LINKAGE AND ASSOCIATION STUDIES OF NON-HLA SUSCEPTIBILITY GENES FOR INSULIN-DEPENDENT DIABETES MELLITUS (IDDM)

By

MARILYN YUANXIN MA BUI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

A dedication to Grandma and Katie:

My own secret inspiration

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to the chairman of my supervisory committee, Dr. Jin-Xiong She, for his most valuable scientific guidance and financial support. I thank the members of my committee, Drs. Noel K. Maclaren, Edward K. Wakeland, Margaret R. Wallace, William E. Winter, and Thomas P. Yang, for their generous consultation not only on my dissertation research but also on my growth as a scientist. I thank Ms. Nana Tian and De-Fang Luo for their technical assistance. I thank Ms. Christy Myrick for her critique on my manuscript. I am also very grateful to all my friends at the Department of Pathology and Laboratory Medicine, whose friendship meant a lot to me throughout these five years.

Last but not the least, a very special appreciation goes to my family, especially my husband Vinh Q. Bui Jr. and my parents, for their unconditional love and support.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

LINKAGE AND ASSOCIATION STUDIES OF NON-HLA SUSCEPTIBILITY GENES FOR INSULIN-DEPENDENT DIABETES MELLITUS (IDDM)

By

Marilyn Yuanxin Ma Bui

August, 1995

Chairman: Jin-Xiong She
Major Department: Pathology and Laboratory Medicine

Insulin-dependent diabetes mellitus (IDDM) is an autoimmune disease of the insulin-producing pancreatic β cells. Susceptibility to IDDM is influenced by a number of genetic as well as environmental factors. Previous studies have indicated that *IDDM1* is located in the HLA Class II region on chromosome 6p, and *IDDM2* is in the insulin gene (*INS*) region on 11p15. These two regions together explain less than 50% of the total familial clustering of IDDM, suggesting the existence of other susceptibility factors.

In this study, the insulin gene region was further investigated as a candidate susceptibility factor by association and linkage studies. The susceptibility interval on 11p15 was narrowed to within a 6.5 Kb region,

which contains the *INS* gene and its associated VNTR. Linkage between *INS* and IDDM was detected only in male meioses using the affected sibpair method. Transmission/disequilibrium test further confirmed the gender-related bias with respect to linkage with *INS*. Even though maternal imprinting was a very attractive hypothesis to explain the observed bias, biallelic expression of the *INS* gene in human fetal pancreatic tissue suggested that the *INS* locus was not imprinted.

In order to search for additional susceptibility genes, several chromosomal regions were screened with 50 highly polymorphic microsatellite markers in up to 25 affected sibpair families. Preliminary linkage evidence was obtained for two chromosomal regions (4q and 6q). Analysis of 104 affected sibpairs confirmed our initial observation. These two regions were then mapped with additional microsatallite markers spaced at 1-5 cM. Linkage evidence for the 4q region ($p=0.028$) was weak in the total data set. In contrast, strong linkage evidence ($p=0.001$) was obtained for the 6q region in the vicinity of *D6S264*. Together with the UK 96 data set, linkage with the 6q region was established and the disease locus has now been designated as *IDDM8*.

CHAPTER 1
INSULIN-DEPENDENT DIABETES MELLITUS (IDDM) IS AN AUTOIMMUNE DISEASE OF INSULIN-PRODUCING PANCREATIC BETA CELLS, AND IS INFLUENCED BY MULTIPLE GENETIC AS WELL AS ENVIRONMENTAL FACTORS

Insulin Dependent Diabetes Mellitus

Insulin dependent diabetes mellitus (IDDM, or Type I diabetes), is characterized by a prolonged, selective and irreversible destruction of insulin-producing pancreatic β cells; an absolute requirement for exogenous insulin; and a young age of onset. IDDM is generally considered to be a disorder of the developed world. Indeed, after asthma, IDDM is the second most common chronic childhood illness in industrialized countries.[1] In the United States, the prevalence of IDDM by the age of 20 years is about 0.26 percent, the lifetime prevalence approaches 0.4 percent,[2] and the average annual incidence of IDDM between 1970 to 1988 under age 15 years was 13.8 per 100,000.[3] Overall, it is estimated that with a population of 250 million, one million Americans have IDDM.[4]

Patients with IDDM depend on a lifelong supply of insulin and medical attention. Although insulin replacement increases life expectancy, the disease is

associated with severe macrovascular and microvascular complications that include blindness and kidney failure. For these reasons, both the quality and quantity of life can be dramatically reduced for IDDM patients. A huge economic burden is placed on the patients, their families and society.[5]

IDDM is also a serious medical problem in the developing world. Although the incidence of the disease is lower in third-world countries, life expectancy is substantially less. One of the main reasons for the reduced life expectancy may be the lack of an insulin supply. Essentially, IDDM is a lethal disease in third-world countries.[6]

Although IDDM is an ancient and worldwide disorder, the etiology and pathogenic mechanisms of β cell destruction are not yet completely understood. Significant progress has been made in the past decade that has advanced our knowledge of the etiopathogenesis of IDDM.

Autoimmune Mechanisms

The guidelines[7] generally accepted for establishing the diagnosis of an autoimmune disease are the following: (1) The disease state can be transferred by the patients' antibodies or T-cells. (2) The disease course can be slowed or prevented by immunosuppressive therapy. (3) The

disease is associated with manifestations of humoral or cell-mediated autoimmunity directed against the target organ. (4) The disease can be experimentally induced by sensitization to an autoantigen present in the target organ, which presupposes knowledge of the target autoantigen. According to these guidelines, there is plentiful evidence[8-10] demonstrating that the destruction of β cells in humans is autoimmune in nature: (1) After allogeneic bone marrow transplantation with a diabetic donor, the recipient acquired diabetes.[11] Similarly, diabetes was observed after pancreas transplantation between identical twins.[12] (2) There are examples of immunosuppressant-dependent survival of pancreatic grafts in diabetic recipients[12] and immunosuppressant augmentation of the length of remission in new-onset IDDM.[13,14] (3) There is immune cells infiltration in the pancreas (called insulitis).[15] There are multiple abnormalities of the immune system,[16] such as changes in the ratios of T-cell subsets,[17] and the appearance of autoantibodies to islet cell components.[18] In spite of the fact that the autoantigens of IDDM remain elusive, because other evidence is overwhelming, it is generally accepted that IDDM is a classic organ-specific autoimmune disease. In this disorder, β cells are destroyed by T-cell mediated mechanisms, and circulating autoantibodies are markers of the ongoing disease process.[19] There is also evidence indicating that, well before the T-cell

mediated amplification and perpetuation phase of β cell destruction, a series of events takes place in a non-lymphocyte-dependent initial phase.[20,21] It remains possible that other pathogenic mechanisms, including direct lysis of β cells by cytokines[22] and macrophage-mediated killing,[23] may participate.

Environmental Factors

Although the environmental factors that may trigger the development of β cell immunity are poorly defined, the importance of the environment has been clearly demonstrated by the following facts: (1) Genetically identical twins are only 36% concordant.[24] (2) There is an increase in IDDM incidence in several countries where there are important changes in the environmental factors[25-27] and among ethnic groups immigrated from lower incidence countries.[28] It remains unclear how environmental factors contribute to IDDM susceptibility. It is speculated that the environmental factors are somehow required in the anti-β cell autoimmunity and allow the expression of IDDM predisposing genes.[27,29]

Genetic Susceptibility

The basic concept of genetic susceptibility is that our body's response to environmental factors triggering the autoimmune process leading to diabetes is genetically

controlled. IDDM has long been known to be a hereditary disease because of its familial clustering: (1) Up to 15% of IDDM patients have a first-degree relative with the disease.[30] (2) The disease concordance rate is 36% in identical twins.[24] (3) The risk for siblings (6%) is much greater than the population prevalence (0.4%). The familial clustering ratio, defined by Risch[31] as λ_S, has been calculated to be 15 for IDDM (average lifetime sibling risk of 6% divided by the population prevalence of 0.4%).

The Role of the MHC

The human major histocompatibility complex (MHC) on chromosome 6p encodes HLA class I molecules that are present on the surface of all nucleated cells. The function of class I molecules is to present antigenic peptides to CD8 (cytotoxic or suppresser) T-cells. The MHC also encodes three HLA class II molecules: HLA-*DP*, *DQ*, and *DR*, that are expressed on the surface of antigen-presenting cells. The function of class II molecules is to present antigens to CD4 (helper) T-cells. Both CD4 and CD8 cells have unique T-cell receptors for antigens on their surface, which are specific for particular complexes of peptide antigens and HLA molecules. Given the major role of MHC molecules in antigen presentation to T cells, MHC genes are obvious candidate

predisposition genes for autoimmune diseases such as IDDM. In fact, genes in the HLA class II complex are by far the most important factors in determining genetic susceptibility or resistance to IDDM.[32] The HLA class II susceptibility was first found associated with *DR3* and *DR4*.[33] Recent studies have demonstrated that IDDM susceptibility is most strongly associated with *DQB1*0201* and *DQB1*0302*, while protection from IDDM is strongly associated with *DQB1*0602*.[32,34,35] Although trans-racial studies have shown that the susceptible molecules and the strength of their susceptibility appear to be different in various populations,[37,38] *DQA1*0301* is found to be significantly associated with IDDM in all ethnic groups and has been considered a candidate susceptibility factor.[36]

Attention has been drawn to the nature of the residue at position 57 of the HLA *DQ*β-chain.[32,39,40] The Asp residue is rarely found in diabetic patients as compared to the general population and almost never in homozygous state (double copy). This observation is particularly interesting with respect to MHC-peptide interactions. It was hypothesized that the *DQ* molecules associated with IDDM susceptibility may preferentially bind and present β cell derived peptides to trigger otherwise anergized T-cells, causing β cell destruction.[32]

The *DQα/β* cis and/or trans heterodimeric complementation hypothesis has been proposed to account for the synergistic effects observed in *DR3/4* and *DR3/9* heterozygous genotypes.[35,41,42] Individuals who are homozygous for the *DR3* or *DR4* are at a much higher risk than those who have only one copy of the susceptibility alleles (eg. *DR3/X* and *DR4/X* heterozygotes). This phenomenon suggests that the dose of susceptibility antigens may influence the degree of disease susceptibility.[41] However the above *DQ* hypothesis is not able to explain the complexity of HLA associations with IDDM. Recently, Huang *et al.* suggested a unified hypothesis for HLA associations and disease prevalence.[43] This hypothesis was based upon the fact that HLA-encoded susceptibility to IDDM is determined by the combined effects of both *DR* and *DQ* molecules (i.e. by both genotypic combinations and linkage disequilibria of *DR* and *DQ* genes). So far, this hypothesis can explain the majority (if not all) of the observed associations between HLA and IDDM, and is fully consistent with the known IDDM incidence rates across ethnic populations.

While the HLA genes seem to be the most important susceptibility factors ($\lambda_S \approx 3.1\text{-}4.5$),[44] they obviously cannot account for the total genetic contribution to the disease ($\lambda_S \approx 15$).[31] This observation suggests that other susceptibility factors must exist. In fact, previous

studies have indicated that the *INS* gene region may be an IDDM susceptibility factor.[45-48,72-75]

The Role of the Insulin Gene (*INS*) Region

The insulin gene (*INS*) region on chromosome 11p15 has received considerable attention as a candidate region for IDDM. The contribution of *INS* region to IDDM susceptibility was initially demonstrated as association using a VNTR polymorphism at the 5′ end of the *INS* gene.[45] Others have since confirmed this association.[46,48,72,73] However, the exact locus that may be responsible for disease susceptibility remains unknown. In addition, the linkage of *INS* to IDDM has been a controversial issue.[46-49] Julier *et al.*[46] reported that in a French population the polymorphisms in the *INS* region were linked to IDDM only in HLA-*DR*-positive individuals, especially in paternal meioses. However, using the same analytical methods described by Julier, different results were obtained in a British population.[48,75] Further studies are required to investigate whether there is a gender-related bias of *INS* in respect to linkage between the *INS* and IDDM.

