exons including the exon-intron boundaries of the selected genes. The primers were designed using Primer3 software (http://frodo.wi.mit.edu/). Resequencing DNA samples were amplified using polymerase chain reaction (PCR) according to their respective PCR primers and standardised conditions were then subjected to sequencing (ABI3700 automated sequencer) in the central instrument facility at UDSC. Analysis of sequencing profiles Chromatograms obtained following the sequencing reaction was analyzed utilizing SeqMan v 5.03 software (DNASTAR inc.). Reference sequences were obtained from GenBank database (NCBI 37). Samples with variants both polymorphism and mutation was confirmed resequencing twice or using both the forward and the reverse primer. In case of Family 3a an attempt to sequence the coding exons of ATR from lymphoblastoid cell lines was carried out using the cDNA made from total RNA.

2.2.7.1. RNA Isolation: The isolation of RNA from the lymphoblastoid cell lines was carried out using the following protocol (Sambrook). The cells were >90% confluent and grown in 25cm² flask. All the glass wares and plastic wares used were DEPC treated.

- The cells were washed with plain RPMI 1640 medium twice at 2000rpm for 5 mins
- The media was discarded and the pellet was washed with 1M PBS at 2000rpm for 5 mins
- The pellet was resuspended in 1ml of Trizol reagent (Sigma). The pellet was mixed thoroughly to form a homogenous suspension using the pipette and was allowed to stand at room temperature for 5-10mins.
- To the suspension 0.2 ml of chloroform was added, mixed well for 15-20mins, followed by centrifugation at 12,000 rpm for 15 mins at 4°C.
- The supernatant was discarded and the pellet was washed in 70% ethanol at 12,000 rpm for 2 mins
- The RNA pellet was air dried and dissolved in DEPC treated double distil water

The RNA thus isolated was checked on the gel for integrity and quantitated using nanodrop at 260/280.

2.2.7.2. cDNA synthesis from total RNA: The high quality RNA obtained was used in cDNA synthesis using Accuscript high fidelity reverse transcriptase (Agilent) according to the manufactures protocol. This kit uses AccuScript reverse transcriptase, which is an RNase H-deficient Moloney murine leukemia virus reverse transcriptase (MMLV RT) engineered to deliver accuracy while promoting cDNA synthesis length and yield. The cDNA was made from the total RNA isolated. Following which the region/exon of interest was amplified from the cDNA synthesized using specific primers.

Table 2: List of components available in the kit and the concentration used for a single reaction.

<table>
<thead>
<tr>
<th>Components</th>
<th>Stock concentration</th>
<th>Working (10µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AccuScript RT Buffer*</td>
<td>10x</td>
<td>1µl</td>
</tr>
<tr>
<td>PfuUltra HF DNA polymerase</td>
<td>2.5 U/µl</td>
<td>1µl</td>
</tr>
<tr>
<td>PCR buffer</td>
<td>10x</td>
<td>1µl</td>
</tr>
<tr>
<td>Deoxynucleotide (dNTP) mix</td>
<td>40 mM (10 mM of each dNTP)</td>
<td>1µl (10mM each)</td>
</tr>
<tr>
<td>Oligo(dT) primer (18 mers)</td>
<td>100 ng/µl</td>
<td>0.6µl</td>
</tr>
<tr>
<td>Random primers (9 mers)</td>
<td>100 ng/µl</td>
<td>0.6µl</td>
</tr>
<tr>
<td>DTT</td>
<td>100 mM</td>
<td>1µl</td>
</tr>
<tr>
<td>RNase-free water</td>
<td></td>
<td>2.8µl</td>
</tr>
<tr>
<td>RNA</td>
<td></td>
<td>1µl (500ng)</td>
</tr>
</tbody>
</table>
*The 10× AccuScript RT buffer contains 0.5 M Tris-HCl (pH 8.3), 0.75 M KCl, 0.03 M MgCl2.

- High quality RNA checked on gel and on nanodrop with OD value at 260/280 between 1.8 and 2 was used
- The template RNA, RNase free water, buffer and primers were denatured for 5mins at 65°C
- The mixture was cooled to room temperature, after which DTT and RT were added and incubated at 25°C for 10mins
- Following this the mixture was incubated at 42°C, the optimum temperature of the RT enzyme for 90 mins
- The first cDNA strand synthesized was used for the amplification of gene-specific primers following the protocol as described earlier for PCR amplification from genomic DNA.

