LOCUS AND GENE SEARCH IN THREE DISORDERS

by

Renin Hazan

B.S. in Molecular Biology and Genetics/Chemistry, Boğaziçi University, 2007

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
in
Molecular Biology and Genetics

Boğaziçi University

2009

LOCUS AND GENE SEARCH IN DISORDERS

APPROVED BY:

Prof. Aslı Tolun……..…….
(Thesis Supervisor)

Prof. Hande Çağlayan……...……..

Assist. Prof. Eda Tahir Turanlı……...……

DATE OF APPROVAL 05.08.2009

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis supervisor, Prof. Aslihan Tolun, whose guidance and support throughout the study enabled me develop myself and gain a wiser understanding of research.

I am heartily thankful to all present and formal KOMMAGENE lab members, especially to Sibel Uğur, Murat Çetinkaya and Tarık Bozoğlu for all their help in learning lab work and comments, Çiğdem Köroğlu for her close friendship and Yeşerin Yıldırım. I would like to send my thanks to my friends in the department, including Emir Tınaztepe, Murat Atasoy and Melek Aslı Kayserili.

I would like to express my thanks and love for all my family members. I must acknowledge the support of my mother.

I would like to acknowledge to The Scientific and Technological Research Council of Turkey BIDEB for providing me financial support throughout the study.

ABSTRACT

LOCUS AND GENE SEARCH IN THREE DISORDERS

Identification of genetic defects associated with hereditary diseases is important in understanding cellular mechanisms and molecular etiology of diseases, developing diagnostic tests and novel therapeutic techniques. In this study, genetic analysis was performed to map genetic loci and identify the responsible gene in three families afflicted with different autosomal recessive hereditary diseases: azoospermia, diabetes mellitus type 1, microhydranencephaly. Gene localizations involved genome scan followed by fine mapping. Localization studies were completed for all diseases, while candidate gene approach led to the identification of the gene responsible for one of them.

Azoospermia is absence of detectable level of spermatozoa in the semen. In a consanguinous family with three affected sibs, the condition was mapped to 17pter-13.1 and 14q11.2-12. Seven candidate genes were analyzed, and the disease gene was identified as *SPAG7*, a homolog of fox sperm acrosomal protein. The mutation R136W resided in a 13 amino acid-cluster conserved in all mammals.

Microhydranencephaly is a rare developmental defect of the brain, where the unusual association of microcephaly and hydranencephaly is observed. In a nuclear family with two affected sibs, genetic linkage analysis and haplotype construction followed by verification with microsatellite markers led to the identification of several candidate loci, with the strongest one at 6p25.3-p25.2. Five candidate genes were analyzed but no mutation was found.

Diabetes Mellitus Type 1 exhibits complex inheritance. A consanguineous family having four of thirteen sibs affected suggested autosomal recessive inheritance. The strongest candidate locus identified by using genome scan data was 8q24.23. The locus harbors gene *FAM135B*.

ÖZET

ÜÇ HASTALIKTA LOKUS VE GEN ARANMASI

Genetik hastalıklardan sorumlu genlerin bulunması, genlerin katıldıkları hücre mekanizmalarının ve moleküler yolaçların bulunmasında, hastalıkların tanı ve sağaltımı için gerekli yeni tekniklerinin geliştirilmesinde önemli rol oynar. Bu çalışma kapsamında, otozomal çekinik geçişli üç kalıtımsal hastalıkta genetik bağlantı analizi kullanılarak hastalıklardan sorumlu genleri içeren bölgeler aranmış, hastalıkların ikisinde anlamlı bölgeler ayrıntılı olarak haritalanarak aday genler incelenmiştir. Çalışılan hastalıklar azospermi, tip 1 diyabet ve mikrohidranensefalidir.

Azospermi semende saptanabilir seviyede spermatozoa bulunmaması durumudur. Çalışmada üç hasta bireyli ve sık akraba evliliği bulunan bir ailede hastalık kromozom 17p13.2-p13.1 ve 14q11.2-q12'e haritalandı. Sonra da hastalık geninin 'sperm associated antigen 7' üreten *SPAG7* geni olduğu gösterildi. Mutasyon (R136W) memelilerde on iki amino asitlik korunmuş bir bölge içerisinde yer almaktadır.

Mikrohidranensefali mikrosefali ve hidranensefalin birlikte gözlemlendiği nadir bir gelişimsel anomalidir. İki hasta kardeşli bir ailenin genom taraması bulguları ile yapılan genetik bağlantı analizi ve ardından mikrosatelit belirteçleriyle yapılan doğrulama sonucunda hastalıkla ilişkilendirilen birden çok bölge ortaya çıkarıldı. En kuvvetli aday bölge olan 6p25.3-p25.2'de beş aday gen incelendi.

Tip 1 diyabet karmaşık geçişli bir hastalıktır. Akraba evliliği yapmış bir ailede onüc kardeşten dördünde hastalığın gözlenmesi otozomal çekinik kalıtıma işaret etmiştir. Mikrosatelit ve SNP belirteçleri ile yapılan genom taramaları sonucunda hastalık 8q24.23 bölgesine haritalandırıldı. Bölgede *FAM135B* bulunmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APS	Ammonium peroxodisulphate
AR	Androgen receptor
AZF	Azoospermia factor
bp	Base pair
CBAVD	Congenital bilateral aplasia of vas deferens
CFTR	Cystic fibrosis transmembrane conductance regulatory protein
cM	Centimorgan
DC	Dendritic cells
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetate
FSH	Follicle-stimulating hormone
GnRH	Gonadotrophin releasing hormone
IDDM	Insulin-dependent diabetes mellitus
KS	Klinefelter's syndrome
LH	Luteinizing hormone
Lod	Log of odds
Mb	Megabase
MCPH	Microcephaly
MHAC	Microhydranencephaly
MCPH	Microcephaly
min	Minute
MODY	Maturity onset diabetes of the young
NCBI	National Center for Biotechnology Information
NOA	Non-obstructive azoospermia
NHLBI	National Heath, Lung and Blood Institute
OA	Obstructive azoospermia
OFC	Occipitofrontal head circumference
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism

rpm	Revolution per minute
SCAFUD	Scanning fluorescent detector
SDS	Sodium dodecyl sulphate
sec	second
SNP	Single nucleotide polymorphism
SRY	Sex determining region
SSCP	Single strand conformation polymorphism
TEMED	N, N, N, N'-Tetramethylethylenediamine
VNTR	Variable number tandem repeats
W	Watt
UTR	Untranslated region

1. INTRODUCTION

This work is comprised of genetic studies that were performed with the aim of identifying the loci of three recessive inherited disorders and subsequently the genes responsible for the diseases: azoospermia, microhydranencephaly (MHAC) and insulin-dependent diabetes mellitus (IDDM).

1.1. Genetic Linkage Analysis

Genetic linkage analysis is utilized to identify genomic regions harboring genetic disease genes. An observation of genetic linkage is the proof that the two loci are inherited from parent to offspring in statistically significantly higher frequencies than what is expected under independent inheritance. Two loci that are close to each other on the chromosome tend to segregate together. During meiosis, crossovers between homologous chromosomes give rise to a new combination of alleles. As the distance between the loci increases, the probability for recombination rises. Recombination frequency, denoted by θ, is the frequency at which recombination will occur between two loci at meiosis. The unit of recombination frequency is centimorgan (cM). One cM corresponds to one per cent probability that the alleles at the two loci will be separated by a recombination event during meiosis. When two loci on a chromosome are sufficiently far apart, at least one recombination event occurs per meiosis. As a result, these loci assort independently and the recombination frequency is 0.5, just as when the loci are on different chromosomes (Teare and Barrett, 2005).

While analyzing recombination events, complications due to the presence of multiple recombination events and interference should be considered. Two or more recombination events can occur between loci that are far apart. When a double exchange takes place between two loci, the recombinant chromosomes are indistinguishable from the non-recombinant chromosomes when analyzed at the two loci, despite the exchange in the middle region. This leads to an underestimation of the true genetic map distance. It is possible to overcome this problem by using recombination data across short subintervals, where multiple crossovers do not take place. Also, the occurrence of one recombination

event tends to interfere with the occurrence of a recombination in its vicinity. This phenomenon is known as interference. It results in a deviation from the true genetic map distance.

Application of genetic linkage analysis to identify disease gene loci relies on polymorphic markers. There are several types of DNA markers. A restriction fragment length polymorphism (RFLP) is a single nucleotide variation that results in either the presence or the absence of a target for a restriction enzyme (Donis-Keller *et al.*, 1987). It has limited informativeness because the maximum number of alleles is two. As the first generation of DNA markers, RFLPs were extensively used in human genetic research in the 1980's. Variable Number Tandem Repeats (VNTRs) are tandem repeats of 14 to 100 nucleotide-long sequences. Despite their highly polymorphic nature, they are not easily amplified by PCR and have uneven distributions throughout the genome. Microsatellite markers are tandem repeats of two to five nucleotide long sequences. They became the basis of second-generation linkage map of the human genome due to their abundance, even distribution in the genome, highly polymorphic nature and convenience in amplification. (Weissenbach *et al.,* 1992). Single nucleotide polymorphisms (SNPs) are single base differences that have at least one per cent frequency in the population. In recent years, they have been extensively used in genome scanning for both linkage and association studies. Although SNP markers have limited informativeness, they allow a dense and uniform coverage of all chromosomes. In addition, haplotyping a large number of SNPs can be performed in automated systems in a convenient and cost-effective manner. SNP markers have a lower mutation rate compared to repeat sequences. As a result, microsatellite markers are generally assumed to show more recent changes in haplotypes of families and used in conjugation with SNP marker data in linkage analysis.

1.1.1. Lod Score Analysis

Parametric (model-based) linkage analysis depends on the investigation of the cosegregation of genetic loci in pedigrees. The genetic model that specifies the mode of inheritance of the disease of interest, the frequency of the disease allele and the degree of penetrance are required for parametric linkage analysis. The major quantity of interest is the recombination fraction (Kruglyak *et al*., 1996).

Morton (1955) introduced lod score analysis that facilitated the search for the chromosomal regions that are linked to the disease in a family. It is based on a comparison of two estimates that depend on two alternate hypotheses: the null hypothesis of no linkage ($\theta = 0.50$) and the alternative hypothesis of linkage ($\theta < 0.5$). The ratio of the two values, namely the odds ratio, is the ratio of the probability of observing the distribution of two traits in a pedigree under linkage at θ to the probability of observing it under absence of linkage. Lod score is the logarithm of odds ratio, and it is shown by function $Z(\theta)$. Lod scores are calculated for a range of θ. The value θ that gives the maximum lod score value is the maximum likelihood estimate.

Various computer programs for linkage analysis have been developed and are available free of charge (Dudbridge, 2003). User-friendly software platforms that are created recently allow usage of various linkage programs with a standardized input data format and present the results in graphical format. EasyLINKAGE (Lindner and Hoffman, 2005) is one of the software platforms, applicable for both SNP and microsatellite marker data.

1.1.2. Homozygosity Mapping

Homozygosity mapping is an efficient way of mapping recessive diseases in consanguineous families. It is based on the fact that the chromosomal region containing the disease locus is homozygous by descent in the affected individuals. It allows identification of a recessive disease locus by detecting a shared homozygous region in all affected individuals that have descended from a common ancestor (Lander and Botstein, 1987). Generally the frequency of a trait is so low that the disease is observed only in families with parental consanguinity. Inbred families with multiple affected siblings are used widely in linkage studies. Such studies have allowed the identification of loci and subsequently the genes responsible for numerous disorders.

1.2. Candidate Gene Approach

Candidate gene approach is utilized to identify the disease gene residing in the identified disease locus. Selection of a candidate gene responsible for a disease is based on

multiple parameters, such as the function of the gene, its spatial and temporal expression and the phenotype of animal models. Those genes that are involved in a pathway likely leading to disease pathogenesis when perturbed, expressed in tissues relevant to disease phenotype or have been inactivated in animal models exhibiting similar disease phenotypes are considered good candidates and selected for mutational screening in patients.

1.3. Azoospermia

Azoospermia is defined as the absence of a measurable level of spermatozoa in the semen of the affected male, and oligozoospermia is characterized by less than 10 million spermatozoa per ml of semen, according to World Health Organization norms. The complete absence of spermatozoa in the semen can be due to either physical blockage in the genital track, known as obstructive azoospermia (OA), or spermatogenic failure, known as non-obstructive azoospermia (NOA).

OA is one of the causes of male infertility, accounting for 7.4 per cent of the cases. The blockage in seminal ducts can be congenital or acquired. External compression (due to neoplasms and cysts), infections and trauma are several causes of acquired OA. Obstruction in vas deferens, epididymis or ejaculatory ducts is associated with distinct syndromes with variable phenotypes. In congenital bilateral aplasia of vas deferens (CBAVD, MIM 277180), degenerative changes in vas deferens occur due to defects in cystic fibrosis transmembrane conductance regulatory protein (CFTR). It accounts for 60-70 per cent of OA cases (Ferlin *et al.*, 2007). It occurs either in isolation or in association with cystic fibrosis. In Young Syndrome (MIM 279000), complete bilateral obstruction of epididymes is observed in the absence of anatomical malformation of seminal ducts. The syndrome shows progression with age and leads to OA in association with bronchopulmonary disease even many years after entering puberty (Meschede *et al*., 1998).

NOS, can be associated with defects in testicular development of germ cells and supporting Sertoli cells, spermatogenesis or hormonal regulation (McLachlan *et al*., 1998). The genetic causes of NOS include chromosomal abnormalities, Y-chromosome microdeletions and mutations in the components of hormonal pathways. Idiopathic azoospermia refers to cases where other genetic abnormalities are suspected.

Both structural chromosomal abnormalities and aneuploidies can lead to NOA or severe oligoozospermia. Klinefelter's syndrome (KS), due to the presence of an extra X chromosome in males, is detected in up to ten per cent of azoospermia and five per cent of severe oligozoospermia (Foresta *et al.*, 2005). KS is characterized by abnormal male breast tissue development (gynecomastia), small firm testes and azoospermia. Other aneuplodies such as 47,XYY are found in higher prevalence in infertile males than in the general male population. In 90 percent of 46,XX male syndromes, sex determining region (SRY) is translocated to the tip of either the X chromosome or an autosome. Those individuals have male phenotype in terms of internal and external anatomy, yet the condition is associated with small testes, short statue, and azoospermia.

Microdeletions at Yq11 are the most common etiologic factors and observed in 10 - 20 per cent of azoospermia or severe oligozoospermia cases. Deletions at distal Yq in azoospermic men suggested the presence of a spermatogenesis gene or azoospermia factor (AZF) (Tiepolo and Zaffari, 1976). One of the three overlapping regions designated AZFa, AZFb and AZFc had been found to be deleted frequently in azoospermic males that are otherwise healthy. Among the several genes in AZF region, USP9Y is the strongest candidate gene to be responsible for the condition.

Genetic mutations that disturb hormonal regulation of reproductive system also lead to NOA. Levels of gonadotrophin releasing hormone (GnRH), luteinizing hormone (LH), follicle-stimulating hormone (FSH), androgen, estrogen and prolactin are screened routinely in infertility clinics to unravel the etiology of azoospermia. Deficiencies in GnRH, LH and FSH as well as related receptor abnormalities and an excess of estrogen, prolactin or glucocorticoid are known to variably affect internal ductal differentiation, external genitalia development and spermatogenesis (Chan and Schlegel, 2000).

Androgens are steroid hormones which are required for normal male sex differentiation, testicular descent and spermatogenesis. Androgen receptor (AR), encoded by a single copy of *AR* gene on the X-chromosome, mediates the effect of androgen by acting as a transcription factor. The complete absence of AR leads to complete androgen insensitivity syndrome, resulting in the complete feminization of male with normal karyotype. *AR* mutations with partial effects on function can lead to a wide spectrum of

milder androgen insensitivity syndromes, from ambiguous genitalia to azoospermia or oligozoospermia. A correlation between the extent of the CAG repeat in *AR* and the risk of impaired spermatogenesis has been established. Amino-acid substitutions in AR were also identified in azoospermia patients (McLachlan *et al*., 1998).

By the help of mouse models, a large number of genes essential for spermatogenesis and male fertility have been identified (Matzuk and Lamb, 2002; Matzuk and Lamb, 2008). Despite the growing knowledge on molecular players of spermatogenesis mechanism, genetic defects have been detected in only a limited number of azoospermia patients. Even in cases where the underlying chromosomal abnormalities or Y chromosome deletions are known, the causative genes can not be easily identified. *SYCP3* and *USP9Y* are two genes in which causative point mutations have been reported. *SYCP3,* a meiosis specific component of synaptonemal complex, is essential for spermatogenesis. Gene defects cause azoospermia due to perturbation of meiosis in an autosomal dominant fashion (Miyamoto *et al.*, 2003). A *de novo* point mutation at the splice-donor site of *USP9Y* that results in exon skipping and subsequent protein truncation was identified in an azoospermia patient (Sun *et al*., 1999). Mutational screening and association studies revealed several genetic polymorphisms predisposing to azoospermia. Functional roles in spermatogenesis were suggested for various genes based on differential expression between azoospermic and healthy individuals.

No autosomal recessive locus or gene has been reported to be responsible for idiopathic azoospermia. In our laboratory, a consanguineous family with three affected sibs without micro Y-deletions had been studied to identify an autosomal recessive genetic locus for idiopathic azoospermia. A genome scan using low density microsatellites (180 in total) and subsequent fine-mapping previously identified a candidate disease locus at 17pter-p13.1. In the framework of this study, further fine-mapping of the identified locus was done and a SNP genome scan was performed to investigate whether there were other candidate loci. Linkage analysis and verification studies identified a second candidate locus at 14q11.2-q12. Using candidate gene approach, the gene responsible for the disorder was identified and a large group of men having azoospermia or oligozoospermia was screened for mutations in the gene.

1.4. MHAC

Microhydranencephaly (MHAC; MIM 605013) is a rare developmental brain defect, in which severe hydranencephaly is associated with microcephaly. Microcephaly (MCPH) is the condition in which the occipitofrontal head circumference (OFC) is at least three standard deviations below the mean value expected for to the age and sex of the individual. MCPH can be found as isolated where it is not complicated by other abnormalities or as syndromic where it is associated with other abnormalities. It can develop due to either genetic or environmental causes. Autosomal recessive, autosomal dominant and X-linked forms of inheritance have been observed in isolated MCPH. It can be present at birth (congenital) or can develop later. Six loci have been identified so far for autosomal recessive MCPH (Table 1.1). Causative genes have been determined at four of the loci. The identification and functional analysis of genes responsible for the condition suggested the presence of a common centrosomal mechanism responsible for regulating the cerebral cortex size (Abuelo, 2007).

Anencephaly (MIM 206500) is the most prevalent congenital malformation of the brain. It is a neuronal tube defect, presenting restricted development of cerebrum and/or other brain parts. Both environmental and autosomal hereditary factors contribute to the condition. In hydranencephaly, spinocerebellar fluid partially replaces brain regions. Almost all cases are sporadic. Its prevalence is about one in 5000 as an isolated defect that is not associated with any other malformation. Early occlusion or developmental defect on carotid or cerebral arteries reduce cerebral hemispheres to gliomatous membrane, while the posterior parts receiving blood through posterior cerebral circulation are well-developed (Csabay *et al.,* 1998).

MHAC exhibiting familial inheritance was described first in a large inbred Anatolian family by Kavaslar *et al*. (2000). The affected individuals had severe mental and motor retardation, very small body size and very small OFC. For instance, the youngest patient who was 4.5 years old at the time of the study had a height of 68 cm, weight of 3.8 kg and OFC of 30.9 cm. By homozygosity mapping, this brain abnormality was mapped to chromosome 16p13.3-p12.1 and called “Familial autosomal recessive

microhydranencephaly" (MIM 605013). The disease gene was identified in our laboratory (Bozoğlu, 2008).

Table 1.1. Loci and Genes for MCPH

MCPH#	Locus	Gene	Description	Reference
MCPH1	8p23	*MCPH1*	Microcephalin 1	Jackson *et al.*, 2005
MCPH2	19p13.1-q13.2	-	-	Roberts *et al*., 1999
MCPH3	9p34	*CDK5RAP2*	Cyclin-dependent kinase 5 regulatory-associated protein 2	Moynihan *et al*., 2000, Bond *et al*., 2005
MCPH4	15q15-q21	-	-	Jamieson *et al*., 1999
MCPH5	1q31	*ASPM*	Abnormal spindle-like microcephaly-associated	Pattison *et al*., 2000, Bond *et al*., 2002
MCPH6	13q12.2	*CENPJ*	Centromere associated protein J	Leal *et al*., 2003, Bond *et al*., 2005
MCPH7	1p32	*STIL*	SLC/TAL1-interrupting locus	Kumar *et al*., 2009

Other familial cases which exhibit phenotypes similar to MHAC were later reported (Alexander *et al*., 1995; Scham *et al.*, 2004). Recently, two Slovak brothers born to healthy parents with no known consanguinity were diagnosed with MHAC (Behunova *et al*., in press). Recurrence of the disease strongly suggested genetic etiology and recessive inheritance. Genetic analysis showed that the disease was not linked to the known MHAC locus 16p13.13 (Behunova *et al*., in press). This exclusion of linkage to MHAC locus indicated genetic heterogeneity for the disease.

In this study, genetic linkage analysis was performed in the Slovak family to find a second disease locus, and subsequent candidate gene approach was used in an attempt to identify the causative gene.

1.5. Diabetes Mellitus

Diabetes Mellitus is a heterogeneous group of diseases characterized by chronic hyperglycemia which is the presence of a high level of glucose in the blood. In healthy

individuals, the rise of glucose level in blood after a meal triggers the release of insulin hormone from pancreatic beta (β) cells residing on islets of Langerhans. Insulin causes muscle and fat cells to remove glucose from blood and liver cells to metabolize glucose, so that glucose level in blood returns to its normal level. There are two major types of diabetes mellitus: diabetes type 1 and diabetes type 2.

Diabetes type 1, also known as insulin-dependent diabetes mellitus, is a multifactorial autoimmune disorder, where immune-mediated destruction of insulin-producing β cells takes place. Multiple genetic and environmental factors are known to contribute to the disease. The disease commonly onsets at a young age and the age of onset can be as early as one year. Despite the fact that the diagnosis is accomplished before the age of 18 in the majority of cases, autoimmune destruction starts much earlier. Thus, β cells lose their capacity to secrete insulin by the time of diagnosis. Liver autopsy data reveal that specific destruction of cells occurs due to the infiltration of islets of Langerhans by dendritic cells (DCs), macrophages and T lymphocytes (CD4+ and CD8+). CD4+ helper cells act by recognizing an extracellular antigen and promoting inflammation through cytokine release. CD8+ cytotoxic T lymphocytes respond to endogenous self antigens and kill the target cells. Insulin, glutamic acid decarboxylase, its 65 kDa isoform and insulin autoantigen-2 are three major autoantigens that act as T-lymphocyte-mediated destruction markers (Mehers and Gillespsie, 2008).

Genome-wide linkage studies using affected sib-pairs and association studies suggest a disease model of multiple genes, each with a minor contribution to the risk of IDDM (Ounissi-Benkalha and Polychronakos, 2008). One major susceptibility locus (IDDM1) at 6p21.3 harbors the major histocompatibility complex (MHAC), while a second locus (IDDM2) at 11p15.5 includes a polymorphic region in the promoter of the insulin gene. In 60-70 per cent of IDDM families, association with one of those two regions has been reported. More than 24 candidate susceptibility loci have been reported, but polymorphisms predisposing to IDDM have been identified in only a fraction of them.

Type 2 diabetes is a heterogeneous metabolic disease, where concomitant or interdependent defects in insulin secretion of pancreatic β cells and insulin sensibility of peripheral tissues result in decompensation of β cell function and chronic hyperglycemia.

A wide range of diabetes phenotypes are observed, depending on the etiology. Polygenic form of type 2 diabetes which is due to the interaction of several susceptibility loci and environmental factors develop in adulthood with insulin resistance and generally associated with overweight and obesity. Monogenic forms of the type 2 diabetes are generally associated with pancreatic β cell dysfunction and have early age of onset.

Maturity onset diabetes of the young (MODY, MIM 606391), the most common form of monogenic diabetes, is observed in one to two per cent of the patients. Defects in insulin secretion arise due to defects in B cells. The condition exhibits an autosomal dominant mode of inheritance and disease onset occurs in the second or third decade of life (Malecki and Mlynarski, 2008). The phenotypes associated with MODY have a wide spectrum, consistent with the genetic heterogenity of the condition. Nine genes, hepatocyte nuclear factor-4α, -1α, and -1ß (*HNF-4α ,-1 α, and -1ß*), glucokinase (*GCK*), insulin promoter factor 1α (*IPF-1α*), neurogenic differentiation factor 1 (*NEUROD1*), Kruppel-like factor 11 (*KLF11*), carboxyl-ester lipase (*CEL*), and Paired box gene 4 (*PAX4*) are identified as responsible genes (Yamagata *et al*., 1996a; Yamagata *et al.*, 1996b; Horikawa *et al.*, 1997; Stoffers *et al*., 1997; Vionnet *et al*.1992; Malecki *et al.*, 1999; Neve *et al*., 2005; Raeder *et al.,* 2006; Plengvidhya *et al*., 2007).

Maternally inherited diabetes with deafness (MIDD, MIM 520000) is caused by mutations in mitochondrial genome and leads to impairment of β cell insulin secretion. The degree of heteroplasmy determines the clinical phenotype, and hence the condition can imitate either type 1 or type 2 diabetes. Numerous mitochondrial mutations have been identified including MTTL1 (Reardon *et al*., 1992), MTTE (Vialettes *et al*., 1997) and MTTK (Kameoka *et al*., 1998). These mutations affect glucosensoric function of β cells and reduce their ability to synthesize insulin. The disease is generally diagnosed in early adulthood. It is associated with conditions, including hearing abnormalities or complex syndromes such as mitochondrial encephalopathy, lactic acidosis and stroke-like episodes (MELAS).

Permanent Neonatal Diabetes Mellitus (PNDM; MIM 606176) is characterized by very early age of onset. Mutations in *KJN11*, encoding a subunit of inwardly rectifying ATP-sensitive potassium channel, are frequently observed in PNDM (Gloyn *et al*., 2004).

Permanent activation and widely opening of potassium channels of β cells in return cause hyperpolarization of cells and closure of calcium channels. Hence, impairment in insulin secretion arises. Severity of PNDM can decrease with age. The ectopic manifestation of the disease in the brain, heart and skeletal muscles is the basis of neurological symptoms such as epilepsy, muscle weakness and developmental delay (DEND). Activating mutations in *ABCC8* encoding SUR1, the regulatory subunit of KATP channel in pancreatic B cell, is observed in a closely related form of PNDM (Babenko *et al*., 2006). Elevation of stimulatory action of SUR1 on Kir channel causes development of transient diabetes mellitus. Another form of PNDM is due to mutations in the glucokinase gene. Patients bearing mutations in the gene have very low birth weights as well as disease onset in the first weeks of life (Njolstad *et al*., 2001).

Wolfram Syndrome (DIDMOAD) is a rare progressive neurodegenerative disorder with autosomal recessive inheritance. Its symptoms include insulin-dependent diabetes mellitus with onset at the first decade of life and bilateral progressive optic nerve atrophy. It arises due to defects in *WLS1* gene, encoding ER Calcium channel. Mutations in *WLS1* gene were also identified in patients that are non-syndromic, non-autoimmune, juvenile onset diabetes (JOD) (Zalloua *et al*., 2008).

