

The Role of Chromosomal Region 16p13.13 in Type 1 Diabetes

Melvin Joseph Tomlinson IV  
Downingtown, Pennsylvania

Bachelor of Science, Millersville University of Pennsylvania, 2009  
Master of Science, University of Virginia, 2012

A Dissertation presented to the Graduate Faculty  
of the University of Virginia in Candidacy for the Degree of  
Doctor of Philosophy

Department of Biochemistry and Molecular Genetics

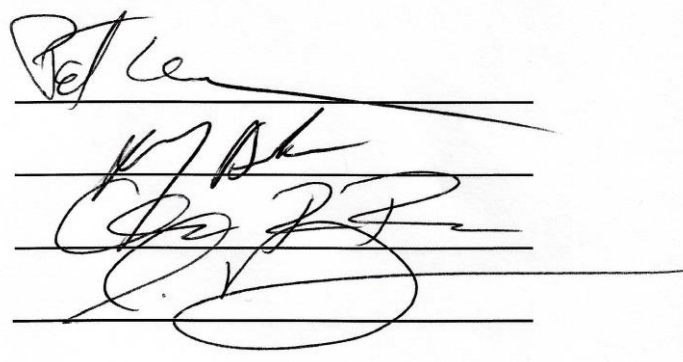
University of Virginia  
December, 2013

Pat Concannon, Ph.D.

Daniel Burke, Ph.D.

Charles Farber, Ph.D.

Ira Hall, Ph.D.

The image shows four handwritten signatures in black ink, each written on a horizontal line. The signatures are stylized and cursive. The first signature is at the top, followed by three more signatures below it, each on its own line.

©Copyright  
Melvin Joseph Tomlinson IV  
All Rights Reserved  
December 2013

### ACKNOWLEDGEMENTS

I would like to thank my graduate advisor Dr. Patrick Concannon for allowing me to join his lab and introducing me to this interesting project about type 1 diabetes. Without his support and guidance the project would not have evolved the way it did. I would also like to thank Dr. Suna Onengut-Gumuscu, who gave me space to work in her lab and provided daily support and guidance for this project. This project was hers to begin with, but she graciously allowed me to take it over. Furthermore, I would like to thank Dr. Sharon Teraoka for her encouragement and advice. She always provided meaningful feedback on presentations and papers and always willing to assist. Whenever anybody asks me about my mentors I always mention I have three; Dr. Patrick Concannon, Dr. Suna Onengut-Gumuscu and Dr. Sharon Teraoka.

I would like to thank my dissertation committee members Dr. Daniel Burke, Dr. Ira Hall and Dr. Charles Farber for their helpful advice and guidance on this project.

I would also like to thank all the individuals in the Concannon and Onengut Labs for their support and assistance on this collaborative project. Dr. Achilleas Pitsillides was essential in helping with data analysis; Dr. Rebecca Pickin assisted me in some the experiments; Matthew Mika helped me run some of the assays and provided stocks and RNA samples; Jessica Bonnie helped me to produce final figures; Dr. Keith Keene helped produce some of the preliminary data; Dr. Yan Ge and Dr. Larry Mesner for helpful advice and feedback on the project; Jo Wright helped keep a well stocked and organized lab that allowed everyone to have a great work environment. Thank you all for your support!

I would also like to thank my mother and sister whose endless love and support helped me get me through college and graduate school, I would not be where I am today without them. And finally, I would like to thank my step-father, Neal Heires, who I have only known for a few years now, but nonetheless gave his love and support as well.

## The Role of Chromosomal Region 16p13.13 in Type 1 Diabetes

Melvin Joseph Tomlinson IV

Advisor: Patrick Concannon Ph.D.

### Abstract

Type 1 diabetes most frequently results from the autoimmune destruction of the insulin producing pancreatic *Beta* cells, resulting in dependence on exogenous insulin. The disease is complex and arises due to the effects of both genetic and environmental factors. With the advent of genome-wide association studies (GWAS) over 40 additional loci have been identified that confer low to moderate risks of developing type 1 diabetes (Odds Ratios, OR,  $\leq 1.3$ ). One of these regions identified is located at chromosome 16p13.13 and is associated with type 1 diabetes and several other autoimmune disorders. The specific region of interest on chromosomal region 16p13.13 implicated in autoimmunity spans approximately 530 KB from 10,937,499-11,467,499 base position (bp) (build Hg19) and encompasses four major genes: *CIITA*, *DEXI*, *CLEC16A* and *SOC1*.

In order to identify specific risk variants for type 1 diabetes in this region and understand the mechanism of their action, we re-sequenced the entire region in 192 individuals (128 patients and 64 controls) identifying 93 novel variants. A panel of 939 SNPs, that included 46 of these novel variants, was genotyped in 3,070 multiplex families with type 1 diabetes to identify potential risk variants. After correcting for multiple testing, 48 SNPs provided statistically significant evidence of association with type 1

diabetes (most significant SNP rs34306440  $P = 5.74 \times 10^{-6}$ ), conditioning on the top SNP found no further significant associations.

The panel of SNPs used for fine mapping was also tested for association with variation in the transcript levels of each reported isoform of all genes in the region. Significant associations were observed only for transcript levels of *DEXI*, a gene originally reported to be unregulated by dexamethasone but whose function is unknown. We also tested novel unreported isoforms of *CIITA* in our study, but found no association.

We examined the relationship between the odds ratio for type 1 diabetes and the magnitude of the effect of *DEXI* transcript levels for each SNP in the region. Among SNPs significantly associated with type 1 diabetes, the common allele conferred an increased risk for disease, and corresponded to lower *DEXI* expression. Our results suggest that the primary mechanism by which genetic variation at *CLEC16A* contributes to risk for type 1 diabetes is through reduced expression of *DEXI*.

### **Dissertation Overview**

The dissertation is divided into five chapters. Chapter 1 (Type 1 Diabetes) focuses on a review of our current knowledge of type 1 diabetes. Chapter 2 (Brief History of Diabetes) examines the magnitude and impact diabetes has had on our society and some of the major historical events that led up to our current understanding of the disease. Chapter 3 (Genetics of Type 1 Diabetes) encompasses a compressive review of the current genetic understanding of type 1 diabetes and the contextual placement of the dissertation project in this rapidly advancing field. Chapter 4 (Chromosome 16p13.13 and Its Role in Type 1 Diabetes) is the main focus of the dissertation and explores the overall project. Chapter 5 (Future Directions) expands upon future experiments of this project and overall future directions of type 1 diabetes research.

## Table of Contents

<b>Chapter 1: Type 1 Diabetes in Today's World .....</b>	<b>1</b>
Introduction .....	2
Incident Rates .....	3
General Features of Diabetes .....	3-4
Type 1 Diabetes Environmental Variables.....	4-5
Pathogenesis of Type 1 Diabetes (What is known) .....	5-8
Important Research Tool- Islet Cell Autoantibodies (ICA) .....	8-9
Conclusion .....	9
Work Cited .....	10-16
 <b>Chapter 2: Brief History of Diabetes .....</b>	 <b>17</b>
Introduction .....	18
Ancient Times--- "To Pass Through" .....	18-19
Diabetes "Mellitus" .....	19-20
Discovery of Insulin .....	20-22
Two Diseases: Type 1 and Type 2 .....	22-23
1970 - Present .....	23-24
Work Cited .....	25-27



<b>Chapter 3: The Genetics of Type 1 Diabetes .....</b>	<b>28</b>
Introduction .....	29-31
Genetic Linkage Studies.....	32
Candidate-Gene Association Studies .....	32-33
GWAS .....	33-37
HLA (IDDM1).....	37-41
Insulin ( <i>INS</i> ) (IDDM2).....	41
<i>CTLA-4</i> (IDDM12) .....	42-43
<i>PTPN22</i> .....	43-44
<i>IL2RA</i> .....	45-46
Conclusion.....	46-47
Work Cited .....	48-64
 <b>Chapter 4: Chromosome 16p13.13 and Its Role in Type 1 Diabetes .....</b>	 <b>65</b>
Introduction .....	66-77
Background Information .....	66-68
<i>CIITA</i> .....	69-70
<i>DEXI</i> .....	71
<i>CLEC16A</i> .....	71-73

<b>Chapter 4: Continued .....</b>	
<i>SOCS1</i> .....	74-75
Project Purpose .....	76-77
Research Design and Methods.....	78-84
Subjects .....	78
Sequencing of Region to Identify SNPs.....	78
SNP Filtering and Selection .....	79
Genotyping .....	79-80
ImmunoChip Dataset .....	80-81
Imputation.....	81
Association Testing.....	81-82
RT-PCR.....	82-83
Expression Quantitative Trait Loci (eQTL) Analysis.....	84
Results .....	85-108
Re-sequencing of the 16p13.13 Region Identifies Novel SNPs .....	85
Fine Mapping Identifies Statistically Significant Association with Type 1 Diabetes .....	85-95
Custom Panel .....	85-87
ImmunoChip .....	88-89

<b>Chapter 4: Continued</b> .....	
Final Merged Dataset.....	90-95
Conditioning on Top SNP .....	96-97
Identification of Novel Splicing Isoforms of <i>CITTA</i> Created by Alternative Splicing .....	98-99
SNPs in <i>CLEC16A</i> are Associated with Variation in Expression Level of <i>DEXI</i> .....	100-106
Relationship of SNPs Associated with Type 1 Diabetes and/or <i>DEXI</i> Expression.....	107-108
Discussion.....	109-111
Acknowledgements for Research Work.....	112
Work Cited .....	113-120
<b>Chapter 5: Future Directions</b> .....	<b>121</b>
Introduction .....	122
Investigating <i>DEXI</i> .....	122-125
Possible Mechanisms Conferring Risk .....	125-127
Future Directions of Type 1 Diabetes Research .....	128-130
Work Cited .....	131-133

## List of Figures

<b>Figure 3-1</b> Study Strengths for Discovering Genes Involved in Complex Diseases .....	31
<b>Figure 3-2</b> Forty-nine Non HLA Gene Regions Associated With Type 1 Diabetes .....	35
<b>Figure 3-3</b> HLA .....	38
<b>Figure 4-1</b> Locations of Autoimmune Related Disease SNPs that Have Reached Genome Wide Significance for a Linkage Disequilibrium Block in Chromosome 16p13.13 .....	67
<b>Figure 4-2</b> Linkage Disequilibrium in Chromosome 16p13.13 .....	68
<b>Figure 4-3</b> The Pathway of CIITA .....	70
<b>Figure 4-4</b> Isoforms of CLEC16A.....	72
<b>Figure 4-5</b> Co-localization of Spinster and <i>Ema</i> .....	73
<b>Figure 4-6</b> The SOCS Activation Pathway (Negative Feedback Inhibition).....	75
<b>Figure 4-7</b> Fine Mapping of a Chromosome 16p13.13 Region with Custom SNP Genotyping Panel for Association with Type 1 Diabetes .....	87
<b>Figure 4-8</b> SNPs from the ImmunoChip Genotyping Panel Associated with Type 1 Diabetes .....	89
<b>Figure 4-9</b> A Comparison of SNP Genotyping Panels Used to Fine Map Chromosome 16p13.13 .....	91

**Figure 4-10** SNPs Associated with Type 1 Diabetes from the

Final Genotyping Panel .....92

**Figure 4-11** Focused View of the 48 Most Statistically Significant SNPs Associated with

Type 1 Diabetes .....93

**Figure 4-12** Conditioning on the SNP Most Significantly Associated with Type 1

Diabetes to Reveal Independently Associated SNPs .....97

**Figure 4-13** Novel Isoforms of *CIITA* .....99**Figure 4-14** eQTL SNPs for *DEXI* .....101**Figure 4-15** Expression Levels of *DEXI* .....106**Figure 4-16** Relationship of Risk for Type 1 Diabetes and *DEXI* Expression.....108**Figure 5-1** Possible Mechanism for Linkage Disequilibrium Block in Chromosome

16p13.13 in Autoimmunity .....127

## List of Tables

<b>Table 4-1</b> Primer Sequences Used for the Expression Quantitative Trait Loci (eQTL)	
Analysis of the Genes and Isoforms in LD Block 16p13.13 .....	83
<b>Table 4-2</b> Functional Annotation of SNPs on Custom Panel .....	86
<b>Table 4-3</b> Functional Annotation of SNPs on Final Genotyping Panel (939 SNPs) .....	90
<b>Table 4-4</b> SNPs in <i>CLEC16A</i> Most Strongly Associated with Type 1 Diabetes	
(total 48) from TDT Analysis Sorted in Order of Base Position .....	94-95
<b>Table 4-5</b> SNPs Significantly Associated with Type 1 Diabetes with their	
Corresponding eQTL Results for <i>DEXI</i> .....	102-103
<b>Table 4-6</b> Top eQTL SNP <i>P</i> -values for Each Gene Tested .....	104

## **Chapter 1**

Type 1 Diabetes in Today's World

## **Introduction**

Type 1 diabetes most frequently results from the autoimmune destruction of the insulin producing pancreatic  $\beta$  cells, ultimately resulting in the affected individual requiring exogenous insulin for survival. Type 1 diabetes is considered a complex disease that arises due to the effects of both genetic and environmental factors evident by the incomplete concordance of 30-70% between monozygotic twins (1-6). For dizygotic twins the concordance rate is estimated to be around 5-15% (2-4,6). Type 1 diabetes was previously known as insulin-dependent diabetes mellitus (IDDM), or juvenile onset diabetes, because it was originally recognized to occur in young children, but these terms are no longer widely used. Type 1 diabetes can occur at any age and it has been found that over 50% of new cases of type 1 diabetes occur in adults (7).

The initial symptoms of type 1 diabetes are excessive thirst, urination, and hunger. These symptoms occur because the body is producing insufficient amounts of insulin resulting in chronic hyperglycemia: a condition which if not treated can result in death (8). Type 1 diabetes has two sub-types; 1A and 1B. Type 1 diabetes 1A refers to the auto-immune driven form of the disease, which is evident by the development of autoantibodies. Type 1 diabetes 1B, also called "idiopathic diabetes," is a form of diabetes with no auto-antibodies present even though the insulin levels are drastically reduced similar to type 1A. Type 1 diabetes 1B mainly occurs in Africans or Asians and the exact mechanism of the disease is poorly understood (8). Type 1 diabetes 1B will not be discussed further here and for simplicity, type 1 diabetes will refer to the sub-type 1A hereafter.



### **Incident Rates**

Based on the Center for Disease Control and Prevention (CDC), approximately 13,000 young people are diagnosed with type 1 diabetes each year. The most recent CDC reports put the prevalence of type 1 diabetes at 1.7 per 1000 people age 0-19 in the United States (9). Type 1 diabetes is predominantly a disease that affects Europeans or populations of European descent. The highest incidence rates (age 0-14) were found in Finland and Sardinia (37/100,000), the USA rate was 16.1/100,000 and the lowest rates of <1/100,000 were found in China and Venezuela (10,11). Globally the rates of type 1 diabetes have been increasing, and the USA has seen a yearly increase of 5.3% (12). A study in Europe has predicted that by 2020 the prevalence of type 1 diabetes among individuals 15 years and younger will increase by 70% (13). Clearly the epidemic of type 1 diabetes is exacerbating and research into the underlying mechanisms is essential in understanding the disease and to help prevent future cases.