2.2.7.3. PCR conditions: The annealing temperatures of the PCR were standardised by setting the reaction at different MgCl2 concentration and temperatures around the calculated Tm. The annealing temperature of the primer was calculated using the formula: Tm (for primers ≈ 20 bps) = [2(A+T) + 4(G+C)] – 5°C. Working concentrations of respective PCR components is provided in Table 2

Table 3: List of PCR components and working concentrations

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Component</th>
<th>Stock concentration</th>
<th>Working concentration</th>
<th>Volume per Reaction</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reaction buffer*</td>
<td>10X</td>
<td>1X</td>
<td>2.0 μl</td>
<td>Supplied with the respective taq polymerase</td>
</tr>
<tr>
<td>2</td>
<td>dNTPs</td>
<td>2.5mM</td>
<td>250μM</td>
<td>1.6 μl</td>
<td>Roche</td>
</tr>
<tr>
<td>3</td>
<td>Primers (F&amp;R)</td>
<td>10μM</td>
<td>5pM each</td>
<td>1.0 μl</td>
<td>Ocimum/ Sigma/Bioserve</td>
</tr>
<tr>
<td>4</td>
<td>Taq polymerase</td>
<td>5U/μl</td>
<td>0.5U/μl</td>
<td>0.05 μl</td>
<td>Biobasic/Biotools/ Prepared in house **</td>
</tr>
<tr>
<td>5</td>
<td>DNA</td>
<td>50ng/ul</td>
<td>100ng/reaction</td>
<td>2.0 μl</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DDW</td>
<td>-</td>
<td>-</td>
<td>13.35 μl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td>20 μl</td>
<td></td>
</tr>
</tbody>
</table>

* Generally the reaction buffer contained 1.5mM MgCl2. For buffers without MgCl2 the reaction mix was supplemented appropriately with 50mM MgCl2 stock (Biotools). DDW-autoclaved double distilled water.
**Taq polymerase prepared in the laboratory (Engelke et al., 1990)
For all amplification reactions, master mix was prepared and an aliquot was transferred to individual PCR tubes or 96 well PCR plates (Axygen). The amplifications were carried out in an MJ PCR machine (Dyad/MJ96 PTC 100/ABI verity) using the general cycling conditions (Table 3).

<table>
<thead>
<tr>
<th>Step</th>
<th>Cycling conditions</th>
<th><strong>Tm</strong>°C, min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Denaturation</td>
<td>95, 5</td>
</tr>
<tr>
<td>2</td>
<td>Denaturation</td>
<td>95, 30 sec</td>
</tr>
<tr>
<td>3</td>
<td>Annealing</td>
<td><strong>Tm</strong>, 30 to 45 sec</td>
</tr>
<tr>
<td>4</td>
<td>Extension</td>
<td>72, 45 sec</td>
</tr>
<tr>
<td>5</td>
<td>Cycle</td>
<td>34 times to step 2 through 4</td>
</tr>
<tr>
<td>6</td>
<td>Final extension</td>
<td>72, 7 min</td>
</tr>
<tr>
<td>7</td>
<td>Storage</td>
<td>4°C ∞</td>
</tr>
</tbody>
</table>

*The amplicons were in the range of 150-700 bp
**Tm was calculated for the specific primers using the formula mentioned above

In the absence of mutations identified in the known candidate genes or putative candidate genes from linked region, deep sequencing of linked region was carried out using next generational sequencing techniques.