In this study, a genetic investigation was performed in a consanguineous IDDM family with four affected sibs to identify a monogenic autosomal recessive locus for the disease. Genome scan data of microsatellite and SNP markers were used together in linkage analysis. A small candidate locus was identified. The onset of the disorder in the sibs were 12.3, 8.94 and 9.6 years.

2. PURPOSE

The aim of this study was to map the disease loci and subsequently determine the underlying genetic defects in three hereditary diseases: azoospermia, IDDM and MHAC. Identification of the causative mutations has become an important part of post genome project area. Azoospermia and IDDM are two disorders with high prevalence in the population, affecting the social quality of life and health of individuals, respectively. Determining the genetic defect in these disorders is important in understanding cellular mechanisms and molecular etiology of the diseases as well as developing diagnostic tests and novel therapeutic techniques. The latest disorder is a very rare disorder of brain development. Identifying the responsible gene would have shed light to understanding this important developmental process.

3. MATERIALS

3.1. Subjects

Informed consent has been obtained from/for all subjects participated in the study. The research was approved by the Committee on Research with Human Participants at Boğaziçi University.

3.1.1. Azoospermia

A consanguineous family afflicted with azoospermia was genetically studied for locus and subsequent gene identification (Figure 3.1). Peripheral blood samples from three affected brothers, their two sibs and mother were obtained by Aslihan Tolun. DNA samples from 5 unrelated azoospermic men and 2 having each an affected brother (A2-8) without karyotype or genetic investigations were provided by Ömer Şanlı, from 5 CBAVD patients (A9-13) by Dr. Ender Altıok and from 49 azoospermic (SA1-49) and 19 oligoozospermic men (O01-19) analyzed for karyotype and genetic investigations by Dr. Mahmut Balkan.

3.1.2. IDDM

A consanguineous family afflicted with IDDM was studied for locus identification (Figure 3.2). Peripheral blood samples from the family members were provided by Drs. Himmet Karazeybek and Ali Ataş at Harran University.

3.1.3. MHAC

DNA samples from the parents and the two affected sibs afflicted with MHAC were provided by Dr. Jana Behunova. Genome scan data was provided by Prof. Bernd Wollnik.

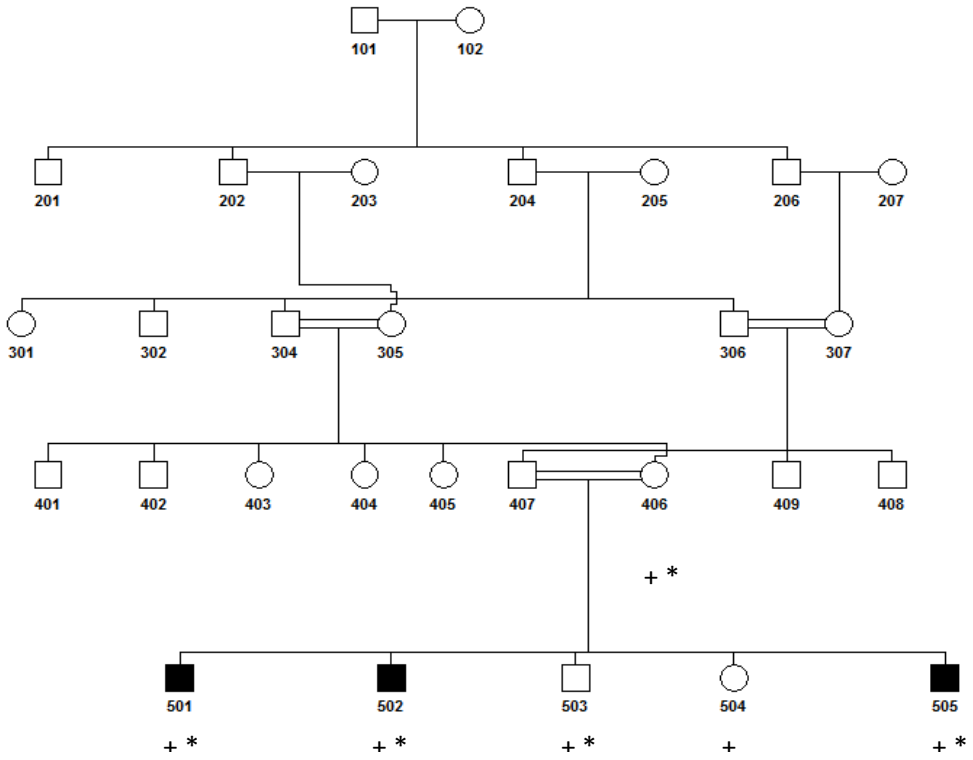


Figure 3.1. Pedigree diagram for the azoospermia family. DNA samples subjected to genome scan with microsatellite markers are marked with + and with SNP markers marked with * .

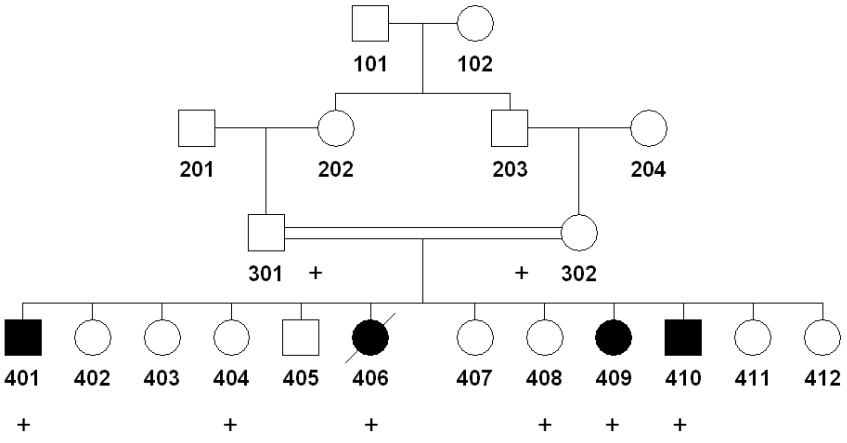


Figure 3.2. Pedigree diagram for IDDM family. DNA samples available for genome scans with microsatellite and SNP markers are marked with +.

3.2. Chemicals

All solid and liquid chemicals used in were purchased from Merck (Germany), Sigma (USA), Riedel da-Haen (Germany), Carlo Erba (Italy) or Biochrom (Germany).

3.3. Buffers and Solutions

All buffers and solutions were prepared in dH_2O unless otherwise stated.

3.3.1. DNA Extraction from Peripheral Blood Samples

Cell Lysis Buffer	:	155 mM NH_4Cl, 10 mM $KHCO_3$ 0.1 mM Na_2EDTA, pH 7.4
Nucleus Lysis Buffer	:	400 mM NaCl, 2 mM Na_2EDTA, 10 mM Tris, pH 8.2
Proteinase K	:	20 mg/ml Proteinase K
Sodiumdodecylsulfate (SDS)	:	10 per cent SDS (w/v)
Ammonium Acetate	:	7.5 M CH_3COONH_4
Ethanol	:	99 per cent Ethanol
TE buffer	:	1 mM EDTA, 20 mM Tris-HCl, pH 8.0

3.3.2 Polymerase Chain Reaction

10 X PCR Buffer 1	:	20 mM $MgCl_2$, 500 mM KCl, 100 mM Tris-HCl, pH 8.3
10 X PCR Buffer 2	:	20 mM $MgSO_4$, 100 mM KCl, 200 mM Tris-HCl, pH 8.8, 100 mM $(NH_4)_2SO_4$, 1 per cent Triton X-100, 1 mg/ml BSA
$MgCl_2$	:	25 mM $MgCl_2$ (Roche, Germany)
dNTP	:	12.5 mM each of dATP, dCTP, dGTP, dTTP (Roche, Germany)
Betaine	:	5 M Betaine (Promega, USA)

3.3.3. Agarose Gel Electrophoresis

Agarose	:	2 per cent Agarose in 0.5 X TBE
10 X TBE Buffer	:	20 mM EDTA, 0.89 M Boric Acid, 0.89 M Trizma base, pH 8.3
6 X Loading Buffer	:	10 mM Tris-HCl, pH 7.6, 50 per cent Glycerol, 60 mM EDTA, 2.5 mg/ml Xylene Cyanol and/or 2.5 mg/ml Bromophenol Blue
Ethidium Bromide	:	10 mg/ml

3.3.4. Polyacrylamide Gel Electrophoresis (PAGE)

40 per cent Acrylamide (Stock)	:	40 per cent Acrylamide-bisacrylamide (19:1)
8 per cent InstaGel (Denaturing)	:	8 per cent Acrylamide-bisacrylamide (19:1), 8.3 M urea in 1 X TBE Buffer
APS	:	10 per cent Ammonium Peroxidisulfate
TEMED	:	N,N,N,N-tetramethylethylenediamine
10 X Loading Buffer	:	95 per cent Formamide, 0.05 per cent Xylene Cyanol, 0.05 per cent Bromophenol Blue, 20 mM EDTA

3.3.5. Single Strand Conformational Polymorphism (SSCP) Gel Electrophoresis

40 per cent Acrylamide (37.5:1)	:	40 per cent Acrylamide-bisacrylamide
8 per cent Acrylamide (37.5:1, non-denaturing)	:	8 per cent Acrylamide-bisacrylamide in 0.6 X TBE Buffer
Glycerol	:	5 per cent glycerol in gel solution

3.3.6. Silver Staining

Staining Buffer	:	0.1 per cent $AgNO_3$
Developing Buffer	:	1.5 per cent NaOH, 0.01 per cent $NaBH_4$, 0.015 per cent formaldehyde

3.4. Enzyme

Taq DNA Polymerase was purchased from Roche, Germany or Qiagen, USA.

3.5. Kits

The commercial kits utilized in the study are given in Table 3.1.

Table 3.1. The commercial kits used, their utilization and manufacturers

Name of Kit	**Used For**	**Company**
Qiagen Taq PCR Core Kit with Q solution	Amplification of difficult templates	Qiagen, USA
GC–Rich PCR System	Amplification of GC-rich templates	Roche, Germany
QIAquick Gel Extraction Kit	Elution of DNA fragments from agarose gels	Qiagen, USA
QIAquick PCR Purification Kit	PCR product purification	Qiagen, USA
Repli-G	Amplification of whole genome	Qiagen, USA

3.6. Oligonucleotide Primers

Program Primer3 was used to design oligonucleotide primer pairs. All primers purchased for the study were custom synthesized in the Massachusetts General Hospital (MGH) DNA Synthesis Core (USA). They were obtained in lyophilized form and dissolved in one ml dH_2O. Dilutions at 10 µM concentration were used for PCR reactions.

3.7. DNA Molecular Weight Markers

Lambda DNA/*Hind*III and pUC19 DNA/*MspI* markers were purchased from Fermentas (Lithuania).

3.8. Equipment

Autoclave	:	Midas 55 (Prior Clave, UK), AMB430T (Astell, UK)
Balance	:	Electronic Balance (Precisa, Switzerland)
Centrifuges	:	MiniSpin Plus (Eppendorf, Germany), Universal 16R (Hettich, Germany), Allegra X-22R, J2-MC (Beckman Coulter, USA)
Deep Freezers	:	- 20 ^{o}C (Bosch, Germany), - 20 ^{o}C (AEG, Turkey), - 80 ^{o}C Ultra Freezer (Thermo Scientific)
Documentation System	:	GelDoc Documentation System with Quantity One 1-D Analysis Software (BioRad, USA)
Electrophoretic Equipment	:	DCode Universal Mutation Detection System, Horizontal Electrophoresis Gel Box, Sequi-Gen Sequencing Cell (BioRad, USA), Primo Minicell Horizontal Gel System (Thermo Scientific, USA)
Incubator	:	Orbital (Gallenkamp, Germany)
Magnetic Stirrer	:	MR3001 (Heidolph, Germany)
Micropipettes	:	Pipetman (Gilson, France)
Minishaker	:	Rotamax 120 (Heidolph, Germany)
Ovens	:	Heraus (Germany), EN400 (Nüve, Turkey)
Refrigerator	:	4 ^{o}C (Arçelik, Turkey)
Spectrophotometers	:	Nano Drop 1000 (Thermo Scientific,USA)
Thermal Cyclers	:	MyCycler (BioRad, USA) PTC-200 (MJ Research, USA)
Vortex	:	Reax vortex mixer (Heidolph, Germany) Lab Dancer Vario (Roth, Germany)
Waterbath	:	Grant LTD 6G Thermostatic Water Bath (Grant, Germany)
Water Purification System	:	Ultra Pure Water Purification System (Watech, Germany)

3.9. Electronic Databases

The following of electronic databases and tools are used in the study:

- Online Mendelian Inheritance In Man (OMIM); http://www.ncbi.nih.gov/Omim
  A database of human genes and phenotypes.
- NCBI genome resources; http://www.ncbi.nlm.nih.gov/genome/guide/human
  A wide database on genome mapping and sequence data.
- USCS Genome Browser; http://genome.uscs.edu/
  Provides a wide range of genome mapping and sequence data.
- Gene Atlas; http://gen.atlas.medecine.univ-paris5.fr
  A database on gene mapping and genetic diseases.
- Gene Cards; http://www.genecards.org
  Provides genomic, proteomic, functional information on human genes.
- İhop; http://www.ihop-net.org
  A gene network for navigating the literature.
- Molecular Modelling & Bioinformatics (MMB); http://mmb2.pcb.ub.es:8080/PMut
  Predicts whether mutations are pathological.
  Output score > 0.5 is pathological, otherwise neutral.
- MutDB; http://www.mutdb.org
  Performs structural analysis of an amino acid substitution to predict its effect on protein function.
- Polymorphism Phenotyping (PolyPhen); http://www.genetics.bwh.harvard.edu/pph
  Predicts impact of an amino acid substitution on structure and function of proteins.
  Output scores start from zero (neutral) and a high positive number is damaging.
- SIFT; http://blocks.fhere.org/sift/SIFT.html/
  Uses sequence homology to calculate position specific scores for amino acids.
  Score range is 0-1, from damaging to neutral.
- SNPs3D; http://snps3d.org
  Predicts functional effects of nonsynonymous SNP through structure sequence analysis. An output score <0 is damaging.

- Biology Workbench http://workbench.sdsc.edu
  Web-based platform providing tools for storing and analyzing input sequence and designing primers.
- Tandem Repeat Finder; http://tandem.bu.edu/trf.trf.html
  Search engine for tandem repeats in a given DNA sequence.
- Alternative Splicing Structural Genomics Project;
  http://moult.umbi.umd.edu/huma2004/
  Models and predicts the effect of SNP on the alternative splicing pattern and function of proteins.
- Laboratory of Statistical Genetics at Rockefeller University;
  http://linkage.rockefeller.edu
  Provides the statistical tools used in genetic linkage analysis.

4. METHODS

4.1. DNA Extraction from Peripheral Blood Samples

Peripheral blood was drawn into sterile evacuated tubes containing K_2EDTA as anticoagulant. Thirty ml of cold cell lysis buffer was added to every 10 ml of blood sample and mixed well. The mixture was kept at 4^{o}C for 15 minutes to lyse the plasma membrane. It was centrifuged at 5000 revolution per minute (rpm) for 10 minutes at 4^{o}C to collect leukocyte nuclei. The supernatant containing the cell debris was discarded. The leukocyte nuclei were washed by suspending in 10 ml cell lysis buffer and centrifuged for 10 minutes at 4^{o}C. The supernatant was discarded, and the nuclei were resuspended in 5 ml of nucleus lysis buffer by vortexing for an extended time. After the pellet dissolved entirely, 50 µl of Proteinase K (20 mg/ml) and 80 µl of 10 per cent SDS were added and mixed gently. In order to digest the nuclear proteins, the sample was incubated at 37^{o}C overnight or at 56^{o}C for 3 hours. NH_4Ac (2.8 ml of 9.5 M) was added, and the solution was shaken to salt out the proteins. It was centrifuged at 10,000 rpm for 25 minutes at room temperature. After the supernatant was transferred to a 50 ml Falcon tube, two volumes of ethanol were added to precipitate out the DNA. DNA threads become first visible and then precipitated out. The precipitate was fished out carefully with a micropipette tip and transferred to a 1.5 ml eppendorf tube. Residual ethanol was let to evaporate; DNA was dissolved in 0.5 ml of TE buffer and stored at 20^{o}C.

4.2. Linkage Analysis

Linkage analysis, a method developed to identify disease loci, requires generation of genome scan data and either parametric linkage analysis or homozygosity mapping. An identified candidate locus is later fine-mapped.

4.2.1. Genome Scan Data

Microsatellite and/or SNP genome scan data were used for linkage analysis for the identification of disease loci.

4.2.1.1. Azoospermia. The first genome scan had been performed in our laboratory using DNA samples from three affected sibs using microsatellite markers from all autosomal and sex chromosomes with approximately 25 cM spacing.

In this study, DNA samples of three affected brothers and their healthy sib were sent to deCODE Institute for whole genome genotyping using Illumina Human 370-Quad Beadchip. The beadchip had been manufactured by Illumina and included 370,000 SNP markers at a median marker spacing of 5.0 kb.

4.2.1.2. IDDM. The initial genome scan data on IDDM family was generated at the National Heart, Lung and Blood Institute (NHLBI) Mammalian Genotyping Service, using Marshfield Screening Set 14 (Contract Number HV48141). The screening set was composed of about 400 microsatellite markers from both autosomal and sex chromosomes with an average spacing of 10 cM. Amplification of microsatellite loci was performed with fluorescently labeled primers, starting with 15-30 μl of DNA per individual. A custom-built scanning fluorescence detector (SCAFUD) and its software automatically processed the SCAFUD output and assign allele sizes. The results were submitted to us in LINKAGE file format. A genotyping error rate of 0.50 per cent was estimated through blind control samples.

In this study, a second genome scan was performed at deCODE Institute using Illumina Human 370-Quad Beadchip.

4.2.1.3. MHAC. The genome wide SNP genotyping data on MHAC family used in this study was generated at the laboratory of Prof. Bernd Wollnik using Affymetrix 250K Sty SNP Chip. The chip covered the entire genome with about 250 000 SNP markers.

4.2.2. Parametric Linkage Analysis and Homozygosity Mapping

All lod score calculations and haplotype constructions were carried out on multiple linkage programs that are implemented on easyLinkage platform. The raw genotyping data, received from NHLBI Mammalian Genotyping Service or deCODE Institute, was

reformatted by Microsoft Excel Macro on Visual Basics and Progeny 7.0, respectively. Table 4.1 lists the linkage programs used in the study and their utilization.

Table 4.1. Parametric linkage analysis programs used in the study

Program	Utilization
PedCheck v1.0	To detect Mendelian inconsistencies
SuperLink v1.6	For two-point parametric lod score calculations
GeneHunter v_5.05	For multipoint parametric lod score calculations and construction of haplotypes in small families (bit size ≤ 19)
Allegro v1.2c	For simulations and multipoint parametric linkage analysis in small families (bit size ≤ 19)
SimWalk v2.91	For multipoint parametric lod score calculations and construction of haplotypes in complex families (bit size > 19)

4.2.2.1. Azoospermia. Due to the complex structure of the pedigree, SimWalk v2.91 was used for multipoint lod score calculation and construction of haplotypes from microsatellite data of azoospermia family. Autosomal recessive inheritance model with full penetrance was assumed. Genome wide SNP scan data was reformatted on excel sheets to identify homozygous informative regions shared only by affected sibs but not the healthy male sib. Two-point and multipoint lod score calculations were conducted on the regions with extended homozygosity. Microsatellite markers used in fine-mapping of two candidate azoospermia loci at chromosomes 17 and 14 are given in Table 4.2.

Table 4.2. Microsatellite markers used in fine-mapping the candidate azoospermia loci

Marker ID	Position			Product size	Repeat type
	Marshfield cM	DeCODE cM	Mb		
D17S578	14.69	19.36	6.76	134-174	TG
D17S1881	14.69	17.38	6.47	216-230	CA
D17S960	15.77	20.49	7.20	127-135	CA
D17S720	14.69	20.96	7.64	221	GAAA
D17S1796	15.77	-	7.73	177-187	CA
D17S1791	17.92	24.41	9.10	232-290	CA
D14S581	21.51	18.55	23.37	191	GAAT
D14S64	22.66	18.97	23.63	126-136	CA
D14S264	22.66	19.62	24.35	216-234	CA

4.2.2.3. MHAC. To map the disease locus in MHAC family, two-point and multipoint parametric linkage analyses of SNP data generated by Affymetrix SNP array were performed using SuperLink and Allegro programs, respectively. The pedigree suggested an autosomal recessive disease model with full penetrance. Parametric linkage analysis was carried out for both autosomal recessive model with full penetrance and dominant model with reduced penetrance. Since there was no known consanguinity in the family, the parents were considered distant cousins. Lod score calculations were performed assuming parents as third, fourth or fifth cousins. The genotyping data for each family member, as ordered according to chromosomal positions, were aligned on excel sheets and reformatted to display homozygous regions that were shared by the affected siblings but not parents in a distinct color. Excel sheets were prepared separately for each chromosome and inspected to identify shared homozygous regions. Both the lod scores and haplotypes of those regions were examined to evaluate their potential as candidate disease loci. Subsequent genotyping with microsatellite markers were used for the verification of the candidate regions (Table 4.3).

Table 4.3. Microsatellite markers used in MHAC

Marker ID	Position		Product size	Repeat type
	cM	Mb		
D6S967	1.40	1.43	339	AGAA
D6S477	9.18	6.08	213-237	GATA
D6S1677	9.18	6.43	248-258	CA
D6S1668	9.18	6.59	136-154	CA
D6S1598	11.89	6.94	217-227	CA
D6S1547	12.62	7.45	290-304	CA
D6S309	14.07	8.17	254-272	CA
D7S1482	126.75	121.03	92	GAAG/GAAA
D7S2847	-	123.08	114-154	-
D18S473	71.32	45.74	231-243	CA
D18S472	71.32	46.26	149-163	CA
D18S363	71.32	46.54	177-247	CA/GA
D19S561	50.81	34.57	300	AAAG
D19S433	51.88	35.11	199-221	GGAA
D19S405	-	35.42	100-120	CA

4.2.2.2. IDDM. In order to identify the disease locus in IDDM family, initially genome wide linkage analysis was performed using microsatellite scan data. SuperLink and GeneHunter linkage programs were used for two-point and multipoint lod score calculation

and haplotype construction, respectively. Both autosomal recessive and dominant inheritance models with varying penetrance were considered as possible disease models. A second parametric linkage analysis was performed excluding the genome scan data of the two healthy siblings. Two loci on chromosomes 2 and 5 with relatively high lod scores were identified and fine-mapped with additional microsatellite markers. Also, other candidate loci that were deduced from family haplotype data were analyzed with microsatellite markers. Table 4.4 lists all the microsatellite markers and their properties used for IDDM family.

Later the data generated by SNP genome scan was used to perform two-point and multipoint parametric linkage analyses on SuperLink and Gene Hunter programs, respectively. Lod scores were calculated assuming either an autosomal recessive or dominant inheritance model, varying the degree of penetrance, setting the disease gene frequency to 0.001. In addition, excel sheets that display homozygous regions shared only by the four affected sibs but not the healthy sib and parents, were prepared for homozygosity mapping.

Table 4.4. Microsatellite markers used in IDDM

Marker ID	Position		Product size	Repeat type
	cM	Mb		
D2S428	103.16	82.84	143-158	GATA
D2S2264	114.42	101.79	241-256	CA
D5S1720	123.45	114.25	212-232	GATA
D5S804	133.65	125.11	228-248	GATA
D5S642	134.72	128.22	183-201	CA
D5S1979	144.06	141.10	157-179	CA
D5S2847	149.48	146.96	162-178	GATA
D5S463	149.48	147.11	178-196	CA
D5S1469	-	149.44	180	GATA
D5S2011	144.06	141.20	120-156	CA
D16S3065	8.16	3.76	152-162	CA
D18S976	12.81	5.24	171-198	GATA
D22S272	45.82	37.415	132-150	CA
D22S1160	53.16	44.807	186-216	CA
D22S283	38.62	35.08	126-152	CA
D22S1140	49.37	43.054	213-231	CA

4.2.3. Fine-mapping of Candidate Loci

After the identification of a candidate disease locus, fine-mapping with microsatellite markers was performed on polyacrylamide gels that were finally stained with silver nitrate. Microsatellite markers on NCBI UniSTS database were primarily selected for typing. In regions where no microsatellite marker was reported, Tandem Repeat Finder Program was used to identify repetitive sequences, and primer pairs were designed to amplify the selected regions.

PCR for each marker was performed in a total volume of 25 µl of 1X PCR buffer, 0.4 µl of each primer, 0.2 mM of each dNTP, 20-100 ng genomic DNA, 0.2 units Taq DNA Polymerase, and adequate dH_2O to adjust the volume. The amplification conditions were as follows: an initial denaturation at 95^{o}C for 5 min followed by 35 cycles of 30 sec denaturation at 94^{o}C, 30 sec annealing at the appropriate temperature and 1 min elongation at 72^{o}C, and a final extension at 72^{o}C for 10 min. In order to prevent shadow fragment synthesis due to replication slippage at dinucleotide repeats, Betain and DMSO were added to PCR reactions with final concentrations of 1.3 M and 1.3 per cent, respectively. Amplification products were run on polyacrylamide gels to resolve alleles, and alleles were visualized by silver staining.

Microsatellite markers from chromosomes 17 and 14 were used in fine-mapping of two candidate azoospermia loci (Table 4.2). Verification of the candidate MHAC loci identified by linkage analysis of SNP genome scan data was performed by microsatellite markers listed in Table 4.3. The microsatellites used in fine-mapping of IDDM are given in Table 4.4.

4.2.3.1. Denaturing Polyacrylamide Gel Electrophoresis. Microsatellite marker alleles were resolved using a Sequi-Gen (Bio-Rad, USA) nucleic acid electrophoresis system. Denaturing instagel was cast in 21 x 40 cm sequencing apparatus that was assembled using 0.4 mm spacers. Thirty-five ml of Instagel was mixed with 300 µl of 10 per cent APS and 30 µl of TEMED and immediately poured between the glass plates of the apparatus. The gel was left to polymerize at least an hour before use. Performing a pre-run at 45 W was essential to heat the buffer and gel to 40-45^{o}C. Urea and other gel particles that were stuck

above the gel were removed by flushing with electrophoresis buffer. DNA products were made ready for loading by mixing with 10X Sample Buffer in a 4:3 ratio. Before loading 6-8 µl of each sample to individual slots, samples were denatured at 95^{o}C in a heat block and immediately chilled on ice. The gel was allowed to run at a constant power of 35 W. The duration of electrophoresis was estimated according to the length of the PCR product.

4.2.3.2. Silver Staining. Following the completion of the electrophoresis, the glass plates were separated gently. The gel, remaining intact on one of the plates, was transferred into staining buffer with the help of a piece of filter paper. It was allowed to stay in staining buffer for 10 minutes and subsequently incubated in developing buffer until the bands appeared. Whenever bands were weakly stained and thus difficult to visualize, the staining procedure was repeated after an extensive washing step.

4.3. Candidate Gene Approach

In this study, either SSCP analysis was used prior to DNA sequence analysis or DNA sequence analysis was performed directly to detect and identify sequence variants in candidate genes.