### **General Features of Diabetes**

In all forms of diabetes, the key identifying trait is hyperglycemia; a condition where there is a high glucose level in the blood stream that can cause very serious medical conditions such as ketoacidosis, kidney failure, heart disease, stroke, and blindness. Hyperglycemia occurs in all four of the classified forms of diabetes: type 1 diabetes, type 2 diabetes, gestational diabetes, and "other" (14). The World Health Organization (WHO) currently estimates that 347 million people in the world have diabetes. In 2004, ~3.4 million people died from hyperglycemia (15). The three other

types of diabetes will be briefly mentioned for contextual understanding. Type 2 diabetes is now the major form of diabetes; it encompasses 90% of all cases and occurs when an individual develops insulin resistance, further risk augments with age and obesity. Generally individuals do not need exogenous insulin for survival and disease can be controlled by diet, exercise and medications (9,16). Gestational diabetes is a form of diabetes that affects pregnant women and usually subsides after pregnancy. However, if the condition persists after pregnancy, the diabetes type needs to be re-classified and treated accordingly. The "other" category of diabetes refers to an underlying identifiable defect in the way the body responds to insulin; it is not related to autoimmune destruction of  $\beta$ -cells (8).

### **Type 1 Diabetes Environmental Variables**

There has been much investigation into "unknown environment factor/factors" that trigger the autoimmune response to type 1 diabetes. More children are diagnosed with diabetes during the winter "cold season" and less during the summer. A plausible explanation for this observation is on the account of the winter season bringing people in closer contact, thus increasing the exposure to viral infections (17,18). No definitive triggering virus strain has been found; however, some evidence hints at the genus of an enterovirus, specifically Coxsackievirus involvement in the development of type 1 diabetes (19-21). It has been reported that when migrants move from a low incidence region for type 1 diabetes to a region of higher risk their incidence rate increases as well (22,23). Various other environmental triggers have also been investigated such as

bacterial gut composition and consumption of certain foods and nutrients (12,24), or lack of exposure to immunogens as postulated in the "hygiene hypothesis". The hygiene hypothesis states that as countries become more developed (cleaner living environments), individuals are exposed to less pathogens and the immune system begins to malfunction due to the "lack of challenge," resulting in autoimmune diseases and allergies (25).

Evidence for this argument is seen in the model organisms for type 1 diabetes such as the non-obese diabetic (NOD) mouse and the inbred BioBreeding (BB) rat, which develop diabetes at a higher rate when animals are kept in a pathogen-free environment (26,27). Further evidence for this hypothesis (as previously mentioned) is seen by the increased incidence of type 1 diabetes among migrants to a region of high incidence (22,23).

However, all these detailed studies have found no conclusive evidence of the causative factor in humans. Ongoing collaborative studies such as The Environmental Determinants of Diabetes in the Young (TEDDY) (28) and Diabetes Autoimmunity Study in the Young (DAISY) (29) aim to determine environmental triggers for type 1 diabetes in children who are at risk based on genetics.

### **Pathogenesis of Type 1 Diabetes (What is known)**

It is predicted that type 1 diabetes occurs due to an accumulation of multiple immune responses that ultimately result in the destruction of the insulin producing  $\beta$ -cells. Type 1 diabetes will occur when 70-90% of the  $\beta$ -cell mass is destroyed. This cascade of immune response can occur very quickly (e.g. juvenile onset of disease) or take years to develop (12,30). Deciphering what events precede diagnosis of type 1

diabetes in humans is very complicated due to the "unknown trigger" and difficulty in examining the pancreas in living individuals. Based on examinations of pancreas from cadavers from long term type 1 diabetes patients, it appears that  $\beta$ -cell destruction in the pancreas is very slow and incomplete (31). Studies also found that during insulinitis, macrophages and  $CD8^+$  T-cells were the most abundant cell types present,  $CD4^+$  T-cells were present as well but at a lower rate (32,33). Due to the difficulty in examining the pathogenesis in humans, we turn to genetic studies to shed light on biochemical pathways leading to type 1 diabetes that can be tested in animal models such as the NOD mouse (34) and BB rat (35).

In the NOD mouse model there are four stages of disease progression; pre-insulinitis, peri-insulinitis, intra-insular insulinitis and finally complete islet destruction (30). In pre-insulinitis no damage or immune response is evident in the pancreatic islets.  $CD11c^+$  dendritic cells and  $ER-MP23^+$  macrophages are the first cells that invade the pancreas in three week old mice (30). Around the same time (or shortly after) pathogenic T-cells, majority  $CD4^+$  T-cells,  $CD8^+$  T-cells, and other immune cells are also present and have been identified surrounding the pancreas. This period of development is referred to as peri-insulinitis (34). It is postulated that these T-cells react to  $\beta$ -cells' antigens based on the high turnover of  $\beta$ -cells in the mouse islet (36). At this point, lymphocytes invade the islets in the stage referred to as intra-insular insulinitis, resulting in more inflammation preempting total islet destruction. In comparison to the animal models, the amount of insulinitis that occurs in humans is much less and incomplete (30,31,37)

In the NOD mouse, it has been found both CD4<sup>+</sup> and CD8<sup>+</sup> T-cells are needed for the development of diabetes (34). For example, when the NOD mouse does not have a functional  $\beta_2$ -microglobulin ( $\beta_2M$ ), resulting in a non-functional MHC class I molecules (CD8<sup>+</sup> T cells absent in periphery); they do not develop diabetes (38-40). When the MHC class II molecule is not expressed as in the CIITA knockout mouse (CD4<sup>+</sup> T cells absent), islet infiltration in the NOD mouse occurs but diabetes does not develop (41). However, if an autoaggressive CD8<sup>+</sup> T cell clone is introduced into NOD CIITA knockout or NOD SCID mice, diabetes will develop in the absence of CD4<sup>+</sup> T cells (41,42). Based on these findings both MHC class I & II molecules are needed in the development of diabetes. The absence of either will inhibit diabetes, but with a slightly greater emphasis on the requirement for CD8<sup>+</sup> T-cells as key to initial development.

It has also been determined that developing self-antigens of insulin is essential in developing diabetes. This finding was based on observations of antibodies against insulin appearing first in young children that develop type 1 diabetes (43,44). Mice have two insulin genes that express proteins differing by two amino acids, at position B9 and B27 (45,46). In insulin 1 knockout NOD mice diabetes was prevented, but in insulin 2 knockout NOD mice diabetes was accelerated.(46,47). The specific epitope triggering the autoimmune reaction in the NOD mice was narrowed down to amino acid sequence B9-B23 in insulin 1 (48-50).

After examining human and animal cases of type 1 diabetes, a pathway of pathogenesis of diabetes can be postulated. Some "unknown trigger" or lack thereof causes the development of autoreactive T-cells (CD4<sup>+</sup> and CD8<sup>+</sup>). It can be speculated

that CD8<sup>+</sup> T-cells are recruited first to pancreas and then after this initial wave of attack CD4<sup>+</sup> T-cells are activated. However, the end result is that T-cells invade the islets and cause  $\beta$ -cell destruction. Over time this causes more inflammation and even greater destruction which destroys an individual's ability to produce insulin and ultimately results in diabetes.

### **Important Research Tool- Islet Cell Autoantibodies (ICA)**

In type 1 diabetes, the development of autoantibodies is indicative of an autoimmune response occurring. The targets of some of the autoantibodies are insulin, glutamate decarboxylase (GAD65), insulinoma-associated antigen-2 (IA2), islet cells, and zinc transporter 8 (ZNT8) (43,51-53). The presence of one of these autoantibodies does not necessarily indicate an individual will develop diabetes because levels of autoantibodies have been found to fluctuate (43). An individual's risk for developing type 1 diabetes does increase as the number of autoantibodies increases rather than the titre of autoantibodies (43,51,52). Also antibodies appear to develop in a sequential order and as previously mentioned antibodies against insulin appear first (43,44). Autoantibodies are important for diabetes research due to the fact they are an easily measurable quantity for investigators that can be indicative of a pending autoimmune or occurring autoimmune destruction of  $\beta$ -cells in the pancreas (44). However, antibodies are not the only specific epitope targeting mechanism in the immune system and it was recently found that cytotoxic T-cells from a type 1 diabetes patient were targeting  $\beta$ -cells through a glucose regulated preproinsulin epitope (54). Future testing to determine an individual's risk for

type 1 diabetes should also include T-cells screening, although they are more difficult to measure.

### **Conclusion**

Type 1 diabetes is a complex disease that is influenced by many different factors, some known and many unknown. Currently, why diabetes occurs (destruction of insulin producing  $\beta$ -cells) is known, but what triggers it and exactly what leads up to this final precipitating event is unknown. Researchers and doctors have a few tools to try to identify the pathogenesis of diabetes by examining the presence of autoantibodies and the genetics of an individual. There are certain combinations of genetic risk factors that greatly increase an individual's likelihood of developing diabetes. But, in order to better understand the present direction of research, understanding the past history of the disease is important.

### **Work Cited**

1. Redondo MJ, L. Y, Hawa M, *et al.*: Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia* 2001;44:354-362
2. Kumar D, Gemayel NS, Deapen D, *et al.*: North-American twins with IDDM: genetic, etiological, and clinical significance of disease concordance according to age zygosity, and the interval after diagnosis in first twin. *Diabetes* 1993;42:1351-1363
3. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J: Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs. *Diabetes* 2003;52:1052-1055
4. Kyvik KO, Green A, Beck-Nielsen H: Concordance rates of insulin dependent diabetes mellitus: a population based study of Danish twins. *BMJ* 1995;311:913-917
5. Hirschhorn JN: Genetic epidemiology of type 1 diabetes. *Pediatric Diabetes* 2003;4:87-100
6. Rich SS, Concannon P: Section 1: In Search of Diabetes Genes. *Diabetes* 2002;51:S288-S294
7. Thunander M, Petersson C, Jonzon K, *et al.*: Incidence of type 1 and type 2 diabetes in adults and children in Kronoberg, Sweden. *Diabetes research and clinical practice* 2008;82:247-255
8. Kahn CR, Weir GC, King GL, *et al.*: *Joslin's Diabetes Mellitus 14th Edit.* Boston, MA, Lippincott Williams & Wilkins, 2005
9. Diabetes Public Health Resource: Children and Diabetes - More Information [article online], 2012. Available from <http://www.cdc.gov/diabetes/projects/cda2.htm>. Accessed May 20, 2013



10. Karvonen M, Viik-Kajander M, Moltchanova E, *et al.*: Incidence of Childhood Type 1 Diabetes Worldwide. 2000;23:1516-1526
11. Soltesz G, C. PC: Worldwide childhood type 1 diabetes incidence- what can we learn from epidemiology? *Pediatric Diabetes* 2007;8:6-14
12. Van Belle TL, Coppieters KT, Von Herrath MG: Type 1 Diabetes: Etiology, Immunology and Therapeutic Strategies. *Physiol Rev* 2011;91:798-118
13. Patterson CC, Dahlquist GG, Gyurus E, *et al.*: Incidence trends for childhood type 1 diabetes in *Lancet* 2009;373:2027-2033
14. (WHO) WHO: Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications- Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva, Switzerland, WHO, 1999, p. 1-66
15. Diabetes [article online], 2013. Available from <http://www.who.int/mediacentre/factsheets/fs312/en/>. Accessed April 28, 2013
16. Ahmad LA, Crandall JP: Type 2 diabetes prevention: a review. *Clinical Diabetes* 2010;28:53-59
17. Kimpimake T, Kupila A, Hamalainen AM, *et al.*: The first signs of B-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the finnish type 1 diabetes prediction and prevention study. *journal of Clinical Endocrinology & Metabolism* 2001;86:4782-4788
18. Knip M, Veijola R, Virtanen SM, *et al.*: Environmental Triggers and Determinants of Type 1 Diabetes. *Diabetes* 2005;54:S125-S136

19. Gamble DR, Path MC, Kinsley ML, *et al.*: Viral Antibodies in Diabetes Mellitus. *British Medical Journal* 1969;3:627-630
20. Clements GB, Galbraith DN, Taylor KW: Coxsackie B virus infection and onset of childhood diabetes. *The Lancet* 1995;346:221-223
21. Filippi CM, von Herrath MG: Viral trigger for type 1 diabetes: pros and cons. *Diabetes* 2008;57:2863-2871
22. Bodansky HJ, Staines A, Haigh D, Cartwright R: Evidence for an environmental effect in the aetiology of insulin dependent diabetes in a transmigratory population. *British Medical Journal* 1992;304:1020-1022
23. Burden AC, Samanta A, Chaudhuri KR: The prevalence and incidence of insulin dependent diabetes in White (W) and Indian (I) children in Leicester city (UK). *Intl J Diab Dev Countries* 1990;10:8-10
24. Akerblom HK, Knip M: Putative Environmental Factors in Type 1 Diabetes. *Diabetes Metab Rev* 1998;14:31-67
25. Okada H, Kuhn C, Feillet H, Bach JF: The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clinical and experimental immunology* 2010;160:1-9
26. Like AA, Guberski DL, Butler L: Influence of environmental viral agents on frequency and tempo of diabetes mellitus in BB/Wor Rats. *Diabetes* 1991;40:259-262
27. Wen L, Ley RE, Volchkov PY, *et al.*: Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* 2008;455:1109-1113

28. Hagopian WA, Lernmark A, Rewers MJ, *et al.*: TEDDY--The Environmental Determinants of Diabetes in the Young: an observational clinical trial. *Annals of the New York Academy of Sciences* 2006;1079:320-326
29. Stene LC, Oikarinen S, Hyoty H, *et al.*: Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). *Diabetes* 2010;59:3174-3180
30. Bluestone JA, Herold K, Eisenbarth G: Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature* 2010;464:1293-1300
31. Gianani R, Campbell-Thompson M, Sarkar SA, *et al.*: Dimorphic histopathology of long-standing childhood-onset diabetes. *Diabetologia* 2010;53:690-698
32. Bottazzo GF, Dean BM, McNally JM, *et al.*: In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulinitis. *The New England Journal of Medicine* 1985;313:353-360
33. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG: Analysis of islet inflammation in human type 1 diabetes. *Clinical and experimental immunology* 2009;155:173-181
34. Anderson MS, Bluestone JA: The NOD mouse: a model of immune dysregulation. *Annual review of immunology* 2005;23:447-485
35. Mordes JP, Bortell R, Doukas J, *et al.*: The BB/Wor rat and the balance hypothesis of autoimmunity. *Diabetes/Metabolism Reviews* 1996;12:103-109