### 2.2.8. Targeted using next generation sequencing techniques: resequencing

Genomic DNA of individuals can be analysed using either whole genome sequencing or by exome sequencing. Whole genome sequencing is more comprehensive, exome sequencing is less expensive, targeted to the coding regions of the genome (which encompass mutations in most of the Mendelian diseases identified till date). Target capture (selective regions of the genome of interest) and exome sequencing (complete coding region of the genome) allows enrichment of selective regions/coding regions by hybridizing genomic DNA to oligonucleotide probes complementary to the region of interest. The selective genomic regions thus enriched are sequenced using high-throughput DNA sequencing technology. In this study, both custom-targeted resequencing of the unambiguously linked region on the X chromosome (F #3) and exome sequencing (targeting the genic region both coding and non coding) in F #4 have been employed. Target sequencing enables deep sequencing of specific regions of interest. There are various commercial platforms which are available for capture and for high throughput sequencing. Each of these platforms employs different
chemistry, probe density and coverage. We have used Agilent technologies for target re-sequencing and Nibelen technology for exome re-sequencing. Following the capture of targeted regions of interest the samples were re-sequenced using Illumina GAIIx and Hiseq respectively. In general the methodology of target sequencing consists of the following steps.

- Preparation of the capture library (custom made or commercially available kit)
- Preparation of the sample DNA
- Hybridisation of the capture library (oligonucleotide probes) on to the processed DNA sample
- Selection of hybridized genomic fragments
- Purification and amplification of the hybridized targets
- Quality check of the targets
- Resequencing of the target region
- Analysis of the data obtained

2.2.8.1. Custom target resequencing (using Agilent Sure Select technologies):

2.2.8.1.1. Designing of probes: Probe selection plays a critical role in enrichment strategies. The genes and non-coding RNA present in the 70 Mb linked region obtained (UCSC 19) for F #3 was selected for probe designing. The co-ordinates of the exons and promoters for the 17 known XLID genes were uploaded to the e-array software (Agilent). The probes were designed to have 2x tiling and the target region was repeat masked for efficient amplification. The probes are 120mer long – biotinylated cRNA baits. The software selects known validated probes available for the region of interest and also designs new probes for the other target regions. More details specific to the study are detailed in Chapter x.

2.2.8.1.2.: Sample Preparation: The sample was prepared using the Agilent protocol outlined in "SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library" (Version 1.0, May 2010). Briefly, the sample was sheared to an average size of 150-200bp using sonication. DNA fragment ends were repaired and phosphorylated using Klenow, T4 DNA Polymerase and T4 Polynucleotide Kinase.
Next, an 'A' base was added to the 3' end of the blunted fragments, followed by ligation of Illumina Paired-End adapters via T-A mediated ligation. The library was PCR-amplified using 4 cycles and 500ng of the library was incubated with the SureSelect custom exome baits for 24 hours at 65C. The targeted region is captured using magnetic streptavidin beads, DNA was washed and then eluted. The eluted DNA was PCR amplified using SureSelect GA primers for a total of 14 cycles. The amplified capture DNA library size and concentration were determined using an Agilent Bioanalyzer.

2.2.8.1.3. GAIIx Run Conditions: The libraries were combined and seeded onto the flowcell at 8pM yielding approximately 529K pass-filter clusters per tile (120 tiles per lane). The library was sequenced using 2x76 cycles of chemistry and imaging.

Figure 3: Sure select overall sequencing workflow. Reproduced from Agilent manual.
2.2.8.2. Exome resequencing (Roche NibleGen technologies): Target capture of primarily the exonic regions of the genome in F #3, F #3a and F #3b was carried out using commercially available SeqCap EZ exome library v.3 (Roche NibleGen technologies). It is a solution-based method to selectively amplify the target regions of interest with 2.1-million-long oligonucleotide probes covering 64Mb of genomic region with target enrichment performance (Clark MJ et al, 2011) to discover variants including both single nucleotide variants and indels. The probes used vary in size over 55-105 bp. The probes cover genes from RefSeq RefGene CDS, CCDS, and miRBase v14 databases, plus coverage of 97% Vega, 97% Gencode, and 99% Ensembl databases. The samples were processed according to the manufacturers instruction, detailed in NimbleGen SeqCap EZ Library SR User’s Guide v3.0. The brief version of the protocol is as follows.

1. Genomic DNA: SeqCap EZ Oligo pool or an array is made against target regions in the genome.

2. Library Preparation: Standard shot-gun sequencing library is made from genomic DNA, ligated with sequence platform specific linkers and enriched using Illumina Truseq sample preparation kit. The enrichment is assessed using genome analyzer. The sample is amplified by ligation mediated PCR using primers complementary to the linkers. The amplified sample is checked for quality using Nanodrop.