4.3.1. PCR Amplification

Primer pairs flanking exonic regions, splice junctions and in some cases also regulatory regions were designed by Primer3 software for candidate azoospermia genes *P2RX1*, *NYD-SP20*, *SPAG7*, *INCA1*, *TEKT1*, *GSG* and *YBOX2* and for candidate MHAC genes *TUBB2A*, *TUBB2B*, *LOC10132153*, *NRN1* and *EXOC2*. Specificity of primer targets was checked by electronic PCR at NCBI. The primer sequences and optimized PCR conditions for the candidate genes are given in Table 4.5.

Unless stated otherwise, PCR reaction mixtures were prepared with 1X PCR buffer, 400 nM of each primer, 0.2 mM of each dNTP, 20-100 ng genomic DNA, 0.2 units Taq DNA Polymerase, and adequate dH_2O to adjust the volume to 25 µl. The amplification conditions were as follow: an initial denaturation at 95^{o}C for 5 min followed by 35 cycles

of 30 sec denaturation at 94^{o}C, 30 sec annealing at the appropriate temperature and 1-2 min elongation at 72^{o}C, and a final extension at 72^{o}C for 10 min.

4.3.2. Analysis of PCR products

To check whether sufficient amounts of PCR product had been synthesized, the products were assayed on agarose gels. Five µl of PCR products were mixed with 1 µl of 6X of loading dye each and loaded on a two per cent agarose gel containing 15 µg ethidium bromide. After electrophoresis in 0.5X TBE at 150 volts for 15 minutes, bands were visualized over an UV light transilluminator.

4.3.3. SSCP Analysis

SSCP analysis is widely employed to detect sequence variants that migrate aberrantly. The exonic and regulatory regions of *P2RX1* and *NYD-SP20* were analyzed by SSCP in the azoospermia family and 13 azoospermic patients. All seven exons of *SPAG7* except exon 5 were initially analyzed for mutations in the same group. All amplified fragments of one affected family member were sequenced together with the azoospermic patient fragments exhibiting aberrant mobility patterns in gels. The regions harboring novel variants were amplified in 100 control samples to estimate its prevalence in the population and to assess whether it is a polymorphism or a mutation. In our study, following the identification of a mutation by direct resequencing on *SPAG7* exon 5, all family members and 100 control individuals were screened for the mutation by SSCP. Additionally, 49 azoospermic and 19 oligozoospermic individuals were analyzed for mutations in all exons of *SPAG7*. Each fragment was subjected to electrophoresis in polyacrylamide gels with a crosslinking ratio of 2 per cent and either with or without glycerol.

4.3.3.1. Preparation of SSCP Gels. SSCP gels were cast between glass plates of 20 cm x 20 cm, assembled with 0.75 mm spacers. The gel solution was prepared with 6 ml of 40 per cent acrylamide stock (37.5:1), 1.8 ml 10X TBE and sufficient dH_2O to adjust the volume to 30 ml. Then 1.5 ml of glycerol was added whenever the gel was to contain 5 per cent glycerol. The gel solution was poured after addition of 250 µl of ten per cent APS and

25 µl TEMED and gental mixing. A 15- or 20-well comb was inserted to form slots, and the gel was left to polymerize for at least an hour.

Samples were prepared for loading by mixing with 10X Stop buffer in 4:3 ratio, denaturing at 95^{o}C for 5 min and immediately cooling on ice. Seven to nine µl of each sample mix was loaded, and the samples subjected to electrophoresis in 0.6X TBE buffer at a constant power of 10 W for 10 to 15 hours, either at room temperature or 4^{o}C in a BioRad DCode Universal Mutation detection system or BioRad Protean IIxi Cell. The gels were silver stained to visualize the DNA samples.

4.3.4. DNA Sequence Analysis

Amplified exonic fragments of candidate azoospermia genes *INCA1*, *TEKT1* and *YBOX2* and candidate MHAC genes *TUBB2A*, *TUBB2B*, *LOC10132153*, *NRN1* and *EXOC2* were directly subjected to resequencing.

A fragment to be sequenced was amplified in four tubes, each containing 25 µl of reaction mix. QIAquick PCR Purification Kit was used to clean the PCR product from primers, nucleotides, polymerase, and salts. Samples were sequenced at Iontek (Istanbul, Turkey), Refgen (Ankara, Turkey), MGH Sequencing Core (MA, USA) or Macrogen Inc. (Korea).

Table 4.5. Primers used in candidate gene analysis in azoospermia

Exon	Primer Name	Primer Sequence (5'→ 3')	Size (bp)	Buffer & Additives	T_m (^{0}C)
P2RX1					
P1	P2RX1p1-F	TTCCTCTCACATCACCACCA	446	Buffer 2	58
	P2RX1p1-R	AACACCCAGAACAAAGGCAG			
P2	P2RX1p2-F	AGTCTGGCAGGGTTTTCCTC	417	Buffer 2	58
	P2RX1p2-R	AAGGAGACCTGTTGGAACCC			
P3	P2RX1p3-F	AAAATGAGACCTCCATCCCC	415	Buffer 2	58
	P2RX1p3-R	GGAACAGCCCTTATTACCCC			
1a	P2RX15UTRaF	CTCCATCTTGCCAGGTTTTG	433	Buffer 2	58
	P2RX1.5UTRaR	GCCTCTGCTCTCAGGGTG			
1b	P2RX1-1F	GAAGCCCCCAGAAGCTCTAC	318	Buffer 1 DMSO	55
	P2RX1-1R	ACGCAAGGGATGTCAGAGTC			

Table 4.5. Primers used in candidate gene analysis in azoospermia (Continued)

2	P2RX1-2F	AAGCGTGTGACAGGTCCAG	275	Buffer 1 DMSO	55
	P2RX1-2R	AGAGTCCCACCCAACTCTGA			
3	P2RX1-3.2F	CTCTGGACCTCTGTGAATGC	332	Buffer 1	56
	P2RX1-3.2R	CCTGGAGACTTTCTCTGCAC			
4	P2RX1-4F	AGAGGGGCTCTTCTCAGGAC	230	Buffer 1 DMSO	55
	P2RX1-4R	GGTGTGGAGAGAGGCAGGT			
5	P2RX1-5F	TTCTGTCCTTGGCTCCTTTG	310	Buffer 2	55.5
	P2RX1-5R	GGGGTCAGAAAAAGGGGTAA			
6	P2RX1-6F	GCTGTGCCAAGCTCTGTGT	271	Buffer 1	
	P2RX1-6R	TTGCCCCTTTGATTCCTTC			
7	P2RX1-7.2F	CAATGTGGAGCCTGTGAG	244	Buffer 2	58
	P2RX1-7.2R	CCATACCTTCTCAGCCAG			
8	P2RX1-8.Bf	CAGGGGTTTTTGATCCTGAA	273	Buffer 2	55
	P2RX1-8.Br	TGGGGACACTTTCTCAGACC			
9	P2RX1-9F	CCTGATGCGCTAAGAACTCC	281	Buffer 1 DMSO	54
	P2RX1-9R	AAACTGCTGGATGAACCCAC			
10	P2RX1-10F	CATGTAGAGTCCGAGGTGGG	273	Buffer 2	55.5
	P2RX1-10R	GAGGGACGTCTTGCTGCTAC			
11	P2RX1-11F	CCTCCAGTGAGGTCAGATCC	283	Buffer 1 DMSO	55
	P2RX1-11R	CACCCCAACACAGACACAAG			
12a	P2RX1-12.2F	CGCTTGTGTCTGTGTTGG	295	Buffer DMSO	55
	P2RX1-12.2R	CTGAGCTTCTGGCAAACTG			
12b	P2RX1.3UTRaF	GACTTGTATCCCCCAACCCT	389	Buffer 2 DMSO	58
	P2RX1.3UTRaR	AAGCTAGGGTGCAGGTGTGT			
12c	P2RX1.3UTRbF	GAGCCCTACCCTAGCCTCAG	432	Buffer1 Bet+DMSO	58
	P2RX1.3UTRbR	TTTTGGTTCAGGAGGTCTGG			
12d	P2RX1.3UTRcF	GGGAACTTATCGACAATGCAA	396	Buffer 2	58
	P2RX1.3UTRcR	GCTTGGGGGTGTAGATGTGT			
12e	P2RX1.3UTRdF	CACACAGCATCCTTCCACAC	359	Buffer 2	58
	P2RX1.3UTRdR	GTCTGGAAAGAGAGCACGGA			
SPATA22					
P1	SPATA22P1_F	ATCAACGCACAGTCTTTCCC	417	Buffer 2	55
	SPATA22P1_R	GGGGCACACTTTTTAGACGA			
P2	SPATA22P2_F	AGAGATTGACTCGAAGGCCA	275	Buffer 2	55
	SPATA22P2_R	ACAGTCCTCTTTGCCAGGAA			
P3	SPATA22P3_F	GCACACAGTAAATGCCCAGT	331	Buffer 2	55
	SPATA22P3_R	TGGCATTCAAGGTCTGCTATT			
1	nydex9_R	GATTCCCGCGCTTTCTTC	231	Buffer1 Bet+DMSO	58
	nydex9_F	GGGAGAGAAGAGGCTCCAGT			
2	nydex8_R	AATCACTGGGCTGAGGATTG	315	Buffer 2	55
	nydex8_F	TTGGACCTAAAAAGTAAAGAGGC			
3	nydex7_R	TTTACAACTAGATCTGCTTATTTGT	308	Buffer 2	55
	nydex7_F	TGTGGATAATGGAAAGAAGAAA			
4	nydex6_R	TCAAGTTTTTAAACCCTCCAAA	251	Buffer 2	55
	nydex6_F	TTGCTTTCAAATTGAGAAATTAAGA			

Table 4.5. Primers used in candidate gene analysis in azoospermia (Continued)

5	nydex5_R	CACCGGGTATGTGGAGAAAT	300	Buffer 2	55
	nydex5_F	GGGAAGAATGAAGAGGAAGACA			
6a	nydex4b_R	TTAAGAAACCATATAGAAACAAA	292	Buffer 1	55
	nydex4b_F	CTAGCCCTCTCATTGTGCAG			
6b	nydex4a_R	ATCTCGCAACAAAGAAACCG	276	Buffer 2	55
	nydex4b_F	TGGAATTTATTTCAGGCGCT			
7	nydex3_R	TTTTACACACTTGTCCATTGTAAGG	299	Buffer 2	55
	nydex3_F	CATTAAAACCTGTCCTCCAGAGA			
8	nydex2_R	TTTGCAGTTTACCAACATTCATTT	190	Buffer 2	55
	nydex2_F	CTGTTGACATAGGCTTCGCA			
9a	nydex1b_R	TTTTGAGGTGGCACATGTTT	350	Buffer 2	55
	nydex1b_F	AATTAAGCAATAACAGAAAACTGCT			
9b	nydex1a_R	AAACTTTCCAGGCATTTGTCA	278	Buffer 2	55
	nydex1a_F	GGATGGTTGTGGAATTTTAGC			
SPAG7					
P1	SPAG7_P1F	ATCTCGGCTCACTGCAGATT	353	Buffer 2	68>58
	SPAG7_P1R	TGCAGCTTGGGATGATACTG			
P2	SPAG7_P2F	CCTTGAATACTTGGAAGGCG	471	Buffer 2	68>58
	SPAG7_P2R	GCCTGTTCGACCTTCCTTC			
P3	SPAG7_P3F	AATCAACACTACGATCCGCC	476	Buffer 2	68>58
	SPAG7_P3R	CCTAAAAGGGCTGCTGACAC			
P4	SPAG7_P4F	GTGCGTCTTAAAGGGAAGGG	582	Buffer 2	68>58
	SPAG7_P4R	CGTCTCTGAATGTATCCGCC			
1	SPAG7-1F	AGCCTCCGGCCTCTGTCT	338	Buffer 1 DMSO	58
	SPAG7-1R	AGAGGGAGAAGTATGGTCCCG			
2	SPAG7-2F	CCAGTGGACACCACACTTCA	276	Buffer 1	58
	SPAG7-2R	CTTCCATGCGTCACCTCTCT			
3	SPAG7-3F	AGAGAGGTGACGCATGGAAG	322	Buffer 2	58
	SPAG7-3R	CCTAGTGCCCTCCTCTGGA			
4	SPAG7-4F	AGGAAGGAAATGGGTGGCTA	318	Buffer 1	58
	SPAG7-4R	AGAGCAGAATGATGGGGTCA			
5	SPAG7-4F	AGGAAGGAAATGGGTGGCTA	304	Buffer 2	68>58
	SPAG7-6R	GTGAGCTATGCTTTGCAGCC			
6	SPAG7-6F	GTGAGCTATGCTTTGCAGCC	345	Buffer 1 DMSO	58
	SPAG7-6R	AAATCCGTACCTCAGCCTCC			
7a	SPAG7-7aF	TGGGAGAGGAGCATAGGATG	322	Buffer 1	58
	SPAG7-7aR	CTGAGGGACAGCTGGTAGGA			
7b	SPAG7-7bF	ATTAGAGCCCATCCTGGAGC	358	Buffer 1	58
	SPAG7-7bR	TGGCCATCTTTAACCTACACAA			
TEKT1					
1	TEKT1_E1F	GTACAGCCGGGAAGTCTGAG	621	Buffer 3	68>58
	TEKT1_E1R	TTCCTAGCCCAGTCCCTTTT			
2	TEKT1_E2F	TTGTTTCTTGGTTTGGGTGC	547	Buffer 3	68>58
	TEKT1_E2R	CAGGCTGGGTTACCTTGTGT			
3	TEKT1_E3F	AGATGCCCCAAATAGCACAG	552	Buffer 3	68>58
	TEKT1_E3R	GCTACCCACAGCCAAACCTA			

Table 4.5. Primers used in candidate gene analysis in azoospermia (Continued)

4	TEKT1_E4F	TCCATTTGGCATGTTTCAGA	798	Buffer 2	68>58
	TEKT1_E4R	TCTGGTGCAGGATGATGGTA			
5	TEKT1_E5F	GCACTCAGGGTAAGAGCTGG	708	Buffer 3	68>58
	TEKT1_E5R	AACCTGGAGAATCAGGAGCC			
6	TEKT1_E6F	GTCTATAAGGTGCCAGCCCA	846	Buffer 2	68>58
	TEKT1_E6R	AACTGGCTGCATCAACTCCT			
7	TEKT1_E7F	TTCCAGAGGACAGAGCTGGT	780	Buffer 3	68>58
	TEKT1_E7R	TCCATTTTCTCCAAGGCAAC			
8	TEKT1_E8F	GATAGTGACCAGCAGCGTGA	841	Buffer 3	68>58
	TEKT1_E8R	TTAGGCAAACACCACACCAG			
INCA1					
1	INCA_E1F	CTCTTGTTAAAGGGGCAACG	480	Q solution	65>55
	INCA_E1R	GGCCTGAAGGGGTTGTTT			
2	INCA_E2F	CTCTTGTTAAAGGGGCAACG	1147	Q solution	65>55
	INCA_E2R	GAGAGTGCACATAGGCTCCC			
3	INCA_E3F	TGTCGCTTAACTGCCATCTG	1084	Buffer 3	68>58
	INCA_E3R	CATTGCATTTCCCGTTCTTT			
4&5	INCA_E4-5F	TCAACCCAGCTTCAATTTCC	739	Buffer 2	68>58
	INCA_E4-5R	GACCCCTGTGATCTCGAAAA			
6	INCA_E6F	CTGTCCTGTCCCAAGCTCTC	603	Buffer 3	68>58
	INCA_E6R	GCAGTTCCCTCCTTCTTTCC			
7	INCA_E7F	GAAAATTGGGAGTGGAGGGT	701	Buffer 3	68>58
	INCA_E7R	GCATCGGTGGTTTCTCCTTA			
8	INCA_E8F	TATCCACAGGAAGAGGACCG	819	Buffer 3	68>58
	INCA_E8R	CAAGACTCCTTGTCTGGGGA			
Haspin					
1a	Haspin_E1F	CTGAAACGAGAGCCATGACC	700	Buffer 3	68>58
	Haspin_E1R	CTGAACAGGGAGGCACTGAT			
1b	Haspin_E2F	CCTAACCGTGACCCCAAGAC	671	Buffer 3	68>58
	Haspin_E2R	CCAGCCTGTCTATTCCCCTT			
1c	Haspin_E3F	GTCCAGGTCTGTCAAGCACA	795	Buffer 3	68>58
	Haspin_E3R	CACCGGATAAGAGGCTCAAC			
1d	Haspin_E4F	TCACACACCCGTAGCCATAA	630	Buffer 3	68>58
	Haspin_E4R	TCACCGGTAAACAGGTCCTC			
1e	Haspin_E5F	ACTTACACTGGGGGAACGTG	1080	Q solution	69>60
	Haspin_E5R	GTGAGGAGTTCGAGACCAGC			
YBX2					
7	YBX2_E7F	ACCAACAGCAGGGAGATGAG	593	Buffer 2	68>58
	YBX2_E7R	GGGAGGAGCCAAAGGATAGA			
8	YBX2_E8F	CTGAGGGGTGTGATTGAGGT	567	Buffer 2	68>58
	YBX2_E8R	ACAACAGCGACTTGTCATCG			
9	YBX2_E9F	GCCACCGCTGTTCTTTAGAC	749	Buffer 2	68>58
	YBX2_E9R	GATAGGGGCACTCTTCACCA			

Table 4.6. Primers used in candidate gene analysis in MHAC

Exon	Primer Name	Primer Sequence (5'→ 3')	Size (bp)	Buffer & Additives	T_m (^{0}C)
TUBB2A					
1	TUBB2AE1_2F	AAAGGGTACTCGACGCCTC	843	Q solution	65>55
	TUBB2AE1_2R	CAAATACACCCAGGGGAATG			
2	TUBB2A-E2F	AATAGGGGGTGCAGCTCTTT	393	Buffer 1 DMSO	58
	TUBB2A-E2R	ATGTTTTTCCAGCAGTTGGC			
3	TUBB2A-E3F	ATGCCAACTGCTGGAAAAAC	452	Buffer 1 DMSO	58
	TUBB2A-E3R	ATGGACTGCATGGAGCTAGG			
4a	TUBB2A-E4aF	GATCCCCAGTGTTTTTGAGG	790	Buffer 1 Bet+DMSO	58
	TUBB2A-E4aR	GAGTCGAACATCTGCTGGGT			
4b	TUBB2AE4b_2F	ACGAGGCCCTGTATGACATC	1145	Buffer 3	65>55
	TUBB2AE4b_2R	GGGCCAGTAAAGGTGAACAA			
TUBB2B					
1	TUBB2B-E1F	GCTCGGAGCATTACGTCAG	465	Buffer 1 Bet+DMSO	58
	TUBB2B-E1R	GAAAGTCACCTCCTAGCCCA			
2	TUBB2B-E2F	CTGAGTCGGGGACATTTCAT	356	Buffer 1 Bet+ DMSO	58
	TUBB2B-E2R	GGCTCTGTTGGCATAAGGAA			
3	TUBB2B-E3F	CAGAGCCCTTTGAGAACCTG	399	Buffer1 Bet+DMSO	58
	TUBB2B-E3R	TAGCAGCTGAAAGGCAAACA			
4a	TUBB2B-E4aF	TAACGTTTGGCCATTTTGGT	995	Buffer 1 Bet+DMSO	55
	TUBB2B-E4aR	CTTGAACAGCTCCTGGATGG			
4b	TUBB2B-E4bF	TGCAGAACAAGAACAGCAGC	940	Buffer 1 DMSO	58
	TUBB2B-E4bR	CTTCAGAATTTTGTGCTGTGA			
LOC10132153					
1&2	TUBBLOCE1-2F	GCACTGGGTGGAGGCTATAA	995	Buffer 3	68>58
	TUBBLOCE1-2R	CGTGTTTTTCCAGTAGTTGGC			
3	TUBBLOC_E3F	AAACACGCTTCATTGGTTCC	689	Buffer 3	68>58
	TUBBLOC_E3R	ATGGACTGCATGGAGCTAGG			
4a	TUBBLOC_E4aF	TGAGGTCAGGAGTTTGGGAC	697	Buffer 3	68>58
	TUBBLOC_E4aR	GTCATACAGGGCCTCGTTGT			
4b	TUBBLOC_E4bF	CATGAACTCCTTCAGCGTCA	657	Buffer 3	68>58
	TUBBLOC_E4bR	CCACCATGAAGCAGGATTTT			
5	TUBBLOC_E5F	CCGATCTTGAGACAGAGGCT	564	Q solution	65>55
	TUBBLOC_E5R	GAGCGCAGGATCGCTAAG G			
NRN1					
1a	NEURITIN-E1F	CTGCTGGGTTGGAGAAAAAG	460	Buffer 2	65>55
	NEURITIN-E1R	GGGATGACACGGAATGATTT			
1b	NEURITIN-E2F	GGAGGCAAAATTTCGTTCAG	637	Buffer 2	65>55
	NEURITIN-E2R	TAGTCCCCAGGAACTGATGC			
2	NEURITIN-E3F	CACCTTTAACACTGGCGAGC	432	Buffer 2	65>55
	NEURITIN-E3R	GAGCCAGAAGGCAAGAGTGA			
3a	NEURITIN-E4aF	CTTTAGACAGCCGCTTCTGG	650	Buffer 2	65>55
	NEURITIN-E4aR	TTGCTTGCTCACTGATTGGT			
3b	NEURITIN-E4bF	AGTTCTTTGGGGACGTTGTG	595	Buffer 1 Bet+DMSO	65>55
	NEURITIN-E4bR	CCTCAGTGACCATAGCAGCA			
3c	NEURITIN-E4cF	CCACTGCACATTTCTCCTCA	531	Buffer 1 Bet+DMSO	65>55
	NEURITIN-E4cR	CATCCTGTGTCCCTTCAGGT			
EXOC2					
1	EXOC2_E1F	ATCCTTCTTTCCCTCCCTCA	771	Buffer 3	68>58
	EXOC2_E1R	ACGCCTAGGCAGACAGGTAA			
2	EXOC2_E2F	GTTTCTCTCTGCCCTTGTGC	686	Buffer 3	68>58
	EXOC2_E2R	TTCAGAGGCAAATCTCAGCC			

Table 4.6. Primers used in candidate gene analysis in MHAC (Continued)

3	EXOC2_E3F	CCAGTAATGCCTGAGCACAC	896	Buffer 3	68>58
	EXOC2_E3R	ACCAACATCCCAATCCTGTC			
4	EXOC2_E4F	AGTCAGACAATAGGCCAGAC	305	Buffer 3	68>58
	EXOC2_E4R	TTCCTGGGATGTTTCTGGAG			
5	EXOC2_E5F	CAGCCATGATGTGGGAAATA	450	Buffer 3	68>58
	EXOC2_E5R	AGATGTTCACGGGTAAACGC			
6	EXOC2_E6F	GAGGGCTCCGAAAGTCTTGT	514	Buffer 3	68>58
	EXOC2_E6R	GAGATGCCAGCCTACTCTGC			
7	EXOC2_E7F	TGGTGTCTGCTGGGTACCTT	542	Buffer 3	68>58
	EXOC2_E7R	TTCCTCGCTTTGAAGATGCT			
8&9	EXOC2_E8_9F	GGGCATGGTAAATATGGTGC	796	Buffer 3	68>58
	EXOC2_E8_9R	TGGTTGCTCATTTTCACTGC			
10	EXOC2_E10F	GCAGTGAAAATGAGCAACCA	825	Buffer 3	68>58
	EXOC2_E10R	TGTCCTACGCCAGATAACCC			
11	EXOC2_E11F	CAGGAGCCATCTCACCTAGC	675	Buffer 3	68>58
	EXOC2_E11R	GTAACTCTGGGAGGCAGCAG			
12	EXOC2_E12F	CATTATCCTCCCACCACCAC	574	Buffer 3	68>58
	EXOC2_E12R	ATAAGCACTGACCTGCACCC			
13	EXOC2_E13F	CGCAGACTCAGGCTATTCCT	562	Buffer 3	68>58
	EXOC2_E13R	CAATTCGGAGTCCCTGAAAA			
14	EXOC2_E14F	TTGACAGTGTTGTCGTGGGT	565	Buffer 3	68>58
	EXOC2_E14R	CCAGGAACAAGCATAACCGT			
15	EXOC2_E15F	TCACAGGGTAGGGTTTTTCC	598	Buffer 3	68>58
	EXOC2_E15R	ATTCCTCACGGAGCTTACGA			
16	EXOC2_E16F	TTTTGTCCTTTTGCCTCCAC	472	Buffer 3	68>58
	EXOC2_E16R	AAAGGAGGTGGCACATTCAG			
17	EXOC2_E17F	TCCAACGTAGAATTCCCCAA	659	Buffer 3	68>58
	EXOC2_E17R	AGCCAGCAACGTCATCTCTT			
18	EXOC2_E18F	GGAGACTCAGCTCATCCCAG	614	Buffer 3	68>58
	EXOC2_E18R	GGAATCGGAAATCAGCACAT			
19	EXOC2_E19F	CAGCCAGTGCTCATTTTTGA	725	Buffer 3	68>58
	EXOC2_E19R	AGGAGAGGGGTTGACTGACA			
20	EXOC2_E20F	TTTCTGTCATTAATGGGGCA	589	Buffer 3	68>58
	EXOC2_E20R	TCTAGCAAGAGCAATGGTGC			
22	EXOC2_E22F	AAGCAGCCAAGAAAAGCAAA	556	Buffer 3	68>58
	EXOC2_E22R	GCCTTTCCCTCCTCATTTTC			
23	EXOC2_E23F	TTGCCTTCCACCAAGACTGT	779	Buffer 3	68>58
	EXOC2_E23R	ATCACAGGAAGCAGGAATGC			
24	EXOC2_E24F	CTGTGTGTGTGTGCGTGAAG	531	Buffer 3	68>58
	EXOC2_E24R	GACAGACAGTGCGAGCAAGA			
25	EXOC2_E25F	GGCAAGAGGCAAAAGAAATG	575	Buffer 3	68>58
	EXOC2_E25R	AATGCTTCATGGGTATGGCT			
26	EXOC2_E26F	ATGGTGGTATTACGCCAACC	533	Buffer 3	68>58
	EXOC2_E26R	TGCAGAGAAACCACAAGGCT			
27	EXOC2_E27F	TTTGAGTTTTTCTGACCCCG	543	Buffer 3	68>58
	EXOC2_E27R	TCACTTGAAACGAGAGCCAA			
28a	EXOC2_E28aF	ATCTTGGGAGGGGATAGGAA	1042	Q solution	68>58
	EXOC2_E28aR	CACGTGCATGTGTAGTGACG			
28b	EXOC2_E28bF	GGGCTAAGCATGAGCTTTCA	1027	Q solution	68>58
	EXOC2_E28bR	GCAGGACCTGACAGATGGTT			

5. RESULTS

5.1. Azoospermia

In this study, candidate locus 17pter-p13.1, identified previously in our laboratory, was fine-mapped and SNP data generated by a new genome scan was utilized to search for other candidate loci for azoospermia. Subsequently, disease gene was searched for by candidate gene approach. An autosomal recessive model with full penetrance was assumed.

5.1.1. Fine-mapping at 17p

Candidate azoospermia locus at 17pter-p13.1, previously identified in our laboratory by microsatellite genome scan and linkage studies, was fine-mapped to a 7.20 Mb interval between pter and D17S960 in this study using the linkage analysis and haplotype data. The locus was delineated as 21.05 cM and did not yield statistically significant lod scores: the maximum two-point lod score was 2.01 (Table 5.1), and the maximum multipoint lod score was 2.38 (Figure 5.1). The haplotypes of azoospermia family members demonstrated shared homozygous genotypes only in the affected individuals, indicating identity by descent of the haplotype (Table 5.2).