36. Turley S, Poirot L, Hattori M, Benoist C, Mathis D: Physiological beta cell death triggers priming of self-reactive T cells by dendritic cells in a type-1 diabetes model. *J Exp Med* 2003;198:1527-1537
37. Dotta F, Censini S, van Halteren AG, *et al.*: Coxsackie B4 virus infection of beta cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:5115-5120
38. Serreze DV, Leiter EH, Christianson GJ, Greiner D, Roopenian DC: Major histocompatibility complex class I-deficient NOD-B2m<sup>null</sup> mice are diabetes and insulitis resistant. *Diabetes* 1994;43:505-509
39. Katz J, Benoist C, Mathis D: Major histocompatibility complex class I molecules are required for the development of insulitis in non-obese diabetic mice. *Eur J Immunol* 1993;23:3358-3360
40. Wicker LS, Leiter EH, Todd JA, *et al.*:  $\beta$ 2-Microglobulin-deficient NOD mice do not develop insulinitis or diabetes. *Diabetes* 1994;43:500-504
41. Mora C, Wong FS, Chang C, Flavell RA: Pancreatic infiltration but not diabetes occurs in the relative absence of MHC class II restricted CD4 T cells: studies using NOD/CIITA-deficient mice. *Journal of Immunology* 1999;162:4576-4588
42. Wong FS, Visintin I, Wen L, Flavell RA, Janeway Jr. CA: CD8 T cell clones from young nonobese diabetic (NOD) islets can transfer rapid onset of diabetes in NOD mice in the absence of CD4 cells. *J Exp Med* 1996;183:67-76

43. Ziegler AG, Hummel M, Schenker M, Bonifacio E: Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes. *Diabetes* 1999;48:460-468
44. Barker JM: Clinical review: Type 1 diabetes-associated autoimmunity: natural history, genetic associations, and screening. *The Journal of clinical endocrinology and metabolism* 2006;91:1210-1217
45. Eisenbarth GS: Chapter 3 Animal Models of Type 1 Diabetes: Genetics and Immunological Function. In *Type 1 Diabetes: Cellular, Molecular & Clinical Immunology*, Online Edition Version 3.0 ed. Eisenbarth GS, Ed. Aurora, Colorado Barbara Davis Center, 2011
46. Moriyama H, Abiru N, Paronen J, *et al.*: Evidence for a primary islet autoantigen (preproinsulin 1) for insulitis and diabetes in the nonobese diabetic mouse. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:10376-10381
47. Thébault-Baumont K, Dubois-Laforgue D, Krief P, *et al.*: Acceleration of type 1 diabetes mellitus in proinsulin 2–deficient NOD mice. *Journal of Clinical Investigation* 2003;111:851-857
48. Steck AK, Pugliese A, Eisenbarth GS: Chapter 7 Diabetes Mellitus of Man: Genetic Susceptibility and Resistance. In *Type 1 Diabetes: Cellular, Molecular & Clinical Immunology* Online Edition Version 3.0 ed. Aurora Colorado, Barbara Davis Center, 2012

- 49.** Nakayama M, Abiru N, Moriyama H, *et al.*: Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice. *Nature* 2005;435:220-223
- 50.** Nakayama M, Beilke JN, Jasinski JM, *et al.*: Priming and effector dependence on insulin B:9-23 peptide in NOD islet autoimmunity. *The Journal of clinical investigation* 2007;117:1835-1843
- 51.** Orban T, Sosenko JM, Cuthbertson D, *et al.*: Pancreatic islet autoantibodies as predictors of type 1 diabetes in the Diabetes Prevention Trial-Type 1. *Diabetes Care* 2009;32:2269-2274
- 52.** Gardner SG, Gale EA, Williams AJK, *et al.*: Progression to diabetes in relatives with islet autoantibodies. *Diabetes Care* 1999;22:2049-2054
- 53.** Wenzlau JM, Juhl K, Yu L, *et al.*: The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:17040-17045
- 54.** Skowera A, Ellis RJ, Varela-Calvino R, *et al.*: CTLs are targeted to kill  $\beta$ -cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. *Journal of Clinical Investigation* 2008;118:3390-3402

## **Chapter 2**

Brief History of Diabetes:

## **Introduction**

Diabetes is a disease that has been documented for thousands of years. Many of the big breakthroughs for understanding diabetes did not occur until the last one hundred years, but there were numerous small steps that led up to these important discoveries. To understand what led up to the current Genome Wide Association Studies (GWAS) that have revolutionized the field, understanding the discoveries that came before is essential to fully appreciate the state of the field today.

## **Ancient Times---"To Pass Through"**

In the Ebers Papyrus, an ancient Egyptian text from ~1500 B.C.E., a disease (presumed to be diabetes) is referenced that causes "great emptying of the urine" (1) and excessive thirst (2). The term "diabetes," a literal translation from Greek as "to pass through," was first used by the Apolonius of Memphis (230 B.C.E.) (1). The naming most likely referred to the continuous urination and wasting away of affected individuals (3). The Greek physician Aretaeus (from 80-138 A.C.E. Cappadocia) noted the sweet taste of diabetic urine (2) in his work "On the Causes and Indications of Acute and Chronic Diseases," (1) an exhaustive analysis of the disease.

Diabetes is a dreadful affliction, not very frequent among men, being a melting down of the flesh and limbs into urine. The patients never stop making water and the flow is incessant, like the opening of the aqueducts...The patients are affected by nausea, restlessness and burning thirst, and within a short time they expire. (1,4)

Numerous allusions to diabetes are referenced in cultures of India, China and Middle East (1,3); for instance in 5<sup>th</sup> century India, physicians Sushruta and Charaka noted the appearance of two types of diabetes: one afflicting younger individuals and the other



affecting the heavier and older (1). Furthermore in 7<sup>th</sup> century China, physician Li Hsuan observed the susceptibility of diabetic patients to boils and lung infections. Although misleadingly concluding as treatment, an abstinence from sex and wine (1,3), these original observations lay the foundation for the discoveries of the 18<sup>th</sup> and 19<sup>th</sup> century.

### **Diabetes "Mellitus"**

In 1670 Thomas Willis re-discovered the “sweet” tasting of urine in affected individuals (1). This “sweet” tasting compound was identified as sugar in 1776 by Matthew Dobson. This was deduced by procuring urine from a diabetic patient and performing various experiments on it: letting sit, boiling, evaporating and mixing with mineral acids.

Two quarts of this urine were, by a gentle, evaporated to dryness...There remained, after the evaporation, a white cake which weighed 3iv. 3ij. and 3ij. This cake was granulated, and broke easily between the fingers ; it smelled sweet like brown sugar. (5)

Matthew Dobson also noted that the level of glucose was elevated in the urine of patients with the disease and that the blood of patients was sweeter too (1,5). In 1788 Thomas Cawley, a surgeon, suggested there was connection between diabetes and the pancreas after he noted individuals that sustained pancreatic injuries developed diabetes (6,7).

John Rollo (1749-1809), a surgeon general in the British Army, was the first person to use the phrase diabetes “mellitus,” (Latin and Greek for "honey") when referring to diabetes. This was done to differentiate the two types of “diabetes” identified at the time; diabetes mellitus (sweet) and diabetes insipidus (tasteless) (1,3). Diabetes insipidus refers to a disease that manifests itself as excessive urination due to hormonal imbalance of vasopressin and is not at all related to diabetes mellitus (8). John Rollo with

assistance from William Cruickshank monitored the amount of sugar excreted by a Captain Meredith who had diabetes and severe glycosuria (sugary urine) (4). They monitored the amount of sugar in his urine by boiling it and weighing the sugar residue leftover. They also meticulously recorded the types of foods he ate. They noticed that breads, grains and fruits increased the sugar content in the urine and meats lowered it. After these observations they concluded the stomach was overproducing sugar from certain foods and recommended a diet low in carbohydrates and high in fat and protein. (4,9). This recommendation became the standard for future treatments of diabetes until the discovery of insulin (4). It should be noted for the treatment of type 2 diabetes this method worked well, but for type 1 diabetes it was lethal due to starvation (1).

### **Discovery of Insulin**

The next big breakthrough in diabetes did not come until 1889 when Joseph von Mering and Oskar Minkowski found that by removing the pancreas of dogs they could trigger a diabetic state (excessive urination, intense thirst, hunger and weight lost), resulting in death of the dogs. However, their attempts to isolate this pancreatic compound failed (2,7). Jean de Meyer in 1910 postulated that this pancreatic chemical that was not being produced in diabetic patients be called "insulin." The term was derived from "insula" which means island in Latin referencing to its origin in the "islets of Langerhans" (10).

### Diabetes

Ligate pancreatic ducts of the dog. Keep dog alive till acini degenerate leaving Islets. Try to isolate the internal secretions of these to relieve glycosurea

Frederick Grant Banting's lab notebook October 31, 1920, University of Toronto, Canada (1,11)

In 1921 Frederick Banting and graduate student Charles Best, utilizing John MacLeod's laboratory, began experiments into extracting pancreatic hormones from dogs with ligated pancreas and re-injecting the extract into dogs with removed pancreas (diabetic dogs). Banting and Best's experiments were effective in relieving diabetic symptoms, prolonging the diabetic dogs' lives. Later experiments used pancreas from fetal calves (no proteolytic enzymes present and were more readily available) and later adult bovines' pancreas when extraction techniques improved (12). The finding that from these pancreas extracts prolonged a diabetic dog's life from fourteen days to seventy days and prompted further experimentation. Banting enlisted James Collips, a biochemist, to help further purify the compound they called "isletin" (1,4,12). The compound "isletin" was later renamed by MacLeod to "insulin" to recognize de Meyer's original naming, which was not known at the time (1).

Human diabetic patients were first tested in January 1922, with this compound called "insulin" with very positive results: immediate effects on lowering blood glucose levels (12). Leonard Thompson, a fourteen year old boy, was the first diabetic patient tested and his life was prolonged 13 years with ongoing insulin therapy. He died at the age of 27 from pneumonia (1).

Interestingly, the group then had difficulties isolating the compound again, but with assistance from a small company named Eli Lilly, the compound was successfully

isolated and produced on a large scale. The breakthrough in large scale production came from the observation of a chemist, George Walden, at Eli Lilly who noticed maintaining insulin at its isoelectric point (no charge) could allow for maximum isolation and extraction from beef or pork pancreas (4,13). In 1923 Banting and Macleod received the Nobel Prize for discovering insulin. (4).

Insulin immediately became a "wonder" drug, drastically changing the lives of diabetics who would have certainly faced premature deaths without treatment (4,14). In the years following the discovery of insulin, breakthroughs occurred in improved production, knowledge of its structure and variations in the types used (4). Even though a successful treatment for diabetes was found, a reclassification of the disease itself was needed to better understand and study the mechanism underlying diabetes.

### **Two Diseases: Type 1 and Type 2**

Over the course of years after insulin was discovered various classification schemes were adopted to describe diabetes. Prior to 1976 the classification of diabetes was based on phenomenological observations and was extremely confusing (15). The terms juvenile onset and maturity onset diabetes were also in popular usage. These terms would be replaced by insulin dependent diabetes (IDDM) and non-insulin-dependent diabetes mellitus to make the classification less dependent on age at onset and more representative of the disease mechanism (15). In 1976 Andrew Cudworth reintroduced the terms type 1 and type 2 diabetes (John Lister had previously used these designations in 1951) (15,16). This time the designations caught on and in 1979 these terms were

officially adopted and standardized by the National Diabetes Data Group Committee to the present day usage: type 1 diabetes (previously IDDM or juvenile onset diabetes) and type 2 diabetes (NIDDM or adult onset diabetes) (4). Interestingly H.P Himsworth originally postulated in 1939 that diabetes had two types: lack of insulin or the resistance to insulin, however it took 40 years for this idea to finally catch on (4).

### **1970 - Present**

Since 1979 there have been updates to the official classification of diabetes to include rarer forms (4). As diabetes was in the midst of re-classification, the scientific field was changing rapidly as well, laying the foundation for a better understanding of type 1 diabetes. The first genetic locus identified to be involved in type 1 diabetes was the HLA region in 1973 in a candidate gene study through serotyping (17). This was followed by a candidate gene association study that identified insulin in 1984(18), a linkage study identifying *CTLA-4* in 1996 (19) and in more recent years *PTPN22* (20) and *IL2RA* (21) were identified in candidate gene association studies.

In the 1990's and early 2000's genome wide linkage studies were performed using microsatellite markers. These relatively infrequent markers were statistically underpowered for detecting new genes affecting type 1 diabetes and only could re-validate known regions (22). Microsatellite markers were replaced with single nucleotide polymorphisms (SNPs) to investigate disease association with much deeper coverage on a greater genomic scale (although at moderate effect sizes) (23-25). The first comprehensive Genome Wide Association Study (GWAS) for type 1 diabetes using

SNPs was reported in 2007 and completely changed the field (26,27). The major genetic findings from these studies for type 1 diabetes will be discussed to a greater extent in chapter three.

### **Work Cited**

1. Poretsky LE: *Principles of Diabetes Mellitus 2nd Edit.* New York, NY, Springer, 2010
2. Polonsky KS: The past 200 years in diabetes. *N Engl J Med* 2012;367:1332-1340
3. Lasker SP, McLachlan CS, Wang L, Ali SMK, Jelinek HF: Discovery, treatment and management of diabetes. *Journal of Diabetology* 2010;1:1-8
4. Kahn CR, Weir GC, King GL, *et al.*: *Joslin's Diabetes Mellitus 14th Edit.* Boston, MA, Lippincott Williams & Wilkins, 2005
5. Dobson M: Nature of the urine in diabetes. *Medical observations and enquiries* 1776;5:298-316
6. Medvei VC: *The History of Clinical Endocrinology.* NY, New York, Parthenon Publishing Group Limited, 1993
7. Minkowski O: Historical development of the theory of pancreatic diabetes. *Diabetes* 1989 (1928);38:1-6
8. Saborio P, Tipton GA, Chan JCM: Diabetes Insipidus. *Pediatrics in Review* 2000;21:122-129
9. Rollo J, Cruickshank W: An account of two case of Diabetes mellitus: the nature of the disease. *C Dilly* 1797;1:1
10. Rosenfeld L: Insulin: Discovery and Controversy. *Clinical Chemistry* 2002;48:2270-2288
11. Bliss M: *The Discovery of Insulin: Twenty-fifth Anniversary Edition.* Chicago, IL, University of Chicago Press, 2007
12. Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA: Pancreatic extracts in the treatment of diabetes mellitus: preliminary report, 1922. *CMAJ* 1991;145:1281-1286

13. Madison JH: *Eli Lily: A Life, 1885 - 1977 (2nd Edition)*. Indiana Historical Society, 2006
14. Geyelin HR, Harrop G, Murray MF, Corwin E: The use of insulin in juvenile diabetes. *Journal of Metabolic Research* 1923;2:767-791
15. Edwin EA: Perspectives in Diabetes: The discovery of type 1 diabetes. *Diabetes* 2001;50:217-226
16. Gale EA: Review: Fifty years of type 1 diabetes. *The British Journal of Diabetes & Vascular Disease* 2002;2:441-445
17. Singal DP, Blajchman MA: Histocompatibility (HLA-A) Antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 1973;22:429-432
18. Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 1984;33:176-183
19. Nistico L, Buzzetti R, Pritchard LE, *et al.*: The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Human molecular genetics* 1996;5:1075-1080
20. Bottini N, Musumeci L, Alonso A, *et al.*: A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature genetics* 2004;36:337-338
21. Vella A, Cooper JD, Lowe CE, *et al.*: Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *American journal of human genetics* 2005;76:773-779