The total number of loci contributing to IDDM susceptibility is unknown. A theoretical calculation indicates that HLA ($\lambda_S \approx 3.1\text{-}4.5$)[44] may account for less than one-third of the familial clustering of IDDM ($\lambda_S \approx$

15);[31] while *INS* ($\lambda_s \approx 1.3\text{-}1.5$)[50] and HLA together ($\lambda_s \approx 4.4\text{-}6.0$) can only explain less than 50% of the total genetic influence. It appears that genetic factors unlinked to the HLA and the *INS* are required to fully account for the total familial clustering of the disease.[51] In fact, the β cell destruction in NOD mice (a model of human IDDM) is controlled by at least ten genes not linked to the MHC H-2 region.[52,53] This provides further support for the speculation of additional susceptibility loci outside the HLA and *INS* regions.

Significance of Genetic Studies of IDDM

Identification of the IDDM susceptibility genes is extremely important, because it might lead better prediction, prevention and treatment. If doctors were able to identify people at risk for IDDM according to their genetic profiles, they could possibly modify the patients' exposure to environmental factors to prevent or delay the onset of the disease. They could closely monitor the patients and treat them at the first sign of disease to postpone the progression to full-blown diabetes so that the quality and quantity of the patients' life could be improved.

Difficulties in Mapping IDDM Susceptibility Genes

A simple genetic disease is genetically controlled by one gene, and is inherited according to Mendelian Laws. In contrast, IDDM is clinically very heterogeneous and is a complex and multifactorial disease which does not follow Mendelian inheritance patterns. Factors that contribute to the difficulties in mapping IDDM genes are: (1) Substantial genetic heterogeneity (identical clinical symptoms are caused by defects at two or more genetic loci). (2) Unknown mode of inheritance and incomplete penetrance of the disease. (3) Lack of large pedigrees with multiple affected members. Finally, mapping of the remaining polygenic susceptibility factors is difficult because each has a small effect and requires the development of more effective mapping strategies.

Strategies for Gene Mapping Studies

One strategy is to first study an analogous form of IDDM in an animal model. Comparative mapping has demonstrated that there are some regions of synteny (two or more homologous genes are located on the same chromosome region in two different species) in mouse and humans. However, because of large differences in the biology of mouse and humans, the effectiveness of gene mapping based on syntenic regions is limited. Recently, Todd and colleagues[54] demonstrated that the magnitude of

the gene effect in an experimental backcross of NOD is likely to correlate only weakly, at best, with the expected magnitude of effect in humans. The reason is that in humans the gene effect will depend more heavily on disease allele frequencies than on the observed penetrance ratios, while such allele frequencies are variable.[54] Hence, the major benefit from animal studies may be a better understanding of the disease process itself, rather than identification of susceptibility regions through comparative mapping.

The second is a candidate gene strategy, in which one selects candidate genes to seek association and linkage between their polymorphisms and the disease. When a candidate gene is implicated in the disease, the coding sequences can be characterized and functional studies can be carried out to shed light on the pathological mechanism. Virtually any gene that affects β cell function or the operation of the immune system is a potential candidate, such as the T-cell receptor, MHC molecules, insulin, and cytokines. Other regions in the human genome that may hold candidate genes are those chromosomal segments homologous to IDDM regions of the mouse genome.[55] Historically, the candidate gene strategy has been extremely successful in the study of the genetics of diabetes. In fact, the involvement in IDDM of both the HLA and *INS* genes were discovered using this strategy. Another successful example was the discovery

of linkage of the glucokinase gene with early-onset non-insulin-dependent diabetes mellitus (MODY) in several European pedigrees.[56,57] In at least one family, a nonsense mutation in the glucokinase gene causes disease.[58]

The third strategy is positional cloning. The location of a disease gene is first identified by association and linkage analyses using anonymous genetic markers. Then, attempts to clone the gene can be followed without any knowledge of the function of the disease gene. Several disease genes, such as, the Huntington's disease gene on chromosome 4,[59] the cystic fibrosis gene on chromosome 7[60] and the neurofibromatosis 1 gene on chromosome 17[61] were successfully mapped using positional cloning. These successes have a major impact on risk prediction, counseling for prevention, and ultimately gene therapy. Positional cloning thus has great potential in identifying genes contributing to IDDM susceptibility.

Mapping IDDM Susceptibility Genes by Association Studies

Association studies identify genetic markers close to the disease genes. They are also important for investigating the interactions between the disease genes and for assessing the relative risks of various genotypic combinations of disease genes in human populations.

There are two kinds of generally-applied association studies. One is case-control analysis, and the other is family-based linkage disequilibrium analysis. The principle of a case-control association study involves the comparison of the frequency of a genetic marker in patients (cases) with the frequency of that marker in normal controls from the same ethnic population. If an association between a marker and a disease exists, the genotypic frequencies will differ between the two study groups.[62] However, the marker should not have a selective effect on the individual, which is an spurious association between the disease and the marker.[63] Candidate genes (by their nature of having some importance in the pathway of disease) may have selective effect. In this case, it is important to differentiate a true association from a spurious association.

The transmission/disequilibrium test (TDT) evaluates the transmission of presumably disease-associated alleles from heterozygous unaffected parents to affected children. The statistical properties of the family-based TDT have been investigated by Spielman *et al.*[64] This analysis has been used in several studies[46,48] and has proven to be more sensitive than the affected sibpair method for detecting linkage.[50] TDT has the advantage of not requiring families with multiple affected members. Thus, simplex families can be included in a study. Since a case-control association study may give a false

positive result due to population stratification, TDT is often used as an alternative association analysis. This analysis can narrow the genetic intervals that contain the susceptibility genes identified by linkage studies.

Mapping IDDM Susceptibility Genes by Linkage Studies

A linkage study maps genes by analyzing the cosegregation of a genetic marker with the disease. The principle of the approach is simple: in an affected family, if the disease locus and another polymorphic locus (often called the marker locus) are closely located on the same chromosome, they are preferentially passed on together rather than independently assorted at meiosis. However, the application of this principle is complicated.

The statistical techniques used in current linkage analysis are mostly based on maximum likelihood estimation and likelihood ratio testing, which requires extended affected families, known mode of inheritance, known penetrance values and disease frequency. Unfortunately, for IDDM most of these parameters are unknown and only few large pedigrees are available. Due to the obvious heterogeneity of IDDM, it would be impossible to attempt a classic linkage study by adding together numerous small families. Thus the affected sibpair method becomes a practical alternative. This

analysis only requires nuclear families of at least two affected children and unaffected parents. It reflects the idea that if two affected siblings share a given allele more often than expected by chance, it supports the hypothesis that the disease is linked to that particular locus. This method has been widely used in family-based epidemiological studies for detecting linkage in non-Mendelian disorders.[65] In fact, it was successful in detecting linkage of the HLA region to IDDM.[66]

The affected sibpair analysis can identify linkage between a marker and a disease (or a disease trait) even if the recombination distance is as large as 10-15 cM. It thus allows us to localize genomic intervals that contain susceptibility genes. Association studies can then further narrow the susceptibility intervals. Once one or more markers are found at a distance of less than 1 cM of the disease gene, they can be used as starting points for positional cloning of the gene, or for identification of candidate genes found in that interval.

Microsatellite Genetic Markers

An essential requirement for mapping IDDM susceptibility genes is the availability of highly polymorphic genetic markers. In general, the most useful markers should be maximally informative and easiest to

genotype. Before 1988, DNA polymorphisms were limited to restriction-fragment-length polymorphism (RFLPs) which are based on nucleotide substitution. RFLPs are not very informative, because they usually have a small number of alleles[67] and their polymorphism information content (PIC) value is low. In addition, RFLPs are routinely genotyped using restriction enzyme digestion, blotting, and hybridization. This process is tedious, expensive, labor intensive, uses a lot of DNA, and is time consuming. The introduction of the polymerase chain reaction (PCR) using thermostable DNA polymerase, provided entirely new means of analyzing polymorphisms and made practical the analysis of highly polymorphic length variations in simple-sequence tandemly repeated DNA. Because simple sequence repeats (SSRs) occur frequently and randomly throughout the human genome and are polymorphic, these elements have shown great utility as genomic markers for genetic mapping. SSRs include minisatellites/variable number tandem repeats (VNTRs) and microsatellites. Microsatellites are oligonucleotide tandem repeats, such as CA repeats and CT repeats. The repeated unit of VNTRs is relatively longer than in microsatellites. The informativeness of microsatellites and VNTRs are very similar. The average PIC value for a CA marker is 0.61, which is about twice the average PIC for RFLPs.[69,70] Microsatellites, however, have more important advantages than VNTRs: (1) They are abundant

and uniformly distributed throughout the human genome.[69] For example, there are an estimated to be 50,000 copies of $(TG)_n$ repeat (n=10-60) sequences interspersed through the human genome.[69] Because of the advances in the Human Genome Project, an international effort to first map and eventually sequence the entire human genomes, microsatellites of very high heterozygosity (70-90%) are easily accessible. (2) They are usually less than 100 bp in length and, therefore are easy to clone, sequence and develop into a PCR assay. In genotyping these by PCR, typically the forward primer is labeled using kinase; the PCR products are detected on a polyacrylamide gel after electrophoresis and radiographed. The potential of automating the entire microsatellite typing process, including data analysis, has made it feasible to analyze the human genome to map IDDM susceptibility genes. (3) microsatellite PCR primers are commercially available. For example, Research Genetics currently offers over 4,000 markers and new markers are constantly being added. These primers are ready to use, come with recommendations for reaction conditions, and are reasonably priced. For the above reasons, PCR-based highly polymorphic microsatellites are obviously the markers of choice for gene mapping.

Specific Aims of This Research

The aim of this research is to map non-HLA genomic intervals containing IDDM susceptibility genes by association and linkage studies. Previous studies have demonstrated that genes in the human major histocompatibity complex appear to have the greatest effect on diabetogenesis. The literature suggests that other promising loci are present on chromosome 11p in the vicinity of the insulin gene. My study was designed to achieve the following aims:

1. To identify the susceptibility locus on chromosome 11p15 using case-control association analysis.

2. To investigate whether there is a gender-related difference with respect to the linkage between the *INS* region and IDDM, and if so, what is the molecular basis.

3. To perform a limited genome-wide search for IDDM genes with highly polymorphic microsatellite markers using affected sibpair analysis.

4. To confirm and replicate potential linkages with a large number of affected sibpair families as well as additional microsatellite markers.

CHAPTER 2
ANALYSIS OF THE INSULIN GENE (*INS*) REGION

Introduction

The *INS* region on chromosome 11p15 is a 19 kb interval spanning the tyrosine hydroxylase gene (*TH*), the insulin gene (*INS*) and the insulin-like growth factor II gene (*IGF-2*). Association between the *INS* region and IDDM was first demonstrated using a VNTR polymorphism at the 5′ of the *INS* gene.[45] The association was then confirmed in many populations using additional polymorphisms in the *INS* region.[46,48,72,73] However, the exact locus responsible for IDDM susceptibility remains unknown.

Linkage of *INS* to IDDM has been demonstrated using the affected sibpair analysis and the TDT test.[46,48,74] Julier *et al*.[46] studied a French population and first reported that the polymorphisms in the *INS* region were linked to IDDM only in HLA-*DR*-positive individuals, suggesting an interaction between HLA and *INS*. This effect was strongest in paternal meioses, suggesting a possible role for maternal imprinting. However, using the same analytical methods described by Julier, transmission distortion (linkage) was observed in both

maternal and paternal meioses in a British population.[48,75]

Therefore, in order to assess the strength of association and potential interactions between the *INS* and the HLA-*DQB1* loci, I studied five polymorphisms in the *INS* gene and surrounding loci in a Caucasian diabetic population ascertained from the South-Eastern United States. My results indicate that the risks conferred by *INS* are not significantly different according to HLA genotypes, suggesting that there is no interaction between the two genetic systems in my study group. Furthermore, my analyses of the polymorphisms around the *INS* gene region suggest that a 6.5 Kb interval on 11p, which contains the *INS* gene and its associated VNTR, is responsible for IDDM susceptibility.