3. Hybridization: The sequencing library is hybridized to the SeqCap EZ oligo pool.

4. Bead Capture: Streptavidin beads are used to pull down the complex of capture oligos and DNA fragments.

5. Washing: Unbound fragments are removed by washing.

5. Target Fragment Elution: The enriched fragment pool is eluted and recovered from the array.

6. Amplification: Enriched fragment pool is amplified by PCR. The concentration, size distribution and quality of the DNA is assessed by Agilent DNA 1000 chip using the genome analyzer.

7. Enrichment QC: The success of enrichment is measured by qPCR at control loci using Syber green technology.
8. Sequencing-Ready DNA: The end product is a sequencing library enriched for target regions, ready for high throughput sequencing.

2.2.8.3. Genome Analyzer Pipeline Analysis: The custom target enriched samples were resequenced using Illumina GAIIx and the exome sequencing samples were pooled and run on the illumine Hiseq 2000. The samples were sequenced either as 76bp paired end (Custom target) or by 101bp paired end (Exome) sequencing. Initial data processing and base calling, including extraction of cluster intensities, were done using RTA 1.8.70.0 (SCS 2.8). Sequence quality filtering script was executed in the Illumina CASAVA software (ver 1.8.0, Illumina, Hayward, CA). Data yield (Mbases), %PF (passfilter), # of reads, % of raw clusters per lane, and quality %Q30 were examined.

2.2.8.4. Data Analysis:

2.2.8.4.1. Quality check: Following the preliminary QC using the CASAVA software the reads are evaluated using Fast QC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) or NGS toolkit (Patel RK and Jain M, 2012). The Phred scores of the individual bases, GC content of the reads, sequence length distribution, sequence duplication levels, filtering of low quality reads with >30% of bases with Phred score <30, Number of reads etc. Based on the statistics obtained, the reads were filtered and used for downstream analysis.

2.2.8.4.2. Alignment and Mapping of Reads: The short nucleotide reads obtained after QC were aligned to the reference genome (UCSC 19) using Burrow wheelers algorithm (BWA) using default parameters. The default parameter allows 2% error per read. It allows for gapped global alignment and supports paired end reads. It assigns mapping quality to every read mapped. Base quality is not considered while assigning hits, thus preliminary QC of the data is essential to obtain optimum performance using BWA. It further performs Smith-Waterman alignment for unmapped reads with mates mapped to rescue mapped mates, and for high-quality anomalous pairs to fix potential alignment errors (Li H and Durbin R, 2009).

2.2.8.4.3. Variant detection: Variant detection of the mapped reads was obtained using SAMtools (0.1.17). It includes utilities for manipulating alignments in SAM format including sorting, merging, indexing and generating alignments in a per-
Materials and Methods

position format. The tool is also used call the variants (difference in sample sequence compared to that of the reference used) - both single nucleotide variants and small indels. Other than variants and genomic co-ordinates, the output also contains SNP quality, mapping quality and read depth which can be used for prioritisation and filtering of the variants (Li H et al., 2009).

2.2.8.4.4. Variant Filtering: False positives and false negatives are critical issues in genomic re-sequencing. Distinguishing pathogenic variants from rare variants is the main challenge in analyzing the variants obtained from next generation sequencing. The high throughput technology is cheap and fast and thus attractive. However, detailed exhaustive coverage and the quality of the starting material is critical for obtaining the true causal variant. The major challenge of next generation sequencing is to analyse the vast number of variants/alterations obtained to identify the true causal variant. The primary filtering used is for the depth of sequencing and for variants already been reported to be polymorphic in the population. The cut-off for the depth is determined based on the average depth of coverage obtained and also literature. Generally variants called with depth greater than 7x are utilized for analysis. Both coverage of the target region and average depth of coverage obtained will affect variant calling and thus further analysis. The variants which were polymorphic were filtered by comparing with dbSNP and 1000 genome project. All variants with frequency > 0.001 were excluded. The genome annotation in this study was done using SeattleSeq (http://snp.gs.washington.edu/SeattleSeqAnnotation/). The variants were further filtered based on family data, in silico prediction software, and pathway analysis based on known disease genes.