- 22.** Concannon P, Erlich HA, Julier C, *et al.*: Type 1 Diabetes Eividence for Susceptibility Loci from Four Genome-Wide Linkage Scans in 1,435 Multiplex Families. *Diabetes* 2005;54:2995-3001
- 23.** Livak KJ, Marmaro J, Todd JA: Towards fully automated genome-wide polymorphism screening. *Nature genetics* 1995;9:341-342
- 24.** Wang DG: Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science* 1998;280:1077-1082
- 25.** Kruglyak L: The use of genetic map of biallelic markers in linkage studies. *Nature genetics* 1997;17:21-24
- 26.** Consortium WTCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-678
- 27.** Hakonarson H, Grant SF, Bradfield JP, *et al.*: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007;448:591-594

## **Chapter 3**

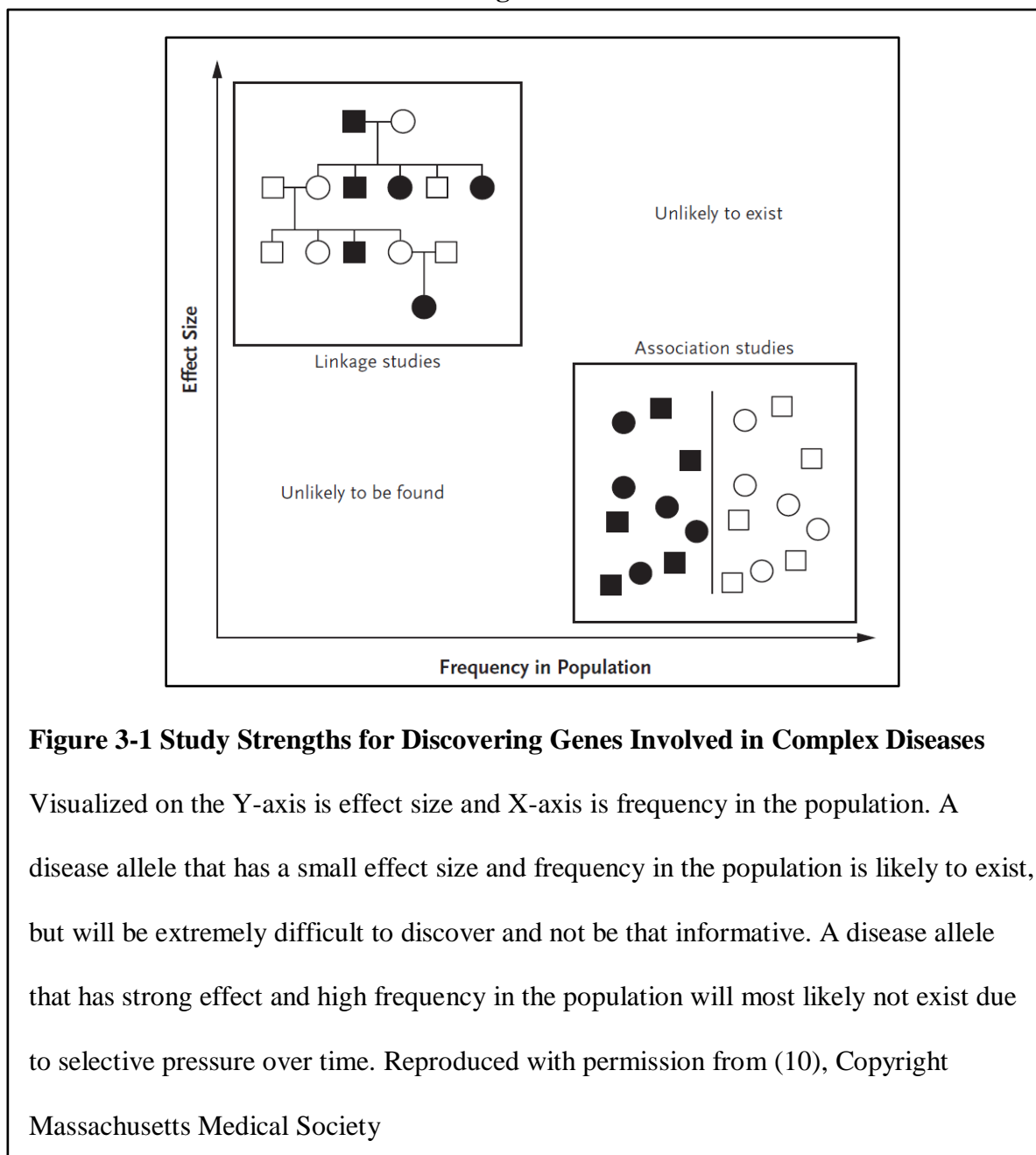
### The Genetics of Type 1 Diabetes

## **Introduction**

Type 1 diabetes is considered a complex disease that arises due to the effects of both genetic and environmental factors evident by <100% concordance between monozygotic twins (1-6). The most prevalent form of type 1 diabetes is thought to occur due the interaction of multiple genes (polygenic) and does not appear to show any clear pattern of inheritance though it does aggregate in families (6,7). It should be noted there are less common forms of type 1 diabetes that are monogenic and are immune-mediated, but these types will not be discussed further (8,9). Due to the ambiguity about the pathogenesis of type 1 diabetes and the exact genes involved, genetic studies have been performed that make few assumptions about the disease (model-free). Two study approaches that are commonly utilized in studying type 1 diabetes are linkage studies and association studies (10).

In linkage studies, affected related individuals (such as siblings) are investigated for identical chromosomal regions that are shared more frequently than expected by Mendelian inheritance. Linkage of the marker and the locus that confers susceptibility for the disease is identified by the accumulation of data across numerous families investigated. In general, linkage studies work well when the overall effect of the risk allele is large and the overall frequency of the allele is low in a population (Figure 3-1). However, linkage studies implicate regions of DNA that are many megabases long, which then require intensive investigation using fine mapping techniques to identify the causative risk variant (10,11). In association studies, case (affected) and control (unaffected or general population) groups are compared based on frequencies of specific

marker alleles within candidate genes, pathways or chromosomal regions. Association studies work well at discovering alleles that have a moderate to high frequency in the population, even if they have a small effect size (Figure 3-1) (10). An issue with association studies is carefully matching the control population for comparison; however, this can be avoided by utilizing the transmission disequilibrium test (TDT) and an affected sib-pair population. The TDT test investigates the number of times a marker allele is transmitted and not transmitted in the cases population (11,12). Association studies take advantage of linkage disequilibrium (LD) in the human genome, exploiting the situation where an allele (testable polymorphic marker) is in linkage disequilibrium with a causative variant. LD is the situation where specific allelic combinations that are relatively close to each other are recombined at a lower frequency than would be predicted (11). Linkage disequilibrium can span from a few hundred bases up to a few hundred kilobases, making identification of the causative variant much simpler than linkage study. (10).

**Figure 3-1**

### **Genetic Linkage Studies**

The first genome wide genetic linkage study for type 1 diabetes was performed in 1994 (13). This study reported linkage with human leukocyte antigen (HLA), insulin and 18 other regions. This was followed up with other genome wide linkage studies in different populations in 1998 (14,15), 2001 (16) and ultimately ended with a comprehensive study that combined all prior studies in 2005 (17). These linkage studies further validated the HLA region as being the major risk locus for type 1 diabetes (10). Many of the other regions identified in linkage studies have been inconsistently seen because they are either false positives or when combining many different families, population heterogeneity is masking the true effects (10). However, a small scale candidate gene linkage study did identify *CTLA-4* gene in 1996 as a risk for type 1 diabetes (18).

### **Candidate-Gene Association Studies**

As previously mentioned, the first gene region identified to be involved with type 1 diabetes was HLA in 1973 (19) through a candidate gene study using serotyping. The next couple of genes discovered were *INS* insulin in 1984 (20), *PTPN22* in 2004 (21) and *IL2RA* in 2005 (22). However, candidate gene association studies require prior knowledge about a gene or region. The field of association studies was completely revolutionized by the development of genome wide association studies (GWAS) which utilized tagging single nucleotide polymorphisms (SNPs) for high-throughput screening of the entire genome and requiring no prior knowledge of a region to implicate it (23-25).

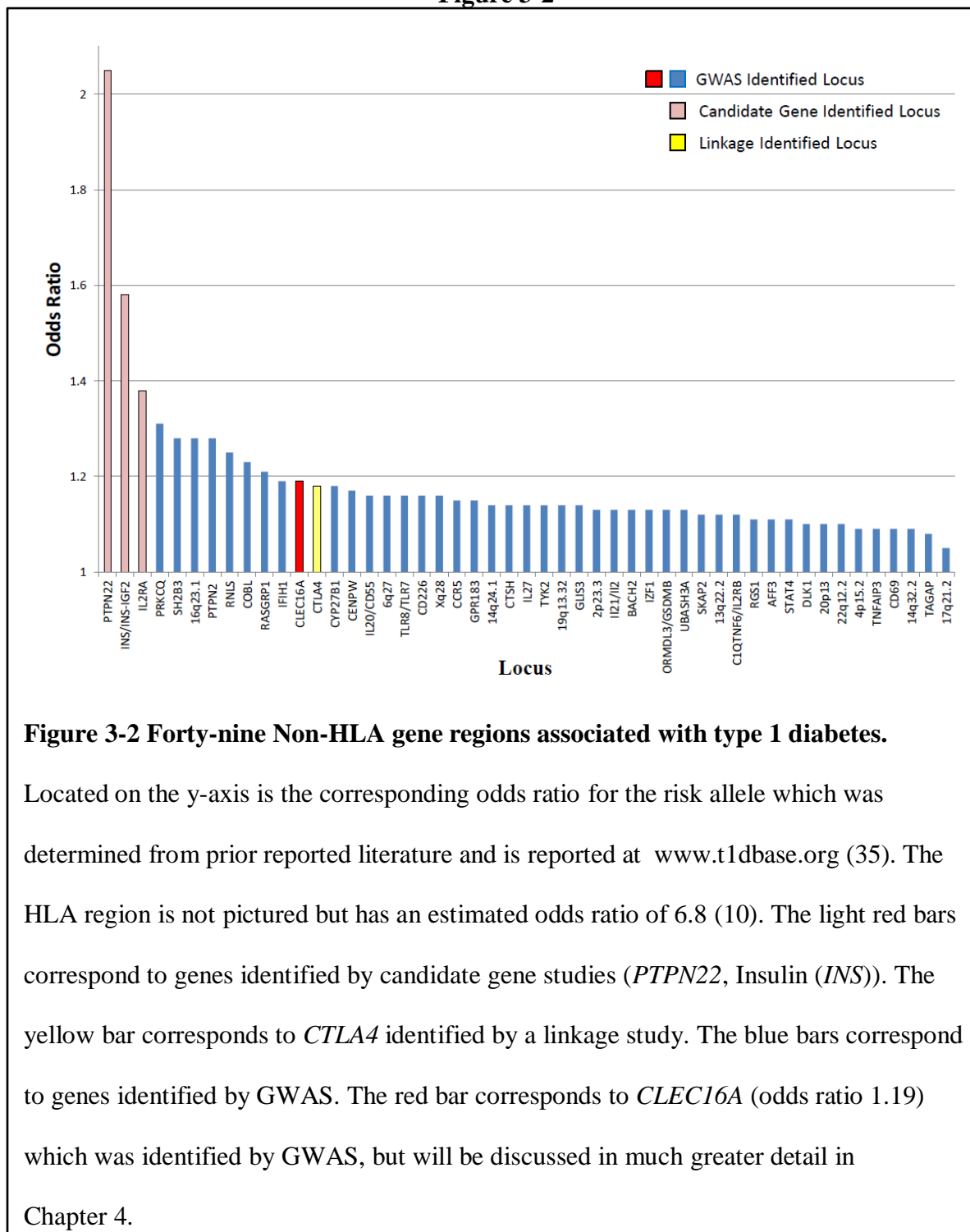
GWAS became feasible due to the availability of fine resolution SNP maps of the whole genome. In addition, knowledge of the linkage disequilibrium patterns in human populations decreased the number of SNPs that needed to be genotyped in order to cover the genome.

### **GWAS**

One of the first GWAS was performed in 2005 on patients with age related macular degeneration (AMD) (26). The first type 1 diabetes GWAS, performed in 2006 using non-synonymous SNPs, identified *IFIH1* in association, however SNP coverage and statistical power was very poor (27). The first comprehensive GWAS that interrogated the entire human genome for type 1 diabetes was performed in 2007 (28). Since that initial study another GWAS was performed with much greater SNP coverage (29) along with a comprehensive meta-analyses (30,31) and numerous follow-up studies that interrogated the identified regions more thoroughly (32-34). GWAS studies have greatly increased the number of type 1 diabetes loci to over fifty reported loci that confer low to high risks. In Figure 3-2, the odds ratios for forty-nine of these loci that are most significantly associated with type 1 diabetes are graphed (note: HLA is excluded but confers an odds ratio of 6.8) (10,35). An odds ratio of one means the event can occur with equal likelihood in both the affected group and controls. An odds ratio greater than one means the event has a greater likelihood to occur in the affected group (36). An odds ratio less than one means the event is less likely to occur in the affected group. When

analyzing alleles (major/minor) switching to the risk allele allows for correction of the odds ratio, so that all values are greater than one.



**Figure 3-2**

GWAS have helped re-confirm known loci such as *HLA*, *INS*, *PTPN22*, *CTLA4* and *IL2RA* whose specific genetic involvement in type 1 diabetes has been elucidated to some extent by prior studies. The current dilemma is that instead of having a shortage of additional loci to investigate (problem in the 1990's), there are now a multitude of possible candidate regions. To better focus time and resources, loci that are involved in specific biochemical pathways or shown to overlap with other autoimmune disorders are currently being prioritized for further interrogation.

For example, the initial non-synonymous GWAS study identified *IFIH1* gene in association with type 1 diabetes (27,34). *IFIH1* codes for the protein melanoma differentiation-associated gene 5 (mda5), which is involved in the recognition of double stranded RNA (dsRNA) from picornaviruses (37,38). This feature of mda5 is exciting because, as previously mentioned, one of the possible environmental triggers of type 1 diabetes is suspected to be a virus. Specifically, it has been hypothesized that the Cocksackievirus B4, a virus in the picornavirus family, is involved in type 1 diabetes (39-41). The SNP rs1990760 in *IFIH1* codes for a non-synonymous mutation of an alanine to threonine (A946T) (27). Recently it was found that the SNP was associated with different frequencies of enterovirus RNA in the blood of healthy children, with heterozygous individuals having the highest amount (42).

Another way to investigate these new loci is to determine whether any of these regions overlap with other autoimmune related disease associated regions, which may hint at a common pathogenic mechanism. For example SNPs in the *CLEC16A* region were identified in association with both type 1 diabetes (28) and multiple sclerosis (43).