In order to investigate the controversy of the gender-specific effect, I analyzed the *INS Pst I* +1127 polymorphism[46] in 123 multiplex families. Linkage was only detected in male meioses using either the affected sibpair analysis or the TDT test. In order to test the maternal imprinting hypothesis, RT-PCR analysis was used to reveal the expression of the *INS* gene in human fetal pancreatic tissues. The biallelic expression, found by this study, indicated that *INS* is not imprinted in the human pancreas, suggesting that the observed gender-related effect cannot be accounted for by maternal imprinting.

Materials and Methods

Patients and Controls for Association Study

All patients and controls used in the association study were unrelated US Caucasians of Northern European descent. The patients had IDDM clinically confirmed using the criteria of the National Diabetes Data Group.[76] They were phenotyped for autoimmune endocrine diseases and the associated relevant autoantibodies. The healthy control subjects were negative for islet cell autoantibodies (ICA) and had no immediate family history of diabetes.

Samples for Linkage Study.

A total of 123 Caucasian families with two or more affected sibs were used for haplotype sharing analysis. In this data set, 53 families were from the Human Biological Data Interchange (HBDI), 8 were from Dr. Spielman at the University of Pennsylvania and 62 were from the South-Eastern USA (mostly Florida). These multiplex families and 15 additional simplex families from North-Central Florida were used for the transmission/disequilibrium test.

DNA Preparation

Lymphocytes were purified from 10-20 ml of whole blood using Ficoll-Hypaque. DNA was purified using proteinase K digestion, phenol/chloroform extraction, and isopropanol precipitation.

PCR Amplification

All PCR amplifications were performed with a template of 50-100ng of genomic DNA in a 25-50 μl reaction volume containing 50 mM KCl, 10 mM Tris-Cl pH 8.3, 1.5 mM $MgCl_2$ and 60 μM of all four dNTPs, 0.2 ng of each primers and 0.5 u of *Taq* polymorase (Boeheringer). Samples were subjected to 35 cycles of 30 seconds at 94 ° C for denaturing, 30 seconds at optimum temperatures for annealing and 30 seconds at 72 °C for extension, using an automated thermal cycler (9600 Perkin-Elmer-Cetus, California). An additional 2 minutes were added to the denaturing step of the first cycle as well as the extension step of the last cycle.

Genotyping of Polymorphisms in the *INS* Region

The five primers used to analyze polymorphisms in the *INS* region are listed in Table 2-1. These

Table 2-1. List of PCR primers used in association study.

Polymorphisms	Detection Method	Primers	Tm (°C)
-4217 (T,C)	Pst I	TH5/TH6	66
+1127 (C,T)	Pst I	INS3/INS2	64
+1428	Fok I	INS3/INS2	64
+2336 (5bp del)	6% acrylamide	INS55/INS41	66
+3580	Msp I	IGF2-1/IGF2-2	64

Primer sequences (5'-3'):

TH5:	GTG ACG CCA AGG ACA AGC TCA
TH6:	ACC CAG CAG CCC CAG TCC T
INS3:	GGA ACC TGC TCT GCG CGG C
INS2:	AGC CCA GCC TCC TCC CTC CA
INS55:	ACC TTT CCT GAG AGC TCC AC
INS44:	GGT GAG CTC CTG GCC TCG A
IGF2-1:	CCC CAT GTG AGC CAG GCC CA
IGF2-2:	GGG AGA CTT GGG GAG CAG CT

polymorphisms were detected using restriction digestion with appropriate enzymes, followed by agarose gel electrophoresis and staining with ethidium bromide.

RNA Extraction and RT-PCR analysis

RNA was extracted from pancreatic tissue of 4 aborted human fetuses between the ages of 55 and 113 days using a protocol modified from Chomczynski and Sacchi.[67] The tissues were briefly homogenized in solution D (4M guanidinium isothiocyanate, 0.75M Na citrate pH 7, 0.5% sarcosyl). RNA was then purified with phenol/chloroform extraction and precipitated with isopropanol. Total RNA (2 μg) was used for cDNA synthesis using reverse transcriptase and oligo-dT priming. An aliquot of cDNA (2 μl, 1/20 volume) was then used as template for PCR amplification of the insulin cDNA. The forward primer (INS7: 5'- CTACACACCCAAGACCCGC-3') is located at the 3' end of exon 1 and the reverse primer (INS8: 5'-TGCAGGAGGCGGCGGGTGT-3') is located in the 3' untranslated region. PCR was done using conditions described above. The optimum annealing temperature was 60 °C. These two primers amplify a fragment of 227 bp from cDNA and a fragment of 1003 bp from genomic DNA (including 786 bp of intron 1 sequences). Thus, the 227 bp product amplified from cDNA should not contain any contamination from amplified genomic DNA, if any was present in the RNA

preparations. Since the amplified fragment contains the *Pst I* +1127 polymorphic site, digestion of RT-PCR products allowed me to distinguish the two *INS* alleles.

Association Analysis

χ^2 tests were used to reveal the statistical significance of the observed genotypic frequency differences between patient and control groups. A p value of less than or equal to 0.05, indicates significant association between the marker and the disease of interest. Relative risks (RR) were calculated by the method of Woolf.[71]

Affected Sibpair Analysis

The inheritance of different alleles at a given locus by affected children from their heterozygous parents was analyzed using identity by descent (IBD). One ibd was scored when the same alleles were shared by the affected sibs. Zero ibd was counted when different alleles were inherited by the affected siblings. Under the hypothesis of no linkage, the random expectation should be 50% for 1 ibd and 0 ibd respectively. A χ^2 test was performed by comparing the observed sharing of the *INS* alleles in affected sibs with random expectation. When deviation from random expectation is statistically

significant, linkage of the *INS* polymorphism and the disease is indicated.

Transmission/disequilibrium Test (TDT)

TDT evaluates the transmission of the presumably disease-associated *INS* allele from heterozygous parents to their affected offspring. If there is linkage of *INS* with IDDM, statistically more disease-associated *INS* alleles should be transmitted.

Results

There Is Association Between *INS* and IDDM

A total of 343 IDDM patients (220 sporadic cases and 123 probands in multiplex families) and 272 normal controls were genotyped for the *Pst I* +1127 polymorphism 3′ of the *INS* gene. The frequencies of the *INS* +/+ homozygous genotype were found to be significantly increased in both sporadic patients and probands of multiplex families above controls (Table 2-2). These results confirmed association between *INS* and IDDM. The disease-associated allele is the *INS* + allele.

The relative risk (RR) conferred by the *INS* gene was 2.1, suggesting that individuals with the *INS* +/+ are twice as likely to develop the disease as those with the *INS* +/- or -/- genotypes.

Table 2-2. Genotypic frequencies of the *Pst I* +1127 polymorphism and relatives risks conferred by the *INS* +/+ genotype in sporadic patients and probands.

	INS genotypes		RR	χ^2	p
	+/+	+/-,-/-			
Controls	167 (61.4%)	105 (38.6%)			
Sporadics	167 (75.9%)	53 (24.1%)	2.0	11.7	0.0006
Probands	98 (79.7%)	25 (20.3%)	2.5	12.8	0.0004
Combined	265 (77.3%)	78 (22.7%)	2.1	18.3	0.00002

A 6.5 Kb Genomic Interval on 11p Confers IDDM Susceptibility

Five distinct genomic polymorphisms within the *INS* gene and the surrounding region were analyzed (Table 2-3) to define the susceptibility interval on chromosome 11p15. 159 normal controls and 197 unrelated diabetic patients were genotyped using the polymerase chain reaction and restriction enzyme digestion. Two polymorphisms within *INS* (+1127 *Pst I* and +1428 *Fok I*) were in complete linkage disequilibrium and demonstrated significant associations with IDDM (RR = 2.0, $P < 0.005$). However, the -4217 *Pst I* polymorphism in the *TH* gene (5' of the *INS* VNTR) was not significantly associated with IDDM, defining the 5' boundary of the susceptibility interval on chromosome 11p. Similarly, the +2336 5 bp deletion and + 3580 *Msp I* polymorphisms were also not significantly associated with IDDM, thus defining the 3' boundary of the susceptibility interval. The -4217 *Pst I* site and the +2336 5 bp deletion site encompass a genomic region of 6.5 Kb including the *INS* gene and its associated VNTR (Figure 2-1), but excluding the *TH* and the *IGF2*.

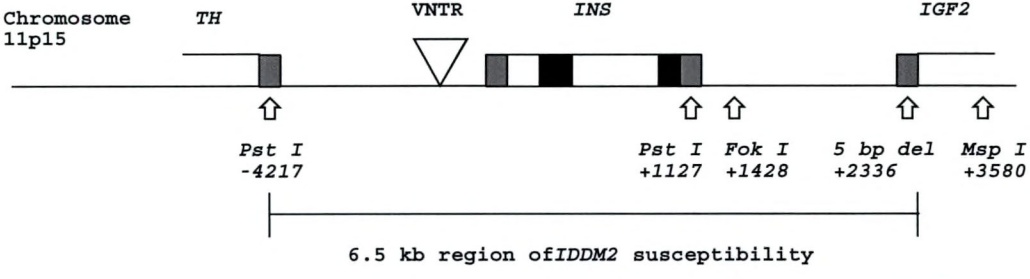
There Is No Interaction Between HLA and *INS*

To investigate the possible interactions between the *INS* and HLA genes, the relative risks conferred by *INS* were calculated according to their *DQB1* genotypes.

Table 2-3. Fine mapping of the IDDM susceptibility interval on 11p.

Polymorphisms	Controls				Diabetics				RR	95% CI	p
	++		+-,--		++		+-,--				
-4217 Pst I	43	(23%)	145	(77%)	32	(20%)	126	(77%)	0.9		ns
+1127 Pst I	97	(61%)	62	(39%)	148	(75%)	49	(25%)	2.0	1.2-3.0	0.005
+1428 Fok I	54	(60%)	36	(40%)	68	(76%)	22	(24%)	2.1	1.1-3.8	0.05
+2336 5bp del	56	(62%)	34	(38%)	58	(62%)	36	(38%)	1.0		ns
+3580 Msp I	67	(44%)	85	(56%)	66	(39%)	105	(61%)	0.8		ns

Figure 2-1. Diagrammatic presentation of the five polymorphisms at the *TH-INS-IGF2* region on 11p15 which define the 6.5 kb interval of *IDDM2* susceptibility. The open, closed and hatched boxes represent introns, exons and untranslated regions, respectively.


Chromosome
11p15
TH
VNTR
INS
IGF2
Pst I
-4217
Pst I
+1127
Fok I
+1428
5 bp del
+2336
Msp I
+3580
6.5 kb region of IDDM2 susceptibility

When 197 diabetic patients were subdivided into four *DQβ1* genotype categories (*0201/0302, *0302/0302 or *0302/X, *0201/0201 or *0201/X, and X/X), the relative risks of the *INS* +/+ homozygotes ranged from 1.6 to 2.4 (Table 2-4). These results are very similar for the entire patient population (RR = 2.1). Since IDDM susceptibility is most strongly associated with **0201* and **0302* (the relative risks conferred by **0201/0302* and **0303/0302* are 20.9 and 12.9, respectively),[73] these results suggest that there is no interaction between the HLA and the *INS* loci.

Affected Sibpair Analysis Reveals Weak Linkage Between *INS* and IDDM in Male Meioses

The *Pst I* + 1127 polymorphism was analyzed in 123 families containing at least two affected siblings (ASPs). There were 42 informative parents (22 fathers and 20 mothers) who were heterozygous for *INS* and whose transmission of *INS* alleles to their affected children can be unambiguously determined. In this data set, 27 affected sibpairs inherited identical *INS* alleles (scored as 1 ibd) and 19 inherited different alleles. Under the hypothesis of no linkage, 1 ibd and 0 ibd should be equal (i.e. 23). In fact, there was no significant difference in observed and expected ibd values in total meioses $\chi^2 = (27-19)^2/(27+19) = 1.4$. However, there were significantly

Table 2-4. Relative risks in diabetic patients conferred by *INS* according to their HLA-*DQB1* status.