5.1.2. Homozygosity Mapping

The data from the SNP genome scan were used to search for shared homozygosity in affected brothers only. The reason why a genome linkage analysis was not performed was that the family was too complex for linkage analysis tools employed. The data were formated on Excel sheets to achieve homozygosity mapping. The analysis verified the candidate locus at 17pter-p13.1, and delineated it as the interval pter to rs3744405 (approximately 7.20 Mb and 21.05 cM). A second locus was identified at 14q11.2-q12 between rs8010057 and rs2332572 (approximately 2.8 Mb and 6 cM). Genotyping with three microsatellite markers within the latter interval confirmed the shared homozygosity in affected brothers (Table 5.3). Two-point and multipoint lod scores were calculated

assuming autosomal recessive inheritance with full penetrance. Maximum two-point and multipoint lod scores were 1.245 and 2.38, respectively, at both candidate loci chromosome 17 (Figure 5.2) and chromosome 14 (Figure 5.3).

Table 5.1. Two-point lod scores for azoospermia at 17pter-p13.1. Twenty-seven microsatellite markers were used.

MARKER	Position (cM)	Z_{max}[1]	Θ_{MLE}[2]	Lod Score at Θ =					
				0	0.05	0.1	0.2	0.3	0.4
D17S1866	0.14	0.7479	0.0500	0.7351	0.7479	0.6854	0.4815	0.2617	0.0873
D17S1308	0.63	1.1392	0.0000	1.1392	1.0063	0.8699	0.5945	0.3321	0.1150
D17S926	0.64	0.7479	0.0500	0.7351	0.7479	0.6854	0.4815	0.2617	0.0873
D17S695	1.15	2.0073	0.0000	2.0073	1.7558	1.5041	1.0077	0.5447	0.1769
D17S1529	3.02	1.5817	0.0000	1.5817	1.3626	1.1492	0.7473	0.3956	0.1280
D17S1533	4.76	2.0073	0.0000	2.0073	1.7558	1.5041	1.0077	0.5447	0.1769
D17S525	7.49	1.1392	0.0000	1.1392	1.0063	0.8699	0.5945	0.3321	0.1150
D17S379	7.60	1.5817	0.0000	1.5817	1.3626	1.1492	0.7473	0.3956	0.1280
D17S1152	9.48	0.4049	0.0000	0.4049	0.3441	0.2842	0.1744	0.0873	0.0292
D17S919	10.61	1.5817	0.0000	1.5817	1.3626	1.1492	0.7473	0.3956	0.1280
D17S1828	10.87	1.1392	0.0000	1.1392	1.0063	0.8699	0.5945	0.3321	0.1150
D17S1584	13.25	2.0073	0.0000	2.0073	1.7558	1.5041	1.0077	0.5447	0.1769
D17S559	13.53	2.0073	0.0000	2.0073	1.7558	1.5041	1.0077	0.5447	0.1769
D17S1854	14.20	0.7479	0.0500	0.7351	0.7479	0.6854	0.4815	0.2617	0.0873
D17S1537	16.53	0.4049	0.0000	0.4049	0.3441	0.2842	0.1744	0.0873	0.0292
D17S513	16.77	0.7479	0.0500	0.7351	0.7479	0.6854	0.4815	0.2617	0.0873
D17S1832	16.86	2.0073	0.0000	2.0073	1.7558	1.5041	1.0077	0.5447	0.1769
D17S796	17.60	0.4049	0.0000	0.4049	0.3441	0.2842	0.1744	0.0873	0.0292
D17S1881	17.88	1.5817	0.0000	1.5817	1.3626	1.1492	0.7473	0.3956	0.1280
D17S906	19.02	1.5817	0.0000	1.5817	1.3626	1.1492	0.7473	0.3956	0.1280
D17S960	21.05	0.1922	0.1500	-1.6331	0.0233	0.1671	0.1732	0.0913	0.0183
D17S720	21.52	0.7072	0.1000	-0.8399	0.6318	0.7072	0.5614	0.3103	0.0867
D17S1791	25.02	0.6366	0.1000	-1.0767	0.5324	0.6366	0.5368	0.3142	0.0956
D17S974	30.95	0.0554	0.2500	-6.4100	-0.6921	-0.2532	0.0242	0.0535	0.0106
D17S1303	32.00	0.0554	0.2500	-6.4100	-0.6921	-0.2532	0.0242	0.0535	0.0106

[1]Maximum two-point lod score with Superlink (autosomal recessive, full penetrance)

[2]Maximum likelihood estimate of recombination fraction (Θ_{MLE})

Table 5.2. Haplotypes of azoospermia family at 17pter-17p13.1. Markers in bold were utilized in this study. Father's haplotypes were deduced. Individual designations are as in pedigree in Figure 3.1.

Marker ID	Position (Mb)	Individual						
		407 F	406 M	501 P1	502 P2	503 B	504 S	505 P3
D17S1866	0.08	*2 2*	1 2	2 2	2 2	2 2	2 1	2 2
D17S1308	0.57	*1 2*	1 1	1 1	1 1	2 1	2 1	1 1
D17S926	0.58	*2 2*	1 2	2 2	2 2	2 2	2 1	2 2
D17S695	0.69	*2 3*	1 2	2 2	2 2	3 2	3 1	2 2
D17S1529	1.00	*2 1*	1 2	2 2	2 2	1 2	1 1	2 2
D17S1533	1.49	*3 1*	2 3	3 3	3 3	1 3	1 2	3 3
D17S525	1.89	*1 2*	1 1	1 1	1 1	2 1	2 1	1 1
D17S379	2.37	*2 1*	1 2	2 2	2 2	1 2	1 1	2 2
D17S1152	3.25	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1
D17S919	3.64	*2 1*	1 2	2 2	2 2	1 2	1 1	2 2
D17S1828	3.75	*2 1*	2 2	2 2	2 2	1 2	1 2	2 2
D17S1584	4.29	*3 2*	1 3	3 3	3 3	2 3	2 1	3 3
D17S559	4.69	*3 2*	1 3	3 3	3 3	2 3	2 1	3 3
D17S1854	5.61	*2 2*	1 2	2 2	2 2	2 2	2 2	2 2
D17S1537	5.84	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1
D17S513	5.87	*1 1*	2 1	1 1	1 1	1 1	1 1	1 1
D17S1832	5.96	*3 1*	2 3	3 3	3 3	1 3	1 3	3 3
D17S796	6.19	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1
D17S1881	6.47	*2 1*	1 2	2 2	2 2	1 2	1 2	2 2
D17S906	6.64	*1 2*	2 1	1 1	1 1	2 1	2 1	1 1
D17S578	6.76	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1
D17S960	7.20	*1 2*	1 2	1 2	1 2	2 2	2 2	1 2
D17S720	7.64	*3 4*	1 2	3 2	3 2	4 2	4 2	3 2
D17S973	8.92	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1
D17S1791	9.10	*1 3*	1 2	1 2	1 2	3 2	3 2	1 2
D17S974	10.46	*2 1*	1 3	2 3	2 1	1 3	1 3	2 3
D17S1303	10.80	*3 2*	2 1	3 1	3 2	2 1	2 1	3 1

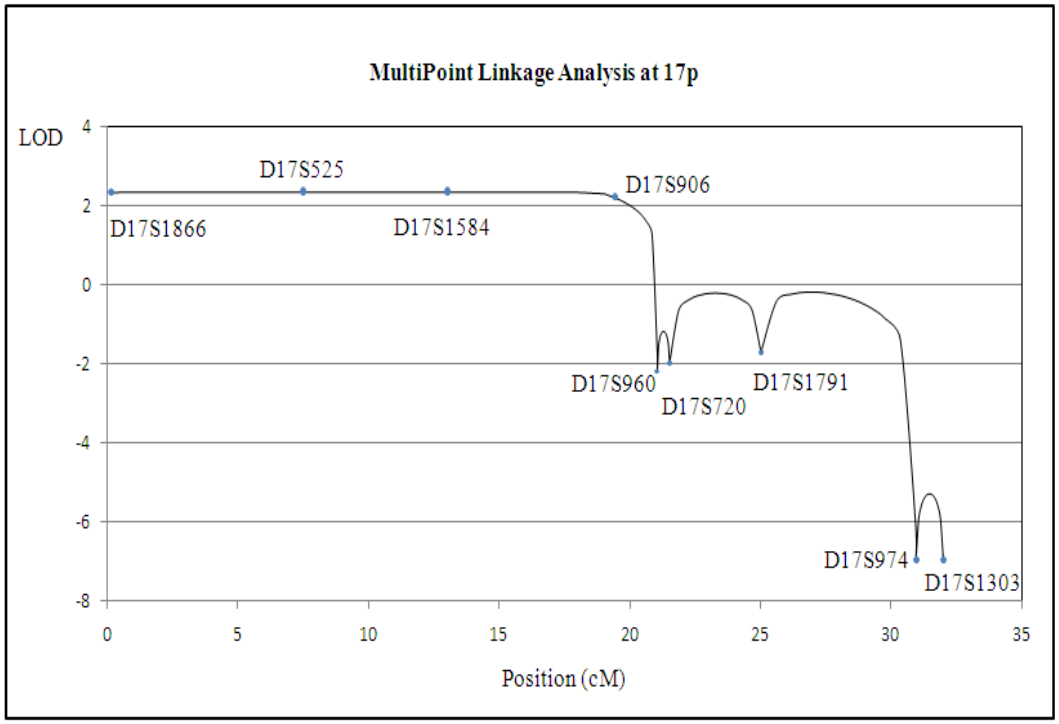


Figure 5.1. Multipoint lod score curve for azoospermia at 17pter-p13.1 using microsatellite data. It was obtained by using genotypes with 27 microsatellite markers.

Table 5.3. Haplotypes of azoospermia family at 14q11.2-q12. Markers in bold are microsatellites. Father's haplotypes were deduced. Individual designations are as in pedigree given in Figure 3.1.

Marker ID	Position (Mb)	Individual					
		407 F	406 M	501 P1	502 P2	503 S1	505 P3
rs8010057	22.57	*A A*	B B	A B	A B	A B	A B
rs10143875	22.65	*A B*	B A	A A	A A	B A	A A
D14S581	23.37	*1 2*	2 1	1 1	1 1	2 1	1 1
D14S64	23.63	*3 2*	1 3	3 3	3 3	2 3	3 3
D14S264	24.35	*2 1*	1 2	2 2	2 2	1 2	2 2
rs17278202	25.40	*B A*	B B	B B	B B	A B	B B
rs2332572	25.41	*A B*	A A	A A	A A	B A	A B

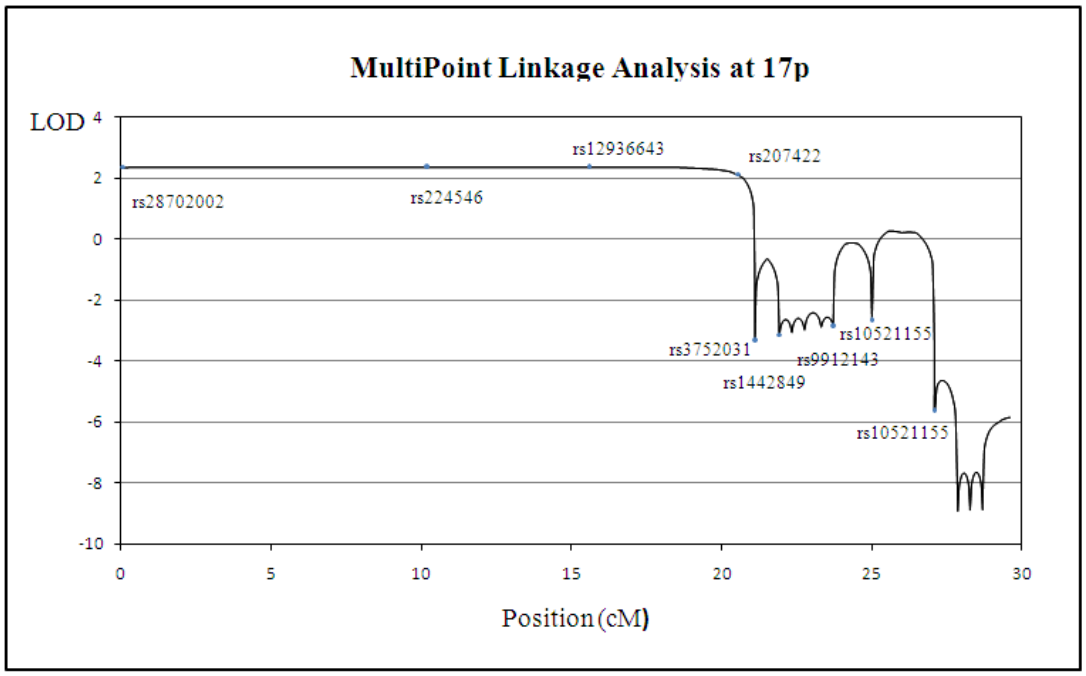


Figure 5.2. Multipoint lod score curve for azoospermia at 17pter-p13.1 using SNP data. It was obtained by using 951 informative SNP markers.

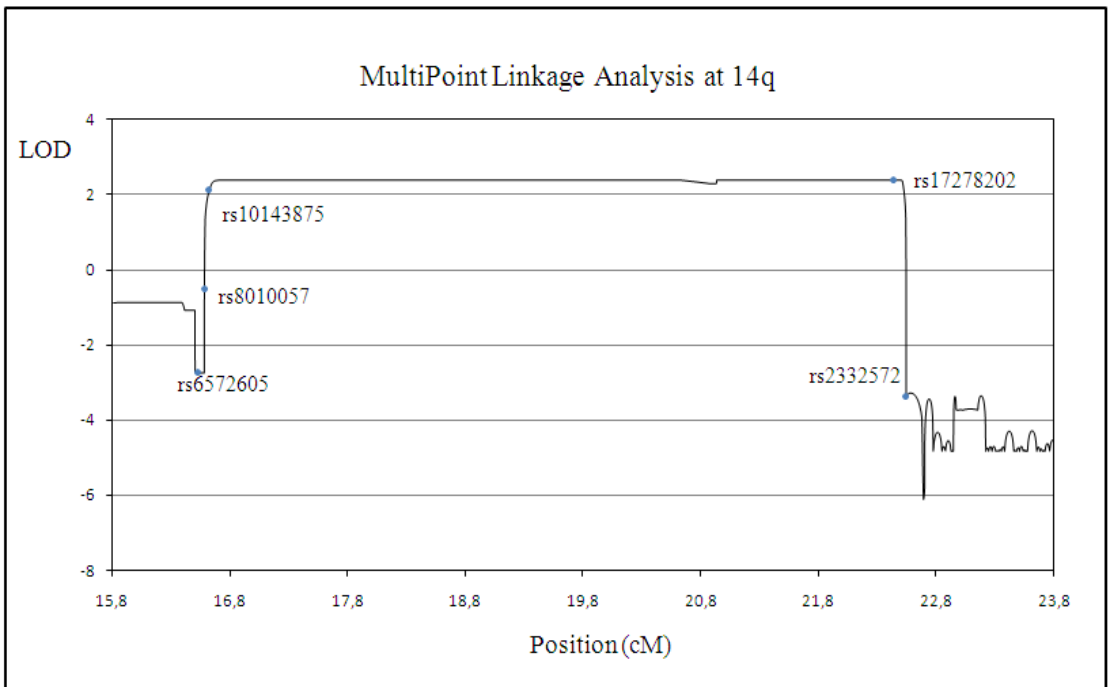


Figure 5.3. Multipoint lod score curve for azoospermia at 14q11.2-q12. It was obtained by using 978 informative SNP markers.

5.1.2. Candidate Gene Approach

Candidate gene analysis was carried out solely at candidate locus 17pter-p13.1, because the homozygosity was longer there than at 14q, making it a stronger candidate. According to NCBI Build 36.3, the region included 189 genes. Seven candidate genes *P2RX1*, *SPATA22*, *SPAG7*, *INCA1*, *TEKT1*, *GSG2* and *YBX2* were analyzed for mutations in one of the azoospermic brothers. A novel variant identified in exon 5 of *SPAG7* was assigned as the causative mutation, because it was not observed in the general population and computational prediction tools assigned it as harmful to protein function with high reliability. The details of the analysis of candidate genes are given below.

P2RX1 (Purinergic receptor *P2X*, ligand gated channel 1) was considered the strongest candidate gene. Initially the coding regions in the twelve exons of the gene, later 5'UTR and 3'UTR and finally the putative regulatory regions were screened for mutations by SSCP in an affected member of the azoospermia family and 14 unrelated azoospermic male. No aberrant mobility pattern was observed in affected sib and the mother in coding regions (Figure 5.4) or UTR (Figure 5.5).

All exonic regions in one affected sib (501) were resequenced to detect any mutation that could have possibly escaped detection by SSCP, and again no sequence variation was found. Extent of resequencing of the gene is given in Table 5.4. Aberrant SSCP mobility patterns were observed in unrelated individuals studied. Those variant fragments also were subjected to DNA resequencing. Table 5.5 lists the 15 sequence variants detected in this study. One of them (c.1033-30C>T) was novel. Observation of this variant particularly in a CBAVD patient implies a possible role for this variant in the etiology of CBAVD. It was not found in the other 12 unrelated azoospermic male analyzed. Human Splicing Finder program that predicts the effect of sequence variants on splicing (Desmet *et al*., 2009) showed that the variant disrupts three splicing enhancer motifs for SRp40 and SF2/ASF proteins and leads to the formation of one enhancer motif for SRp55 and a silencer motif, indicating the variant could be pathogenic.

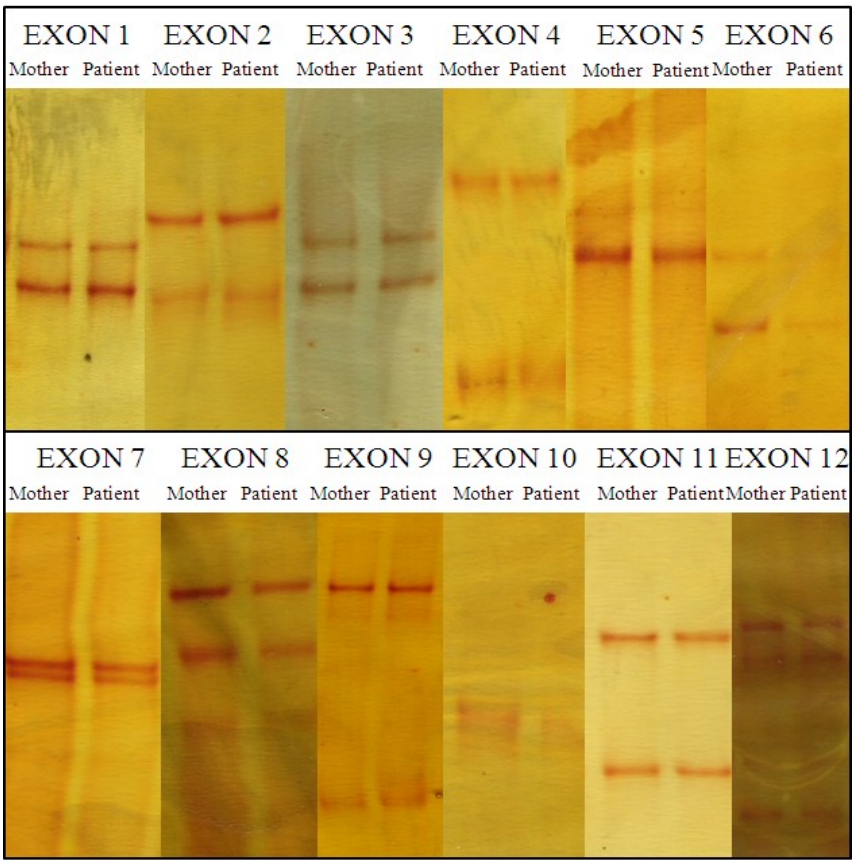


Figure 5.4. SSCP gels for *P2RX1* coding regions

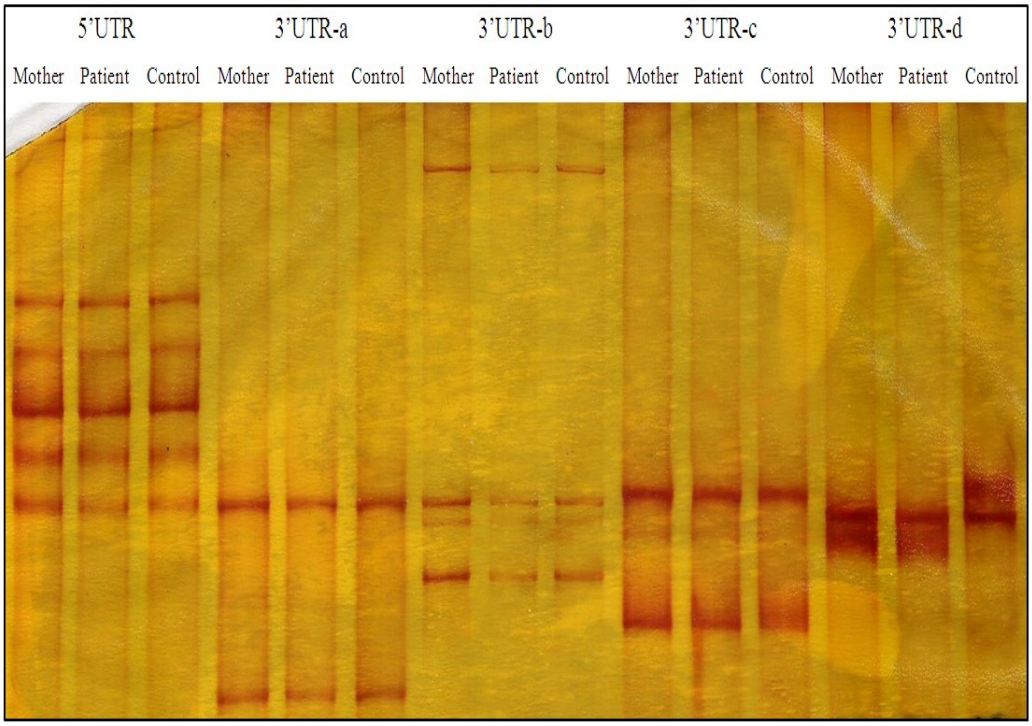


Figure 5.5. SSCP gel for *P2RX1* 5'UTR and 3'UTR

Table 5.4. Extent of resequencing in *P2RX1* in azoospermia family

Gene Region	Sequence Read	Gene Region	Sequence Read	Gene Region	Sequence Read
P1	-1757 to -2162	Exon 3	-125 to +52	Exon 8	-33 to +37
P2	-2248 to -2624	Exon 4	-11 to +43	Exon 9	-54 to +60
P3	-124 to -498	Exon 5	-57 to +78	Exon 10	-67 to +60
Exon 1	-11 to +57	Exon 6	-	Exon 11	-69 to +43
Exon 2	-17 to +43	Exon 7	-58 to +74	Exon 12	-66 to +93

Table 5.5. *P2RX1* sequence variants detected/identified in azoospermia family and single subjects

Subjects	Variant	Location	Allele freq.	dbSNP/novel
501, A11	c.-2759A>C	P2	0.697	rs1516801
A2, A11	c.-2812A>G	P2	-	rs3760473
501, A7	c.-2550A>G	P1	0.19	rs60539718
501, A7	c.-2469A>C	P1	0.576	rs59283857
A7	c.-2448A>T	P1	0.5	rs8068167
501, A4, A7	c.-2400C>T	P1	0.663	rs4995289
A7	c.-2434T>C	P1	0.415	rs4995288
501, A4, A7	c.-2321T>C	P1	-	rs4995290
501, A7	c.-720G>C	P3	0.682	rs11078476
501, A7	c.-813 A>G	P3	0.692	rs11078477
A7	c.-645 A>C	P3	-	rs8080074
A5, A6, A7	c.606-27T>A	Intron 6	-	rs7219732
A13	c.1033-30C>T	Intron 10	-	Novel
A7	c.*890G>A	3'UTR	-	rs28930668
A7	c.*1190C>G	3'UTR	0.207	rs6502752

SPATA22 (Spermatogenesis associated 22), also known as testis development-related *NYD-SP20*, stood out as the second strongest candidate disease gene. All nine exons of the gene and three putative regulatory regions were screened for mutations using SSCP analysis in the affected sib of the study family and unrelated azoospermic male group. All amplified fragments in affected sib 501 were later resequenced together with samples from other subjects showing aberrant mobility patterns on SSCP gels. Table 5.6 gives the extent of resequencing. Table 5.7 lists *SPATA22* sequence variants detected in this study. Three of the seven variants, g-288C>A, g.-278A>G and c.922A>C were novel. Their allele frequencies in all chromosomes studied in unrelated azoospermia group were 0.077 (2/26), 0.077 (2/26) and 0.038 (1/26), respectively. Chromatograms of the variants located at the exons are presented in Figure 5.6.

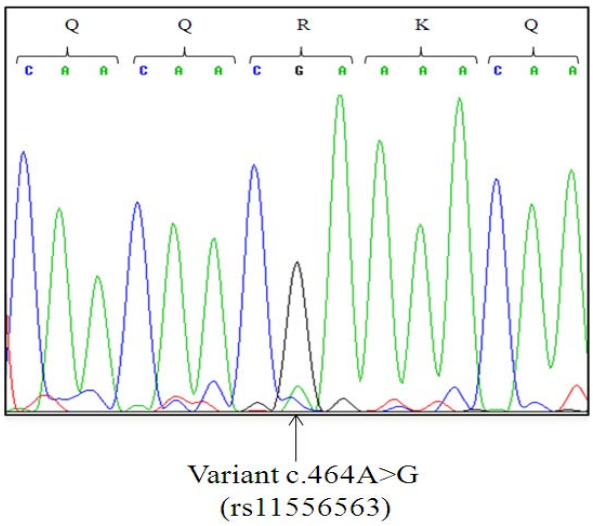
Table 5.6. Extent of resequencing in *SPATA22* in azoospermia family

Gene Region	Sequence Read	Gene Region	Sequence Read
P3	-832 to -1122	Exon 4	-58 to +53
P2	-476 to -652	Exon 5	-98 to +46
P1	-322 to +54	Exon 6	-11 to +49
Exon 1	-39 to -13	Exon 7	-60 to +35
Exon 2	-48 to +94	Exon 8	-25 to +23
Exon 3	-32 to +76	Exon 9	-81 to +39

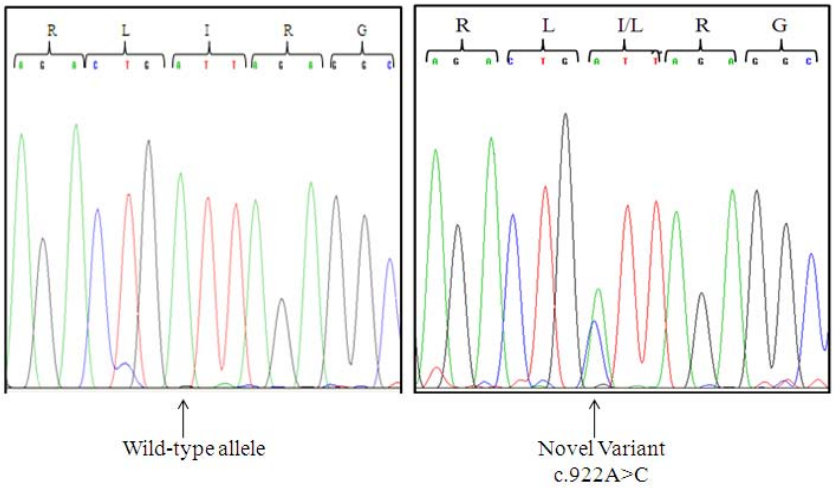
Table 5.7. *SPATA22* sequence variants detected/identified in azoospermia family and single subjects

Subjects	Variants	Location	Effect on Protein	Allele Frequency		dbSNP/novel
				Observed in Azoospermia	Reported	
A5, A7	g-288C>A	P1	-	0.077 (2/26)	-	Novel
501	g.-278A>G	P1	-	0.077 (2/26)	-	Novel
501, A4 A6, A7	g.-186T>G	P1	-	0.31(5/26)	-	rs62071295
501, A4 A6, A7	g.-55C>A	P1		0.31(5/26)	-	rs62071294
501	c.464A>G	Exon 6	Q155R	0.077 (2/26)	0.54	rs11556563
A11	c.922A>C	Exon 9	I308L	0.038 (1/26)	-	Novel
A3,A11	c.999G>A	Exon 9	P333P	0.077 (2/26)	0.017	rs9901726

a.



b.



c.