The association with type 1 diabetes risk maps to a locus that is composed of four genes; *CIITA*, *DEXI*, *CLEC16A* and *SOCS1*. Two of these genes are good candidates for an immune related disorder; *CIITA* is a positive regulator of MHC class II molecules (44) and *SOCS1* is involved in the negative regulation of cytokines (45). The other two genes in this region; *DEXI* and *CLEC16A* currently have unknown functions, making this region even more perplexing and important to decipher. Further exploration of this entire region will be the focus of chapter 4.

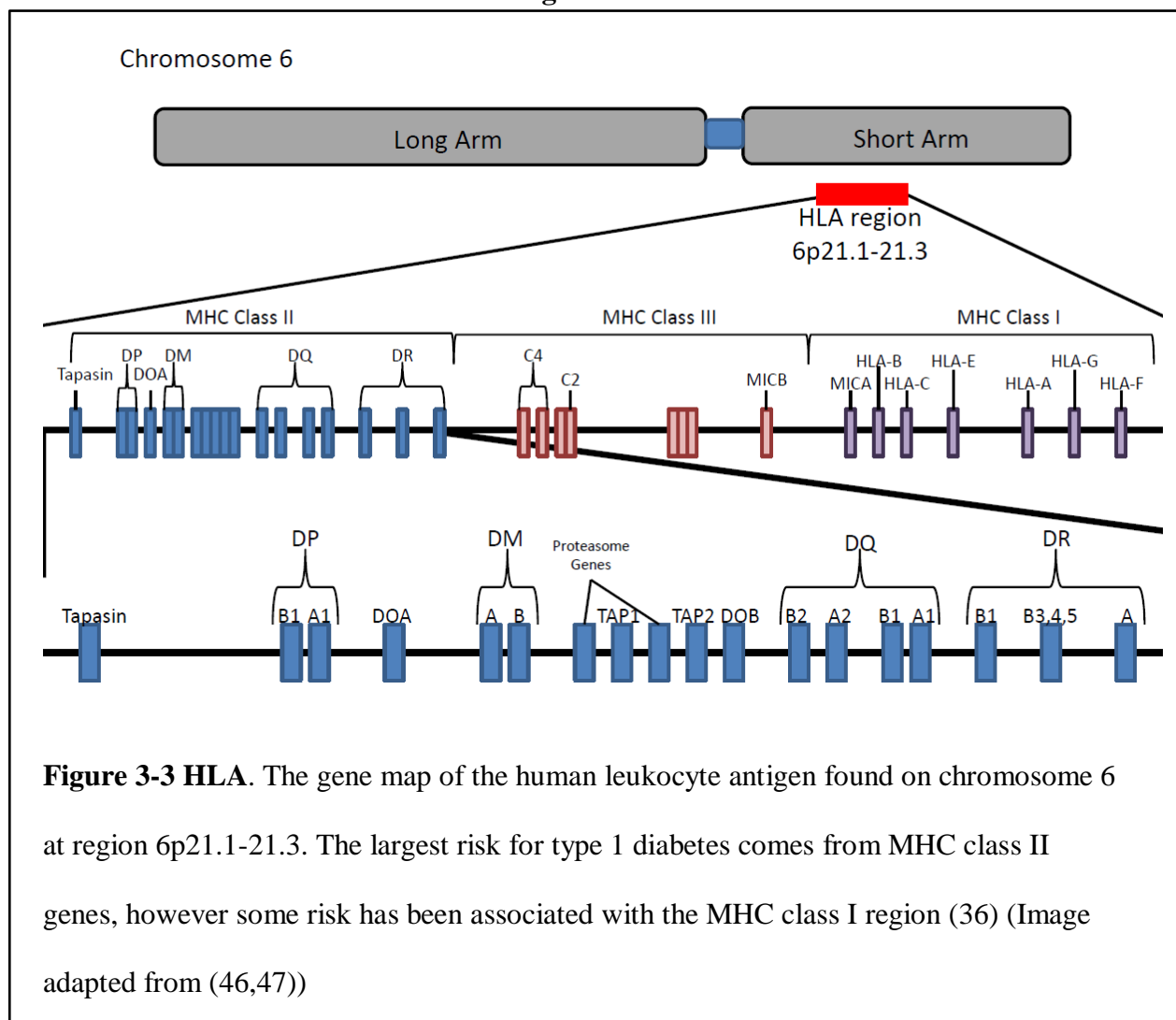
Clearly understanding which of these loci truly affect type 1 diabetes is important in understanding the disease. The next section explores how *HLA*, insulin, *CTLA-4*, *PTPN22* and *IL2RA* confer genetic risk in the disease.

### **HLA (IDDM1)**

The Human Leukocyte Antigen (HLA) region was the first region identified in association with type 1 diabetes and confers the largest risk. As shown in Figure 3-2, the odds ratio for the association of the HLA region with type 1 diabetes is 6.8, which is significantly larger than any other locus (odds ratios usually between 1.2 - 1.3, excluding *PTPN22* and insulin) (10). The HLA region located on chromosome 6 at the region 6p21.1-21.3 codes for three classes of major histocompatibility complex (MHC) molecules (MHC class I, MHC class II, and MHC class III), with each class consisting of several genes (Figure 3-3). MHC class I molecules are expressed on all nucleated cells and present antigens derived from the cytoplasmic proteins and activate CD8<sup>+</sup> T cells against abnormal cells. MHC class II molecules are expressed on certain antigen

presenting cells (APC), such as macrophages and dendritic cells that present external antigens to CD4<sup>+</sup> T cells. CD4<sup>+</sup> T cells activated by specific peptides in the context of APCs are involved in B cell proliferation and antibody production. MHC class II molecules are also involved in antigen processing such as tapasin, DM, DO and proteasome. MHC class III molecules code for molecules involved in other immune related processes like complement, cytokines and heat shock proteins (46).

**Figure 3-3**



The nomenclature for the region starts with a four character alphanumeric code which refers to the locus, followed by an asterisk and four digits which refer to the allele at the locus. For example in DQB1\*0302, the 0302 refers to the allele at the highly polymorphic locus DQB1 which is located in MHC class II (48).

The HLA region was first implicated in association with type 1 diabetes in the 1970's through a candidate gene study that was based on prior studies implicating HLA in autoimmune disorders. Originally the MHC class I antigens (B8 and B15) were reported to be in association with type 1 diabetes (19,49,50). Much stronger associations were then found with the MHC class II molecules HLA-DR and HLA-DQ (36,51-56). The prior association with B8 and B15 was detected due to high linkage disequilibrium among MHC molecules (B8 and DR3, and B15 and DR4) (51,57,58). However, MHC class I molecules do confer some risk in type 1 diabetes (discussed later).

Since the original association of MHC class II with type 1 diabetes was discovered, studies have shown a hierarchy of gene/haplotype combinations within MHC II that range from protective to high risk for development of type 1 diabetes (59,60), differing slightly among different ethnicities (61-65). For example in Caucasians of European descent the haplotype DRB1\*1501-DQA1\*0102-DQB1\*0602 was found to be strongly protective and occurring in ~20% of the population, but is only seen in ~1% of diabetes patients (60). The highest susceptibility for type 1 diabetes in Caucasians is found in individuals who are heterozygous for the haplotypes DR3/DR4-DQ8 (note: this can also be written as DR3/DR4-DQ2/DQ8 or DR3/DR4) (36,60,66). Currently 30-50% of new type 1 diabetes cases have the DR3/DR4-DQ8 genotype, which only occurs in

2.4% of the general population (67). One hypothesis for the increased risk from the haplotypes is that slight changes in the amino acid codes change how the MHC II molecules present antigen peptides that favor T-cell activation and result in the attack against self antigens in the pancreatic  $\beta$  cells (60,66).

It would appear MHC class II molecules confer the largest risk in development of type 1 diabetes, however the MHC class I molecules have been re-implicated in association with type 1 diabetes, but to a much lesser extent than MHC class II molecules (68-73). After adjusting for the strong association of MHC class II molecules, MHC class I molecules HLA-B and HLA-A genes were identified to confer a risk in disease (74). The HLA-B and HLA-A genes appear to modify the severity of an individual's disease risk; with certain alleles being associated with age of diagnosis and overall risk of developing type 1 diabetes (67,75-79).

On the functional side there is evidence that  $CD8^+$  T-cells, which are activated by APC presenting MHC class I molecules, drive the progression of the  $\beta$ -cell destruction in the pancreas (80,81). This finding was supported by studies showing that MHC I molecules (allele HLA-A\*02) loaded with beta cell antigens react strongly with  $CD8^+$  cytotoxic T-cells (CTLs) from type 1 diabetes patients' peripheral blood (82-84). The selection of HLA-A\*02 allele (also written as HLA-A2 allele) was based on the fact that it is found in >60% of type 1 diabetes patients (67,69). This allele is also correlated with increased risk of developing type 1 diabetes when found in individuals with the high risk haplotype DR3/DR4-DQ8 (67,69). In transgenic NOD mice (mouse model of type 1 diabetes) expressing HLA-A\*02, the onset of diabetes was accelerated (85). These

studies support a role for MHC class I molecules in risk of type 1 diabetes, adding to the risk conferred by MHC class II molecules (36).

### **Insulin (*INS*) (IDDM2)**

In 1984 a polymorphism located in the 5' region of the insulin gene was identified in a candidate gene study to be associated with type 1 diabetes in a Caucasian population (20). The association was with a "variable number of tandem repeats" (VNTR) region located in the promoter of insulin (86-89). The VNTR region could be classified into three different groups based on the size of the VNTR region (20,90,91). Individuals with the shortest VNTR (class I) were found to be at the highest risk for type 1 diabetes and individuals with the longest VNTR (class III) were protected from type 1 diabetes (20,92,93). The medium length VNTR (class II) rarely occurs in Caucasians (20,91). It was found that the short VNTR (class I) risk allele decreased expression of insulin in the thymus and increased expression in the peripheral tissue, while the protective, longest VNTR (class III) showed higher expression in the thymus and lower expression in the peripheral tissue (66,86,93-95). It is hypothesized that individuals with the risk associated "short" class I VNTR have less expression of insulin in the thymus and therefore negative selection of autoreactive T cells for insulin is less efficient. Whereas in the "long" class III VNTR individuals the selection is more efficient, preventing the production of autoreactive T-cells for insulin (94,95).

### **CTLA-4 (IDDM12)**

Cytotoxic T lymphocyte-associated protein 4 (*CTLA-4*) gene was first identified as a susceptibility locus for type 1 diabetes in a 1996 linkage study (18) and was confirmed in a multi-ethnic association study in 1997 (96). *CTLA-4* is involved in negative regulation of T-cells. The exact mechanism of how it regulates T-cells is unknown, but it is speculated that it functions by directly binding B7 (on APC cells) and competitively blocking CD28 from binding, or it binds B7 and sends an inhibitory signal (phosphatases) damping the TCR and CD28 signal, blocking T-cell activation (46,97). When the *CTLA-4* gene was knocked out in a mouse it resulted in severe lymphoproliferation (over production of immune cells) that resulted in early lethality (98). This finding hints that *CTLA-4* is an essential molecule in the negative regulation of the immune system. The risk associated with *CTLA-4* has also been replicated in other autoimmune related diseases like multiple sclerosis (MS) (99), systemic lupus erythematosus (SLE) (100), rheumatoid arthritis (101), Grave's disease (102) and Addison's Disease (103). Currently three polymorphism types have been reported that could influence *CTLA-4*: SNPs in the promoter of *CTLA-4*, a non-synonymous SNP in exon 1, and microsatellite repeats in the 3'UTR (97,104-106). All three polymorphisms are in linkage disequilibrium with each other and deciphering the true causative variant is difficult because the three variants almost always co-occur (107), though a meta-analysis from 2005 suggests the SNP in exon 1 is the true risk variant (108). One of the SNPs in the promoter region of *CTLA-4* (318 C→T) was found to increase promoter activity and increase *CTLA-4* expression (104). The exon 1 polymorphism (SNP rs231775 49A→G)



results in a non-synonymous amino acid change from threonine to alanine at residue 17 (T17A). In-vitro, the non-synonymous variant, resulted in CTLA-4 being processed incorrectly in the endoplasmic reticulum resulting in less surface expression of the protein (105). In a study of myasthenia gravis patients (autoimmune neuromuscular disease) the 3'UTR polymorphism (longer AT)<sub>n</sub> was found to affect mRNA stability and reduced CTLA-4 expression (106). The risk variants appear to reduce the amount of CTLA-4 expressed (either through reduced mRNA expression or reduction of final protein expression) increasing T-cell activation and increasing the risk of autoimmunity. However, further investigation is necessary due to conflicting reports for both effects on expression and for association with autoimmune disease.

### **PTPN22**

A SNP (rs2476601) in the *PTPN22* gene, which produces the lymphoid protein tyrosine phosphatase (LYP), was originally identified in a candidate gene study in 2004 to be associated with type 1 diabetes (21,109). *PTPN22* (LYP) forms a complex with C-terminal Src kinase (CSK) to negatively regulate the T-cell receptor (TCR) (110). LYP functions by dephosphorylating the proteins involved in the TCR signaling cascade like Lck (Src family kinase), the ITAMs of TCR $\zeta$ /CD3 complex, Zap70, Vav and valosin (111).

The SNP (rs2476601) is located in a coding exon (1858C→T) and results in non-synonymous amino acid change (arginine to tryptophan) at residue 620 in the LYP protein (R620W) (21,29). This same SNP has also been confirmed to be associated with

numerous other autoimmune diseases like rheumatoid arthritis and systemic lupus erythematosus (110,112,113). It has been found that the variant allele disrupts the interaction of LYP and CSK (21). Currently there are two theories (loss of function/gain of function) that predict how the mutations could affect function of a T-cell with evidence that supports both models (36). If the mutation causes a loss of function, it could mean that autoreactive T-cell in the periphery could become more active (hyperactivity) and cause more damage. This loss of function hypothesis is consistent with a mouse model where the mouse orthologue of LYP, PEST-enriched tyrosine phosphatase (PEP) (110), was knocked out and an expanded T-cell population occurs (114). If the mutation causes a gain of function, it could mean that T-cell activation is reduced which could allow autoreactive T-cells to escape negative selection in the thymus (110). This gain of function model was seen in T cells from heterozygous carriers of the allele who had less phosphorylated proteins targeted by LYP and also showed less T-cell signaling (115,116). Interestingly, there also appears to be a dose dependent factor with the *PTPN22*'s SNP rs2476601 risk allele (T) in some autoimmune diseases, meaning the number of copies determines overall risk (10,109,117,118). Also, the polymorphism R620W appears to affect B cells. The polymorphism affected B cell receptor (BCR) signaling (signal reduction) and also favored production of autoreactive B cells and antibodies (66,116,119,120).

