Chapter 6
Summary, leads and limitations
Chapter 6: Summary, Leads and Limitations

Intellectual disability, also referred to as Mental Retardation (ID/MR), is the most common developmental disorder, with a prevalence of 1-3%, and includes a highly diverse group of cognitive disorders. The condition is present as an isolated finding or as a part of a syndrome or disorder and represents an important health burden in the population, especially in the paediatric group. The known aetiology of ID is diverse and includes both genetic and environmental factors. 50% of MR cases have a genetic basis of which 8-10% are X-linked and majority are thought to be autosomal genes. For monogenic causes, genes have mainly been found on the X chromosome than on any other comparable segment on the genome. This is partially related to the greater ease in pointing out X-linked genetic disorders, especially in males. Genetic heterogeneity underlies both syndromic and non-syndromic MR. Of date, 102 genes on the X chromosome have been associated with ID with 85 of them implicated in syndromic XLID and 35 genes in NS-XLID. Further, more than 140 genes involved in syndromic autosomal ID have been reported for both dominant and autosomal recessive forms. Several gross chromosomal anomalies including both recurrent and rare copy number variants have been elucidated to cause both X-linked and autosomal forms of ID.

In spite of the exponential increase in the knowledge of the genetic aetiology, less than 50% of inherited ID is demonstrated. There are 27 XLID syndromes mapped on the X- chromosome and 13 loci for autosomal NSID wherein pathogenic genes remain unknown. Further, there are multiple reports of syndromic cases where no causal variants could be identified in the already reported genes. Identification of the genetic basis is further complicated by individually occurring rare disease variants and genetic/phenotypic heterogeneity. Thus the genetic analysis of ID including identification of new gene(s), their characterization and subsequent utilization for DNA based diagnosis continues to be one the most challenging areas of biomedical research. XLID/AID pedigrees with syndromic and non-syndromic forms wherein mutations in known genes have been ruled out are potentially useful for novel gene identification. Comparatively large numbers of nuclear/extended families available in our country are a rich resource for such efforts in novel disease gene identification.
Based on this background, the main objective of this study was to identify novel gene(s) for ID.

To attain this objective, an attempt was made to identify the potential pathogenic loci/variant(s) in the four informative families recruited. The three putative X-linked (F#1, 2 & 3) were screened systematically and ruled out for chromosomal anomalies and \textit{FMR1} mutation (the most common cause of inherited ID). Analysis of array CGH in probands of F #1 and F #2 identified potential copy number variants which might underlie their disease aetiology.

\textbf{F #1}: Duplication of 1.8Mb region at 2q13 previously reported to be involved in ID was observed in the affected sibs in one arm of the pedigree. The duplication was not maternally inherited, thus assumed to be inherited from apparently normal father. \textbf{Limitations}: (i) unavailability of paternal DNA to confirm the origin of duplication; (ii) absence of shared haplotype (microsatellite (MS) markers) on the X- chromosome between the affected maternal cousins; (iii) difficulty in explaining the etiology of the affected male individual in the second arm of this putative X-linked pedigree. \textbf{Way forward}: Additional analysis is warranted to evaluate the penetrance phenomenon and positional effect of the duplication. In the absence of identified causal gene and low frequency of the segmental duplication in the controls, it is likely to be contributory and not disease causing. Resequencing of the X chromosomal genes in the affected cousins might shed light regarding the aetiology of the disease. However, the possibility of different causal variants in the two arms also cannot be ruled out.

\textbf{F #2}: Duplication of \textasciitilde0.5Mb genomic region at Xp11.4 encompassing \textit{BCOR}, a known ID gene in all the three affected sibs was observed. It is a novel, non recurrent duplication and maternally inherited. Loss of function mutations reported in ID and literature on dosage effect during neurodevelopment substantiates the role of \textit{BCOR} in disease aetiology in this family. \textbf{Limitations}: (i) Haplotyping using MS markers on the X chromosome has shown the chromosomal segment encompassing the duplication to be likely passed on from the maternal grand father (I-3) to the affected grand children (sibs, III-1, 2 and 3); (ii) Anonymous origin of duplication, whether it was \textit{de novo} in the germline of I-3 or inherited from I-3. \textbf{Way forward}: In the absence of DNA from I-3, quantitative expression analysis of the gene in the lymphoblastoid cell lines of the affected sibs could be utilised to confirm the dosage...
of BCOR; (ii) the position effect of the segmental duplication also has to be validated by FISH.