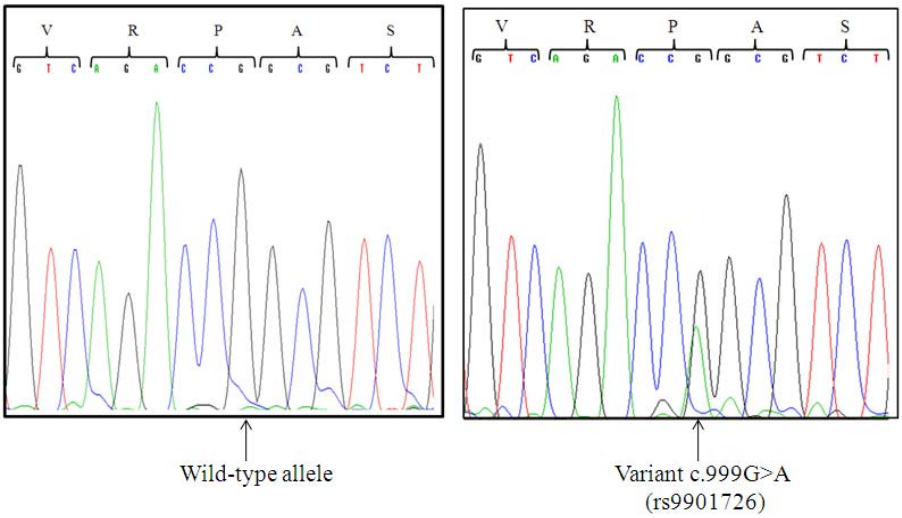


Figure 5.6. Chromatograms showing *SPATA22* variants and wild-type alleles (a-c).

SPAG7 (Spermatogenesis associated antigen 7) was the third candidate gene. All exons except exon 5 were analyzed for mutations by SSCP in an affected sib of the family (501) and unrelated azoospermic subjects (Figure 5.7). All exons of the affected sib 501 and any sample exhibiting an aberrant pattern was further investigated by sequence analysis (Table 5.8).

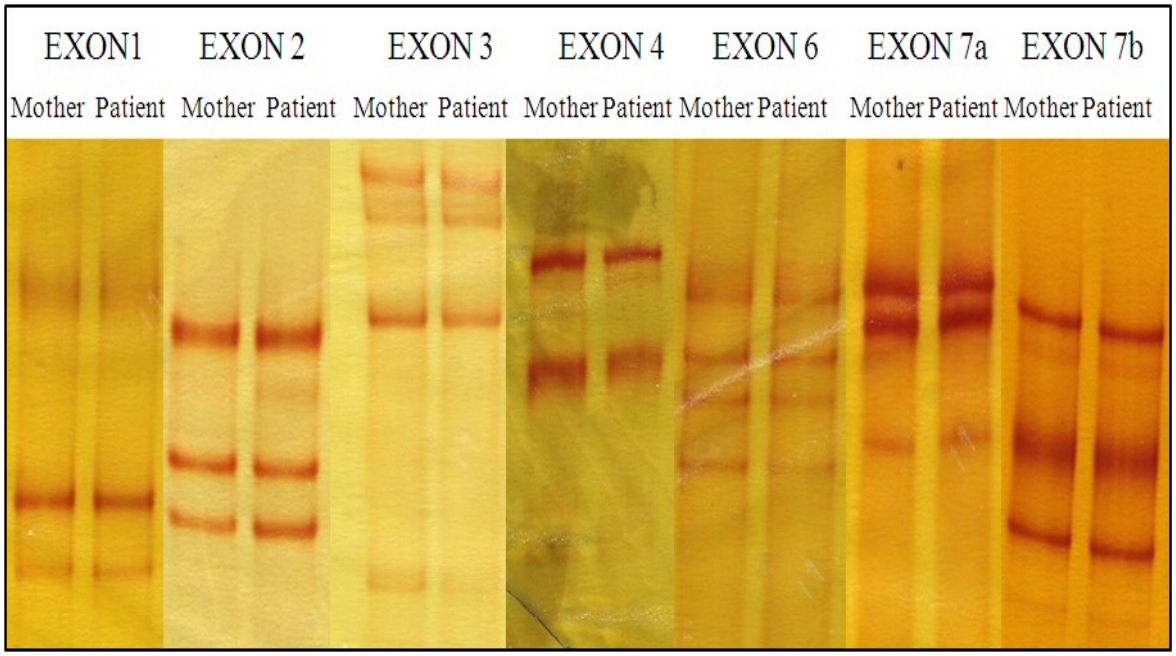


Figure 5.7. SSCP gels for *SPAG7* exons

Table 5.8. Extent of resequencing in *SPAG7* in azoospermia family

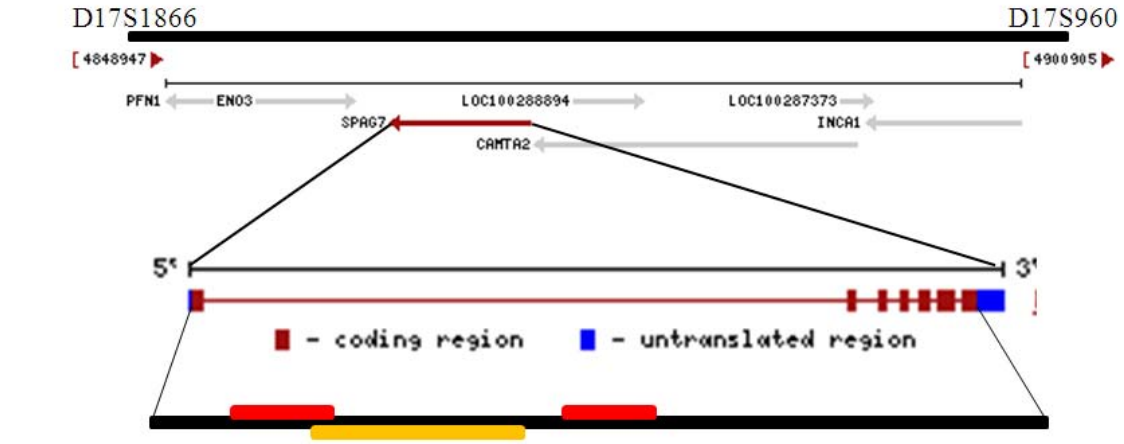
Exon	Sequence Read
Exon 1	-116 to +43
Exon 2	-71 to +63
Exon 3	-80 to +63
Exon 4	-80 to +90
Exon 5	-162 to +105
Exon 6	-85 to +53
Exon 7	-53 to +73

Novel variant c.406C→T was identified in the homozygous state in exon 5 in the affected sib (Figure 5.8). Analysis of the family revealed that all three affected brothers were homozygous for this variant while the mother and the healthy brother were heterozygous. Several observations strongly suggested that this novel variant is a mutation and not a polymorphism. At the protein level, the nucleotide transition led to the substitution of tryptophan (W) for arginine (R) at residue 136. One hundred one individuals from the general population were screened by SSCP analysis, and R136W was not found in any. Online tools including PolyPhen, SIFT, and MMB predicted the amino acid substitution as damaging to the protein function by high reliability (Table 5.9). The residue was highly conserved among species (Figure 5.10).

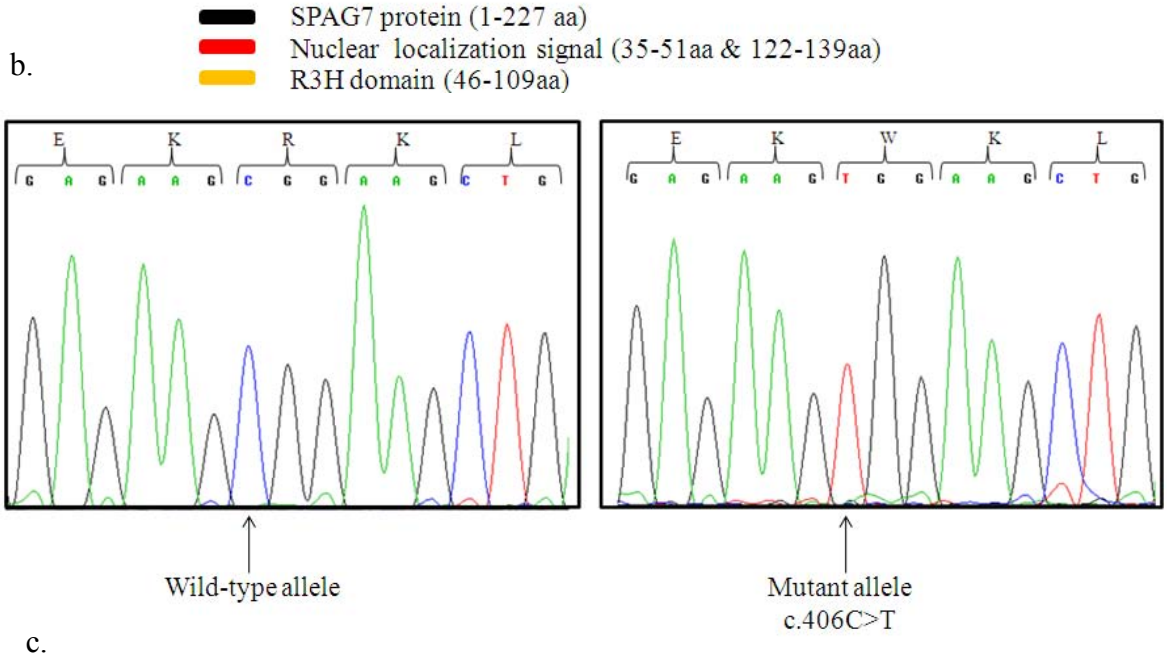
Two novel sequence variants, namely c.327+60C>A, and c.575-23C>T were identified in 14 unrelated azoospermia subjects (Table 5.10). Their frequencies were determined as 0.018 (4/220) and 0.025 (3/224) in the general population. The frequency data suggested that these two novel variants are polymorphisms, rather than mutations.

After the establishment of *SPAG7* as the disease gene in the family, all seven exons of the gene were screened for mutations by SSCP in a wide patient group. The group was composed of 43 azoospermic and 14 oligospermic unrelated azoospermia male, all with normal karyotypes and no Y chromosome microdeletions. Table 5.10 lists all the sequence variants observed in the patient groups. Three novel sequence variants c.417+7T>A, c.669G>A and c.678C>A were identified. Their allele frequencies in all chromosomes studied in unrelated azoospermic subjects was 0.067 (1/150). SSCP gels and chromotograms of sequence variants are presented in Figure 5.10 and Figure 5.11, respectively. c.669G>A and c.678C>A are synonymous mutations. According to human splicing prediction program, c.669G>A leads to the formation of an enhancer motif and disruption of a silencer motif, while c.678C>A does not alter splicing motifs. c.417+7T>A was located near the exon/intron junction. Variant c.417+7T>A is located near the exon intron junction, suggesting a role in splicing.

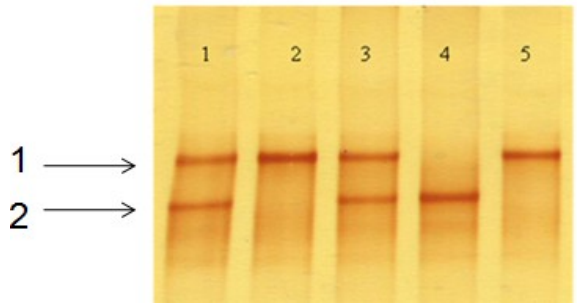
a.



b.



c.



Lane ID	1	2	3	4	5
DNA ID	406	502	503	504	505
Genotype	1/2	1/1	1/2	2/2	1/1

Figure 5.8. Mutation *SPAG7* R136W. a. Genomic structure of *SPAG7* and schematic representation of its protein. b. Chromatograms showing the transition c.406C→T in one affected sib (501) and wild-type allele in a control. c. SSCP results for c.406C→T (allele1) screening in azoospermia family and alleles compiled in a table.

Table 5.9. Effect of R136W substitution on protein function as predicted by online tools

Program	Score	Prediction
MMB	0.9898/1.0	Pathological with high reliability
SIFT	0.01	Affects protein function
PolyPhen	1.95	Possibly damaging

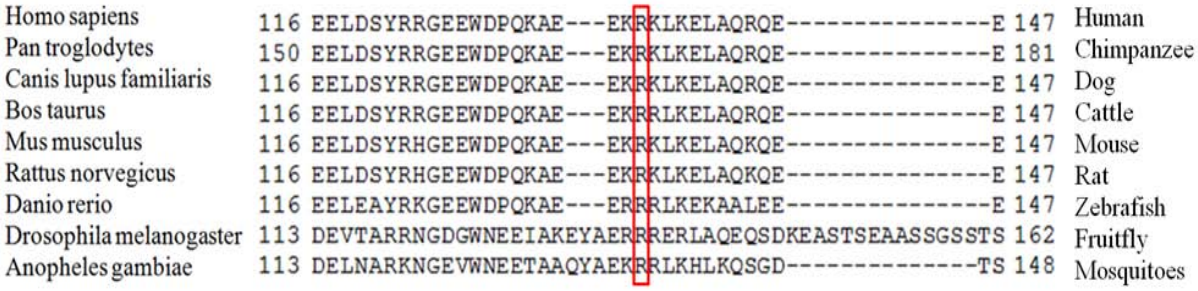


Figure 5.9. Evolutionary conservation of p.R136 amino acid residue

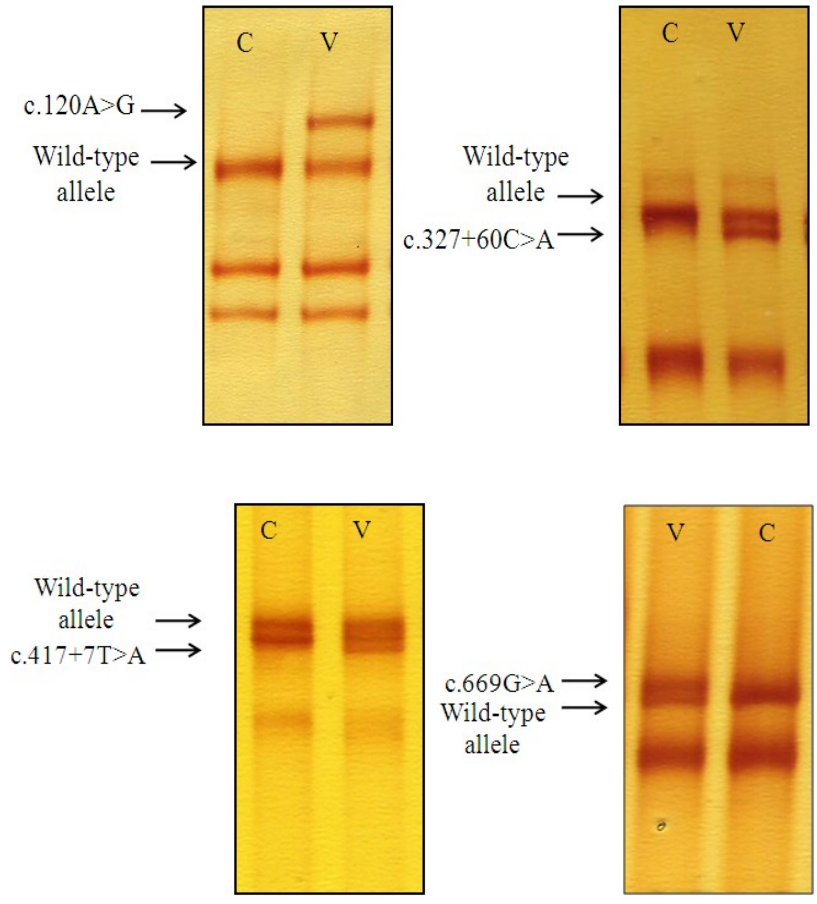


Figure 5.10. SSCP gels for *SPAG7* sequence variants

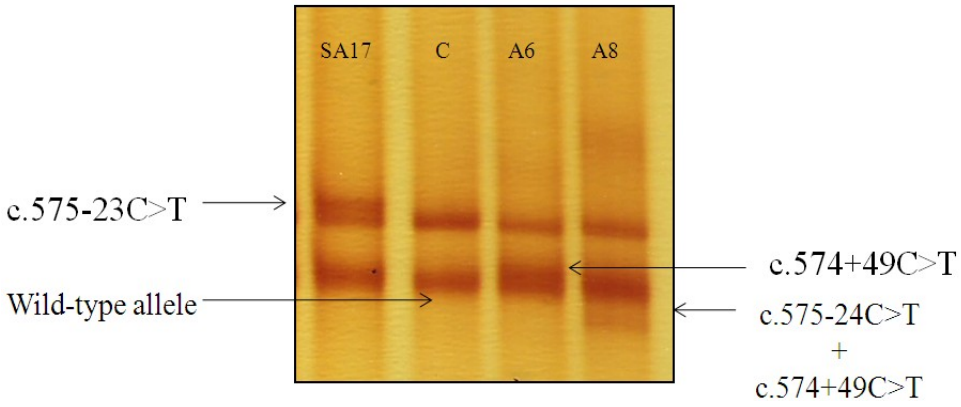
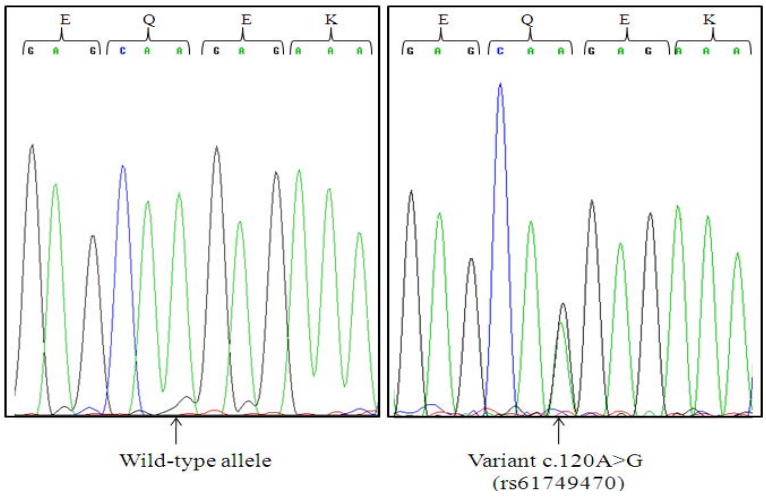


Figure 5.10. SSCP gels for *SPAG7* sequence variants (Continued)

a.



b.

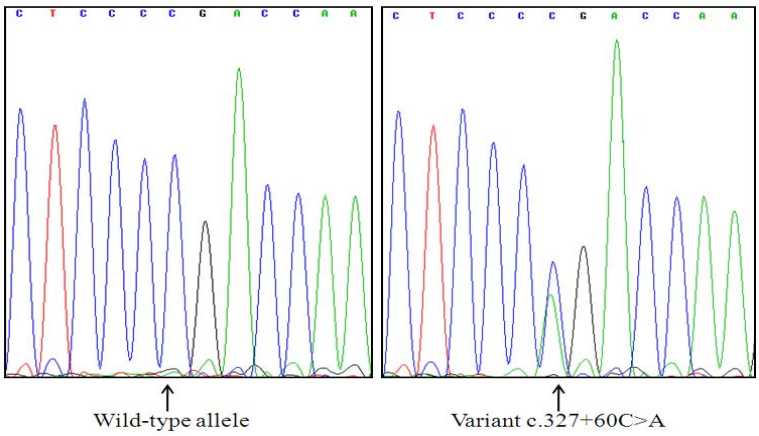
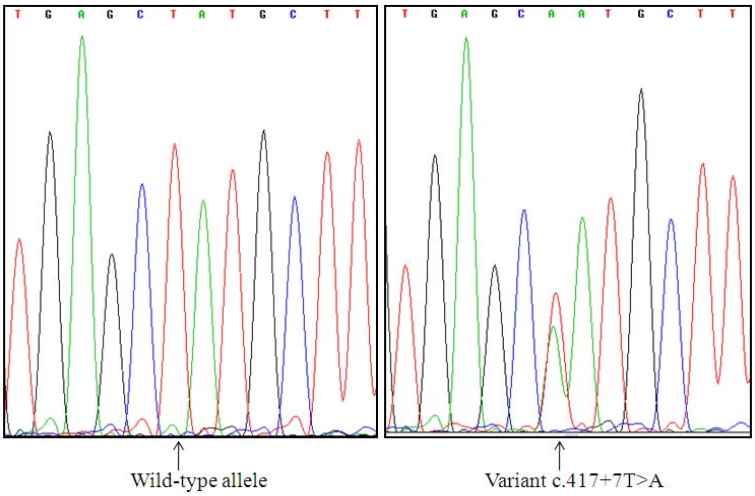
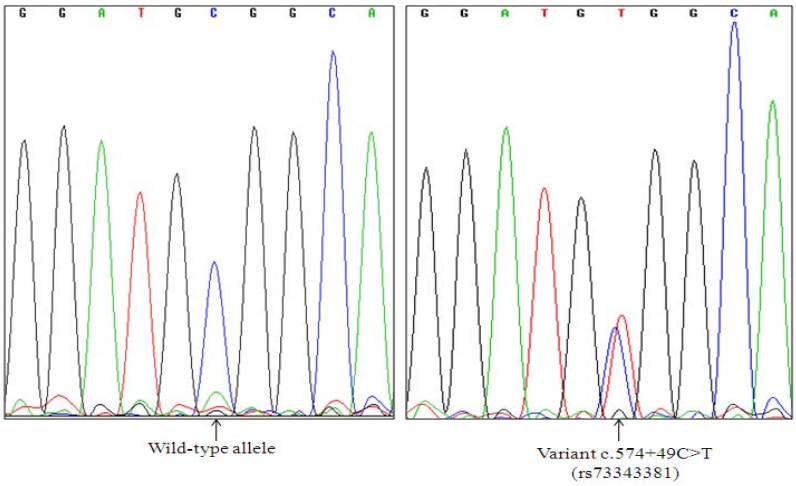


Figure 5.11. Chromatograms for *SPAG7* sequence variants in azoospermia patients and the corresponding wild-type alleles (a-h). The corresponding nucleotide sequences are shown above each chromatogram.

c.



d.



e.

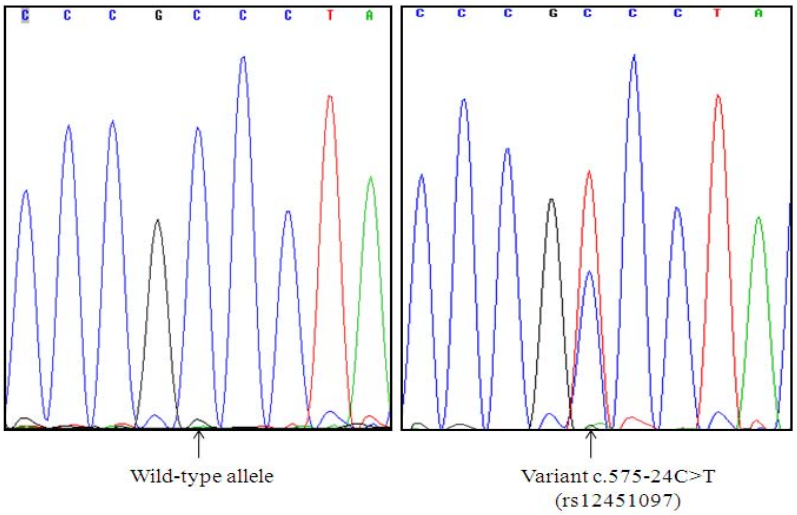
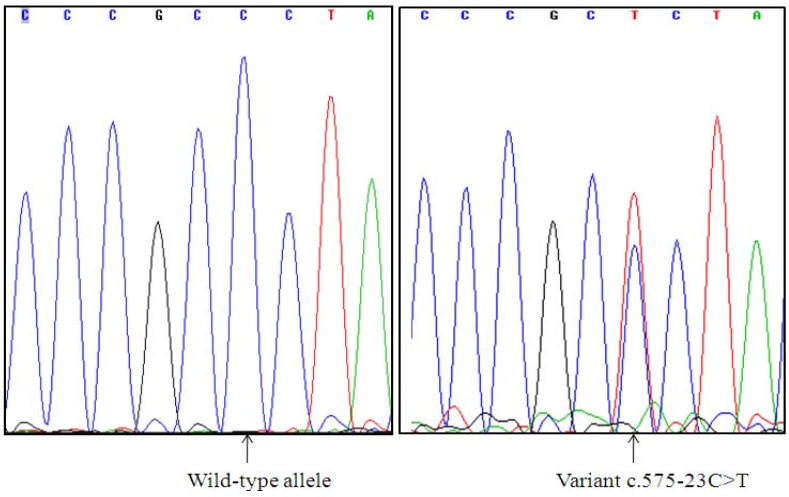
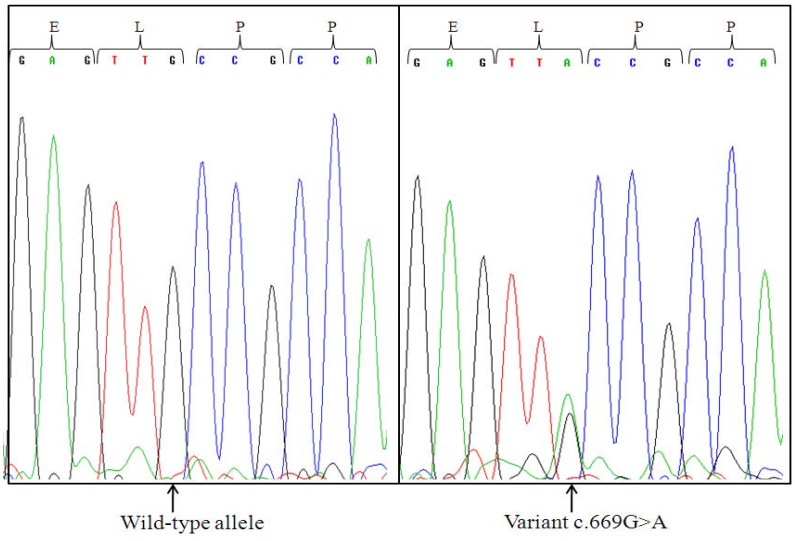


Figure 5.11. Chromatograms showing *SPAG7* sequence variants in azoospermia patients and the corresponding wild-type alleles (a-h). The corresponding nucleotide sequences are shown above each chromatogram (Continued).

f.



g.



h.