	INS Status		RR♣	p
HLA-*DQB1* Status	+/+	+/-,-/-		
0201/0302	49	20	1.6	ns
0302/0302 or 0302/X	41	13	2.0	0.05
0201/0201 or 0201/X	41	11	2.4	0.05
X/X	17	5	2.2	ns
All	148	49	2.0	0.005

♣ The relative risks were computed using 97 (61%) controls with the *INS* +/+ and 62 (39%) controls with the *INS* +/- or -/-.

more (p=0.01) affected sibpairs that inherited identical alleles than different alleles from their heterozygous fathers (19 one ibd versus 6 zero ibd) (Table 2-5). Thus, a weak linkage in male meioses was confirmed using conventional haplotype sharing analysis among affected sibpairs.

TDT Reveals Sex Difference of *INS* Transmission

All 123 multiplex and 15 simplex families were combined for TDT. There were 56 informative heterozygous parents for *INS* (31 fathers and 25 mothers). These parents transmitted 103 alleles (69 allele + and 34 allele -) to their diabetic offspring (Table 2-6). Under the hypothesis of no linkage, the expected number of + and - alleles transmitted is equal (i.e. 51.5). The χ^2 was calculated using the formula $(x-y)^2/(x+y)$, where x is the number of the + alleles and y is the number of - alleles that are transmitted. The difference observed was significant, $\chi^2=(69-34)^2/(69+34)=11.9$, p=0.0006 supporting linkage. In the case of *INS*, this study again demonstrated that TDT is more sensitive than affected sibpair analysis in detecting linkage.

To test whether there is sex difference in *INS* transmission, paternal and maternal transmissions were counted separately. Among 31 fathers heterozygous

Table 2-5. Affected sibpairs analysis at the *INS* locus.

	Fathers	Mothers	Combined
	IBD (1 : 0)	IBD (1 : 0)	IBD (1 : 0)
Observed	19 : 6	8 : 13	27 : 19
Expected	12.5 : 12.5	10.5 : 10.5	23 : 23
χ^2	6.7	1.2	1.4
p	0.01	ns	ns

Table 2-6. Transmission-disequilibrium test of *INS* + and - alleles transmitted from heterozygous (+/-) fathers or mothers to affected children.

	Fathers		Mothers		Combined	
	+	-	+	-	+	-
Observed	44	12	25	22	69	34
Expected	28	28	23.5	23.5	51.5	51.5
χ^2	18.3		0.2		11.9	
p	0.00002		ns		0.0006	

for *INS*, 37 + alleles and 12 - alleles were transmitted to their diabetic children. This is significantly different from random expectation: $\chi^2=(44-12)^2/(44+12)=18.3$, $p < 0.00002$. Among 25 mothers heterozygous for *INS*, 25 + alleles and 22 - alleles were transmitted. This difference is not significant from random expectation. These results suggest that there is a transmission distortion of *INS* from fathers to diabetic children.

There Is No Segregation Distortion of *INS* Transmitted to Unaffected Children

The difference found with the TDT could be due to an "artifact" of meiotic segregation distortion. If it was an artifact, one would expect to see such distortion in both affected and unaffected offspring. The *INS* transmissions from heterozygous parents to unaffected sibs within diabetic families, as well as to normal children in non-diabetic families were analyzed. As shown in Table 2-7, among the 29 informative individuals who inherited *INS* alleles from heterozygous fathers, 13 were unaffected children in diabetic families (6 + alleles and 7 - alleles) and 16 were children in normal families (11 + alleles and 5 - alleles). Among the 32 informative individuals who inherited *INS* alleles from heterozygous mothers, 11 were unaffected children in diabetic families (3 + alleles and 8 - alleles) and 21

Table 2-7. Observed and expected number of *INS* + and - alleles transmitted from heterozygous fathers or mothers (+/-) to normal children.

	Fathers		Mothers		Combined	
Alleles	+	-	+	-	+	-
Observed	17	12	14	18	31	30
Expected	14.5	14.5	16	16	30.5	30.5
p	ns		ns		ns	

were children in normal families (11 + alleles and 10 - alleles). The observed numbers of *INS* alleles transmitted to non-diabetic children were not significantly different from random expectation in male or female meioses. These results do not support the speculation of segregation distortion.

INS Is Biallelically Expressed in Human Pancreatic Tissue

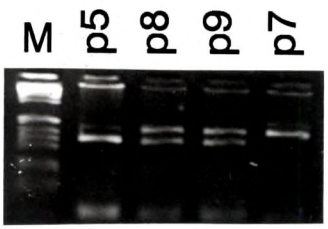
Pancreatic tissue was obtained from four aborted human fetuses. Their genomic DNAs were used as templates to amplify the *INS Pst I* +1127 polymorphism site. *Pst I* digestion of these PCR products revealed that two samples (p8 and p9) were heterozygous for the *INS* + allele, while the other two samples (p5 and p7) were homozygous for - or + alleles respectively. RT-PCR analysis from p8 and p9 mRNA revealed that both *INS* alleles were expressed, at apparently equal level. This biallelic expression of *INS* (Fig. 2-2) suggests that *INS* is not imprinted in human pancreatic tissues.

Discussion

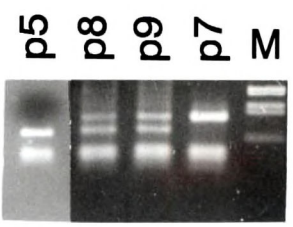
Both association and linkage studies have shown that the genomic region on chromosome 11p spanning the insulin gene contains a susceptibility locus for IDDM.[45,46,48,72,75,77,78] There have been attempts to

Figure 2-2. Genomic polymorphism and expression of *INS* in human pancreas. Genomic PCR: A fragment of 338 bp which contains the *Pst I* +1127 polymorphism was amplified from genomic DNA using primers INS3 and INS6. The products were digested with *Pst I* restriction enzyme and then eletrophoresed in a 3% agarose gel. The + alleles only contain a monomorphic *Pst I* site and were digested into two fragments (163 bp and 75 bp). The alleles which contain a monomorphic site and the polymorphic *Pst I* + 1127 site were digested into three fragments (112, 51 and 75 bp). The samples P8 and P9 were heterozygous for *INS*, as shown in the left panel. RT-PCR: A fragment of 227 bp which contains the *Pst I* +1127 polymorphic site was amplified from cDNA (derived from total RNA of human pancreas) using the primers INS7 and INS8. RT-PCR products were digested with *Pst I*. Digested products of the - alleles produced two fragments (197 bp and 30 bp respectively). Products of + alleles were not digested (227 bp). The samples P8 and P9 were biallelically expressed as shown in the right panel.

genomic PCR



RT-PCR



the IDDM susceptibility factor on 11p. In this study, significant associations with IDDM were found for two polymorphisms within the *INS* gene, while no significant associations were found for the polymorphisms flanking *INS*. A 6.5 Kb genomic region was defined by the *Pst I* -4217 polymorphism in the *TH* gene and the +2336 deletion polymorphism in the *IGF2* gene. Similar observations were obtained by Lucassen *et al.*[78] After analyzing ten polymorphisms in a 4.1 kb region extending from the *INS* 5' VNTR and across the insulin gene, they found significant associations with IDDM. However, it is not possible to specifically identify the IDDM susceptibility site(s) since all of these polymorphisms are in strong linkage disequilibrium. In addition, they were not able to detect associations with IDDM at the *INS* flanking regions, as in this study.

Both Lucassen's and my studies indicate that the susceptibility interval on 11p contains the *INS* gene and its associated VNTR. However, the mechanism by which the *INS* gene and/or its associated VNTR contribute to IDDM susceptibility is unknown.

The possible interaction between HLA and *INS* has been a controversial issue. Analyses of the French population by Julier and Lucassen have suggested that the association of *INS* with IDDM may be stronger in HLA**DR4* positive individuals, indicating interactive effects between the *INS* and the HLA susceptibility loci.

However, my analyses showed that the risk conferred by *INS* was similar in all HLA genotypes. Similar results have also been reported in three other studies.[48,72,75] These observations suggest that there are no interactions between HLA and *INS*.

Risk assessment is an important aspect of genetic studies of IDDM. At the *INS* locus, the absolute risk for general population is 0.0084, which is calculated by the relative risk (2.1) multiplied by the disease prevence (0.004). It seems that the *INS* gene has very minor effect in IDDM susceptibility. In addition, the predictibility of such assessment is limited in IDDM, because the concordance of the disease in identical twin pairs is as low as 36%.[24] Therefore, it may be more feasible to exclude the people who are not at risk rather than to identify the people at risk to IDDM.

Two of the most important issues with respect to linkage of *INS* and IDDM are: (1) is there a gender-related bias, (2) if there is, what is the molecular mechanism responsible for the sex difference. It appears that a sex difference exists in most ethnically heterogeneous populations, such as the French population and the US populations. However, it does not exist in ethnically more homogeneous populations, such as the British population.[48] There are several possible explanations for the sex difference in transmission. Random transmission of *INS* in non-diabetic families is

not consistent with the hypothesis of segregation distortion and thus provides further evidence for linkage. Since the maternal gene did not seem to be important in IDDM susceptibility, the maternal gene may not be expressed, in another word, may be imprinted. Maternal imprinting could account for the observation, and was an very attractive hypothesis because of previously documented maternally imprinted genes in this region.[68,79-83] The *IGF2* gene located 3' of *INS* is known to be imprinted in the mouse[68] and human.[81-83] *INS* is also known to be imprinted in the mouse yolk sac although not in the pancreas.[84] However, our RT-PCR analysis revealed biallelic expression of *INS* in the pancreas of human fetus. Similar results were also obtained from adult pancreas.[85] These results indicate that *INS* is not imprinted in the pancreatic islets. Therefore, other potential mechanisms must be responsible for the observed sex difference.

It remains possible that the *INS* gene may be maternally imprinted in human yolk sac. Another possible mechanism could be mother-fetal interactions. This hypothesis implies that maternal insulin would have an impact on IDDM susceptibility, probably through its effects on β cell mass of the fetus during the early developmental stage. The third possibility is that the neighboring locus *IGF2* could be a candidate gene for IDDM. Supporting evidence for this hypothesis is that

IGF2 is maternally imprinted.[82,83] In addition, *IGF2* encodes insulin-like growth factor 2 which is important in embryogenesis and in β cell development.[83] However, two polymorphisms in the *IGF2* gene (+2336 5 bp del and +3580 *Msp I*) were not associated with IDDM in our population and in a French population.[46] These results did not support the *IGF2* hypothesis. Nevertheless, there may exist other polymorphisms in the *IGF2* gene that are in linkage disequilibrium with the disease-associated *INS* polymorphisms. Alternatively, the polymorphisms in *INS* may affect the expression of the *IGF2* gene, since these two regions are only separated by a few kilo-base pairs. Thus, further studies are required to understand which gene in the *INS-IGF2* region on 11p is involved in IDDM susceptibility, and by what mechanism this gene acts.

Chapter 3
MAPPING OF TWO NOVEL IDDM SUSCEPTIBILITY INTERVALS (4q AND 6q) BY AFFECTED SIBPAIR ANALYSIS

Introduction

As mentioned above, the HLA class II genes and *INS* gene together can only explain a portion of the total genetic influence, suggesting that other IDDM susceptibility factors exist. Indeed, linkage studies have suggested that at least 10 genes are involved in the expression of insulitis and/or diabetes in the nonobese diabetic (NOD) mouse.[52,86] Given the ethnic and genetic heterogeneities of IDDM in humans, the number of susceptibility genes is probably even higher. The candidate gene approach has been successful in limited cases such as *INS*. In the case of the majority of the susceptibility genes, which are likely scattered throughout the genome, linkage studies seem to be more feasible. In fact, several groups have recently reported localization of at least four other non-HLA IDDM susceptibility regions[44,87] using genome-wide linkage mapping. In my mapping studies, a two-stage approach has been applied. The first stage involved an initial genome-wide screen using a subset of 25 Florida affected sibpair families and 50 microsatellite markers located

throughout several chromosomal regions to obtain preliminary linkage evidence. The second stage was to replicate the linkages with 104 affected sibpair families and additional microsatellite markers in those regions. My study demonstrated that there is some evidence for linkage in a novel region on chromosome 4q in the vicinity of marker *D4S1566* (p=0.028). Most importantly, strong linkage evidence for the 6q25-q27 region was obtained. Together with results from a UK data set,[44] linkage to this second region was confirmed. This disease locus has now been designated as *IDDM8*.