(i) Variants called in known disease genes
The variants in the reported MR genes were listed and analysed. In the absence of deleterious variants in known genes, unknown/novel variants obtained were analysed for putative mutations that could result in the observed disease phenotype. The novel variants obtained were prioritised with three approaches a) genetic, b) positional and functional characterisation of the variant and c) pathways or based on literature.

(ii) Genetic approach
• The inheritance pattern of the disease along with the shared chromosomal region among affected family members was utilised for this purpose. The
affected family members usually share chromosomal segments by identity by descent. The sharing of chromosomal regions available by means of multipoint linkage analysis and haplotyping was employed for extracting the variants from the genome data.

- Further, those homozygous variants called both in the affected and the unaffected individuals were excluded.
- Additionally, only those homozygous variants called in the affected and not in the unaffected individuals from the total variants called were also obtained.

(iii) **Positional and functional characterisation of the variant**

The variants obtained were classified as coding and non-coding variants.

- Coding variants are those called in the exons that translate to form the protein. The variants called were classified as synonymous, non-synonymous, stop gain/loss, in frame/not in frame deletion and duplication. Gain/loss in stop codon and frame shift mutations were prioritised over other coding variants.
- Non-coding variants were those called in the untranslated region of the gene (3' and 5' UTR), introns and Intergenic region. These variants were kept second in priority.
- The variants were also prioritised using *in silico* prediction software like SIFT, Polyphen2 etc. Different prediction algorithms employ different parameters and each has its own strength and weakness. Thus multiple prediction software programmes were used to analyse the functional significance of the prioritised putative variants. The list of *in silico* prediction software used for prioritizing the novel/unknown variants obtained are listed below.

- **Mutation taster:** Mutation taster uses naïve Bayes classifier to determine the functional implication of the variant. The prediction is given as a probability value. The sources used for obtaining relevant information are biomart, Uniprot/swissprot, NCBI, Ensemble, dbSNP and hapmap. Analyses comprise evolutionary conservation (Multiple sequence alignment with homologous sequences), splice-site changes (by NNSplice), loss of protein features
(grantham matrix and protein features using Uniprot/Swissprot) and changes that might affect the amount of mRNA (functional stability using polyadq tool, analysis of Kozak consensus). This software cannot analyze insertion-deletions greater than 12 base pairs and alterations spanning an intron-exon border. It has a false positive predictive rate of 15% and has the highest true positive predictive rate (85%) compared to existing in silico tools (Schwarz et al., 2010).

- **SIFT (Sorting Intolerant From Tolerant):** SIFT predicts effects of substitutions in protein-coding regions of the genome. The prediction is based on the degree of conservation of amino acid residues in sequence alignment derived from closely related sequences, collected through PSI-BLAST conservation, nature of amino acid substitution. Scores less than 0.05 are considered deleterious and the false positive rate is 20%. The confidence in the substitution predicted depends on the diversity of the sequences used for alignment. Availability of only closely related sequences along with the conservation of residue would increase the false positive rates (Kumar p et al., 2009; Ng PC and Henikoff S, 2002; 2003; 2006).

- **Polyphen2:** The primary determines the effect of non-synonymous variants and subsequent impact of amino acid allele variants on the structure based on multiple sequence alignment and protein 3D structures. It uses eight sequence-based (Profile score of amino acid (Position Specific Independent Counts score), CpG context, sequence identity and amino acid to the closest homologue) and three structure-based (hydrophobic propensity, access to the surface area, Crystallographic B factor representing conformational mobility, change in residue side chain, Ramachandaran phi-psi map) predictive features. It also estimates if the variation is present in a hypermutable site on the genome by evolutionary comparisons. It has a 5-10% false positive rate for HumDiv trained model and 10-20% false positive rate for HumVar trained model. It uses a naïve Bayes classifier to calculate the probabilistic score (Adzhubei et al., 2010; Sunyaev et al., 1999; Ramensky., 2002).

- **NNSplice:** NNSplice (http://www.fruitfly.org/seq_tools/splice.html) analyses potential splice site changes. A sequence of 60 bases flanking the alteration is
used to compare wild-type and mutant sequence. If the alteration causes any changes (existing splice site becomes stronger/weaker, additional splice site activated or splice site completely lost), this will be displayed with the effect on the splice site (lost, gained, splicing likelihood increased or decreased) is shown as prediction scores. Prediction scores of 0.5 or higher were considered to affect the splicing pattern by the variant.