### **IL2RA**

The interleukin (IL)-2 receptor  $\alpha$  gene (*IL2RA*) was identified in type 1 diabetes candidate gene association studies in 2005 (22) and was re-confirmed and fine-mapped in numerous follow up studies (28,121,122). The region has also been found in association with multiple sclerosis (123). The *IL2RA*, also known as CD25, is one of three non-covalently associated proteins that make up the interleukin-2 receptor. The interleukin-2 receptor is composed of *IL2RA*, *IL-2/15R $\beta$*  and  $\gamma_c$  (46). *IL2RA* is the only subunit that is unique to the interleukin-2 receptor, the other two subunits are found in other cytokine receptors.

When cytokine interleukin-2 (IL-2) is released by a T cell, it binds to the *IL-2R $\alpha\beta\gamma_c$*  complex and signal transduction occurs through *IL-2/15R $\beta$*  and  $\gamma_c$  to activate the Jak3-STAT5 signal transduction pathway (46,124). All three subunits are required for high affinity binding of interleukin-2 (124). Interleukin-2 is important for T-cell proliferation and survival and is essential for Tregs (T regulatory cells) to survive (124). Multiple sclerosis patients have an increased amount of soluble *IL2RA* (s*IL2RA*) in circulation (125). The levels of s*IL2RA* in multiple sclerosis patients correlate with certain SNP variants and disease severity (122,123,126). An explanation for this observation is that s*IL2RA* can bind to circulating IL-2 and impair Treg function and disease variants in multiple sclerosis appear to increase s*IL2RA* levels (36). However, in type 1 diabetes patients the risk variants correlate with lower levels of s*IL2RA* (122,123). This lower level of s*IL2RA* in type 1 diabetes patients was also reported in in-vitro stimulated peripheral blood mononuclear cells (PBMCs) (127). Interestingly, normal

levels of Tregs can be found in type 1 diabetes patients (128), however, their ability to suppress effector T-cells is inhibited(129). A possible explanation is a risk polymorphism in the IL2RA receptor causes a defective receptor causing lower expression of the receptor and/or poorer IL-2 signaling in Tregs, which ultimately results in inhibited Treg function. Supporting this idea are two different studies performed in healthy individuals. In one study individuals with the risk haplotype have lower IL2RA levels (130). Whereas in another study they found individuals with the risk haplotype have reduced interleukin-2 signaling in memory T cells and Tregs and a reduction in Treg ability to suppress effector T-cells (131). Clearly, *IL2RA* is important in T-cell regulation and different autoimmune disease loci affect *IL2RA* in different ways. However, both pathways appear to result in less interleukin-2 being recognized by T-cells (defective production of IL2RA or higher levels of sIL2RA) which drastically affects Tregs and inhibits their ability to function properly (36).

## **Conclusion**

Deciphering the risks involved in type 1 diabetes will take time and resources. In the last couple of years, major strides have been made in identifying multiple genetic risk regions in addition to those identified prior to GWAS. The challenge is to understand how these loci contribute to the risk of type 1 diabetes and other autoimmune disorders.

To understand all these novel gene regions we begin by fine mapping the regions to locate the true causative gene variant. GWAS help to narrow the risk region down to a linkage disequilibrium block that could contain many genes. In the type 1 diabetes

associated loci identified by GWAS, the average number is 8.5 protein-coding genes per region (132). The tagging SNP could be in linkage with the true causative variant and so fine mapping with greater coverage (sequencing to identify additional SNPs in the region) is needed. After locating the variant(s) most significantly associated with risk, deciphering its role in a gene will be important in determining whether it is causative to disease risk: does it influence the protein directly, or does it influence neighboring genes? Risk associated SNPs can be assessed for effects on expression of the gene or neighboring genes. The next step would be determining how the change in expression of these genes increases risk for type 1 diabetes. Such a study as described above will be the focus of Chapter 4 where we investigated chromosomal region 16p13.13 and its role in type 1 diabetes.

### **Work Cited**

1. Redondo MJ, L. Y, Hawa M, *et al.*: Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia* 2001;44:354-362
2. Kumar D, Gemayel NS, Deapen D, *et al.*: North-American twins with IDDM: genetic, etiological, and clinical significance of disease concordance according to age zygosity, and the interval after diagnosis in first twin. *Diabetes* 1993;42:1351-1363
3. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J: Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs. *Diabetes* 2003;52:1052-1055
4. Kyvik KO, Green A, Beck-Nielsen H: Concordance rates of insulin dependent diabetes mellitus: a population based study of Danish twins. *BMJ* 1995;311:913-917
5. Hirschhorn JN: Genetic epidemiology of type 1 diabetes. *Pediatric Diabetes* 2003;4:87-100
6. Rich SS, Concannon P: Section 1: In Search of Diabetes Genes. *Diabetes* 2002;51:S288-S294
7. Rich SS: Perspective in Diabetes Mapping Genes in Diabetes Genetic Epidemiological Perspective. *Diabetes* 1990;39:1315-1319
8. Bennett CL, Christie J, Ramsdell F, *et al.*: The immune dysregulation polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nature genetics* 2001;27:20-21
9. Nagamine K, Peterson P, Scott HS, *et al.*: Positional cloning of the APECED gene. *Nature genetics* 1997;17:393-398

10. Concannon P, Rich SS, Nepom GT: Genetics of Type 1A Diabetes. *The New England Journal of Medicine* 2009;360:1646-1654
11. Strachan T, Read A: *Human Molecular Genetics 4th Edition*. New York, NY, Garland Science, 2011
12. Spielman RS, McGinnis RE, Ewens WJ: Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM). *American journal of human genetics* 1993;52:506-516
13. Davies JL, Kawaguchi Y, Bennett ST, *et al.*: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994;371:130-136
14. Mein CA, Esposito L, Dunn MG, *et al.*: A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nature genetics* 1998;19:297-300
15. Concannon P, Gogolin-Ewens KJ, Hinds DA, *et al.*: A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nature genetics* 1998;19
16. Cox NJ, Wapelhorst B, Morrison VA, *et al.*: Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. *American journal of human genetics* 2001;69:820-830
17. Concannon P, Erlich HA, Julier C, *et al.*: Type 1 Diabetes Evidence for Susceptibility Loci from Four Genome-Wide Linkage Scans in 1,435 Multiplex Families. *Diabetes* 2005;54:2995-3001

18. Nistico L, Buzzetti R, Pritchard LE, *et al.*: The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Human molecular genetics* 1996;5:1075-1080
19. Singal DP, Blajchman MA: Histocompatibility (HLA-A) Antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 1973;22:429-432
20. Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 1984;33:176-183
21. Bottini N, Musumeci L, Alonso A, *et al.*: A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature genetics* 2004;36:337-338
22. Vella A, Cooper JD, Lowe CE, *et al.*: Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *American journal of human genetics* 2005;76:773-779
23. Livak KJ, Marmaro J, Todd JA: Towards fully automated genome-wide polymorphism screening. *Nature genetics* 1995;9:341-342
24. Wang DG: Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science* 1998;280:1077-1082
25. Kruglyak L: The use of genetic map of biallelic markers in linkage studies. *Nature genetics* 1997;17:21-24
26. Klein RJ, Zeiss C, Chew EY, *et al.*: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385-389



- 27.** Smyth DJ, Cooper JD, Bailey R, *et al.*: A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nature genetics* 2006;38:617-619
- 28.** Consortium WTCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-678
- 29.** Hakonarson H, Grant SF, Bradfield JP, *et al.*: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007;448:591-594
- 30.** Barrett JC, Clayton DG, Concannon P, *et al.*: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics* 2009;41:703-707
- 31.** Cooper JD, Smyth DJ, Smiles AM, *et al.*: Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nature genetics* 2008;40:1399-1401
- 32.** Grant SF, Qu HQ, Bradfield JP, *et al.*: Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. *Diabetes* 2009;58:290-295
- 33.** Hakonarson H, Qu HQ, Bradfield JP, *et al.*: A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* 2008;57:1143-1146
- 34.** Todd JA, Walker NM, Cooper JD, *et al.*: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics* 2007;39:857-864

35. Human Type 1 Diabetes Loci [article online], 2013. Available from [http://www.t1dbase.org/page/Regions/display/species/Human/disease/1#locus\\_604](http://www.t1dbase.org/page/Regions/display/species/Human/disease/1#locus_604). Accessed 05-01-2013
36. Van Belle TL, Coppieters KT, Von Herrath MG: Type 1 Diabetes: Etiology, Immunology and Therapeutic Strategies. *Physiol Rev* 2011;91:798-118
37. Kato H, Takeuchi O, Sato S, *et al.*: Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 2006;441:101-105
38. Gitlin L, Barchet W, Gilfillan S, *et al.*: Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103:8459-8464
39. Gamble DR, Path MC, Kinsley ML, *et al.*: Viral Antibodies in Diabetes Mellitus. *British Medical Journal* 1969;3:627-630
40. Filippi CM, von Herrath MG: Viral trigger for type 1 diabetes: pros and cons. *Diabetes* 2008;57:2863-2871
41. Clements GB, Galbraith DN, Taylor KW: Cocksackie B virus infection and onset of childhood diabetes. *The Lancet* 1995;346:221-223
42. Cinek O, Tapia G, Witso E, *et al.*: Enterovirus RNA in peripheral blood may be associated with the variants of rs1990760, a common type 1 diabetes associated polymorphism in IFIH1. *PLoS One* 2012;7:e48409
43. Consortium IMMSG: Risk Alleles for Multiple Sclerosis Identified by a Genomwide Study. *The New England Journal of Medicine* 2007;357:851-862

44. LeibundGut-Landmann S, Waldburger JM, Krawczyk M, *et al.*: Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. *European journal of immunology* 2004;34:1513-1525
45. Palmer DC, Restifo NP: Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. *Trends in immunology* 2009;30:592-602
46. Abbas AK, Lichtman AH, Pillai S: *Cellular and Molecular Immunology*. China, Saunders Elsevier, 2010
47. Mehra NK, Kaur G: MHC-based vaccination approaches: progress and perspectives. *Expert Reviews in Molecular Medicine* 2004;5
48. Kahn CR, Weir GC, King GL, *et al.*: *Joslin's Diabetes Mellitus 14th Edit.* Boston, MA, Lippincott Williams & Wilkins, 2005
49. Nerup J, Platz P, Andersen OO, *et al.*: HLA-A antigens and diabetes mellitus. *The Lancet* 1974;304:864-866 [Abstract]
50. Cudworth AG, Woodrow JC: HLA-System and Diabetes Mellitus. *Diabetes* 1975;24:345-349
51. Farid N, Sampson L, Noel P, *et al.*: HLA-D related (Drw) antigens in juvenile diabetes mellitus. *Diabetes* 1979;28:548-551
52. Wolf C, Spencer KM, Cudworth AG: The genetic susceptibility to type 1 (insulin-dependent) diabetes: analysis of the HLA-DR association. *Diabetologia* 1983;24:224-330 [Abstract]

53. Owerbach D, Lernmark A, Platz P, *et al.*: HLA-D region B-chain DNA endonuclease fragments differ between HLA-DR identical healthy and insulin-dependent diabetic individuals. *Nature* 1983;303:815-817
54. Todd JA, Bell JI, McDevitt HO: HLA-DQ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987;329:599-604
55. Thorsby E, Ronningen KS: Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 (insulin dependent) diabetes mellitus. *Diabetologia* 1993;36:371-377 [Abstract]
56. Undlien DE, Friede T, Rammensee HG, *et al.*: DR4 subtypes may be associated with different degrees of protection. 1997;46:143-149
57. Solow H, Hidalgo R, Singal DP: Juvenile-onset diabetes: HLA-A, -B, -C, and -DR alloantigens. *Diabetes* 1979;28:1-4
58. Leslie RD: *Causes of Diabetes: Genetic and Environmental Factors*. New York, NY, John Wiley & Sons, 1993
59. Noble JA, Valdes AM, Cook M, *et al.*: The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 caucasian, multiplex families. *American journal of human genetics* 1996;59:1134-1148
60. Erlich H, Valdes AM, Noble J, *et al.*: HLA DR-DQ haplotypes and genotypes and type diabetes risk. *Diabetes* 2008;57:1084-1092
61. Cucca F, Muntoni F, Lampis R, *et al.*: Combinations of specific DRB1, DQA1, DQB1, haplotypes are associated with insulin-dependent diabetes mellitus in Sardinia. *Hum Immunol* 1993;37:85-94

62. Erlich HA, Zeidler A, Chang J, *et al.*: HLA class II alleles and susceptibility and resistance to insulin dependent diabetes mellitus in Mexican-American families. *Nature genetics* 1993;3:358-364
63. Bugawan TL, Klitz W, Alejandrino M, *et al.*: The association of specific HLA class I and II alleles with type 1 diabetes among Filipinos. *Tissue antigens* 2002;59:452-469
64. Mijovic C, Jenkins D, Jacobs KH, *et al.*: HLA-DQA1 and -DQB1 alleles associated with genetic susceptibility to IDDM in a black population. *Diabetes* 1991;40:748-753
65. Park YS, Wang CY, Ko KW, *et al.*: Combinations of HLA DR and DQ molecules determine the susceptibility to insulin-dependent diabetes mellitus in Koreans. *Human Immunology* 1998;59:794-801
66. Steck AK, Pugliese A, Eisenbarth GS: Chapter 7 Diabetes Mellitus of Man: Genetic Susceptibility and Resistance. In *Type 1 Diabetes: Cellular, Molecular & Clinical Immunology*  
Online Edition Version 3.0 ed. Aurora Colorado, Barbara Davis Center, 2012
67. Robles DT, Eisenbarth GS, Wang T, *et al.*: Millennium award recipient contribution. Identification of children with early onset and high incidence of anti-islet autoantibodies. *Clinical immunology* 2002;102:217-224
68. Robinson WP, Barbosa J, Rich SS, Thomson G: Homozygous parents affected sib pair method for detecting disease predisposing variants: application to insulin dependent diabetes mellitus. *Genetic epidemiology* 1993;10:273-288

69. Fennesy M, Metcalfe K, Hitman GA, *et al.*: A gene in the HLA class I region contributes to susceptibility to IDDM in the Finnish population. *Diabetologia* 1994;37:937-944 [Abstract]
70. Lie BA, Todd JA, Pociot F, *et al.*: The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *American journal of human genetics* 1999;64:793-800
71. Nejentsev S, Gombos Z, Laine AP, *et al.*: Non-class II HLA gene associated with type 1 diabetes maps to the 240kb region near HLA-B. *Diabetes* 2000;49:2217-2221
72. Noble JA, Valdes AM, Bugawan TL, *et al.*: The HLA class I A locus affects susceptibility to type 1 diabetes. *Human Immunology* 2002;63:657-664
73. Valdes AM, Wapelhorst B, Concannon P, *et al.*: Extended DR3-D6S273-HLA-B haplotypes are associated with increased susceptibility to type 1 diabetes in US Caucasians. *Tissue antigens* 2005;65:115-119
74. Nejentsev S, Howson JM, Walker NM, *et al.*: Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 2007;450:887-892
75. Honeyman MC, Harrison LC, Drummond B, Colman PG, Tait BD: Analysis of families at risk for insulin-dependent diabetes mellitus reveals that HLA antigens influence progression to clinical disease. *Mol Med* 1995;1:576-582
76. Valdes AM, Erlich HA, Noble JA: Human leukocyte antigen class I B and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. *Hum Immunol* 2005;66:301-313