Materials and Methods

Affected Sibpair Families

Genomic DNA from a total of 104 American Caucasian families was obtained. Each family had two affected siblings and normal parents. In this set, forty-seven of the samples were collected and ascertained in our hands from the South-Eastern United States, mostly from North-Central Florida (Florida data set). Forty-nine other families were obtained from the Human Biological Data Interchange (HBDI data set). Eight more were provided generously by Dr. Richard Spielman at the University of Pennsylvania.

Microsatellite Markers

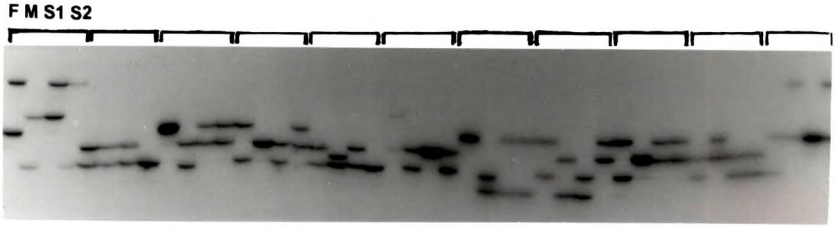
Microsatellite markers were purchased from Research Genetics. Distances between markers are measured in centimorgans (cM). For markers that did not meet our technical specifications, new markers were redesigned and synthesized based on published sequence.

Genotyping

Highly polymorphic microsatellite markers were genotyped using radioactive labeling of PCR primers and denaturing polyacrylamide gel electrophoresis (Figure 3-1). One of the PCR primers was end-labeled using γ^{32}P-ATP and T4 polynucleotide kinase. PCR amplifications were performed on 40 ng of genomic DNA (prealiquoted into a 96-well microtitre plate) in a 12 µl reaction volume containing 50 mM KCL, 10 mM Tris-CL pH 8.3, 1.5 mM $MgCl_2$, and 60 µM of all four dNTPs, 0.2 ng of each primers and 0.5 u of *Taq* polymorase (Boehringer). Samples were subjected to 27-30 cycles of 30 seconds at 94°C for denaturing, 30 seconds at the optimum annealing temperature, and 30 seconds at 72°C for extension using a Perkin-Elmer-Cetus 9600 thermal cycler. After PCR amplification, two volumes of sequencing loading solution (0.3% xylene cyanol, 0.3% bromophenol blue, 10 mM EDTA pH 8.0 and 90% volume of formamide) were added. The samples

Figure 3-1. An example of genotyping *D4S243* using radioactive labeling of PCR primers and denaturing polyacrylamide gel electrophoresis. Eleven affected sibpair families were analyzed. F: Father. M: Mother. S1, S2: Affected siblings.

D4S243



were then heated at 95°C for 10 min to denature the DNA, and 2-4 µl were immediately loaded onto a 6.5 % polyacrylamide DNA sequencing gel. PCR products from 3-4 different markers with non-overlapping allele sizes (amplified in separate reactions) were combined together before loading to genotype multiple markers simultaneously. Alternatively, in some cases products of the same marker (but different samples) were loaded four times (each separated by 30-60 min). Multiplexing of different markers or multiple loading of products from the same marker greatly increased the efficiency of genotyping.

Data Analysis

A χ^2 test was used to determine the statistical significance of the excess of gene sharing by affected sibpairs. The χ^2 was calculated using $(1 \text{ ibd} - 0 \text{ ibd})^2/(1 \text{ ibd} + 0 \text{ ibd})$, with one degree of freedom. A p value less than or equal to 0.05 suggests linkage. In order to detect potential linkages, correction for multiple comparisons was not performed.

The maximum lod score (MLS) statistic T was calculated according to Risch[88] using the following equation: $T = (N_1)[\log_{10}(N_1/0.5N)] + (N_0)[\log_{10}(N_0/0.5N)]$. Where N is the total number of informative meioses (N_1+N_0), N_1 and N_0 are the observed number of affected

sibpairs sharing 1 or 0 alleles respectively. The random expectation for 1 ibd and 0 ibd is 50% respectively. A MLS value of 1.0 indicates linkage.

To increase the informativeness of these families, informative flanking markers were used to deduce the transmission of alleles from homozygous parents (referred to as haplotyping). Haplotyping analyses were performed using markers spaced at less than 5 cM to minimize the possibility of double recombinations. Percent of gene sharing (PGS) was calculated by the formula 1 ibd/(1 ibd + 0 ibd).

Results

Screen for Linkage on Several Chromosomal Regions

Initially, up to twenty-five of the Florida families were analyzed for 50 microsatellite markers randomly chosen throughout several chromosomal regions. Among these regions, some were syntenic to IDDM genes in NOD mouse, some encampass candidate disease genes in humans. As expected, the ibd values drawn from the 25 sibpairs were not sufficient to claim linkage. For example, *IL2RB* on 22q had a p value of 0.01 in the first 25 affected sibpairs, but linkage disappeared when all 104 affected sibpairs were analyzed. However some positive preliminary data were obtained on two markers, *D4S1566* on

4q and *D6S264* on 6q. In addition to the linkage evidence, these markers are in candidate gene regions. The 4q region is syntenic to a mouse chromosome 3 region which contains a IDDM gene (Iddm3) in the NOD mouse. The 6q region is in the neighborhood of the *SOD2* and *IGF2R* genes in human. It was obvious that these two regions were worthy of further investigation.

The rest of 104 affected sibpairs were then genotyped at *D4S1566*. Weak evidence of linkage was obtained in the Florida data set (p=0.026) and the total data set (p=0.028) (Table 3-1). The affected sibpairs in the HBDI families had increased gene sharing compared to random expectation but the excess of gene sharing was not statistically significant. For *D6S264*, linkage evidence was obtained in the Florida data set (p=0.03) and HBDI families (p=0.0073). The combined data set gave a p value of 0.0013 (Table 3-1). At this point, I proceeded to more closely map the 4q and 6q regions to localize the potential IDDM susceptibility genes.

Fine Mapping of Chromosome 4q Region

Seven additional microsatellite markers were analyzed. They are *D4S393*, *D4S1603*, *D4S349*, *D4S1596*, *D4S243*, *D4S1545* and *D4S622* (Table 3-2). Linkage evidence was strongest at *D4S1566* (p=0.028). Since this region

Table 3-1. Linkage Evidence from Genome-wide Screen.

Markers	Data sets	IBD (1:0)	PGS	p	MLS
D4S1566	FL 47	46 : 27	63.0%	0.026	1.1
	HBDI 49	48 : 40	54.5%	ns	
	UF 104	102 : 73	58.3%	0.028	1.1
D6S264	FL 47	35 : 19	64.8%	0.030	1.1
	HBDI 49	52 : 28	65.0%	0.0073	1.6
	UF 104	89 : 51	63.6%	0.0013	2.3

has not been previously reported and is in the vicinity of a candidate region, further studies in other independent data sets will be necessary to confirm this linkage.

Fine Mapping of *IDDM8* on Chromosome 6q

As shown in Figure 3-2, twenty-one markers were analyzed to localize the susceptibility gene on 6q. to be within 1-2 cM of the given locus are flagged with "≈". The first six markers are in the interval of *IDDM5* a These markers encompass a region of 43 cM with an average distance of 3-5 cM. The markers that are estimated round *ESR*, which was first identified by Davies and colleagues.[44] In my study, the linkage at *ESR* was surprisingly weak (MLS=0.9, which was only slightly higher than its flanking markers). The strongest linkage evidence was detected at *D6S446*, which gave a MLS value of 2.8 (1 ibd = 116 and 0 ibd = 68). Since this marker was more than 30 cM telemetric to *ESR*, it was speculated that there may exist another IDDM predisposition gene in the 6q region.

In order to verify this speculation, combining the result from the 96 UK data set[44] with ours, the total MLS values were recalculated (Table 3-3). For *ESR*, the combined results were (95 1 ibd, 59 0 ibd and MLS=1.8),

Table 3-2. Fine mapping of the region around *D4S1566*.

Marker	D (cM)	1 ibd	0 ibd	χ^2	p
D4S393	0	34	38		ns
D4S1603	1	79	59	2.9	0.09
D4S349	2	93	66	4.6	0.03
D4S1566	5	102	73	4.8	0.028
D4S1596	6	83	70	1.1	ns
D4S243	7	101	84	1.6	ns
D4S1545	12	74	71		ns
D4S622	13	63	48		ns

Figure 3-2. Schematic presentation of the locations of *IDDM5* and *IDDM8*. The plot was based on the data in Table 3-4.

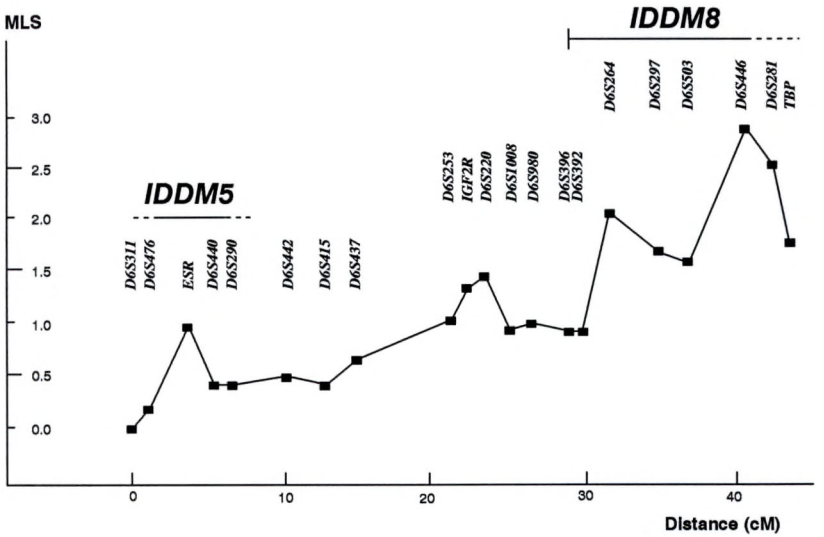

MLS
IDDM8
IDDM5
D6S264
D6S297
D6S503
D6S446
D6S281
TBP
D6S253
IGF2R
D6S220
D6S1008
D6S980
D6S396
D6S392
D6S311
D6S476
ESR
D6S440
D6S290
D6S442
D6S415
D6S437
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
Distance (cM)

Table 3-3. Fine mapping of *IDDM8* on chromosome 6q.

Markers	D (cM)	1 ibd	0 ibd	MLS	MLS(+UK)
D6S311	0	88	85	0.0	
D6S476	2	101	88	0.2	
ESR	4	110	82	0.9	2.5
D6S440	6	109	90	0.4	
D6S290	7	108	90	0.4	
D6S442	10	110	89	0.5	
D6S415	13	107	89	0.4	
D6S437	15	107	82	0.7	
D6S253	22	112	81	1.1	1.8
IGF2R	≈ 22	111	78	1.3	
D6S220	≈ 23	111	76	1.4	
D6S1008	≈ 25	108	81	0.8	
D6S980	≈ 27	109	81	0.9	
D6S396	≈ 29	109	83	0.8	
D6S392	≈ 30	112	85	0.8	
D6S264	32	117	75	2.0	3.4
D6S297	35	113	76	1.6	
D6S503	≈ 37	112	76	1.5	
D6S446	41	116	68	2.8	
D6S281	42	107	63	2.5	2.0
TBP	≈ 43	79	48	1.7	

the total MLS was 2.5. For *D6S264*, a MLS value of 3.4 was achieved. In addition, for this marker, a p value of 0.001 was initially demonstrated in our data set. Together with additional linkage evidence (p=0.01) obtained in the independent UK 96 data set, it was very clear that 6q encompassed another IDDM susceptibility locus besides *IDDM5*. This second disease locus, near *D6S264*, has been officially designated as *IDDM8*.