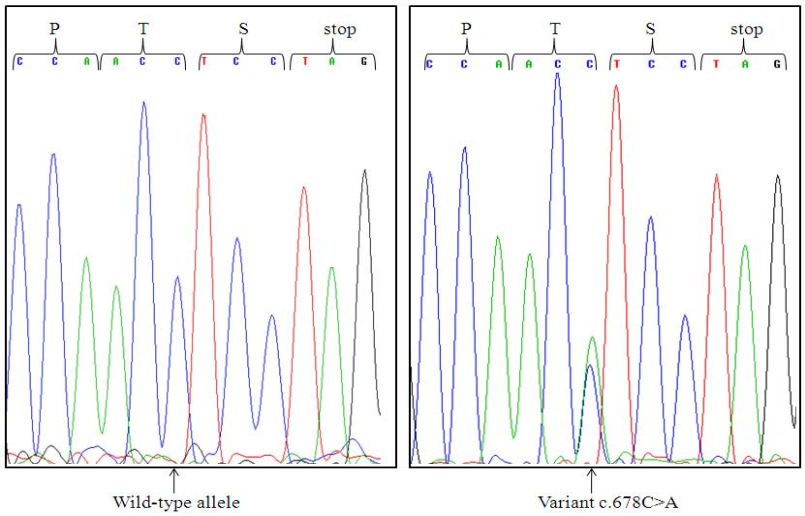


Figure 5.11. Chromatograms showing *SPAG7* sequence variants in azoospermia patients and the corresponding wild-type alleles (a-h). The corresponding nucleotide sequences are shown above each chromatogram (Continued).

Table 5.10. *SPAG7* sequence variants detected/identified in this study

Subjects	Variant (dbSNP/novel)	Location	Effect	Observed Frequency	
				in azoospermia	in popul.
A5,SA17,SA28, SA40, O16	c.120A>G (rs61749470)	Exon 2	Synonymous (p.Q40Q)	0.033 (5/150)	-
A5,SA24,SA35	c.327+60C>A (Novel)	Intron 4	Intronic	0.0203 (3/148)	0.018 (4/220)
501	c.406C>T (Novel)	Exon 5	Missense (R136W)	0.0 (0/148)	0.0 (0/202)
SA45	c.417+7T>A (Novel)	Intron 5	possibly splicing	0.0067 (1/148)	-
A6,A8,SA19, SA20,SA22,SA23, SA24,SA27,SA32, SA45, O08,O10	c.574+49C>T (rs73343381)	Intron 6	Intronic	0.0810 (12/148)	0.103 (23/224)
A8,SA10,SA19, SA20,SA22,SA23, SA24,SA27,SA32, O08	c.575-24C>T (rs12451097)	Intron 6	Intronic	0.067 (10/148)	0.089 (20/224)
A5,SA17,SA28, SA40,O16	c.575-23C>T (Novel)	Intron 6	Intronic	0.0333 (5/150)	0.013 (3/224)
SA12	c.669G>A (Novel)	Exon 7	Synonymous (p.L223L)	0.0067 (1/150)	-
O06	c.678C>A (Novel)	Exon 7	Synonymous (p.T226T)	0.0067 (1/150)	-
A3, A6,A7, A12,SA4,SA6, SA20,SA25,SA29, SA33,SA48,SA42, O04,O05,O06, O07,O14	c.*189_*190del2	3'UTR	Polymorphism	0.108 (16/148)	-

Four other candidate genes residing in the region, namely *INCA1*, *TEKT1*, *GSG* and *YBX2,* were also analyzed for mutations to make sure that they did not harbor a causative mutation. All eight exons of both *INCA1* and *TEKT1* were amplified using intronic primers and subjected to direct resequencing. The extent of resequencing is given in Table 5.11 and Table 5.12, respectively. Seven variants that are detected in noncoding regions in *INCA1* are all reported in SNP databases (Table 5.13). No sequence variation was observed in

TEKT1. These results indicated that *INCA1* and *TEKT1* do not carry causative mutation for azoospermia. Haploid germ cell-specific nuclear protein kinase (*GSG2*) was also assessed as a good candidate. The intronless gene was amplified in five fragments and resequenced in affected sib 501. No variant was detected, implying *GSG2* is not responsible for azoospermia.

Table 5.11. Extent of resequencing of *INCA1* in azoospermia family

Exon	Sequence Read	Exon	Sequence Read
Exon 1	-210 to +336	Exon 5	-222 to +130
Exon 2	-	Exon 6	-237 to +154
Exon 3	-734 to +74	Exon 7	-107 to +371
Exon 4	-174 to +222	Exon 8	-250 to +80

Table 5.12. Extent of resequencing of *TEKT1* in azoospermia family

Exon	Sequence Read	Exon	Sequence Read
Exon 1	-304 to +147	Exon 5	-348 to +156
Exon 2	-109 to +171	Exon 6	-187 to +142
Exon 3	-117 to +211	Exon 7	-380 to +142
Exon 4	-180 to +412	Exon 8	-425 to +98

Table 5.13. *INCA1* sequence variants detected in azoospermia family

Variants	Location	Reported Allele freq	dbSNP
c.-328+14A>G	Intron1	-	rs886582
c.-328+162T>C	Intron1	0.56	rs1004379
c.-38-763C>T	Intron 2	-	rs471064
c.-38-161T>C	Intron 2	-	rs9897114
c.158+90C>G	Intron 4	0.37	rs427703
c.517-107C>T	Intron 7	-	rs426475

The last three exons of *YBX2* that were located at the boundary of the candidate region (between the most centrometic homozygous marker and the closest heterozygous marker) were analyzed by direct sequencing in affected sib 501 (Table 5.14). Sequence variants in the heterozygous state were detected in intron 6 and 7 as well as in the last exon (Table 5.15). This finding facilitated the narrowing down of the candidate locus from 0-7.134Mb to 0-7.133Mb and excluding *YBX2* from the candidate region.

Table 5.14. Extent of *YBX2* resequencing in azoospermia family

Exon	Sequence Read
Exon 7	-263 to +77
Exon 8	-365 to +53
Exon 9	-172 to +124

Table 5.15. Heterozygous *YBX2* sequence variants in azoospermia family

Variants	Location	Reported Allele freq	DbSNP
c.848+32C>T	Intron 6	0.45	rs3744405
c.849-16T>C	Intron 6	-	rs55781423
c.849-13delT	Intron 6	-	rs3833134
c.1045-41A>C	Intron 7	-	rs222841
c.*16G>A	3'UTR	0.644	rs222842

In summary, *SPAG7* was identified as the gene responsible for azoospermia in the family and mutation R136W as the underlying gene defect. In other idiopathic azoospermia/oligozoospermia subjects, novel variants were identified (Table 5.16).

Table 5.16. Novel variants identified in this study for azoospermia

Gene	Variant	Location	Effect	Allele Frequency	
				in azoospermia	in population
P2RX1	c.1033-30C>T	Intron	intronic	0.038 (1/26)	-
SPATA22	g.-288C>A	Promoter		0.077 (2/26)	-
SPATA22	g.-278A>G	Promoter		0.077 (2/26)	-
SPATA 22	c.922A>C	Exon 9	missense (I308L)	0.038 (1/26)	-
SPAG7	c.327+60C>A	Intron 4	intronic	0.0203 (3/148)	0.018 (4/220)

Table 5.16. Novel variants identified in this study (Continued)

SPAG7	c.406C>T (Novel)	Exon 5	missense (R136W)	0.0 (0/148)	0.0 (0/202)
SPAG7	c.417+7T>A (Novel)	Intron 5	possibly splicing	1/148 (0.0067)	-
SPAG7	c.575-23C>T (Novel)	Intron 6	intronic	0.0333 (5/150)	0.013 (3/224)
SPAG7	c.669G>A (Novel)	Exon 7	synonymous (p.L223L)	0.0067 (1/150)	-
SPAG7	c.678C>A (Novel)	Exon 7	synonymous (p.T226T)	0.0067 (1/150)	-

5.2. MHAC

5.2.1. Linkage Analysis

Genome scan for MHAC family was performed with 250 000 SNPs. Two-point and multipoint lod scores were calculated under the models of autosomal recessive inheritance with full penetrance and autosomal dominant inheritance with 80 per cent penetrance. Despite the fact that no parental consanguinity had been declared, the parents were assumed distant relatives, and lod score calculations were performed assuming them third, fourth or fifth cousins. Loci of variable sizes with lod scores of about 3 were observed on 9 chromosomes in an autosomal recessive model even when parents were assumed third cousins (Figure 5.12).

Homozygosity mapping strategy was applied as a complementary method to map the disease locus. Multiple homozygous regions that were shared by affected siblings but not parents were determined. Table 5.17 lists eight homozygous regions on three chromosomes that are larger than 1.5 cM. A maximum lod score of almost 3 was obtained in the multipoint linkage analysis, indicating that all those eight regions were equally almost significant (Figure 5.13). In linkage analysis using a vast number of SNP markers, it is common to observe more than one homozygous locus. In this study, the absence of healthy siblings prevented exclusion of loci that are not linked to the disease. Since the disease locus harboring the causative mutation is generally larger than the homozygous loci that arise due to pure chance, only homozygous regions larger than 3 cM were further investigated with additional microsatellite markers. Only one of them is expected to be the

gene locus. Genotyping with microsatellite markers at those loci verified true homozygosity at only 6p25.3-p23 (Table 5.18) and 19q12 (Table 5.19).

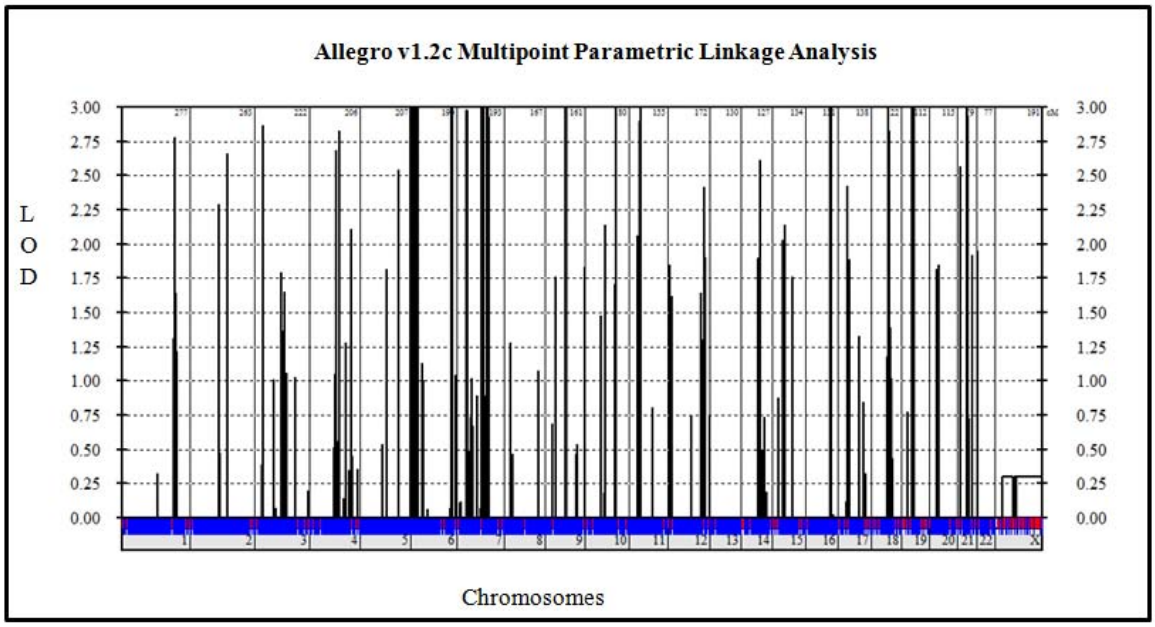
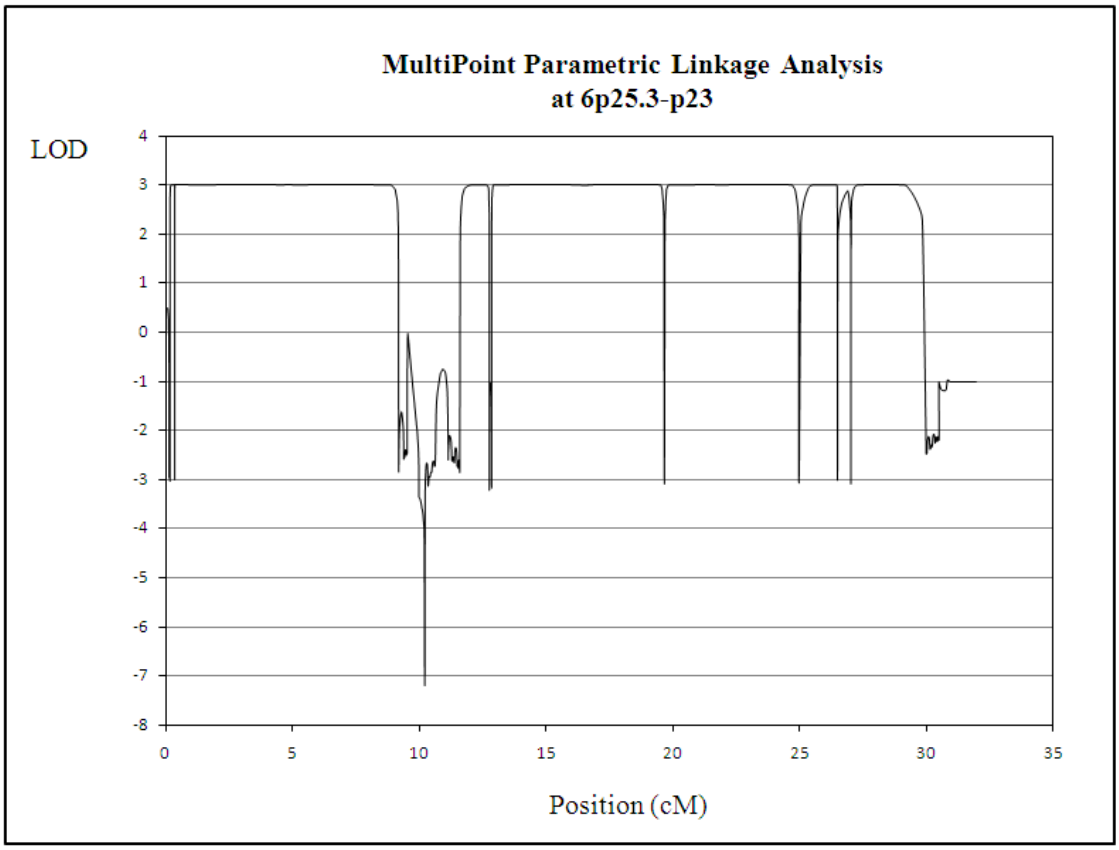
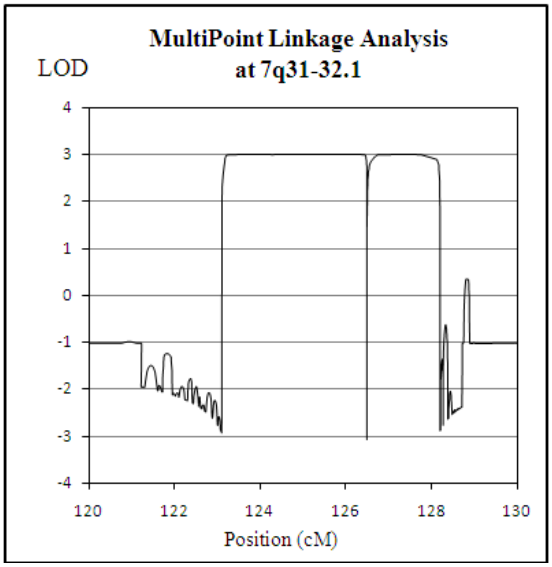


Figure 5.12. Multipoint lod scores for MHAC family. It is obtained by using the SNP genome scan data. Parents were assumed third cousins in a recessive model with full penetrance.

Table 5.17. MHAC candidate loci. Homozygous loci shared only by sibs, their chromosomal locations, the flanking markers and the sizes.

Locus	Location (Mb)	Flanking Markers	Size (cM)
6p25.3-p25.2	0 – 3.43	rs2671415-rs2183307	9
6p25.1	4.96 – 6.43	rs2764120-D6S1677	4.08
6p25.1-p24.3	6.43 – 7.73	D6S1677-rs5009024	2.74
6p24.3-p24.2	7.73 – 10.68	rs5009024-rs609800	5.31
6p24.1	11.58 – 12.89	rs6907890-rs9472551	5
7q31-q31.32	116.15 – 122.80	rs38857-rs4367475	3.3
7q31.32-7q31.33	122.80 – 125.22	rs4367475- rs17148961	1.71
19q12	33.50 – 35.86	rs4805151-rs746668	3.22





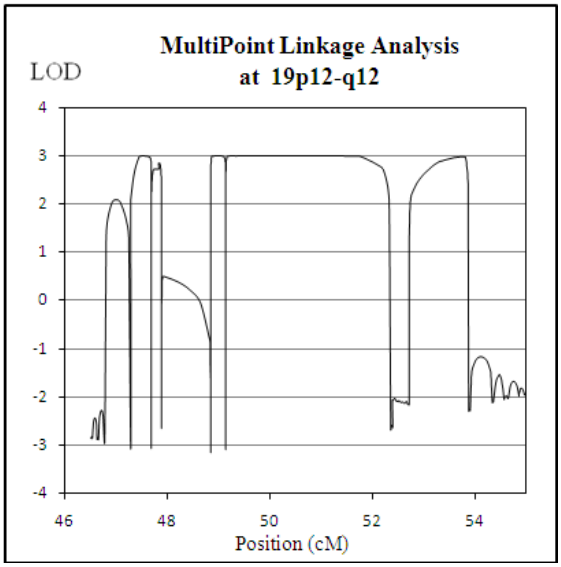


Figure 5.13. Multipoint parametric lod score curves for MHAC candidate loci. Eight homozygous regions shared only by affected siblings, within 6p25.3-p23, 7q31-q32.1 and 19p12-q12.

Table 5.18. Haplotypes at 6p25.3-24.1 including five shared homozygosity region. For regions genotyped by SNP markers only, the flanking heterozygous markers are given only. Markers in bold are microsatellites.

Description	Marker ID	Position (Mb)	F		M		P1		P2	
Region1	rs 2671415	0.31	-	-	A	B	A	B	A	B
	D6S967	1.43	1	3	1	3	1	1	1	1
	D6S3103kb	3.10	1	1	1	2	1	1	1	1
	rs2183307	3.429	A	B	A	A	A	B	A	B
Region 2	rs2764120	4.96	A	B	A	B	A	B	A	B
	D6S477	6.09	2	1	2	2	2	2	2	2
	D6S1677	6.43	1	2	2	1	1	2	1	2
Region 3	**D6S1677**	6.43	1	2	2	1	1	2	1	2
	D6S1668	6.59	3	2	3	1	3	3	3	3
	D6S1598	6.94	1	1	1	1	1	1	1	1
	D6S1547	7.46	1	1	1	1	1	1	1	1
	rs5009024	7.73	A	B	A	B	A	B	A	B
Region 4	rs5009024	7.73	A	B	A	B	A	B	A	B
	D6S309	8.17	1	2	1	3	1	1	1	1
	rs609800	10.68	A	B	A	B	A	B	-	-
Region 5	rs6907890	11.58	A	B	B	B	-	-	A	B
	rs9472551	12.89	B	B	A	B	A	B	A	B

Table 5.19. Haplotypes at 19q12, a shared homozygosity region. For regions genotyped by SNP markers only, only the flanking heterozygous markers are given. Markers in bold are microsatellites.

Marker ID	Position (Mb)	F		M		P1		P2	
rs4805151	33.50	A	B	A	B	A	B	A	B
D19S561	34.57	2	3	1	2	2	2	2	2
D19S405	35.42	1	2	2	1	1	1	1	1
rs746668	52.35	A	A	A	B	A	B	A	B

5.2.2. Candidate Gene Approach

Candidate gene search was initiated at 6p25.3-6p25.2, the largest of the candidate loci identified for the disease in the family. According to NCBI Build 36.3, this locus was approximately 9 cM and included 41 genes.

Two beta II tubulin genes *TUBB2A* and *TUBB2B* and a pseudo gene *LOC10132153* with high degree of similarity to tubulin gene stood out as the strongest MHAC candidate genes. All four exons each of *TUBB2A* and *TUBB2B* and five predicted exons of *LOC10132153* were subjected to direct resequencing in an affected sib. The extent of resequencing are given in Table 5.20, Table 5.21 and Table 5.22. No variant was found in *TUBB2B*. All sequence variations detected in *TUBB2A* (Table 5.23) and *LOC10132153* (Table 5.24) were polymorphisms that were reported in Entrez SNP database

Table 5.20. Extent of resequencing in *TUBB2A*

Exon	Sequence Read
Exon 1	-360 to +245
Exon 2	-130 to +114
Exon 3	-155 to +176
Exon 4	-133 to +180

Table 5.21. Extent of resequencing in *TUBB2B*

Exon	Sequence Read
Exon 1	-131 to +145
Exon 2	-65 to +147
Exon 3	-111 to +136
Exon 4	-87 to +19

Table 5.22. Extent of resequencing in *LOC10132153*

Exon	Sequence Read
Exon 1	-120 to +540
Exon 2	-540 to +125
Exon 3	-390 to +125
Exon 4	-305 to +227
Exon 5	-246 to +125

Table 5.23. *TUBB2A* sequence variants

Variants	Location	Effect	Reported Allele Freq.	dbSNP
c.-278A>G	5' near gene	-	-	rs2326176
c.-112G>A	5' near gene	-	-	rs909965
c.-101C>T	5' near gene	-	-	rs909964
c.57+199G>C	Intron 1	Intronic	-	rs909963
c.57+242T>C	Intron 1	Intronic	0.04	rs909962
c.564T>G	Exon 3	synonymous	-	rs17849443

Table 5.24. *LOC10132153* sequence variants detected

Variants	Location	dbSNP
c.48+245_48+246insA	Intron 1	rs5873862
c.48+260_48+261insGG	Intron 1	rs60117469
c.48+267_48+268insG	Intron 1	rs5873860
c.49-223_49-222insG	Intron 1	rs5873859
c.49-168_49-167insC	Intron 1	rs5873858

EXOC2 (Exocyst complex component 2) was assessed as another good candidate gene. Twenty-four of 28 exons were analyzed for mutations by direct sequencing (Table 5.25). One of the two variants (c.90+319C>A) detected in the analysis was novel (Table 5.26). The variant was located approximately 300 bp away from exon-intron junction, suggesting a possible role in splicing as a part of branch point sequence. Human splicing prediction programs did not assign any role in splicing to this variant.

In addition to the strongest candidate locus at 6p25.3-p25.2, four smaller candidate loci clustered at 6p25.1-p24.2 were analyzed. *NRN1 (*Neuritin 1) stood out as a good candidate gene. All three exons were amplified in six segments in total and subjected to direct resequencing. The extent of resequencing and the variants detected/identified are listed in Table 5.27 and Table 5.28, respectively. One of the three *NRN1* sequence variants detected was novel. The novel variant c.*374A>T was located at 3'UTR suggesting a role in poly adenylation mechanism. The variant was not a part of three known polyadenylation signals implying it is unlikely to be the causative mutation.

Table 5.25. Extent of resequencing in *EXOC2*

Gene Region	Sequence Read	Gene Region	Sequence Read
Exon 1	-240 to +370	Exon 15	-87 to +300
Exon 2	-198 to +234	Exon 16	-266 to +23
Exon 3	-524 to +133	Exon 17	-235 to +297
Exon 4	-34 to +80	Exon 18	-318 to +136
Exon 5	-161 to +109	Exon 19	-576 to +38
Exon 6	-182 to +136	Exon 20	-155 to +313
Exon 7	-152 to +279	Exon 21	-
Exon 8	-170 to +139	Exon 22	-113 to +259
Exon 9	-139 to +173	Exon 23	-382 to +262
Exon 10	-544 to +116	Exon 24	-305 to +160
Exon 11	-140 to +359	Exon 25	-250 to +140
Exon 12	-	Exon 26	-76 to +329
Exon 13	-90 to +289	Exon 27	-179 to +244
Exon 14	-	Exon 28	-

Table 5.26. *EXOC2* sequence variants detected/identified

Variation	Location	Effect	Reported Allele Freq.	dbSNP
c.90+319C>A	Intron 1	intronic	-	Novel
c*1085G>A	3'UTR	-	0.25	rs242914

Table 5.27. Extent of resequencing in *NRN1*

Gene Region	Sequence Read
Exon 1	-89 to +114
Exon 2	-130 to +116
Exon 3	-189 to +60

Table 5.28. *NRN1* sequence variants detected/identified

Variation	Position	dbSNP
c.201-27G>C	Intron 2	rs3749860
c.*374A>T	3'UTR	Novel
c.*883delA	3'UTR	rs3834316

In summary, 6p25.3-p25.2 was the strongest candidate locus for MHAC. Candidate genes *TUBB2A*, *TUBB2B*, *LOC10132153* and *EXOC2* (except for 4 of total 28 exons) were analyzed for mutations by direct sequencing. In addition, *NRN1* gene located at one of the seven other candidate loci (6p25.1-p24.2) was analyzed for mutations. No mutation but two novel variations (c.90+319C>A and c.*374A>T) were identified *NRN1* and *EXOC2.*

5.3. IDDM

Genome scan data generated by microsatellite markers were initially used to search for an IDDM disease locus, assuming the disease was transmitted by an autosomal recessive inheritance pattern in the family. However, no locus linking to the disease was identified. Thus, a second genome scan using SNP markers was performed. Two genome scan data were utilized together to search for candidate loci.

5.3.1. Linkage Analysis Based on Microsatellite Genome Scan Data

Using the microsatellite marker genome scan data, parametric linkage analysis was carried out assuming an autosomal recessive model under two different levels of penetrance. No candidate region was detected when all family members were included in the analysis under full penetrance (Figure 5.14) or 80 per cent penetrance (Figure 5.15).

In parametric linkage analysis of the genome scan data including only affected sibs and parents, multipoint lod scores >2.0 were observed at two loci: 2p14-q14.1 and 5q14.3-q35.1 (Figure 5.16). These two regions on chromosome 2 (Table 5.29) and chromosome 5 (Table 2.30) were excluded after fine-mapping with additional markers, since no shared homozygosity was found.