77. Kobayashi T, Tamemoto K, Nakanishi K, *et al.*: Immunogenetic and clinical characterization of slowly progressive IDDM. *Diabetes Care* 1993;16:780-788
78. Deschamps I, Marcelli-Barge A, Poirier JC, *et al.*: Two distinct HLA-DR3 haplotypes are associated with age related heterogeneity in type 1 (insulin-dependent) diabetes. *Diabetologia* 1988;31:896-901 [Abstract]
79. Tait BD, Harrison LC, Drummond BP, *et al.*: HLA antigens and age at diagnosis of insulin-dependent diabetes mellitus. *Human Immunology* 1995;42:116-122
80. Bottazzo GF, Dean BM, McNally JM, *et al.*: In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulinitis. *The New England Journal of Medicine* 1985;313:353-360
81. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG: Analysis of islet inflammation in human type 1 diabetes. *Clinical and experimental immunology* 2009;155:173-181
82. Panagiotopoulos C, Qin H, Tan R, Verchere CB: Identification of a  $\beta$ -Cell specific HLA class I restricted epitope in type 1 diabetes *Diabetes* 2003;52:2647-2651
83. Panina-Bordignon P, Lang R, Van Ender PM, *et al.*: Cytotoxic T cells specific for glutamic acid decarboxylase in autoimmune diabetes. *J Exp Med* 1995;181:1923-1927
84. Pinkse GG, Tysma OH, Bergen CA, *et al.*: Autoreactive CD8 T cells associated with beta cell destruction in type 1 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:18425-18430
85. Marron MP, Graser RT, Chapman HD, Serreze DV: Functional evidence for the mediation of diabetogenic T cell responses by HLA-A2.1 MHC class I molecules through

transgenic expression in NOD mice. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:13753-13758

**86.** Bennett ST, Lucassen AM, Gough SCL, *et al.*: Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature genetics* 1995;9:284-292

**87.** Julier C, Hyer RN, Davies J, *et al.*: Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* 1991;354:155-159

**88.** Lucassen AM, Julier C, Beressi JP, *et al.*: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nature genetics* 1993;4:305-310

**89.** Owerbach D, Gabbay KH: Localization of a type 1 diabetes susceptibility locus to the variable tandem repeat region flanking the insulin gene. *Diabetes* 1993;42:1708-1714

**90.** Bell GI, Karam JH, Rutter WJ: Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proceedings of the National Academy of Sciences of the United States of America* 1981;78:5759-5763

**91.** Bell GI, Selby MJ, Rutter WJ: the highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* 1982;295:31-35

**92.** Pugliese A, Awdeh ZL, Alper CA, Jackson RA, Eisenbarth G: The paternally inherited insulin gene B allele (1,428 FokI site) confers protection from insulin-dependent diabetes in families. *Journal of autoimmunity* 1994;7:687-694



93. Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nature genetics* 1995;9:293-298
94. Pugliese A, Zeller M, Fernandez A, *et al.*: The insulin gene is transcribed in the human thymus and transcription levels correlate with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nature genetics* 1997;15:293-297
95. Vafiadis P, Bennett ST, Todd JA, *et al.*: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nature genetics* 1997;15:289-292
96. Marron MP, Raffel LJ, Garchon HJ, *et al.*: Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA polymorphisms in multiple ethnic groups. *Human molecular genetics* 1997;6:1275-1282
97. Teft WA, Kirchhof MG, Madrenas J: A molecular perspective of CTLA-4 function. *Annual review of immunology* 2006;24:65-97
98. Waterhouse P, Penninger JM, Timms E, *et al.*: Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* 1995;270:985-988
99. Maurer M, Ponath A, Kruse N, Rieckmann P: CTLA-4 exon 1 dimorphism is associated with primary progressive multiple sclerosis. *Journal of Neuroimmunology* 2002;131:213-215
100. Barreto M, Santos E, Ferreira R, *et al.*: Evidence for CTLA4 as a susceptibility gene for systemic lupus erythematosus. *European journal of human genetics : EJHG* 2004;12:620-626

- 101.** Vaidya B, Pearce SH, Charlton PS, *et al.*: An association between the CTLA4 exon 1 polymorphism and early rheumatoid arthritis with autoimmune endocrinopathies. *Rheumatology* 2002;41:180-183
- 102.** Donner H, Rau H, Walfish PG, *et al.*: CTLA4 alanine-17 confers genetic susceptibility to grave's disease and to type 1 diabetes mellitus. *Journal of Clinical Endocrinology & Metabolism* 1997;82:143-146
- 103.** Blomhoff A, Lie BA, Myhre AG, *et al.*: Polymorphisms in the cytotoxic T lymphocyte antigen-4 gene region confer susceptibility to Addison's disease. *The Journal of clinical endocrinology and metabolism* 2004;89:3474-3476
- 104.** Wang XB, Zhao X, Giscombe R, Lefvert AK: A CTLA-4 gene polymorphism at position -318 in the promoter region affects the expression of protein. *Genes and immunity* 2002;3:233-234
- 105.** Anjos S, Nguyen A, Ounissi-Benkhalha H, Tessier MC, Polychronakos C: A common autoimmunity predisposing signal peptide variant of the cytotoxic T-lymphocyte antigen 4 results in inefficient glycosylation of the susceptibility allele. *The Journal of biological chemistry* 2002;277:46478-46486
- 106.** Wang XB, Kakoulidou M, Giscombe R, *et al.*: Abnormal expression of CTLA-4 by T cells from patients with myasthenia gravis: effect of an AT-rich gene sequence. *Journal of Immunology* 2002;130:224-232
- 107.** Holopainen PM, Patanen JA: Technical note: linkage disequilibrium and disease-associated CTLA4 gene polymorphisms. *Journal of Immunology* 2001;167:2457-2458

- 108.** Kavvoura FK, Ioannidis JP: CTLA-4 gene polymorphisms and susceptibility to type 1 diabetes mellitus: a HuGE Review and meta-analysis. *American journal of epidemiology* 2005;162:3-16
- 109.** Smyth D, Cooper JD, Collins JE, *et al.*: Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes* 2004;53:3020-3023
- 110.** Bottini N, Vang T, Cucca F, Mustelin T: Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Seminars in immunology* 2006;18:207-213
- 111.** Wu J, Katrekar A, Honigberg LA, *et al.*: Identification of substrates of human protein-tyrosine phosphatase PTPN22. *The Journal of biological chemistry* 2006;281:11002-11010
- 112.** Begovich AB, Carlton VEH, Honigberg LA, *et al.*: A missense single-nucleotide polymorphism in a gene encoding a protein tyrosien phosphatase (PTPN22) is associated with rheumatoid arthritis. *American journal of human genetics* 2004;75:330-337
- 113.** Kyogoku C, Langefeld CD, Ortmann WA, *et al.*: Genetic Association of the R620W Polymorphism of Protein Tyrosine Phosphatase PTPN22 with Human SLE. *Am J Hum Genet* 2004;75:504-507
- 114.** Hasegawa K, Martin F, Huang G, *et al.*: PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. *Science* 2004;303:685-689
- 115.** Vang T, Congia M, Macis MD, *et al.*: Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nature genetics* 2005;37:1317-1319

- 116.** Rieck M, Arechiga A, Onengut-Gumuscu S, *et al.*: Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *Journal of Immunology* 2007;179:4704-4710
- 117.** Lee AT, Li W, Liew A, *et al.*: The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. *Genes and immunity* 2005;6:129-133
- 118.** Kyogoku C, Langefeld CD, Ortmann W, *et al.*: Genetic association of the R620W polymorphism of protein tyrosine phosphatase ptpn22 with human SLE. *American journal of human genetics* 2004;75:504-507
- 119.** Habib T, Funk A, Rieck M, *et al.*: Altered B cell homeostasis is associated with type I diabetes and carriers of the PTPN22 allelic variant. *J Immunol* 2012;188:487-496
- 120.** Menard L, Saadoun D, Isnardi I, *et al.*: The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans. *The Journal of clinical investigation* 2011;121:3635-3644
- 121.** Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C: Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes* 2007;56:1174-1176
- 122.** Lowe CE, Cooper JD, Brusko T, *et al.*: Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nature genetics* 2007;39:1074-1082

- 123.** Maier LM, Lowe CE, Cooper J, *et al.*: IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 2009;5:e1000322
- 124.** Malek TR, Bayer AL: Tolerance, not immunity, crucially depends on IL-2. *Nature reviews Immunology* 2004;4:665-674
- 125.** Greenberg SJ, Marcon L, Hurwitz BJ, Waldmann TA, Nelson DL: Elevated levels of soluble interleukin-2 receptors in multiple sclerosis. *The New England Journal of Medicine* 1998;319:1019-1020
- 126.** Maier LM, Anderson DE, Severson CA, *et al.*: Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *Journal of Immunology* 2009;182:1541-1547
- 127.** Giordano C, Panto F, Caruso C, *et al.*: Interleukin 2 and soluble interleukin 2-receptor secretion defect in vitro in newly diagnosed type 1 diabetic patients. *Diabetes* 1989;39:310-315
- 128.** Brusko T, Wasserfall C, McGrail K, *et al.*: No alterations in the frequency of FOXP3<sup>+</sup> regulatory T-cells in type 1 diabetes. *Diabetes* 2007;56:604-612
- 129.** Brusko TM, Wasserfall CH, Clare-Salzler M, Schatz DA, Atkinson MA: Functional defects and the influence of age on the frequency of CD4<sup>+</sup> CD25<sup>+</sup> T-cells in type 1 diabetes. *Diabetes* 2005;54:1407-1414
- 130.** Qu HQ, Verlaan DJ, Ge B, *et al.*: A cis-acting regulatory variant in the IL2RA locus. *J Immunol* 2009;183:5158-5162

**131.** Garg G, Tyler JR, Yang JH, *et al.*: Type 1 diabetes-associated IL2RA variation lowers IL-2 signaling and contributes to diminished CD4+CD25+ regulatory T cell function. *J Immunol* 2012;188:4644-4653

**132.** Human Type 1 Diabetes Loci [article online], 2013. Available from <http://www.t1dbase.org/page/Regions/display/species/human/disease/1/type/assoc>.

Accessed 7-04-2013

## **Chapter 4:**

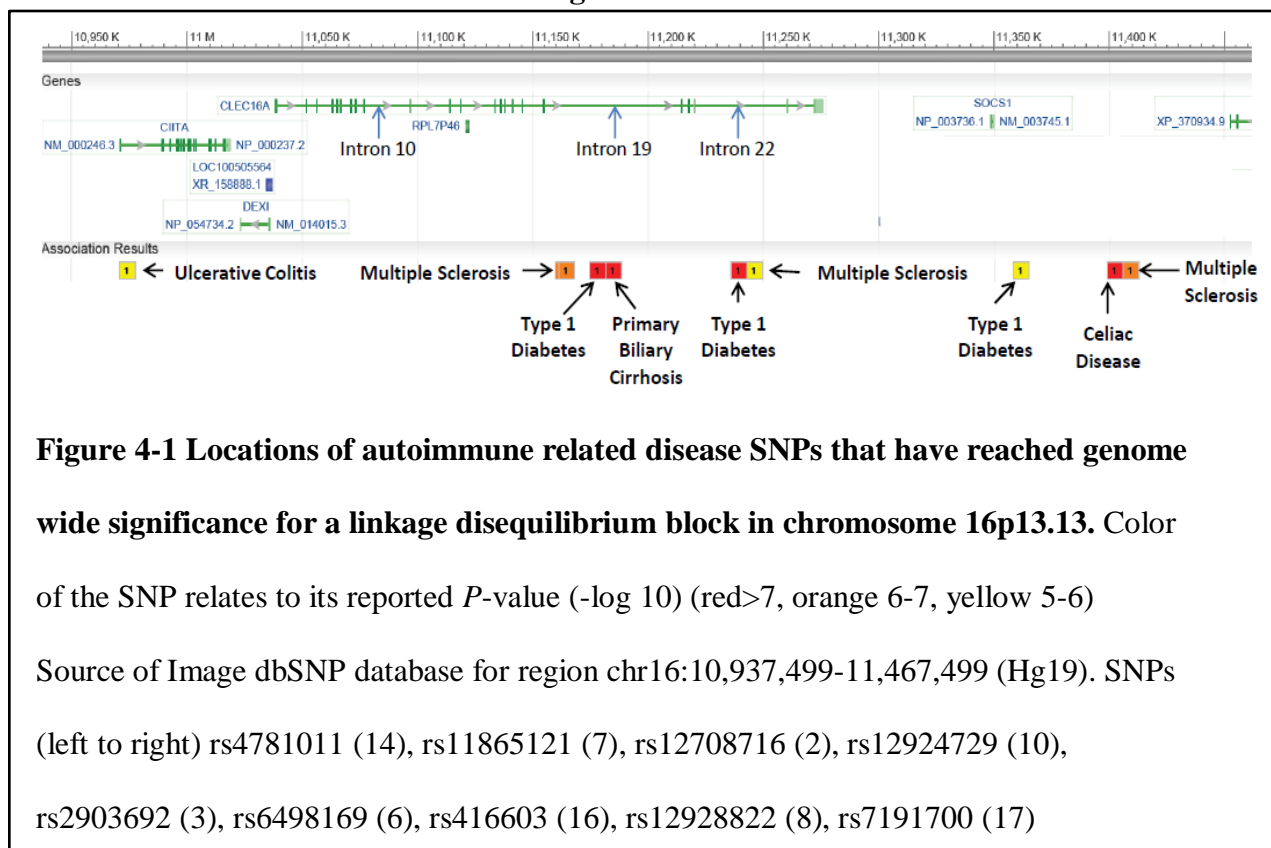
### Chromosome 16p13.13 and Its Role in Type 1 Diabetes