Genetic Heterogeneity According to HLA-DR/DQ Status of the Affected Sibpairs

To test HLA-associated heterogeneity, the identity by decent (ibd) data of affected families were subdivided according to HLA-*DR/DQ* haplotypes: sibpairs who shared 2 identical HLA haplotypes (HLA 2) and sibpairs who shared 1 or 0 HLA haplotype (HLA 1, 0). There were variations in the proportions of genes shared by affected sibpairs between the HLA 2 and HLA 1,0 categories for most marker loci in this study. There were also variations of ibd values in data subsets with different HLA-*DR*. However, none of the comparisons reached statistical significance. Therefore, the differences in ibd values between different HLA categories in most cases is likely due to random chance, or HLA's effect is too weak to be detected.

Discussion

Mapping genes predisposing to complex disorders such as IDDM is a difficult task. Suarez and colleagues[89] have shown by computer simulation that if a number of loci (each with a moderately small effect on disease) are implicated, then linkage will be difficult to detect and to replicate. The difficulty is due to heterogeneity expected between data sets, or even within studies. In monogenic diseases, the generally accepted norm for linkage is a LOD score of 3 ($p<0.001$). Previous studies[44,87,90] have shown that this norm can not be effectively achieved in studies of diseases with substantial genetic heterogeneity. The reason is that weak linkages could easily be missed even with 100 or more affected sibpairs. Lander and Schork have suggested that a p value of $3x10^{-5}$ (or MLS=3.6) is required to claim a true linkage (confident at the 5% level) when the human genome is examined.[91] Such criteria may be difficult to apply to complex diseases such as IDDM, because pooling of different data sets in light of substantial genetic heterogeneity may create serious problems. Alternatively, Davies and colleagues have suggested guidelines for statistic significance: 1) to obtain a p value of 0.001 in the initial data set. 2) to replicate this linkage in another independent data set with a p value of 0.05.[44] However, the false positive rate such

criteria is not yet known. In general, it is accepted that less stringent criteria should be applied for the initial establishment of linkage for complex diseases and more stringent criteria should then be applied to confirm the susceptibility genes. Therefore, I have reported any linkage evidence when p is less or equal to 0.05. Even though such evidence is not strong considering the number of markers tested, any marker that indicates linkage in one data set should be further investigated.

The linkage evidence for *D4S1566* was novel and warrants further studies in other independent families. Linkage evidence for *IDDM8* in my data set (MLS=2.8 for *D6S446* and MLS=2.0, p=0.001 for *D6S264*) and the weak evidence in the UK data set (MLS=1.4, p=0.01 for *D6S264*) together establish the presence of a disease locus in the 6q region using the criteria of Davies *et al*. When the UK data set and my data set were combined, linkage evidence for *D6S264* (MLS=3.4) almost reached the stringent criteria (MLS=3.6) suggested by Lander and Schork. Since *D6S264* is 28 cM more telomeric than *ESR* (*IDDM5*), this study suggests that there are probably two distinct IDDM genes on 6q (*IDDM5* near *ESR* and *IDDM8* near *D6S264-D6S446*). This conclusion is also supported by the UK data set. Since a 95% confidence interval is defined as the region that contains all markers having a MLS value greater than or equal to MLS_{max} - 1.4,[92] all markers that have a MLS of 1.4 (i.e. 2.8-1.4=1.4) are in the 95%

confidence interval of *IDDM8*. Thus, *IDDM8* is probably located in the interval telomeric to *D6S220*.

There are two observations worthy of notice. First, there was a fluctuation of MLS values along the 6q region. This observation is consistent with the allele-sharing of a complex genetic trait.[92] In the situation of a complex trait, the MLS follows a random walk in the neighborhood of its peak, with steps occurring at transitions between sharing and nonsharing. Second, the percentage of genes shared by affected sibpairs was 62.5% in the UK data set, which is very similar to that observed in my USA families (62.2%). If these observations can be confirmed in other independent families, *IDDM8* may be one of the most important susceptibility genes for IDDM in addition to the HLA class II genes. The contribution of a single disease locus to the total λ_s can be estimated from the ratio of the expected proportion of affected sibpairs sharing no alleles (0 ibd=0.25) and the observed proportion.[50] In fact, the λ_s conferred by *IDDM8* was estimated to be 1.8, which was higher than other non-HLA susceptibility genes (λ_s = 1.5, 1.4, 1.6, 1.2 and 1.3 for *IDDM2*, *IDDM3*, *IDDM4*, *IDDM5*, and *IDDM7*).[96] *IDDM8* is thus the most important IDDM susceptibility factor other than HLA.

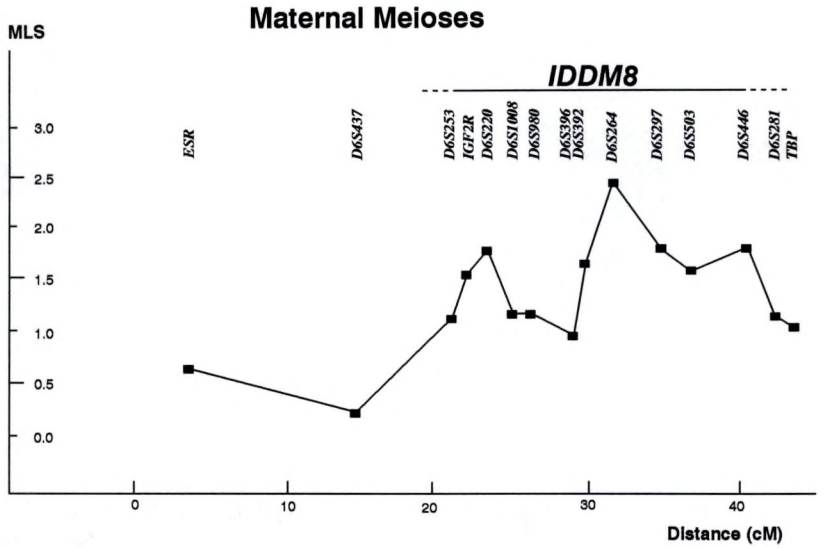
In order to investigate the characteristics of the potential *IDDM8*, the evidence of linkage for *IDDM8* was analyzed according to parent-of-origin status. As shown

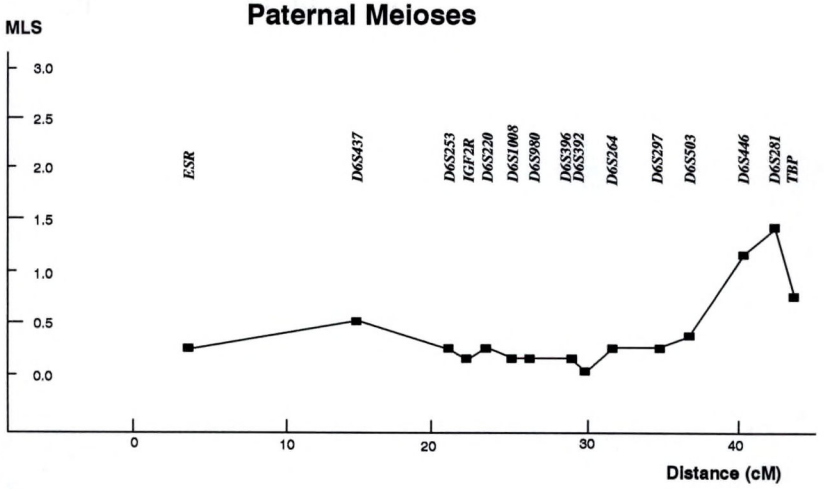
in Table 3-4 and Figure 3-3, it appeared that linkage for *IDDM8* was only detected in maternal meioses but not in paternal meioses. Since the paternal gene did not seem to be important in IDDM susceptibility, the paternal gene may not be expressed, suggesting a possible role for paternal imprinting.

Table 3-4. Evidence of paternal imprinting at *IDDM8*.

Markers	D (cM)	Paternal Meioses			Maternal Meioses		
		1 ibd	0 ibd	MLS	1 ibd	0 ibd	MLS
ESR	4	54	42	0.3	56	40	0.6
D6S437	15	55	40	0.5	52	42	0.2
D6S253	22	53	44	0.2	59	37	1.1
IGF2R	22	53	45	0.1	58	33	1.5
D6S220	23	53	44	0.2	58	32	1.7
D6S1008	25	52	46	0.1	56	35	1.1
D6S980	27	53	46	0.1	56	35	1.1
D6S396	29	52	46	0.1	57	37	0.9
D6S392	30	51	50	0.0	61	35	1.6
D6S264	32	53	43	0.2	64	32	2.4
D6S297	35	53	43	0.2	60	33	1.7
D6S503	37	53	42	0.3	59	34	1.5
D6S446	41	58	36	1.1	58	32	1.7
D6S281	42	54	31	1.4	53	32	1.1
TBP	43	38	24	0.7	41	24	1.0

Figure 3-3. Schematic presentation of evidence for paternal imprinting at *IDDM8*. The plot was based on the data in Table 3-4.


Maternal Meioses
MLS
IDDM8
ESR
D6S437
D6S253
IGF2R
D6S220
D6S1008
D6S980
D6S396
D6S392
D6S264
D6S297
D6S503
D6S446
D6S281
TBP
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
Distance (cM)


Paternal Meioses
MLS
ESR
D6S437
D6S253
IGF2R
D6S220
D6S1008
D6S980
D6S396
D6S392
D6S264
D6S297
D6S503
D6S446
D6S281
TBP
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
Distance (cM)

CHAPTER 4
DISCUSSION

Three years ago, I set out to answer three questions: (1) How many genes may contribute to IDDM susceptibility? (2) Where are they located? (3) How can they be identified? To date, most of these questions have been at least preliminary answered.

Genetic susceptibility to IDDM is complex, with HLA class II genes on chromosome 6p21 (*IDDM1*) as the major locus, with the insulin (*INS*) gene on chromosome 11p15 (*IDDM2*) as a minor locus, and with at least five additional minor loci on chromosomes 15q (*IDDM3*),[90] 11q (*IDDM4*),[87] 6q (*IDDM5*),[44] 2q (*IDDM7*)[50,93] and 6q (*IDDM8*).[96]

For *IDDM1*, the genetic determinants are the polymorphisms within the peptide-binding sites of the HLA-*DQ* and -*DR* molecules, but the identity of other disease-predisposing mutations remain to be identified.

For *IDDM2*, the locus was mapped by this and Lucassen's[78] study to the *INS* gene and its associated VNTR. However, the exact identity of *IDDM2* remained unknown until recently. Bennett *et al.*[85] revealed that *IDDM2* is determined by the VNTR at the 5' of the *INS* gene using a cross-match haplotype analysis. This notion is now generally accepted. Since this polymorphism does not

encode any known gene products so that it must exert its effect on IDDM susceptibility by regulating the expression of other genes. I hypothesize that the VNTR may regulate the transcription of its downstream genes, such as *INS* and *IGF2*.

The *INS*-associated VNTR is a 14 bp repeat sequence located in the promoter of the *INS* gene and is 365 bp from the *INS*'s transcription initiation site. This interesting location suggests that VNTR might be essential in regulating the *INS* gene expression. Since the *INS* gene encodes insulin (which may be an autoantigen in the process of disease development), the effect of the *INS* gene may be derived from increased insulin secretion and thereby lead to an augmentation of the targeted autoantigens expressed on pancreatic beta cells. There is evidence to support this hypothesis. Recently, Kennedy *et al.*[94] demonstrated that the *INS*-associated VNTR could be bound and activated by a transcription factor Pur-1 *in vitro*. The same study was also able to present preliminary evidence that the transcriptional levels of reporter genes are correlated with allelic variation within the VNTR. However, the VNTR-*INS* hypothesis cannot explain the observed gender-related transmission bias of *IDDM2*.