- **dbNSFP:** It is observed that different *in silico* prediction tools tend to give varying results depending on the rate of positive and negative predictive value, parameters employed, algorithm and training sets. The concordance between the results obtained is critical in prioritising the variants for further analysis. The dbNSFP is an integrated database of functional predictions from multiple algorithms for the comprehensive collection of human non-synonymous SNPs (NSs). It compiles the predictive values of non-synonymous alteration from PhyloP, SIFT, Polyphen2, LRT, and MutationTaster. It also imputes missing scores when calculating overall effect of the variant when one or more tools fail to predict the effect (Liu et al., 2011).

(iv) **Pathways and Literature mining**
The variants prioritised from the step above were analysed for their role in normal metabolic pathways and functioning in the human system. The role of the prioritised variants in various pathways was obtained from REACTOME and KEGG. The interacting partners of the variants were also obtained from databases including STRING and their role in normal cellular processes was obtained from literature. By consolidating the data and based on relevance to the syndrome observed in the family, the variants were ranked.

2.2.9. **Validation of the filtered variants:**
The prioritised variants based on the sequential filtering detailed above were validated by

(i) Confirming the presence of the variant(s) by Sanger sequencing in the sample;

(ii) Resequencing the variant(s) in all available family members to assess the co-segregation of the putative causal variant with the disease phenotype in the family;
(iii) Screening of co-segregating variant(s) in the control population (N=200), to
determine the frequency of the mutant allele.

Variant(s) if present in the control population are not likely to be causal for a rare
disease, thus excluded. Variants were screened in the population by PCR followed by
resequencing, as detailed earlier in this chapter. If the variant alters the restriction
enzyme site(s), the population controls were screened by PCR-restriction length
polymorphism.

(iv) The causal variant was further confirmed by functional validation of the mutant
observed where possible

2.2.9.1. PCR-RFLP: Following the standardization of PCR as described earlier,
digestion of the PCR products using respective restriction enzymes were carried out.
Based on the expected DNA fragments, the digested products were run on 2-3% agarose gel. Sequencing confirmed positive controls for both wild type and mutant alleles along with appropriate sizing ladder used [syntax is wrong].

Table 2.3: General components used for digestion of PCR products

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Component</th>
<th>Stock concentration</th>
<th>Working concentration</th>
<th>Volume per sample (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NEB Buffer</td>
<td>10X</td>
<td>1X</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>BSA*</td>
<td>100X</td>
<td>1X</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>Enzyme</td>
<td>5U/μl to 20U/μl</td>
<td>0.5U</td>
<td>Appropriate volume</td>
</tr>
<tr>
<td>4</td>
<td>Water</td>
<td>To make up the volume to 30μl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PCR product</td>
<td></td>
<td>18μl</td>
<td></td>
</tr>
</tbody>
</table>

*wherever recommended

Using the appropriate buffer and restriction enzyme the master mix was prepared and
an aliquot was added to each tube/plate. The tube/plate was incubated at the required
temperature for the appropriate time (varied between 3 hours and overnight for
specific restriction enzymes).

2.2.9.2. Agarose gel electrophoresis: PCR products, after digestion with the
respective restriction endonuclease, were resolved on 2-3% agarose gel. For making a
2.5% agarose gel, 2.5g of agarose (Gibco-BRL/ USB) was dissolved (by heating) in
100ml of 1X TAE buffer, ethidium bromide was added to a concentration of 0.5 μg/ml and poured in a pre-assembled gel casting apparatus. Electrophoresis was carried out at 80-100V. The profiles were recorded on the UV transilluminator and the gel photographs saved for future reference. In each of the PCR reaction controls with known genotype (PCR product carrying the allele which gets digested) or sequence confirmed samples (3-4 PCR products for each of the markers were resequenced to confirm the allele profiles) are included in both PCR as well as digestion experiments as positive controls.

2.2.9.3. Site directed Mutagenesis: Site directed mutagenesis is a powerful technique which is used in making modification in genomic sequencing allowing better understanding of the protein structure and function. The effect of the point mutation in MID2 protein identified in this study was assessed by making a mutant construct using site directed mutagenesis using a kit (Quick change II XL site directed mutagenesis kit, Stratagene #200501).