F #3: Array CGH did not reveal any shared significant aberration among the affected cousins in the large kindred. Linkage analysis using microsatellite markers identified a significant linked region with LOD score of 3.5. Custom target re-sequencing of the linked region (Xq21-Xq24) followed by prioritisation of the novel variants by in silico softwares identified a missense mutation (NM012216.3 c.1040G>A; Q9UJV3 p.Arg347Gln) in MID2, a novel gene hitherto unreported for ID, which encodes ubiquitin ligase E3 as the likely cause of ID in this family. The mutation was observed in all affected and obligate carriers but not in any unaffected males of the family or in the population controls (n=200). When transiently expressed in HEK293T cell line, the mutation was found to abolish the function of the COS domain in the protein. The GFP tagged mutant protein accumulated in the cytoplasm instead of binding to the cytoskeleton resulting in altered sub-cellular localization. Screening of the coding exons of this gene in additional 480 unrelated individuals with idiopathic ID identified another novel variation p.Asn343Ser in one of the probands in the COS domain which was not present in controls. This study highlights the growing role of the ubiquitin pathway in ID and also, the difference in MID2 determined phenotype observed in this family compared to that of its paralogue MIDI reported in literature. Limitation: Functional characterisation of the sporadic mutation p.Asn343Ser not done yet. Way forward: (i) Additional experiments to demonstrate the non functionality of the mutant MID2 protein; (ii) Screening of large cohorts of individuals with NSID to determine the contribution of this gene in the population.

F #4: Seckel is a genetically and clinically heterogeneous disorder with six genes and seven loci implicated. Additionally, it overlaps clinically with MCPH/MOPDII syndrome. The role of known genes in the family was excluded based on MS markers and subsequent resequencing of the coding exons/UTR’s using exome sequencing approach. Multiple shared regions were observed in the family on haplotyping using MS marker based genotypes. The additional families recruited could not be used in fine mapping the multiple regions obtained in F #4. Exome sequencing followed by the analysis of the variants from the shared genomic regions in the affected individuals in the family (assuming autosomal recessive and X linked inheritance) by sequential filtration and in silico prediction revealed multiple putative candidates.
(MLL3, CA13, P2RY4, TBX22, AMOT, RBMXL3, AIFM1, RBMX and ZBTB10).

**Limitations**: (i) Lenient filtering of the variants for quality parameters were employed as the average depth of coverage was moderate. Thus some of the prioritised variants could be false positives; (though better than false negatives); (ii) The UTR regions of many of the genes present in the shared genomic co-ordinates were not covered in the exome resequencing data. **Way forward**: The role of coding/UTR variants in the known genes of Seckel, MOPDII and MCPH has been ruled out except for the UTR in *ATRIP*. However, since the exome data became available by then, haplotypes were constructed using variants called in the region encompassing *ATRIP*. Both the haplotypes were shared by the affected and unaffected individual and thus the role of this gene in the etiology of the disease can be excluded. Further, the putative candidates obtained following the analysis of exome sequencing have to be validated by resequencing in the affected individual and its segregation in the family. Further, the frequency of the variants thus prioritised has to be screened in the population controls to unequivocally determine its pathogenicity.

In summary, next generation sequencing is a powerful tool to discover disease causing mutations in novel genes. The combination of linkage analysis with next generation sequencing technology was successfully employed in identifying *MID2*, a novel ID gene in Family #3. Based on the high pathogenicity, evolutionary conserved scores, interaction with known ID genes as well as functional consideration, mutation in *MID2* was confirmed to be causative for ID in F #3. This is the first report of ID caused by mutation in this gene and also the highlight of the thesis. Further, exome resequencing of Family #4 with genetically heterogeneous Seckel syndrome has identified a few novel, putative causal variants warranting additional analysis. Additionally, the data generated using next generation sequencing tools in this study could serve as a variant repository to facilitate identification of true causal variants and thereby novel ID genes. Results of these four families are excellent examples of genetic heterogeneity/complexity underlying ID and the continuing biomedical challenge the condition poses.