The next downstream gene to the *INS* is the *IGF2* gene which encodes a protein (insulin-like growth factor) that is important in β cell development.[83] In addition, this

gene is known to be maternally imprinted.[82,83] This evidence suggests the potential role for the *IGF2* gene in IDDM pathogenesis. Nevertheless, it remains possible that both the *INS* and *IGF2* genes are involved in the VNTR's effects in IDDM.

The identity of *IDDM8* is still unknown. In this study, the paternal imprinting characteristic of *IDDM8* was first identified. Recently, evidence suggests that an imprinted gene on chromosome 6 may be involved in transient neonatal diabetes mellitus (TNDM).[95] This gene appears to be important for pancreatic β cell development. It remains to be seen whether the TNDM gene is identical or related to *IDDM8* on 6q. Another candidate gene for *IDDM8* is *IGF2R*. Since *IGF2R* exhibits paternal imprinting in mice and in humans, it may be the paternally imprinted factor on 6q. Intriguingly, *IGF2*, a candidate gene for *IDDM2* on 11q15, is maternally imprinted. The above information together suggests that the *IGF2-IGF2R* hypothesis is a very attractive mechanism for IDDM susceptibility and deserves further investigation. In our lab, a microsatellite marker located in the 3'-untranslated region of *IGF2R* was examined by other colleagues using linkage disequilibrium analysis. Although linkage disequilibrium was not demonstrated, this does not exclude *IGF2R* as a candidate for *IDDM8*. Further mutation analysis, especially in the regulatory region, is of great importance.

Thus far seven susceptibility loci (*IDDM1*, *IDDM2*, *IDDM3*, *IDDM4*, *IDDM5*, *IDDM7* and *IDDM8*) have been identified. What is their combined effect on the total familial clustering of IDDM (λ_s=15)? *IDDM1* is the major locus for IDDM susceptibility, with a λ_s of 3.1-4.5.[44] The λ_s for *IDDM2* and *IDDM7* are both 1.3.[50] *IDDM4* and *IDDM5* both have λ_s=1.1.[44,87] For *IDDM3*, the λ_s is 1.4.[96] Finally, the λ_s for *IDDM8* is 1.8. Therefore, the total λ_s is 11.1-12.5, which is about 80% of the total familial clustering of IDDM. Three conclusions can be drawn from this calculation. First, IDDM is definitely polygenically controlled. Second, it seems that most of the IDDM susceptibility genes, if not all, have been localized. The next logical step will be to reveal the identities of these genes and to investigate how they interact with one another and the environment to cause disease. Third, since λ_s for *IDDM8* is 1.8, which accounts for a higher proportion of the familial clustering of IDDM (i.e. higher λ_s value) than other non-HLA susceptibility genes, *IDDM8* may be the most important non-HLA susceptibility factor.

The success in the localization of polygenic factors of IDDM is a big leap for mankind in the journey of conquering this ancient and worldwide disease. The genetic studies of IDDM will ultimately have a great impact on the prediction, prevention and treatment of the disease.

REFERENCE LIST

1. R. David, G. Ledlie, and B. Elliott. Early environmental events as cause of IDDM. Diabetes 43:843-850 (1994).

2. R.E. LaPorte and K.J. Cruickshanks. Incidence and risk factors for insulin-dependent diabetes. In: "National Diabetes Data Group's diabetes data compiled 1984", Department of Health and Human Services., NIH, Bethesda, MD. pp. III-1-III-12 (1985).

3. M. Karvonen, J. Tuomilehto, I. Libman, and R.F. LaPorte. A review of the recent epidemiological data on the worldwide incidence of Type 1 (insulin-dependent) diabetes mellitus. Diabetologia 36:883-892 (1993).

4. M.A. Atkinson and N.K. Maclaren. The pathogenesis of insulin-dependent diabetes mellitus. N Engl J Med 331:1428-1436 (1994).

5. A.L. Notkins. The causes of diabetes. Sci Am 241:62-73 (1979).

6. K.G. Alberti. Insulin dependent diabetes mellitus: A lethal disease in the developing world. BMJ 309:754-755 (1994).

7. H.S. Jun and J.W. Yoon. Initiation of autoimmune type 1 diabetes and molecular cloning of a gene encoding for islet cell-specific 37 kd autoantigen. Adv Exp Med Biol 347:207-220 (1994).

8. J.A. Todd. Genetic control of autoimmunity in type 1 diabetes. Immunology Today 11:122-129 (1990).

9. G.S. Eisenbarth. Type 1 diabetes mellitus: a chronic autoimmune disease. N Engl J Med 314:1360-1368 (1986).

10. R. Wassmuth and A. Lernmark. The genetics of susceptibility to diabetes. Clin Immunol Immunopathol 53:358-399 (1989).

11. E.F. Lampeter, M. Homberg, K. Quabeck, U.W. Schaefer, P. Wernet, J. Bertrams, H. Grosse-Wilde, F.A. Gries, and H. Kolb. Transfer of insulin-dependent diabetes between HLA-identical siblings by bone marrow transplantation. Lancet 341:1243-1244 (1993).

12. R.K. Sibley, D.E. Sutherland, F. Goetz, and A.F. Michael. Recurrent diabetes mellitus in the pancreas iso- and allograft: a light and electron microscopic and immunohistochemical analysis of four cases. Lab Invest 53:132-144 (1985).

13. Canadian-European Randomized Control Trail Group. Cyclosporin-induced remission of IDDM after early intervention: association of 1 year of cyclosporin treatment with enhanced insulin secretion. Diabetes 37:1574-1582 (1988).

14. G. Feutren, L. Papoz, and R.E. Assan. Cyclosporin increases the rate and length of remissions in insulin-dependent diabetes of recent onset: results of a multicenter double-blind trial. Lancet 2:119-124 (1986).

15. A.K. Foulis and A. Clark. "Joslin's Diabetes Mellitus" Lea & Febiger, Baltimore, MD., (1994).

16. D.W. Drell and A.L. Notkins. Multiple immunological abnormalities in patients with type 1 (insulin-dependent) diabetes mellitus. Diabetologia 30:132-143 (1987).

17. D. Faustman, G. Eisenbarth, J. Daley, and J. Breitmeyer. Abnormal T-lymphocyte subsets in type 1 diabetes. Diabetes 38:1462-1468 (1989).

18. Y. Naparstek and P.H. Plotz. The role of autoantibodies in autoimmune disease. Annu Rev Immunol 11:79-104 (1993).

19. J. Nerup, T. Mandrup-Poulsen, S. Helqvist, H.U. Andersen, and T.E. Lorenzen. On the pathogenesis of IDDM. Diabetologia 37:S82-S89 (1994).

20. T. Mandrup-Poulsen. On the pathogenesis of insulin-dependent diabetes mellitus. Dan Med Bull 39:509-542 (1992).

21. J. Molvig. A model of the pathogenesis of insulin-dependent diabetes mellitus. Dan Med Bull 39:509-541 (1992).

22. J. Nerup, T. Mandrup-Poulsen, and J.E. Molvig. Insulin resistance and insulin-dependent diabetes mellitus. Diabetes Care 17:S16-S23 (1987).

23. B. Appels, V. Burkart, and G.E. Kantwerk-Funke. Spontaneous cytotoxicity of macrophages against pancreatic islet cells. J Immunol 142:3803-3808 (1989).

24. P. Olmos, R. A'hern, D. Heaton, B.A. Millward, D. Risley, D.A. Pyke, and R.D.G. Leslie. The significance of the concordance rate for type 1 (insulin-dependent) diabetes in identical twins. Diabetologia 31:747-750 (1988).

25. M.A. Metcalfe and J.D. Baum. Incidence of insulin dependent diabetes in children aged under 15 years in the British Isles during 1988. Br Med J 302:443-447 (1991).

26. J. Tuomilehto, M. Rewers, A. Reunanen, P. Lounamaa, R. Lounamaa, E. Tuomilehto-Wolf, and H.K. Akerblom. Increasing trend in type 1 (insulin-dependent) diabetes mellitus in childhood in Finland. Diabetologia 34:282-287 (1991).

27. C.E. Drykoningen, A.L. Mulder, G.J. Vandrager, R.E. LaPorte, and G.J. Bruining. The incidence of male childhood type 1 (insulin-dependent) diabetes mellitus is rising rapidly in the Netherlands. Diabetologia 35:139-142 (1992).

28. D.P. Singal and M.A. Blajchman. Histocompatibility (HLA) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes 22:429-432 (1973).

29. J. Bach. Insulin-dependent diabetes mellitus as an autoimmune disease. Endocrine Reviews 15:516-542 (1994).

30. E. Wolf, K.M. Spencer, and A.G. Cudworth. The genetic susceptibility to type 1 (insulin-dependent) diabetes: Analysis of the HLA-DR association. Diabetologia 24:224-230 (1983).

31. N. Risch. Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1-14 (1987).

32. E. Thorsby and K.S. Ronningen. Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 (insulin-dependent) diabetes mellitus. Diabetologia 36:371-377 (1993).

33. J.L. Tiwari and P.I. Terasaki. HLA and disease associations. In: "Endocrinology", J.L. Tiwari, Springer-Verlag, New York, pp. 183-231 (1985).

34. D.G. Owerbach and K.H. Gabbay. Primary association of HLA-DQw8 with type I diabetes in DR4 patients. Diabetes 38:942-945 (1989).

35. E. Thorsby and K.S. Ronningen. Role of HLA genes in predisposition to develop insulin-dependent diabetes mellitus. Ann Med 24:523-531 (1992).

36. J.A. Todd, C. Mijovic, J. Fletcher, D. Jenkins, A.R. Bradwell, and A.H. Barnett. Identification of susceptibility loci for insulin-dependent diabetes mellitus by trans-racial gene mapping. Nature 338:587-589 (1989).

37. T. Awata, T. Kuzuya, A. Matsuda, Y. Iwamoto, and Y. Kanazawa. Genetic analysis of HLA class II alleles and susceptibility to type I (insulin-dependent) diabetes mellitus in Japanese subjects. Diabetologia 35:419-424 (1992).

38. C. Mijovic, D. Jenkins, K.H. Jacobs, M.A. Penny, J.A. Fletcher, and A.H. Barnett. HLA-DQA1 and -DQB1 alleles associated with genetic susceptibility to IDDM in a black population. Diabetes 40:748-753 (1991).

39. J.A. Todd, J.I. Bell, and H.O. McDevitt. HLA-DQ beta gene contributes to susceptibility evidence for gene trans-complementation. Nature 329:599-604 (1987).

40. D.J. Charron, V. Lotteau, and P. Turmel. Hybrid HLA-DC antignes provide molecular evidence for gene trans-complementation. Nature 312:157-159 (1984).

41. I. Khalil, I. Deschamps, V. Lepage, R. Al-Daccak, L. Degos, and J. Hors. Dose effect of cis and trans-encoded HLA-DQαβ heterodimers in IDDM susceptibility. Diabetes 41:378-384 (1992).

42. R. Buzzetti, L. Nistico, J. Osborn, C. Giovannini, A. Chersi, and R. Sorrentino. HLA-DQA1 and DQB1 gene polymorphisms in type I diabetic patients from central Italy and their use for risk prediction. Diabetes 42:1173-1178 (1993).

43 H.S. Huang, J.T. Peng, J.Y. She, L.P. Zhang, and J.X. She. HLA-encoded susceptibility to IDDM is determined by DR and DQ genes as well as their linkage disequilibria: a unified hypothesis for HLA association and disease prevalence. Human Immunology (in press).

44. J.L. Davies, K. Yoshihiko, S. Bennett, J.B. Copeman, H.J. Cordell, L.E. Pritchard, P.W. Reed, S.C.L. Gough, C. Jenkins, S.M. Palmer, K.M. Balfour, B.R. Rowe, M. Farrall, A.H. Barnett, S.C. Bain, and J.A. Todd. A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371:130-136 (1994).

45. G.I. Bell, S. Horita, and J.H. Karam. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176-183 (1983).

46. C. Julier, R.N. Hyer, J. Davies, F. Merlin, P. Soularue, L. Briant, G. Cathelineau, I. Deschamps, J.I. Rotter, P. Froguel, C. Boitard, J.I. Bell and G.M. Lathrop. Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature 354:155-159 (1991).