**Primer Designing:** Primers of 35bp was made. Sequences of 17bp on either side of the base position selected for mutation was included in the primer. The primer consisted of the changed nucleotide instead of the wildtype. The annealing temperature of the primer was calculated to of 72°C using the following formula.

\[ T_m = 81.5 + 0.41 \times (%GC) - 675/N - \% \text{mismatch} \]

For calculating Tm:

- N is the primer length in bases.
- values for %GC and % mismatch are whole numbers

The primers synthesized were PAGE purified and used.

**Principle:** The oligo nucleotide primers containing the desired mutation, complementary to the opposite stand of the vector, are extended during temperature cycling. The PfuUltra HF DNA polymerase used allows for efficient amplification of the double stranded plasmid without primer displacement and high fidelity. Extension of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product was treated with Dpn I. The Dpn I endonuclease (target sequence: 5’-Gm6ATC-3’) is specific for methylated and hemimethylated DNA and was used to digest the parental DNA template and to select
for mutation-containing synthesized DNA. Following this, the plasmid was transformed using the XL10-Gold ultracompetent cells. Controls for SDM and transformation were available with the kit to compare the efficiency obtained.

**Figure 4:** This figure illustrates the site directed mutagenesis procedure (Reproduced from Stratagene manual)

**Procedure:**

- As a mutagenesis control, the following reaction was set: 5 μl of 10× reaction buffer; 2 μl (10 ng) of pWhitescript 4.5-kb control plasmid (5 ng/μl); 1.25 μl (125 ng) of oligonucleotide control primer #1[34-mer (100 ng/μl)]; 1.25 μl (125 ng) of oligonucleotide control primer #2 [34-mer (100 ng/μl)]; 1 μl of dNTP mix; 3 μl of QuikSolution reagent; 36.5 μl of double-distilled water (ddH2O) to a final volume of 50 μl. To this 1 μl of PfuUltra HF DNA polymerase (2.5 U/μl) was added.
- Reaction was set with 5 μl of 10× reaction buffer, 30 ng of dsDNA template, 125 ng of oligonucleotide primer #1, 125 ng of oligonucleotide
primer #2, 1 μl of dNTP mix, 3 μl of QuikSolution and ddH2O to a final volume of 50 μl. To this 1 μl of PfuUltra HF DNA polymerase (2.5 U/μl) was added.

- The temperature cycling were 95°C (1 m); 18 cycles of 95°C (50 s), 60°C (50 s) and 68°C (5 m); and a final extension of 68°C (7 m).
- Following temperature cycling, the samples were placed on ice for 2 mins to bring down the temperature to 37°C. Then, 1 μl DpnI (10U/μl) was added and incubated at 37°C for 1hr to digest the parental methylated and hemi-methylated DNA strands.

The plasmids thus obtained were transformed using XL gold competent cells.

- 45 μl of the XL10 ultra competent cells were aliquoted to prechilled Faclons
- 2 μl of the β-ME mix provided with the kit was added to the 45 μl of cells and incubated in ice for 10 mins
- 2 μl of Dpn I treated DNA from control and sample was transferred to separate aliquots and incubated for 30 mins
- Heat shock was given for 30 s at 42°C in water bath and transferred back to ice
- To the cells, 0.5 ml of preheated (42°C) NZY+ broth (components are detailed in the Appendix) was added and incubated at 37°C for hour with shaking at 225-250rpm
- The cells (250 μl) were plated on agar plates with 50mg/ml of Kanamycin
- For mutagenesis (250μl) and transformation control (5 μl) cells were spread on LB–ampicillin agar plates containing 80 μg/ml X-gal and 20 mM IPTG.

The plates were incubated at overnight at 37°C for the colonies to grow. The Mutagenesis efficiency of the pWhitescript control plasmid is calculated by dividing the number of blue colony formed by the total number of colonies observed on the plate. All the colonies formed by the control pUC18 had blue phenotype.

**Confirmation of the Mutation**: The individual colonies selected from the plate were grown in LB broth, and their plasmid isolated using commercially available kit (Qiagen). The plasmids thus isolated were checked by digestion for the presence of the insert and resequenced using gene specific primers to confirm the presence of the expected mutation in the constructs.