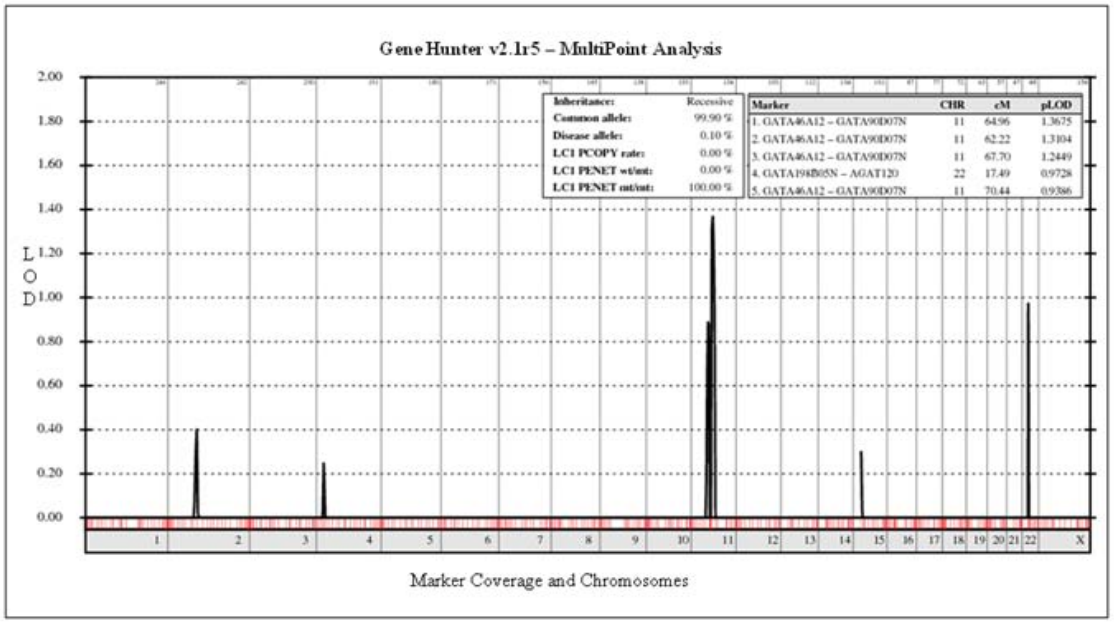


Figure 5.14. Multipoint parametric lod scores for IDDM with full penetrance. The total data set generated by microsatellite genome scan was used in recessive model with full penetrance.

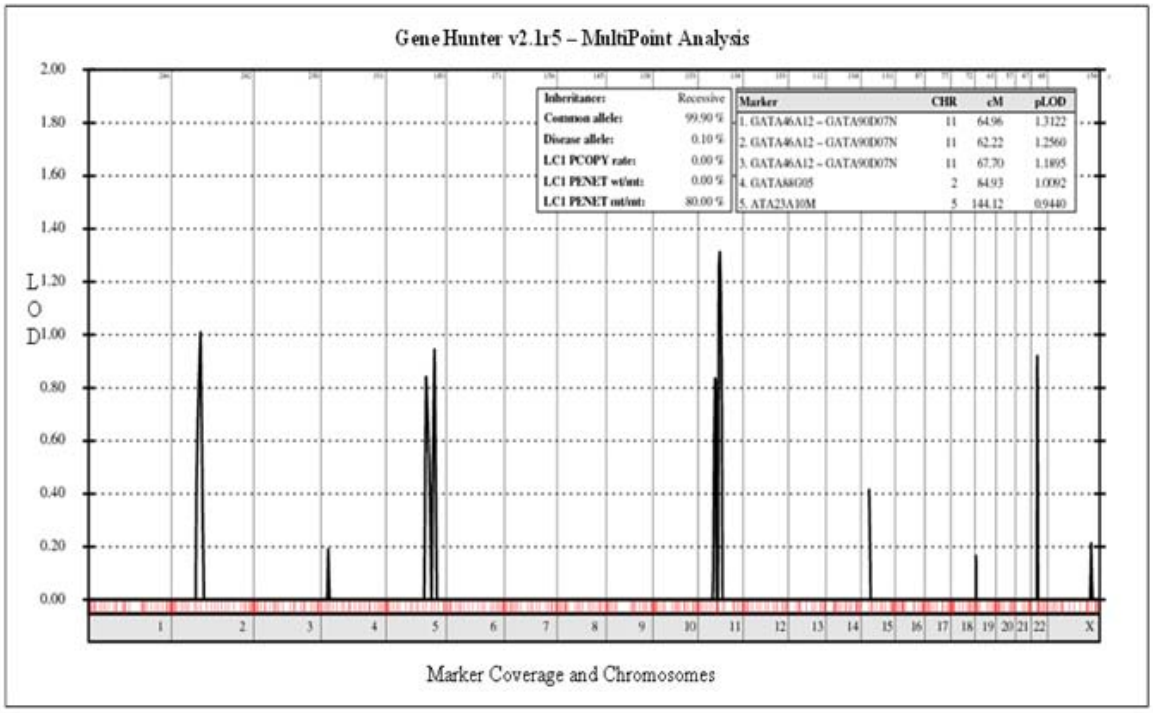


Figure 5.15. Multipoint parametric lod score for IDDM with 80 per cent penetrance. The total data set generated by microsatellite genome scan was used in a recessive model with 80 per cent penetrance.

Haplotype segregation analysis revealed other candidate regions. The locus at 16pter, where genome scan microsatellite markers were non-informative, was genotyped with one other marker (Table 5.31). Linkage to disease gene was excluded since shared homozygosity was not found in sibs. Two other regions at 4p15.33-p15.2 (Table 5.32) and 22q13.1-q13.32 (Table 5.33), with shared heterozygous genotypes in affected sibs, were analyzed with additional markers to investigate whether they harbored a hidden homozygosity, but no such homozygosity was found.

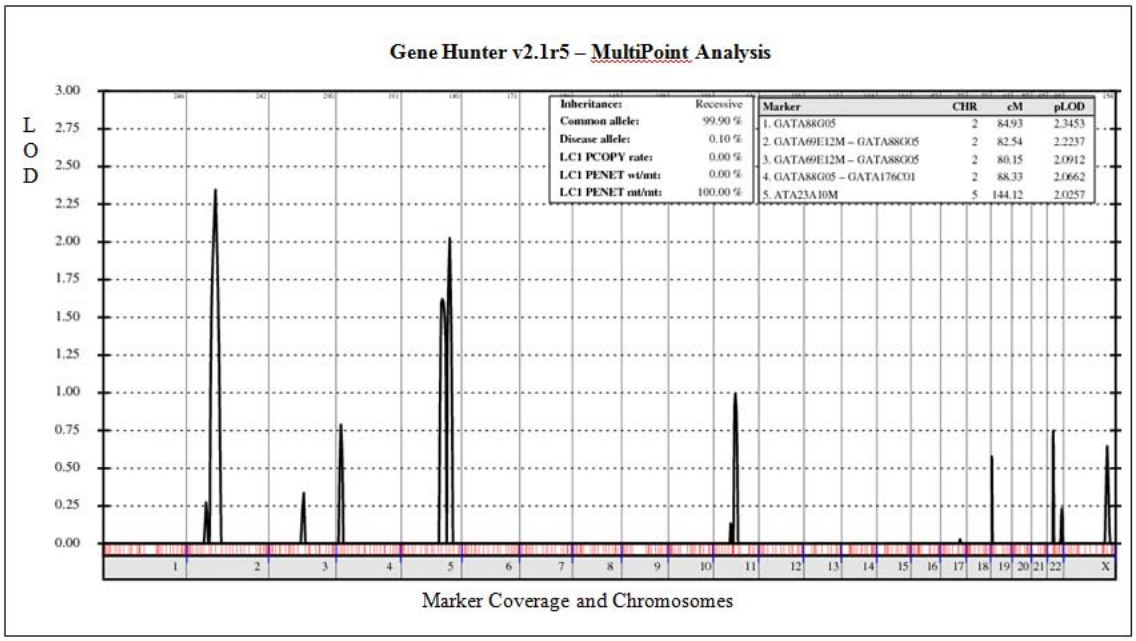


Figure 5.16. Multipoint parametric lod scores for IDDM of affected sib and parents. Microsatellite genome scan data set was used in a recessive model with full penetrance.

Table 5.29. Haplotypes of IDDM family at 2p14-q14.1. Markers in bold were genotyped in this study. Genotypes in italics are deduced. Individual IDs are as in the pedigree.

Marker ID	Position (Mb)	Individual							
		301 F	302 M	401 P1	404 S1	406 P2	408 S2	409 P3	410 P4
GATA66D01	66.90	*1 2*	2 1	2 2	2 2	1 2	2 1	1 2	2 2
GATA69E12M	72.99	*1 1*	1 1	1 1	1 1	1 1	1 1	1 1	1 1
D2S428	82.84	1 2	2 3	2 2	2 2	1 2	2 3	1 2	2 2
GATA88G05	84.93	*1 1*	1 2	1 1	*1 1*	1 1	*1 2*	1 1	1 1
D2S2264	101.79	3 2	2 1	2 2	2 2	3 1	2 2	3 2	2 2
GATA76C01	101.94	*1 3*	3 2	3 3	3 3	1 2	3 3	*1 3*	3 3
GATA4E11	115.96	*2 1*	1 2	1 1	1 1	2 2	1 1	2 1	1 1

Table 5.30. Haplotypes of IDDM family at 5q14.3-q35.1. Markers in bold were genotyped in this study. Genotypes in italics are deduced. Individual IDs are as in the pedigree.

Marker ID	Position (Mb)	Individual															
		301 F		302 M		401 P1		404 S1		406 P2		408 S2		409 P3		410 P4	
GATA89G08Z	89.20	*2*	*3*	3	1	2	3	2	1	3	1	3	1	3	1	2	3
GATA3H06M	96.41	*1*	*3*	3	2	3	3	1	2	3	2	3	2	3	2	1	3
GATA68A03	110.06	*1*	*3*	4	2	3	4	1	2	3	2	*3*	*2*	3	2	3	4
D5S1720	114.25	2	1	3	3	1	3	2	3	1	3	1	3	1	3	1	3
D5S1505	119.13	2	3	1	3	3	3	2	3	3	3	3	3	3	3	3	3
D5S804	125.11	1	2	1	1	2	1	1	1	2	1	2	1	2	1	2	1
D5S642	128.22	2	1	1	2	1	2	2	2	1	2	1	2	1	2	1	2
GATA2H09	135.33	-	*2*	1	1	2	1	2	1	2	1	*2*	*1*	2	1	2	1
D5S1979	141.10	3	1	4	2	1	2	1	4	1	2	1	2	1	2	1	2
D5S2011	141.20	1	2	3	2	2	2	2	3	2	2	2	2	2	2	2	2
ATA23A10M	144.12	-	*2*	1	2	2	2	2	1	2	2	2	2	2	2	2	2
D5S2847	146.96	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1
D5S463	147.11	3	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1
D5S1469	149.44	1	2	3	2	2	2	2	3	2	2	2	2	2	3	2	2
GATA6E05	156.05	-	*2*	3	1	2	1	2	3	2	1	2	1	2	3	2	3
GATA11A11P	168.97	*2*	*3*	1	1	3	1	3	1	3	1	3	1	3	1	2	1

Table 5.31. Haplotypes of IDDM family at 16p13.3. Markers in bold were genotyped in this study. Genotypes in italics are deduced. Individual IDs are as in the pedigree.

Marker ID	Position(Mb)	Individual															
		301 F		302 M		401 P1		404 S1		406 P2		408 S2		409 P3		410 P4	
TTTA028	0.61	*1*	*1*	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D16S3065	3.76	1	3	1	2	1	3	1	2	2	3	1	1	1	2	1	3
ATA41E04	6.21	*1*	*2*	2	2	1	2	2	2	1	2	2	2	*2*	*2*	*1*	*2*

Table 5.32. Haplotypes of IDDM family at 4p15.33-p15.2. Markers in bold were genotyped in this study. Genotypes in italics are deduced. Individuals are as in the pedigree.

Marker ID	Position (Mb)	Individual															
		301 F		302 M		401 P1		404 S1		406 P2		408 S2		409 P3		410 P4	
ATT015	13.54	*1*	*1*	1	2	1	1	1	1	1	1	1	2	*1*	*1*	1	1
D4S2639	18.45	3	2	2	1	3	2	2	1	2	1	2	1	3	2	3	2
GATA70E01	20.72	*2*	*3*	3	1	2	3	3	1	2	3	3	1	2	3	2	3
D4S3013	23.51	3	2	2	1	3	1	2	1	3	1	2	1	3	1	3	2
ATA27C07P	26.87	*2*	*1*	1	1	2	1	1	1	2	1	1	1	*2*	*1*	2	1

Table 5.33. Haplotypes of IDDM family at 22q13.1-q13.32. Markers in bold were genotyped in this study. Genotypes in italics are deduced. Individuals are as in the pedigree.

Marker ID	Position (Mb)	Individual															
		301 F		302 M		401 P1		404 S1		406 P2		408 S2		409 P3		410 P4	
GATA11B12	34.84	*3*	*1*	1	2	1	1	3	2	1	1	1	1	1	2	1	2
D22S283	35.08	2	3	3	1	3	3	2	1	3	3	3	3	3	1	3	1
ATA37D06	35.87	*2*	*1*	1	1	1	1	2	1	1	1	*1*	*1*	1	1	1	1
D22S272	37.42	2	1	1	3	1	1	2	3	1	1	1	1	1	3	1	3
D22S1140	43.05	1	1	1	2	1	1	1	2	1	1	1	1	1	2	1	2
UT7136	44.50	-	*1*	1	1	1	1	-	1	1	1	1	1	1	1	1	1
D22S1160	44.81	2	1	1	2	*1*	*1*	2	2	1	1	1	1	1	2	1	2
TCTA015M	47.93	*3*	*2*	2	1	3	2	3	1	2	2	2	2	2	2	2	1

5.3.2. Linkage Analysis based on SNP genome scan data

Parametric linkage analysis was performed for both autosomal recessive and dominant disease models using genome scan data of 370 000 SNP markers. No region with significant lod scores was identified in parametric linkage analysis assuming autosomal dominant model with reduced penetrance (80 per cent). Lod score calculations under autosomal recessive model assuming full penetrance yielded several regions with lod scores >2 on chromosome 8 (Figure 5.17), 10 (Figure 5.18) and 17q22 (Figure 5.19). Homozygosity mapping strategy showed that those regions were in fact homozygous in affected siblings and not in unaffected siblings or parents. The homozygous region at 8q24.23 was about 1.5 cM and the one at 10q23.2 was one cM. The locus at chromosome 17 was not investigated further due to relatively smaller size of approximately 0.5 cM. A large homozygous shared region was expected as the true disease locus in a consanguineous family where parents are first cousins. The small sizes of the homozygous regions suggested that homozygosity might have arisen due to pure chance. The haplotypes generated by microsatellite genome scan data were utilized to investigate this possibility. The affected children had not inherited the same maternal allele at 10q23.2. The microsatellite genome scan data revealed known crossovers at the region but pointed out to a candidate locus at 8q24.23 between GATA50D10 (136.36 Mb) and D8S1836 (143 Mb) (Table 5.34). According to SNP data, the candidate locus was between markers rs1519372 and rs11166844. It harbors part of two genes; *COL22A1* and *FAM135B*. *FAM135B* is a strong candidate because it is

expressed in pancreas. Only one exon (exon 1) of *FAM135B* is located in the homozygous region.

Table 5.34. Haplotypes of IDDM family at 8q24.23

Marker ID	Position(Mb)	Individual							
		301 F	302 M	401 P1	404 S1	406 P2	408 S2	409 P3	410 P4
GATA7G07	125.98	3 3	1 2	3 1	3 2	3 2	3 2	3 1	3 1
GATA21C12	128.66	2 1	2 1	- -	2 1	2 1	- -	2 1	- -
GATA50D10	136.36	2 1	1 1	2 1	2 1	2 1	1 1	2 1	2 1
D8S1836	143	1 2	3 2	1 3	1 2	1 3	1 2	1 3	1 2
UT721M	144.29	4 2	3 1	4 3	4 1	4 3	4 1	2 3	4 1

In summary 8q24.23 was found to be the only candidate locus. *FAM135B* gene, one of the two genes in the loci, is a strong candidate for IDDM. The region was too small for a family with first cousin parents. Therefore, mitochondrial inheritance needed to be ruled out. In order to check if the mitochondrial genome of IDDM patient harbored any mitochondrial mutation, DNA sample of the youngest affected sibling was sent to DNA sequencing. No mitochondrial mutation was identified, but only polymorphisms that are registered in databases were observed in the mitochondrial genome.

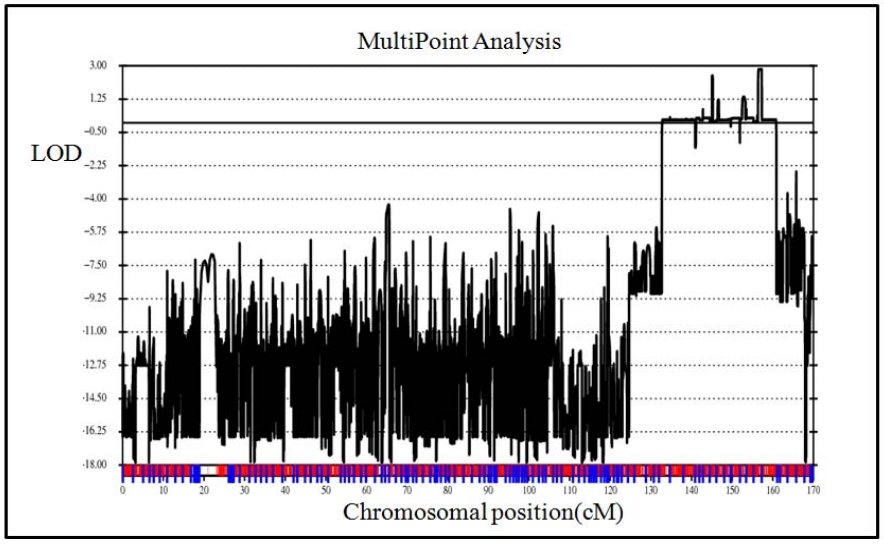


Figure 5.14. Multipoint lod score for IDDM at 8p24.23. The total data set generated by SNP genome scan was used in recessive model with full penetrance.

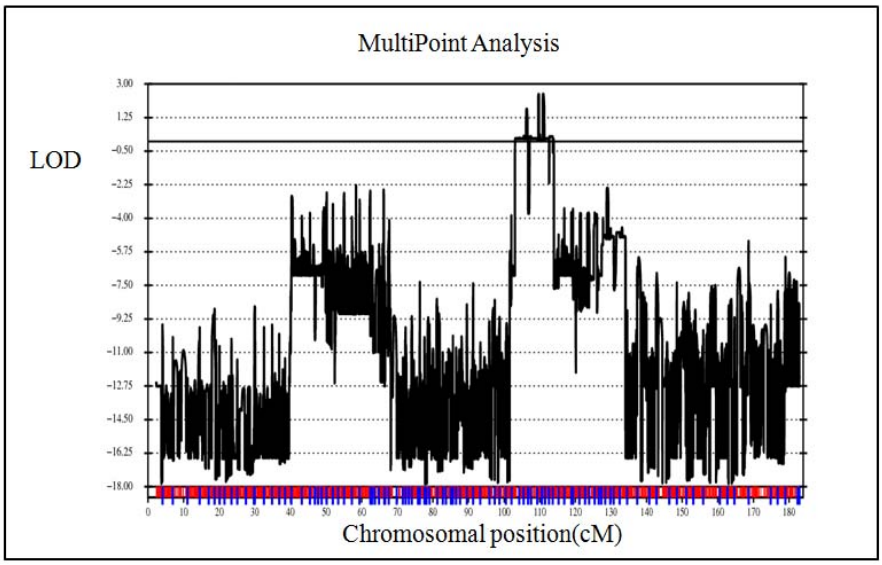


Figure 5.15. Multipoint lod score for IDDM at 10q23.2. The total data set generated by SNP genome scan was used in recessive model with full penetrance.

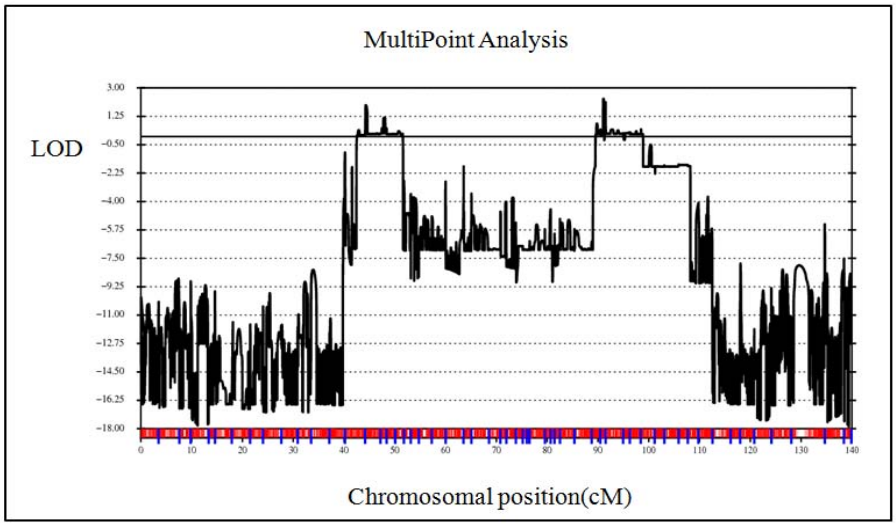


Figure 5.16. Multipoint lod score for IDDM at 17q22. The total data set generated by SNP genome scan was used in recessive model with full penetrance.

6. DISCUSSION

In the framework of this study, genetic analyses on azoospermia, MHAC and IDDM were performed. Our laboratory has been utilizing genome scan data generated by microsatellite markers as the basis of candidate locus and gene search. Recent advances in technology allowed genotyping of large numbers of SNP markers in arrays rapidly in a cost-effective manner. This study employed genome-wide lod score analysis and haplotyping utilizing both SNP and microsatellite genome scan data. SNP genome scan data were used in conjugation with the previously generated microsatellite genome scan data in azoospermia and IDDM. The study on MHAC was based on SNP genome scan data alone.

6.1. Azoospermia

Homozygosity mapping based on data generated by a low density microsatellite genome scan and further fine-mapping previously identified a 10-Mb candidate azoospermia locus. Genome-wide parametric linkage analysis was not performed because only the three affected brothers had been genotyped using markers at about 25 cM-density. In this study, fine mapping with additional markers confined the candidate locus to a 7 Mb-interval at 17pter-p13.1. The two-point and multipoint lod scores for the locus were 2.01 and 2.38, respectively. In order to investigate whether there were other candidate disease loci that had been missed using the data generated by the previous low-density scan, an SNP genome scan was performed. This scan was expected to give additional information for two reasons: In addition to three affected siblings, unaffected brother and mother were included and SNP markers were spaced at about 0.01 cM in contrast to the 25 cM spacing of microsatellite markers used in the previous scan. Homozygosity mapping using the SNP genome data verified the candidate locus at 17p13.3-p13.1 and identified a second locus at 14q11.2-q12. Two-point and multipoint lod scores of 1.24 and 2.38 were obtained for both loci.

The lod scores did not reach 3.0, the threshold value accepted for significant linkage to the disease. The two possible underlying reason for the low value was the absence of

genotype data of the father and the high degree of consanguinity in the family. The three consanguineous marriages in the pedigree increase the probability of homozygous haplotypes by chance; hence a decrease in the lod score is expected. Simulation programs such as Slink and Allegro are commonly used to calculate the maximum multipoint lod score attainable for a pedigree. Due to computational limitations, simulation could not be performed for our complex pedigree.

Candidate gene approach was used to find the disease gene expected to reside at the larger candidate locus 17p13.3-p13.1. According to NCBI Build 36.3, the region was approximately 28.8 cM and 7.13 Mb. It harbored 189 genes. Data on the expression profiles, cellular localizations and functions for each gene were obtained through both literature search and evaluation of the information mainly in databases Entrez Gene, USCS, OMIM, Gene Atlas and Gene Card. The genes that were screened for mutations were *P2RX1*, *SPATA22*, *SPAG7*, *INCA1*, *TEKT1*, *GSG2* and *YBOX1*.

P2RX1 (Purinergic receptor *P2X*, ligand gated channel 1) was assigned as the strongest candidate gene by function. The family members of P2X receptors are ATP-gated nonselective cation channels that are expressed in a wide range of excitable, epithelial/endothelial, and immune effector cell types. Homomeric and heteromeric channels are composed of three subunits, encoded by a combination of seven P2XR genes in vertebrate genomes (reviewed in North *et al.*, 2002). P2RX1 is a nucleotide-gated ion channel having relatively high calcium ion permeability and mediates synaptic transmission both between neurons and from neurons to smooth muscles through binding extracellular ATP (Valera *et al.*, 1994). P2X1 receptor is involved in vas deferens contraction in response to sympathetic nerve stimulation that propels sperm into ejaculate. In *P2RX1* deficient male mice, there was a 90 per cent reduction in fertility due to the decrease of sperm quantity in the ejaculate, while no impairment in spermatogenesis was observed (Mulryan *et al.*, 2000). In studies of animal models, the utilization of selective P2RX1 receptor antagonists as male contraceptive agents was suggested. Due to the essential role of *P2X1* in normal male reproductive function and the mucoid nature of family subjects' semen, mutational screening of the 12 exons and the putative regulatory elements were performed by SSCP in the azoospermia family and 12 unrelated azoospermia subjects. A causative mutation was not identified.

Spermatogenesis associated 22 (*SPATA22*), also known as testis development-related *NYD-SP20*, stood out as the second strongest candidate disease gene. cDNA microarray databases reported four alternatively spliced forms of the gene. While all spliceoforms were expressed in human testis and spermatozoa, one of the spliceoforms was not expressed in fetal and elderly testis tissue (Huang *et al.*, 2005). This developmentally regulated expression pattern suggested that the gene has a role in spermatogenesis. All exons and putative regulatory regions of *SPATA22* were screened for mutations in the family and 14 azoospermia subjects by SSCP analysis. A sequence variant c.464A→G that leads to the substitution of a polar amino acid having neutral charge (glutamine, Q) by another polar amino acid with positive charge (arginine, R) in residue 155 (p.Q155R) near the carboxyl end of the 163 residue protein was identified in the proband. The variant was a known single nucleotide polymorphism denoted as rs11556563 in NCBI SNP database, and its reported allele frequency was higher than 0.5 in the European population.

In one of the unrelated azoospermic subjects, novel c.922A→C was identified in the heterozygous state. This nucleotide transition in exon 9 resulted in the substitution of the isoleucine at residue 308 by leucine (p.I308L). Similar chemical properties of those two amino acids suggested that the variation would not have a deleterious effect on protein function. In accordance, amino acid substitution prediction programs PMut and Polyphen described the variation as neutral. Another program SNP3D predicted the variation to be harmful for the protein based on its location on a highly conserved segment of the protein. Second variant c.999G→A located in again exon 9 was observed in the heterozygous state in two unrelated subjects. This is a known polymorphism denoted rs9901726. At the protein level, this synonymous variation does not result in an amino acid change. Our findings showed that *SPATA22* was not the gene responsible for the condition in azoospermia family.

Sperm associated antigen 7 (*SPAG7*), another gene at the locus, is a homolog of fox sperm acrosomal protein-1 (*FSA*-1) (Mao *et al.*, 1998). Fox sperm protein (FSA-1) was shown to localize to the inner acrosomal compartment of the fox sperm. Immunofluorescent studies reported its expression in round and elongating fox spermatids on the developing acrosome. The expression profile of the gene and localization of its protein suggested a structural function for the protein at the acrosome (Beaton *et al.*, 1994).

A novel variant c.406C→T was identified in exon 5 of the gene in all affected sibs of the family. The residue is within a 12-aa cluster that has been conserved in all mammals. The residue itself is conserved also in zebrafish, *Drosophila* and mosquito. This variation was asigned as a mutation because it was not observed in a total of 350 chromosomes tested, implying a population frequency lower than 0.01 with 90 per cent power and it resulted the substitution of an amino acid with long basic chain (arginine, R) for an amino acid with bulky aromatic ring (tryptophan, W) in the nuclear localization signal domain (p.R136W). Online prediction tools MMB, SIFT and Polyphen predicted this amino acid change as pathological with high reliability. The findings established that the gene was responsible for the condition in the family. However, it is likely not a predominant gene responsible for azoospermia, because 68 azoospermic/oligozoospermic subjects from Eastern Turkey, where the prevalence of recessive conditions is high due to frequent parental consanguinity, were found not to carry mutations in the gene.