47. N.J. Cox, L. Barker, and R.S. Spielman. Insulin gene sharing in sub-pairs with insulin-dependent diabetes mellitus: No evidence for linkage. Am J Hum Genet 42:167-? (1988).

48. S.C. Bain, J.B. Prins, C.M. Hearne, N.R. Rodrigues, B.R. Rowe, L.E. Pritchard, R.J. Richie, J.R. Hall, D.E. Undlien, K.S. Ronningen, D.B. Dunger, A.H. Barnet and J.A. Todd. Insulin gene region-encoded susceptibility to type 1 diabetes is not restricted to HLA-DR4-positive individuals. Nat Genet 2:212-215 (1992).

49. W.E. Winter, T. Chihara, and D. Schatz. The genetics of autoimmune diabetes: Approaching a solution to the problem. Am J Dis Child 147:1282-1290 (1993).

50. J.B. Copeman, F. Cucca, C.M. Hearne, R.J. Cornall, P.W. Reed, K.S. Ronningen, D.E. Undlien, L. Nistico, R. Buzzetti, R. Tosi, F. Pociot, J. Nerup, F. Cornelis, A.H. Barnett, S.C. Bain, and J.A. Todd. Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. Nature Genetics 9:80-85 (1995).

51. J.A. Todd and S.C. Bain. A practical approach to identification of susceptibility genes for IDDM. Diabetes 41:1029-1034 (1992).

52. J.A. Todd, T.J. Aitman, R.J. Cornall, S. Ghosh, Hall, C.M. Hearne, A.M. knight, J.M. Love, M.A. McAleer, J. Prins, N. Rodrigues, G.M. Lathrop, A. Pressey, N. DeLarato, L.B. Peterson, and L.S. Wicker. Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature (Lond) 351:542-547 (1991).

53. M. Prochazka, E.H. Leiter, D.V. Serreze, and D.L. Coleman. Three recessive loci required for insulin-dependent diabetes in nonobese diabetic mice. Science 237:286-289 (1987).

54. N. Risch, S. Ghosh, and J.A. Todd. Statistical evaluation of multiple -locus linkage data in experimental species and its relevance to human studies: application to nonobese diabetic (NOD) mouse and human insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 53:702-714 (1993).

55. J.H. Nadeau. Maps of linkage and synteny homologies between mouse and man. Trends Genet 5:82-86 (1989).

56. P. Froguel, M. Vaxillaire, F. Sun, G. Velho, H. Zouali, M.O. Butel, S. Lesage, N. Vionnet, K. Clement, F. Fougerousse, Y. Tanizawa, J. Weissenbach, J.S. Beckmann, G.M. Lathrop, P. Passa, M.A. Permutt, and D. Cohen. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature (Lond) 356:162-164 (1992).

57. A. Hattersley, R.C. Turner, A.M. Permutt, P. Patel, Y. Tanizawa, K.C. Chiu, S. O'Rahilly, P. Watkins, and J.S. Wainscoat. Type 2 diabetes is linked to the glucokinase gene in a large pedigree. Lancet 340:54-55 (1992).

58. N. Vionnet, M. Stoffel, J. Takeda, K. Yasuda, G.I. Bell, H. Zouali, S. Lesage, G. Velho, F. Iris, P. Passa, and P.D.C. Froguel. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature (Lond) 356:721-722 (1992).

59. J.F. Gusella, N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, A.B. Young, I. Shoulson, E. Bonilla, and J.B. Martin. A polymorphic DNA marker genetically linked to Huntington's disease. Nature (Lond) 306:234-238 (1983).

60. L.C. Tsui, M. Buchwald, D. Barker, J.C. Braman, R. Knowlton, J.W. Schumm, H. Eiberg, J. Mohr, D. Kennedy, and N. Plavsic. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. Science 230:1054-1057 (1985).

61. M.R. Wallace, D.A. Marchuk, L.B. Anderson, R. Letcher, H.M. Odeh, A.M. Saulino, J.W. Fountain, A. Brereton, J. Nicholson, A.L. Mitchell, B.H. Brownstein, and F.S. Collins. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science 249:181-186 (1990).

62. S.S. Rich. Mapping genes in diabetes-genetic epidemiological perspective. Diabetes 39:1315-1319 (1990).

63. R.C. Elston. The use of polymorphic markers to detect genetic variability. In: "Phenotypic Variation in Populations: Relevance to Risk Assessment", A.D. Woodhead, Plenum, New York, pp. 105-112 (1988).

64. R.S. Spielman, R.E. McGinnis, and W.J. Ewens. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506-516 (1993).

65. J. Ott. Methods of linkage analysis. In: "Analysis of Human Genetic Linkage", The Johns Hopkins University Press, Baltimore and London, pp. 54-80 (1991).

66. J.A. Todd. Genetic control of autoimmunity in type 1 diabetes. Immunol Today 11:122-129 (1990).

67. P. Chomoczynski and N. Sacchi. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156-159 (1987).

68. T.M. DeChiara, E. Robertson, J., and A. Efstratiadis. Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64:849-859 (1991).

69. J.L. Weber. Human DNA polymorphisms based on length variations in simple sequence tandem repeats. In: "Genome Analysis", K.E. Davies, CSHL, pp. 159-181 (1990).

70. J.L. Weber. Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524-530 (1990).

71. J.L. Tiwari and P.I. Terasaki. The data and statistical analysis. In: "HLA and disease associations", J.L. Tiwari and P.I. Terasaki, Springer Verlag, New York, pp. 18-27 (1985)

72. D. Owerbach and K.H. Gabbay. Localization of a type I diabetes susceptibility locus to the variable tandem repeat region flanking the insulin gene. Diabetes 42:1708-1714 (1993).

73. J.X. She, M.M. Bui, X.H. Tian, A. Muir, E.K. Wakeland, B. Zorovich, L.P. Zhang, M.C. Liu, G. Thomson, and N.K. Maclaren. Additive susceptibility to insulin-dependent diabetes conferred by HLA-DQB1 and Insulin genes. Autoimmunity 18:195-203 (1994).

74. D. Owerbach and K.H. Gabbay. Linkage of the VNTR/insulin-gene and type I diabetes mellitus: Increased gene sharing in affected sibling pairs. Am J Hum Genet 54:909-912 (1994).

75. B.J. Van-der-Auwera, H. Heimberg, A.F. Schrevens, C. Van-Waeyenberge, J. Flament, and F.C. Schuit. 5' insulin gene polymorphism confers risk to IDDM independently of HLA class II susceptibility. Diabetes 42:851-854 (1993).

76. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus. Diabetes 28:1039-1057 (1979).

77. G. Thomson, W.P. Robinson, M.K. Kuhner, S. Joe, and W. Klitz. HLA and insulin gene associations with IDDM. Genet Epidemiol 6:155-160 (1989).

78. A.M. Lucassen, C. Julier, J.P. Beressi, C. Boitard, P. Froguel, M. Lathrop, and J.I. Bell. Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. Nature Genetics 4:305-310 (1993).

79. P. Heutink, A.G.L. Van Der Mey, L.A. Sandkuijl, A.P.G. Van Gils, A. Bardoel, G.J. Breedveld, M. Van Vliet, G.J.B. Van Ommen, C.J. Cornelisse, B.A. Oostra, J.L. Weber, and P. Devilee. A gene subject to genomic imprinting and responsible for paragangliomas maps to chromosome 11q23. Human Molec Genet 1:7-10 (1992).

80. Y. Zhang and B. Tycko. Monoalleleic expression of the human H19 gene. Nature Genetics 1:40-44 (1992).

81. S. Rainier, L.A. Johnson, C.J. Dobry, A.J. Ping, P.E. Grundy, and A.P. Feinberg. Relaxation of imprinted genes in human cancer. Nature 362:747-749 (1993).

82. N. Giannoukakis, N. Deal, J. Paquette, C.G. Goodyer, and C. Polychronakos. Paternal genomic imprinting of the human IGF2 gene. Nature Genetics 4:94-97 (1993).

83. R. Ohlsson, A. Nystrom, S. Pfeifer-Ohlsson, V. Tohonen, F. Hedborg, p. Schofield, F. Flam, and T.J. Ekstrom. IGF2 is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. Nature Genetics 4:94-97 (1993).

84. S.J. Giddings, C.D. King, K.W. Harman, J.F. Flood, and L.R. Carnaghi. Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. Nature Genetics 6:310-313 (1994).

85. S.T. Bennett, A.M. Lucassen, S.C.L. Gough, E.E. Powell, D.E. Undlien, L.E. Pritchard, M.E. Merriman, M.J. Kawaguchi, F. Dronsfield, J. Pociot, N. Nerup, A. Bouzekri, A. Cambon-Thomsen, K.S. Ronningen, A.H. Barnett, S.C. Bain, and J.A. Todd. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nature Genetics 9:284-292 (1995).

86. S. Ghosh, S.M. Palmer, N.R. Rodrigues, H.J. Cordell, C.M. Hearne, R. Cornall, J., J.B. Prins, P. McShane, G.M. Lathrop, L.B. Peterson, L.S. Wicker, and J.A. Todd. Polygenic control of autoimmune diabetes in nonobese diabetic mice. Nature Genetics 4:404-409 (1993).

87. L. Hashimoto, C. Habita, J.P. Beressi, M. Delepine, C. Besse, A. Cambon-Thomsen, I. Deschamps, J.I. Rotter, S. Djoulah, M.R. James, P. Froguel, J. Weissenbach, G.M. Lathrop, and C. Julier. Genetic mapping of susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371:161-164 (1994).

88. N. Risch. Linkage strategies for genetically complex traits. Am J Hum Genet 46:229-241 (1990).

89. B.K. Suarez, C.L. Hampe, and P. Van Eerdewegh. "Genetic Approaches to Mental Disorders." American Psychiatric Press, Washington, D.C. (1994).

90. L.L. Field, R. Tobias, and T. Magnus. A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus. Nature Genetics 8:189-194 (1994).

91. E.S. Lander and N.J. Schork. Genetic dissection of complex traits. Science 265:2307-2048 (1994).

92. L. Kruglyak and E. Lander. High-resolution genetic mapping of complex traits. Am J Hum Genet 56:1212-1223 (1995).

93. D. Owerbach and K. Gabbay. The HOXD8 locus (2q31) is linked to type 1 diabetes. Diabetes 44:132-136 (1995).

94. G.C. Kennedy, M.S. German, and W.J. Rutter. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. Nature Genetics 9:293-298 (1995).

95. R.S. James, J.A. Crolla, F.L. Sitch, P.A. Jacobs, W.M. Howell, P. Betts, J.D. Baum, and J.P.H. Shield. An imprinted gene(s) for diabetes. Nature Genetics 9:110-112 (1995).

96. D.F. Luo, M.M. Bui, A. Muir, N.K. Maclaren, G. Thomson, and J.X. She. Affected sibpair mapping of a novel susceptibility gene to insulin-dependent diabetes mellitus (IDDM8) on chromosome 6q25-q27. Am J Hum Genet (in press) (1995).

BIOGRAPHICAL SKETCH

Marilyn Yuanxin Ma Bui was born in Beijing, China, on August 10, 1962. She grew up with two brothers in a family of career diplomats. After earning her medical doctor degree in 1986, she joined the faculty of the same medical school--Capital Institute of Medicine in Beijing, China. In 1989, she came to the States. In 1990, she entered the graduate program in the Department of Pathology and Laboratory Medicine at the University of Florida College of Medicine. Her Ph.D. degree was conferred on August 12, 1995.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jin-Xiong She, Chair
Assistant Professor of
Pathology and Laboratory
Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Noel K. Maclaren
Professor of Pathology and
Laboratory Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward K. Wakeland
Professor of Pathology and
Laboratory Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Margaret R. Wallace
Assistant Professor of
Biochemistry and Molecular
Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William E. Winter
Associate Professor of
Pathology and Laboratory
Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas P. Yang
Associate Professor of
Biochemistry and Molecular
Biology

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1995

Dean, College of Medicine

Dean, Graduate School