All exons of three other candidate genes, *INCA1, TEKT1* and *GSG2* were analyzed for mutations in azoospermia family by direct resequencing. *INCA1* (inhibitor of *CDK*, cyclin A1 interacting protein 1) is an evolutionary conserved gene that is highly expressed in testis. The gene product was identified as the interaction partner of CDK interacting cyclin A1, which is crucial for spermatogenesis (Liu *et al.* 1998; Diederichs *et al.*, 1994). Seven *INCA1* sequence variants were detected in the intronic or untranslated regions in the homozygous state. All variants were reported in NCBI SNP database. *TEKT1* is a member of filament-forming protein-encoding gene family that is predominantly expressed in testis. The gene product coassembles with tubulins to form ciliary and flagellar microtubules (Xu *et al*,. 2004). No *TEKT1* sequence variant was observed in the affected sibs. Haploid germ cell-specific nuclear protein kinase, Haspin (*GSG2*) is required for cell cycle arrest and differentiation of haploid germ cells (Tanaka *et al*., 1999) and metaphase chromosome alignment during mitosis (Dai and Higgins, 2005). No *GSG2* sequence variant was observed in the affected sibs of azoospermia family.

YBX2 (Y box binding protein 2) was considered another strong candidate, since it is a germ-cell specific nucleic acid-binding protein from Y box gene family (Gu *et al.*, 1998; Tekur *et al*., 1999). Polymorphisms in the gene were found to be associated with impaired spermatogenesis, and the gene was proposed to play a role in susceptibility to idiopathic

spermatogenic impairment (Deng *et al*, 2008; Hammoud *et al*., 2009). Its three exons are located between the most centrometic homozygous marker and the closest heterozygous marker at candidate locus 17p13.3-p13.1. Those exons were analyzed by resequencing. Five sequence variants that were reported in databases as polymorphisms were observed in heterozygous state in the affected sibling. Thus, the candidate region was narrowed down to 7.11Mb.

In summary, a total of four genes were analyzed for mutations in the family only and further three also in unrelated azoospermic individuals. The only predictably pathogenic mutations was c.406C→T. *SPAG7* was considered as the disease gene. To my knowledge, it is the first gene identified as responsible for recessive azoospermia.

6.2. MHAC

In this study, a Slovak MHAC family that does not link to MHAC locus 16p13.3-p12.1 was studied with the purpose of mapping the disease locus and subsequently identifying the disease gene. There was no known consanguinity in the family. Nevertheless, the occurrence of this rare condition in two sibs of a healthy couple strongly implied that the mutant alleles were homozygous by descent. Considering the small size of the population in the small geographical area of Slovakia, the probability that the disease had arisen due to parental consanguinity was considered high. Thus, the parents were taken as distant cousins and the disease was assumed to segregate in an autosomal recessive fashion with full penetrance. While the identification of a very small gene locus was expected, several regions that are homozygous in the sibs but not parents were identified upon alignment of haplotypes. This was a strong evidence for high parental consanguinity. Such consanguinity generally does not allow locus identification, because several loci would be IBD. Multipoint lod scores of all homozygous regions were 3. According to simulations conducted by Yang et. al. (2008), observation of multiple homozygous regions reaching the maximum achievable lod score is expected in SNP linkage analysis. However, the IBD region that contains the causative mutations is generally larger than the IBD regions arising due to random chance. As a result, the largest homozygous region at 16p13.3 that was approximately 9 cM was selected primarily for candidate gene approach. It harbored 41 known or predicted genes.

In the region of interest, there were three genes from tubulin gene family: *TUBB2A*, *TUBB2B* and *LOC10132153*. Tubulins are the building blocks of microtubules which form the mitotic spindles and specify the location of contractile ring, hence the plane of symmetry in dividing cell (Jackson, 2009). Since asymmetrical cell division on the ventricular zone is required for neurogenesis and cortex development and its disruption can lead to developmental brain defects, tubulin genes stood out as strong candidates. Upregulation of tubulin gene expression in the embryonic stages when extensive proliferation of neural precursor cells takes place supports this hypothesis. Six variants detected in *TUBB2A* and fix variants detected in *LOC10132153* were known polymorphisms.

Another candidate gene that was analyzed for mutations was Exocyst Complex Component 2 (*EXOC2*). It encodes one of the components of the octomeric exocyst complex, which is implicated in the establishment of polarity by targeting of secretory vesicles to specific regions on the plasma membrane. The interaction of EXOC2 with other molecules that are regulating actin cytoskeletal remodeling and vesicle transport integrates the secretory and cytoskeletal pathways (Sugihara *et al*., 2001). One of the two variants detected in EXOC2 was novel. Its position 230 base pair away from exon intron junction suggested that the sequence doesn't have a play a role in alternative splicing.

The second candidate locus at 6p25.1 was also investigated by candidate gene approach. Neuritin 1 stood out as a good candidate. This gene is expressed in post-mitotic differentiating neurons and the encoded protein promotes neurite outgrowth and arborization (Naeve *et al*., 1997). One of the three *NRN1* sequence variants (c.*374A>T) detected was novel. The location of the variant in respect to polyA signaling sites was investigated. It was found to be away from the essential motifs suggesting that this is not the causative mutation for MHAC.

The findings showed that the genes analyzed were most likely not the disease genes. There remains the low probability that the mutation lies in the gene regions not analyzed, such as the regulatory regions and the introns. In future work, other candidate genes need

to be analyzed for mutations. The study did not turn out to be fruitful because of the unpredicted close parental consanguinity.

6.3. IDDM

A consanguineous IDDM family with four affected sibs and healthy parents was investigated in this study in an attempt to identify a candidate locus under the assumption of a monogenic autosomal recessive inheritance. Initially, parametric linkage analysis was performed using the data generated by a genome scan using 400 microsatellite markers. Fine-mapping with additional microsatellites led to the exclusion of all candidate loci. Therefore, genotyping was repeated with a high density SNP array that covers the genome more densely and uniformly. Parametric lod score calculation and homozygosity mapping lead to the identification of three loci at chromosomes 8q24.23, 10q23.2 and 17q22 with highest lod scores. The sizes of homozygous region were approximately 1.5 cM , one cM. and 0.5 cM, respectively. The locus at chromosome 17 was not investigated further due to relatively smaller size. A large homozygous shared region was expected as the true disease locus in a consanguineous family. The observed small homozygous regions might have arisen due to pure chance. The haplotypes generated by microsatellite genome scan data were used to investigate this possibility. The affected children had not inherited the same maternal allele at 10q23.2, resulting in exclusion of this candidate locus. The microsatellite genome scan data revealed known crossovers at the region but pointed out to a candidate locus at 8q24.23 between GATA50D10 (136.36 Mb) and D8S1836 (143 Mb) (Table 5.34). This locus yielded a significantly high lod score of 2.818. The region includes part of two genes *COL22A* (collagen type XXII, alpha) and *FAM135B* (family with sequence similarity 135 member B). *FAM135B* is a good candidate for IDDM because it is expressed in pancreas. In the future studies, this gene needs to be analyzed in the patients.

REFERENCES

Abuelo, D., 2007, "Microcephaly Syndromes", *Seminars in Pediatric Neurology*, Vol.14, No. 3, pp. 118-127.

Alexander, I. E., G. P. Tauro and A. Bankier, 1995, "Fetal brain disruption sequence in sisters", *European Journal of Pediatrics*, Vol. 154., pp. 654-657.

Babenko, A. P., M. Polak, H. Cavé, K. Busiah, P. Czernichow, R. Scharfmann, J. Bryan, L. Aguilar-Bryan, M. Vaxillaire and P. Froguel, 2006, "Activating Mutations in the ABCC8 Gene in Neonatal Diabetes Mellitus", *The New England Journal of Medicine*, Vol. 355, No. 5, pp. 507-510.

Beaton, S., A. Cleary, J. ten Have, and M. P. Bradley, 1994, "Cloning and Characterization of a fox sperm protein FSA-1", *Reproduction, Fertility and Development*, Vol. 6, No. 6, pp. 761-770.

Behunova, J., E. Zavadilikova, T. M. Bozoglu, A. Gunduz, A. Tolun and C. Yalcinkaya, 2008, "Recessive Microcephaly Not Mapping to 16p13.13-12.2; Differential Diagnosis to Fetal Brain Disruption Sequence and Atelencephalic Microcephaly", *Clin Dysmorphol*, awaiting publication.

Bond, J., E. Roberts, G. H. Mochida *et al.*, 2002, "ASPM is a Major Determinant of Cerebral Cortical Size", *Nature Genetics*, Vol. 32, pp. 316-320.

Bond, J., E. Robert, K. Springell *et al*., 2005, "A Centrosomal Mechanism Involving CDK5RAP2 and CENPJ Controls Brain Size", *Nature Genetics*, Vol. 37, pp. 353-355.

Bozoglu, T., 2008, Genetic Analysis in Three Inherited Disorders, M.S. Thesis, Boğaziçi University.

Chan, P.T.K. and P. N. Schlegel, 2000, "Nonobstructive Azoospermia", *Current Opinion in Urology*, Vol. 10, pp. 617-624.

Csabay, L., I. Szabó, C. Papp, E. Ttóh-Pál and Z. Papp, 1998, "Central Nervous System Anomalies", *Annals of New York Academy of Science*, Vol. 847, pp. 21-45.

Dai, J., and J. M. Higgins, 2005, " Haspin: a Mitotic Histone Kinase Required for Metaphase Chromosome Alignment", *Cell Cycle*, Vol. 4, No. 5 , pp. 665-668.

Deng, Y., W. Zhang, D. Su, Y. Yang, Y. Ma, H. Zhang, and S. Zhang, 2008, "Some Single Nucleotide Polymorphisms of MSY2 Gene Might Contribute To Susceptibility to Spermatogentic Impairment in Idiopathic Infertile Men", *Urology*, Vol. 71, No.5, pp. 878-882.

Desmet, F. O., D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustre and G. Beroud, 2009, "Human Splicing Finder: an Online Bioinformatic Tool to Predict Splicing Signals", *Nucleic Acid Research*, Vol. 37, No. 9, pp. e57.

Diederichs, S., N. Baumer, P. Ji, S.K. Metzelder, G.E. Idos, T. Cauvet, W. Wang, M. Möller, S. Pierschalski, J. Gromoll and C. Müller-Tidow, 2004, "Identification of Interaction Partners and Substrates of the Cyclin A1-CKD2 Complex", *The Journal of Biological Chemistry*, Vol. 279, No. 32, pp. 33727-33741.

Donis-Keller, H., P. Green, C. Helms, S. Cartinhour, B. Weiffenbach, K. Stephens, T. P. Keith, D. W. Bowden, D. R. Smith and E. S. Lander, 1987, "A Genetic Linkage Map of the Human Genome", *Cell*, Vol. 51, pp. 319-37.

Dudbridge, F., 2003, "A Survey of Current Software for Linkage Analysis", *Human Genomics*, Vol. 1, No. 1, pp. 63-65.

Ferlin, A., F. Raicu, V. Gatta, D. Zuccarello, G. Palka and C. Foresta, 2007, "Review – Male Infertility: Role of Genetic Background", *Reproductive Biomedicine Online*, Vol. 14, No. 6, pp. 734-745.

Gloyn, A. L., E. R. Pearson, J. F. Antcliff *et al.*, 2004, "Activating Mutations in the Gene Encoding the ATP-sensitive Potassium-Channel Subunit Kir6.2 and Permanent Neonatal Diabetes", *The New England Journal of Medicine*, Vol. 351, No. 18, pp. 1838-1849.

Gu, W., S. Tekur, R. Reinbold, J. J. Eppig, Y. C. Choi, J. Z. Zheng, M. T. Murray, and N. B. Helcht, 1998, "Mammalian Male and Female Germ Cells Express a Germ Cell-Specific Y-box Protein, MSY2", *Biology of Reproduction*, Vol. 59, No. 5, pp. 1266-1274.

Hammoud, S., B. R. Emery, D. Dunn, R. B. Weiss and D. T. Carrell, 2009, "Sequence Alterations in the YBX2 Gene are Associated with Male Factor Infertility", *Fertility and Sterility*, Vol. 91, No. 4, pp. 1090-1095.

Horikawa, Y., N. Iwasaki, M. Hara, H. Furuta, Y. Hinokio, B. N. Cockburn, T. Lindner, K. Yamagata, M. Ogata, O. Tomonaga, H. Kuroki, T. Kasahara, Y. Iwamoto, and G. I. Bell, 1997, "Mutation in Hepatocyte Nuclear Factor-1 beta gene (TCF2) Associated with MODY", *Nature Genetics*, Vol. 17, pp. 384-385.

Huang, X., J. Li, L. Lu, M. Xu, J. Xiao, L. Yin, H. Zhu, Z. Zhou and J. Sha, 2005, "Novel Development-related Alternative Splices in Human Testis Identified by cDNA Microarrays" *Journal of Andrology*, Vol. 26, No. 2, pp. 189-196.

Jackson, A.P., 2009, "Diversifying Microtubules in Brain Development", *Nature*, Vol. 41, No. 6, pp. 638-640.

Jackson, A. P., D. P. McHale, D. A. Campbell, H. Jafri, Y. Rashid, J. Mannan, G. Karbani, P. Corry, M. I. Levene, R. F. Mueller, A. F. Markham, N. J. Lench and C. G. Woods, 1998, "Primary Autosomal Recessive Microcephaly (MCPH1) Maps to Chromosome 8p22-pter", *American Journal of Human Genetics*, Vol. 63, pp. 541-546.

Jamieson, C. R., C. Govaerts and M. J. Abramowicz, "Primary Autosomal Recessive Microcephaly: Homozygosity Mapping of MCPH4 to Chromosome 15", *American Journal of Human Genetics*, Vol. 65, pp. 1465-1469.

Kameoka, K., H. Isotani, K. Tanaka, K. Azukari, Y. Fujimura, Y. Shiota, E. Sasaki, M. Majima, K. Furukawa, S. Haginomori, H. Kitaoka and N. Ohsawa, 1998, "Novel Mitochondrial DNA Mutation in tRNA-lys (8296A-G) Associated with Diabetes" *Biochemical and Biophysical Research Communications*, Vol. 245, No. 2, pp. 523-527.

Kavaslar, G.N., S.Önengüt, O. Derman, A. Kaya and A. Tolun, 2000, "The Novel Genetic Disorder Microhydranencephaly Maps to Chromosome 16p13.3-12.1", *American Journal of Human Genetics*, Vol. 66, pp. 1705-1709.

Kruglyak, L., M. J. Daly, M. P. Reeve-Daly and E. S. Lander, 1996, "Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach", *American Journal of Human Genetics*, Vol. 58, pp. 1347-1363.

Kumar, A., S. C. Girimaji, M. R. Duvvari and S. H. Blanton, 2009, Mutations in STIL, Encoding a Pericentriolar and Centrosomal Protein, Cause Primary Microcephaly", *American Journal of Human Genetics*, Vol. 84, pp. 286-290.

Lander, E. S. and D. Botstein, 1987, "Homozygosity Mapping: A Way to Map Human Recessive Traits with the DNA of Inbred Children", *Science*, Vol. 236, pp. 1567-1570.

Leal, G. F., E. Roberts, E. O. Silva, S. M. R. Costa, D. J. Hampshire and D. J. Woods, 2003, "A Novel Locus for Autosomal Recessive Primary Microcephaly (MCPH6) Maps to 13q12.2", *Journal of Medical Genetics*, Vol. 40, pp. 540-542.

Lindner, T. H. and K. Hoffmann, 2005, "easyLINKAGE: a PERL Script for Easy and Automated Two-/Multi-point Linkage Analyses", *Bioinformatics*, Vol. 21, No. 3, pp. 405-407.

Liu D., M. M. Matzuk, W. K. Sung, Q Guo, P. Wang and D. J. Wolgemuth, 1998, "Cyclin A1 is Required in the Male Mouse", *Nature Genetics*, Vol. 20, No. 4, pp. 377-380.

Malecki, M. and W. Mlynarski, 2008, "Monogenic Diabetes: Implications for Therapy of Rare Types of Disease", *Diabetes, Obesity and Metabolism*, Vol. 10, pp. 607-616.

Malecki, M. T., U. S. Jhala, A. Antonellis, L. Fields, A. Doria, T. Orban, M. Saad, J. H. Warram, M. Montminy, and A. S. Krolewski, 1999, "Mutations in NEUROD1 are Associated with the Development of Type 2 Diabetes Mellitus" *Nature Genetics*, Vol. 23, No. 3, pp. 323-328.

Mao, M., G. Fu, J. S. Wu, Q. H. Zhang, J. Zhou, L. X. Kan, Q. H. Huang, B. W. Gu, Z. G. Han, Y. Shen, J. Gu, Y. P. Yu, S. H. Xu, Y. X. Wang, S. J. Chen and Z. Chen, 1998, "Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning.", *Proceedings of the National Academy of Science of the United States of America*, Vol. 95, No. 14, pp. 8175-8180.

Matzuk, M.M., and D.J. Lamb, 2008, "The Biology of Infertility: Research Advances and Clinical Challenges", *Nature Medicine*, Vol. 14, No. 11, pp. 1197-1213.

Matzuk, M.M., and D.J. Lamb, 2002, "Genetic Dissection of Mammalian Fertility Pathways", *Nature Cell Biology & Nature Medicine*, Vol. 4, pp. 41-48.

McLachlan, R.,I., C. Mallidis, K. Ma, S. Bhasin and D.M. de Kretser, 1998, "Genetic Disorders and Spermatogenesis", *Reproduction, Fertility and Development*, Vol. 10, 97-104.

Mehers K.L. and K. M. Gillespie, 2008, "The Genetic Basis for Type 1 Diabetes", *British Medical Bullettin*, Vol. 88, pp. 115-129.

Meschede, D., B. Dworniczak, E. Nieschlag, and J. Horst, 1998, "Genetic Diseases of the Seminal Ducts", *Biomedicine and Pharmacotherapy*, Vol. 52, No. 5, pp. 197-203.

Miyamoto, T., S. Hasuike, L. Yogev, M. R. Maduro, M. Ishikawa, H. Westphal and D. J. Lamb, 2003, "Azoospermia in Patients Heterozygous for a Mutation in SYCP3", *Lancet*, Vol. 362, pp. 1714-1719.

Morton, N. E., 1955, "Sequential Tests for the Detection of Linkage", *American Journal of Human Genetics*, Vol. 7, pp. 277-318.

Moynihan, L., A. P. Jackson, E. Roberts, G. Karbani, I. Lewis, P. Corry, G. Turner, R. F. Mueller, N. J. Lench and C. G. Woods, 2000, "A Third Novel Locus for Primary Autosomal Recessive Microcephaly Maps to Chromosome 9q34" *American Journal of Human Genetics*, Vol. 66, pp. 724-727.

Mulryan, K., D. P. Gitterman, C. J. Lewis, C. Vial, B.J. Leckie, A. L. Cobb, J. E. Brown, E.C. Conley, G. Buell, C. A. Pritchard and R. J. Evans, 2000, "Reduced Vas Deferens Contraction and Male Infertility in Mice Lacking P2RX1 Receptors", *Nature*, Vol. 403, No.6765, pp. 86-89.

Naeve, G. S., M. Ramakrishnan, R. Kramer, D. Hevroni, Y. Citri, and L.E. Theill, 1997, "Neuritin: a Gene Induced by Neural Activity and Neutrophins That Promotes Neuritogenesis", *Proceedings of the National Academy of Science of the United States of America*, Vol. 18, No. 94, pp.1648-2653.

Neve, B., M. E. Fernandez-Zapico, V. Ashkenazi-Katalan *et al.*, 2005, "Role of Transcription Factor KLF11 and its Diabetes-Associated Gene Variants in Pancreatic Beta Cell Function" *Proceedings of the National Academy of Science of the United States of America*, Vol.102, No. 13, pp. 4807-4812.

Njolstad, P. R., O. Sovik, A. Cuesta-Munoz, L. Bjorkhaug, O. Massa, F. Barbetti, D. E. Undlien, C. Shiota, M. A. Magnuson, A. Molven, F. M. Matschinsky and G. I. Bell, 2001, "Neonatal Diabetes Mellitus due to Complete Glucokinase Deficiency", *The New England Journal of Medicine*, Vol. 344, No. 21, pp. 1588-1592.

North, R.A., 2002, "Molecular Physiology of P2X Receptors", *Physiological Reviews*, Vol. 82, pp. 1013-1067.

Ounissi-Benkalha, H. and C. Polychronakos, 2008, "The Molecular Genetics of Type 1 Diabetes: New Genes and Emerging Mechanism", *Trends in Molecular Medicine*, Vol. 14, No. 6, pp. 268-275.

Pattison, L., Y. J. Crow, V. J. Deeble, A. P. Jackson, H. Jafri, Y. Rashid, E. Roberts and C. G. Woods, 2000, "A Fifth Locus for Primary Autosomal Recessive Microcephaly Maps to Chromosome 1q31", *American Journal of Human Genetics*, Vol. 67, pp. 1578-1580.

Plengvidhya, N., S. Kooptiwut, N. Songtawee, A. Doi, H. Furuta, M. Nishi, K. Nanjo, W. Tantibhedhyangkul, W. Boonyasrisawat, P. T. Yenchitsomanus, A. Doria, and N. Banchuin, 2007, "PAX4 Mutations in Thais with Maturity Onset Diabetes of the Young", *The Journal of Clinical Endocrinology and Metabolism*, Vol. 92, No. 7, pp. 2821-2826.

Raeder, H., S. Johansson, P. I. Holm, I. S. Haldorsen, E. Mas, V. Sbarra, I. Nermoen, S. A., Eide, L. Grevle, L. Bjørkhaug, J. V. Sagen, L. Aksnes, O. Søvik, D. Lombardo, A. Molven, and P. R. Njølstad, 2006, "Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction", *Nature Genetics*, Vol. 38, No. 1, pp. 12-13.

Reardon, W., R. J. M. Ross, M. G. Sweeney, L. M. Luxon, M. E. Pembrey, A. E. Harding and R. C. Trembath, 1992, "Diabetes Mellitus Associated with a Pathological Point Mutation in Mitochondrial DNA", *Lancet*, Vol. 340, pp. 1376-1379.

Roberts, E., A. P. Jackson, A. C. Carradice, V. J. Deeble, J. Mannan, Y. Rashid, H. Jafri, D. P. McHale, A. F. Markham, N. J. Lench and C. G. Woods, 1999, "The Second Locus for Autosomal Recessive Primary Microcephaly (MCPH2) Maps to Chromosome 19q13.1-13.2", *European Journal of Human Genetics*, Vol. 7, pp. 815-820.

Scham, A., H. Y. Kroes, K. Sollie, B. Timmer, P. Barth and T. Van Essen, 2004, "Hereditary fetal brain degeneration resembling fetal brain disruption sequence in two sibships", *Am. J. Med. Genet*., Vol. 127A, pp. 172-182.

Stoffers, D. A., J. Ferrer, W. L. Clarke, and J. F. Habener, 1997, Early-onset type-II diabetes mellitus (MODY4) linked to IPF1, *Nature Genetics*, Vol. 17, No.2, pp. 138-139.

Sugihara, K., S. Asana, K. Tanaka, A. Iwamatsu, K. Okawa and Y. Ohta, 2002, "The Exocyst Complex Binds the Small GTPase RalA to Mediate Filopodia Formation", *Nature Cell Biology*, Vol. 4, No. 1, pp.73-78.

Sun, C., H. Skaletsky, B. Birren, K. Devon, Z. Tang, S. Silber, R. Oates and D.C. Page, 1999, "An Azoospermic Man with a de novo Point Mutation in the Y-chromosomal gene USP9Y", *Nature Genetics*, Vol. 23, pp. 429-432.

Tanaka, H., Y. Yoshimura, M. Nozaki *et al*., 1999, "Identification and Characterization of a Haploid Germ Cell-specific Nuclear Protein Kinase (Haspin) in Spermatid Nuclei and its Effect on Somatic Cells", *The Journal of Biological Chemistry*, Vol. 274, No. 24, pp.17049-17057.

Teare, M.D. and J.H. Barrett, 2005, "Genetic Epidemiology 2 – Genetic Linkage Studies", *Lancet*, Vol. 366, pp.1036-44.

Tekur, S., A. Pawlak, G. Guellaen, and N. B. Hecht, 1999, "Contrin, the Human Homologue of a Germ-cell Y-box-binding Protein: Cloning, Expression, and Chromosomal Localization", *Journal of Andrology*, Vol. 20, No. 1, pp. 135-144.

Tiepolo, L. and O. Zuffardi, 1976, "Localization of Factors Controlling Spermatogenesis in the Nonfluorescent Portion of the Human Y Chromosome Long Arm", *Human Genetics*, Vol. 34, No. 2, pp. 119-124.

Valera, S., N. Hussy, R. J. Evans, N. Adami, R. A. North, A. Surprenant and G. Buell, 1994, "A New Class of Ligand-gated Ion Channel Defined by P2X Receptor for Extracellular ATP", *Nature*, Vol. 371, pp. 516-518, October.

Vialettes, B. H., V. Paquis-Flucklinger, J.F. Pellissier, D. Bendahan, H. Narbonne, P. Silvestre-Aillaud, M. F. Montfort, M. Righini-Chossegros, J. Pouget, P. J. Cozzone and C. Desnuelle, 1997, "Phenotypic Expression of Diabetes Secondary to a T14709C Mutation of Mitochondrial DNA: Comparision with MIDD Syndrome (A3243G mutation): a case report., *Diabetes Care*, Vol. 20, pp. 1731-1737.

Vionnet, N., M. Stoffel, J. Takeda, K. Yasuda, G. I. Bell, H. Zouali, S. Lesage, G. Velho, F. Iris, Ph. Passa, Ph. Froguel and D. Cohen, 1992, "Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus", *Nature*, Vol. 356, pp.721-722.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix and M. Lathrop, 1992, "A Second Generation Linkage Map of the Human Genome", *Nature*, Vol. 359, pp. 794-801.

Xu M., Z. Zhou, C. Cheng, W. Zhao, R. Tang, Y. Huang, W. Wang, J. Xu, L. Zeng, Y. Xie and Y. Mao, 2001, "Cloning and Characterization of a novel human TEKTIN1 gene" *The International Journal of Biochemistry and Cell Biology*, Vol. 33, No.12, pp.1172-1182.

Yamagata, K., H. Furuta, N. Oda, P. J. Kaisaki, S. Menzel, N. J. Cox, S. S. Fajans, S. Signorini, M. Stoffel and G. I. Bell, 1996, "Mutations in the Hepatocyte Nuclear Factor-4alpha gene in maturity-onset diabetes of the young (MODY1)", *Nature*, Vol. 384, pp.458-460.

Yamagata, K., N. Oda, P. J. Kaisaki *et al.*, 1996, "Mutations in the Hepatocyte Nuclear Factor-1alpha gene in maturity-onset diabetes of the young (MODY3)", *Nature*, Vol. 384, pp.455-458.

Zalloua, P. A., S. T. Azar, M. Delepine, N. J. Makhoul, H. Blanc, M. Sanyoura, A. Lavergne, K. Stankov, A. Lemainque, P. Baz and C. Julier, 2008, "WFS1 Mutations are Frequent Monogenic Causes of Juvenile-Onset Diabetes Mellitus in Lebanon", *Human Molecular Genetics,* Vol. 17, No. 24, pp. 4012-4021.