## **Introduction**

### **Background Information**

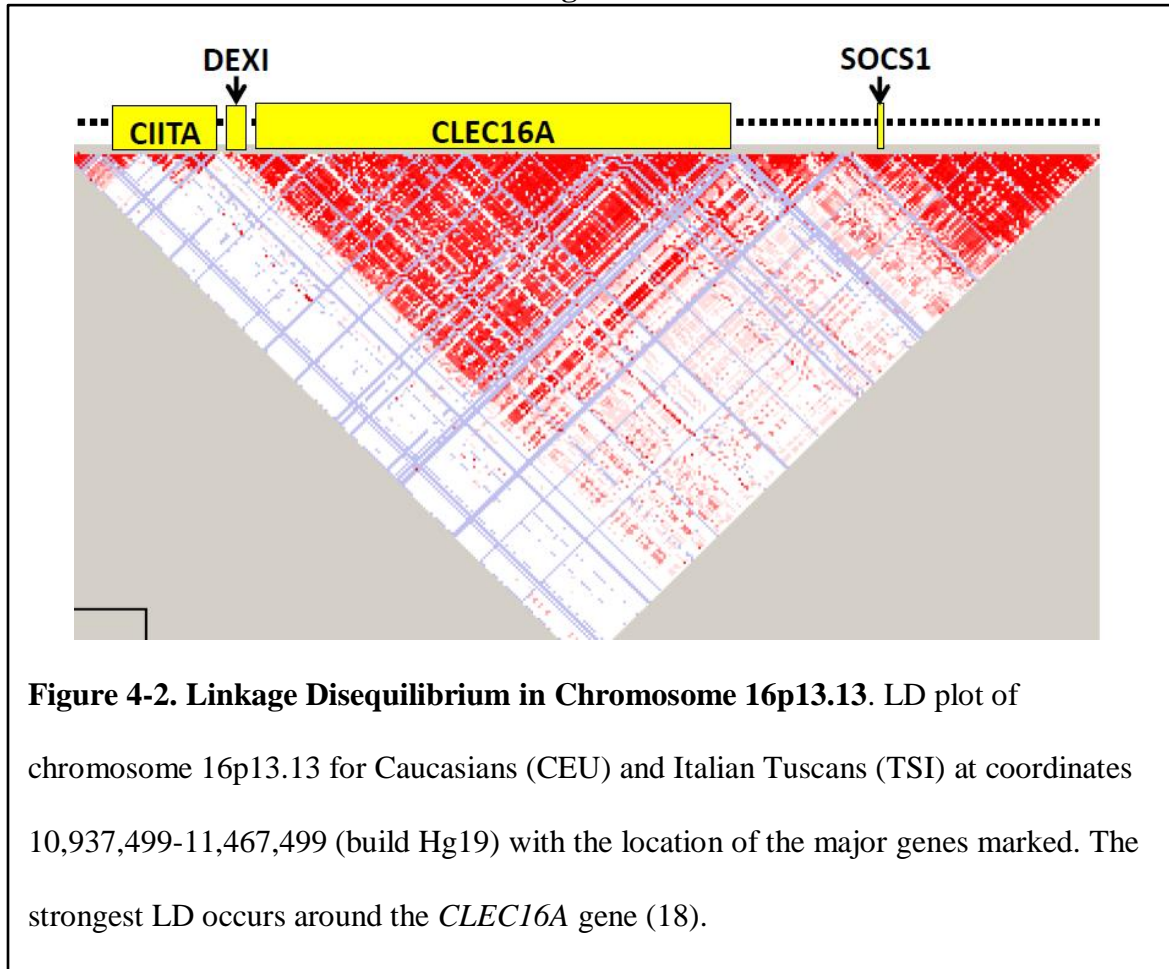
As mentioned in previous chapters prior to 2007, a small number of genetic loci that contributed to type 1 diabetes risk were identified through linkage and candidate gene studies including *HLA*, *INS*, *IL2RA*, *CTLA4* and *PTPN22* (1). With the advent of genome-wide association studies (GWAS) over 40 additional loci have been identified that confer low to moderate risks of developing type 1 diabetes (Odds Ratios, OR,  $\leq 1.3$ )(2). One of these regions is located at chromosome 16p13.13 and is associated with type 1 diabetes and several other autoimmune disorders, including Multiple Sclerosis (MS), celiac disease, Primary Biliary Cirrhosis (PBC), Systemic Lupus Erythematosus, Crohn's Disease, Ulcerative Colitis and primary adrenal insufficiency (Figure 4-1) (3-15).



**Figure 4-1**

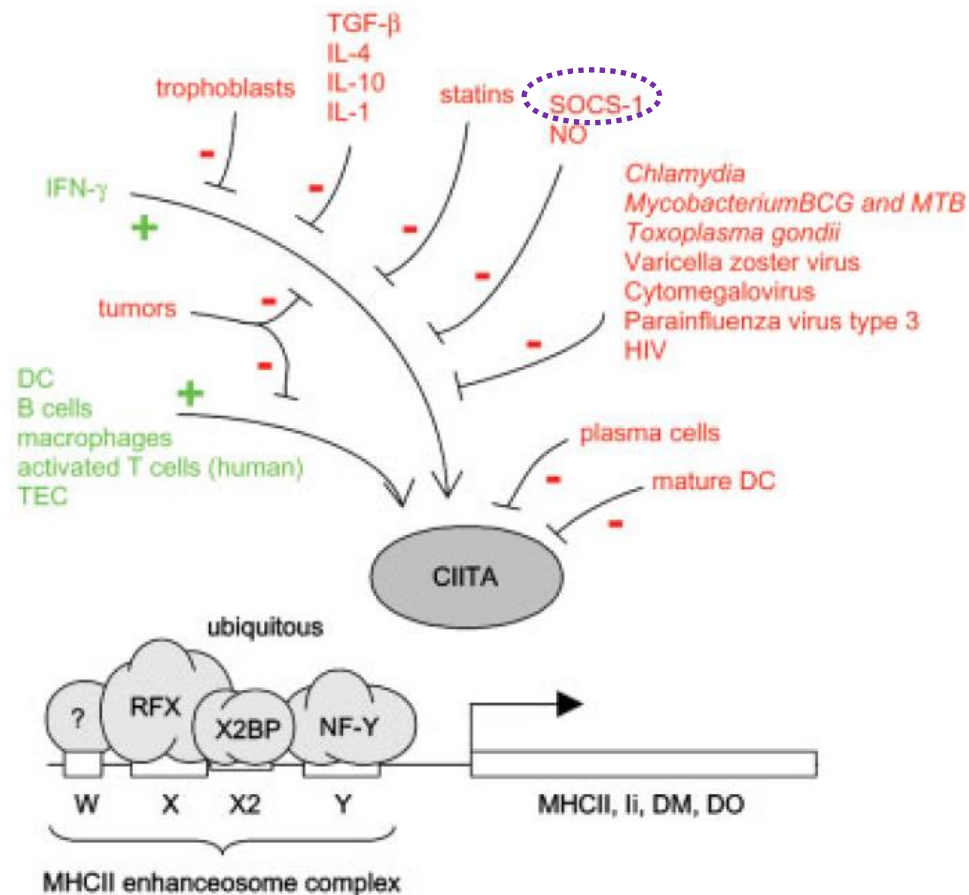
The specific region on chromosomal region 16p13.13 implicated in autoimmunity spans approximately 530 KB from position 10,937,499-11,467,499 (build Hg19) (Figure 4-2). The most significant evidence of association with type 1 diabetes was with two SNPs, rs12708716 and rs2903692, located in the different introns of a gene of only inferred function, *CLECI6A*, previously known as *KIAA0350* (2,3). Complicating the interpretation of these results is the fact that neighboring genes appear to be stronger candidates for playing a role in type 1 diabetes pathogenesis based on their known functions. The region contains 4 genes *CIITA*, *DEXI*, *CLECI6A* and *SOCS1*.

Figure 4-2



## ***CIITA***

*CIITA* is a ~48 KB gene that is controlled by four different promoters (pI-IV) that produce three isoforms of *CIITA* (pII promoter function is currently unknown). The three isoforms of *CIITA* are cell type specific: isoform I is found in myeloid cells, isoform III is found in lymphoid cells and isoform IV in cells from non-hematopoietic origin. *CIITA* is involved in the regulation of HLA class II gene transcription and has been extensively studied due to its important immunological role (19). Mutations in the *CIITA* regulation pathway can cause bare lymphocyte syndrome, a severe immune deficiency disease (20). SNPs specifically in *CIITA* have been found to increase susceptibility for multiple sclerosis, arthritis and myocardial infarction (21). Down regulation of *CIITA* has also been implicated in various forms cancer (22). *CIITA* is constitutively expressed in dendritic cells (DC), B cells, macrophages and activated T Cells, while in other cell types it is induced by IFN- $\gamma$ . *CIITA* functions by binding to the MHC II enhanceosome complex promoter and acts as master regulator of the MHC II transcription complex (Figure 4-3). It is inhibited by various cytokines (examples: TGF- $\beta$  & IL-4), negative feedback inhibitor (SOCS1) and by various pathogens (19). In an interesting development it was found that regulation of *CIITA* also occurs through chromatin remodeling that regulates distal enhancer sites. One of these distal enhancer sites is located in the intronic region of adjacent gene *DEXI* (23,24).

**Figure 4-3**

**Figure 4-3. The Pathway of CIITA.** CIITA is activated by a variety of factors (green) and inhibited by others (red). It is constitutively expressed in dendritic cells, B-cells, macrophages and activated T-Cells, while in other cells types it is induced by IFN- $\gamma$ . CIITA functions by binding to the MHC II promoter (specifically the enhanceosome complex) and act as master regulator of MHC II transcription. CIITA is inhibited by various factors like cytokines (examples: TGF- $\beta$  & IL-4, negative feedback inhibitor like SOCS1 (purple dashed circle) and by various pathogens. (Copyright 2004, with permission from John Wiley and Sons (19))

### ***DEXI***

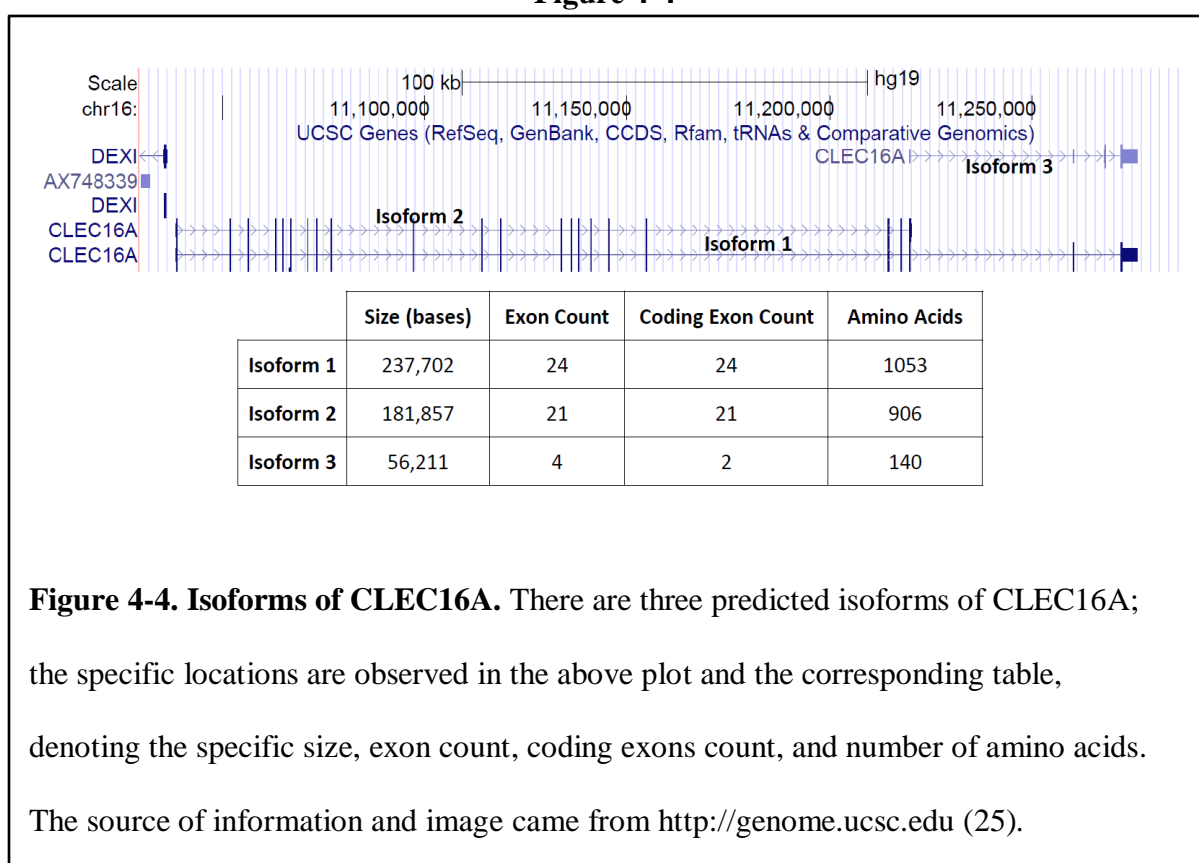
*DEXI* is a 14 KB gene that is transcribed in the opposite direction to *CIITA* and *CLEC16A*. *DEXI* is composed of two exons (only 1 coding) and codes for a protein 95 amino acids long (25). *DEXI* is up-regulated by dexamethasone, a glucocorticoid (a steroidal hormone) but little is known about its function. Protein analyses have identified a central transmembrane domain, carboxy-terminal leucine zipper motif and a negatively charged carboxy-terminal. It is speculated that *DEXI* could be a membrane protein and not a nuclear protein because of the central transmembrane domain. It is found to be highly expressed in the heart, brain and liver (26). In a yeast two-hybrid screen *DEXI* was found to interact with Golgi membrane protein 1 (GOLM1) and tubulin-specific chaperone D protein (TBCD) (27). GOLM1, also known as GP73, has been found to be up-regulated in a viral infection (28), implicating *DEXI* in a possible immune response mechanism. Whereas TBCD is a protein involved in helping beta-tubulin fold correctly and seems like a unlikely target for an auto-immune disease (29).

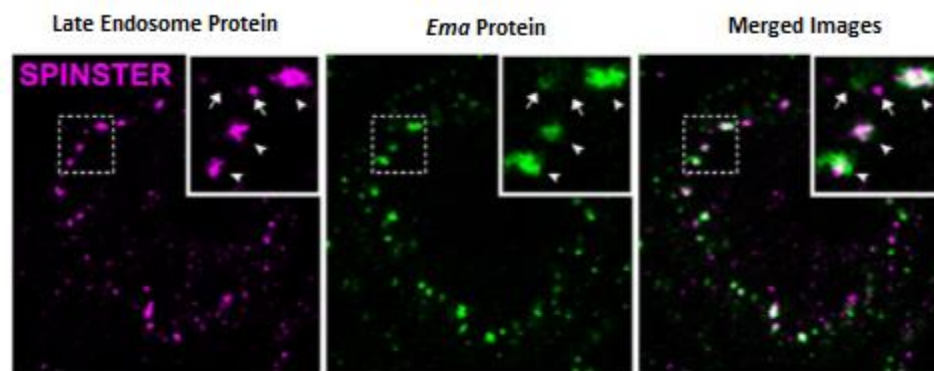
### ***CLEC16A***

*CLEC16A* also referred to as *KIAA0350*, is a possible C-type lectin that is ~240 KB with mostly unknown function. C-type lectins normally function as cell surface receptors in the immune system. Based on available cDNA, there are three predicted isoforms of *CLEC16A* that vary greatly in size. Their specific genomic location, size, exons and number of amino acids can be seen in Figure 4-4. There also appears to be an ITAM sequence located in the middle region of the protein (4). ITAM sequences function

as targets in tyrosine phosphorylation and in the transmission of cell surface signals (30). CLEC16A is highly expressed in immune cells like T cells, B cells and NK cells (31). However, in a more recent study with *Drosophila* orthologue of CLEC16A - *ema* - was reported to be involved in late endosomal maturation (Figure 4-5) (32).

**Figure 4-4**



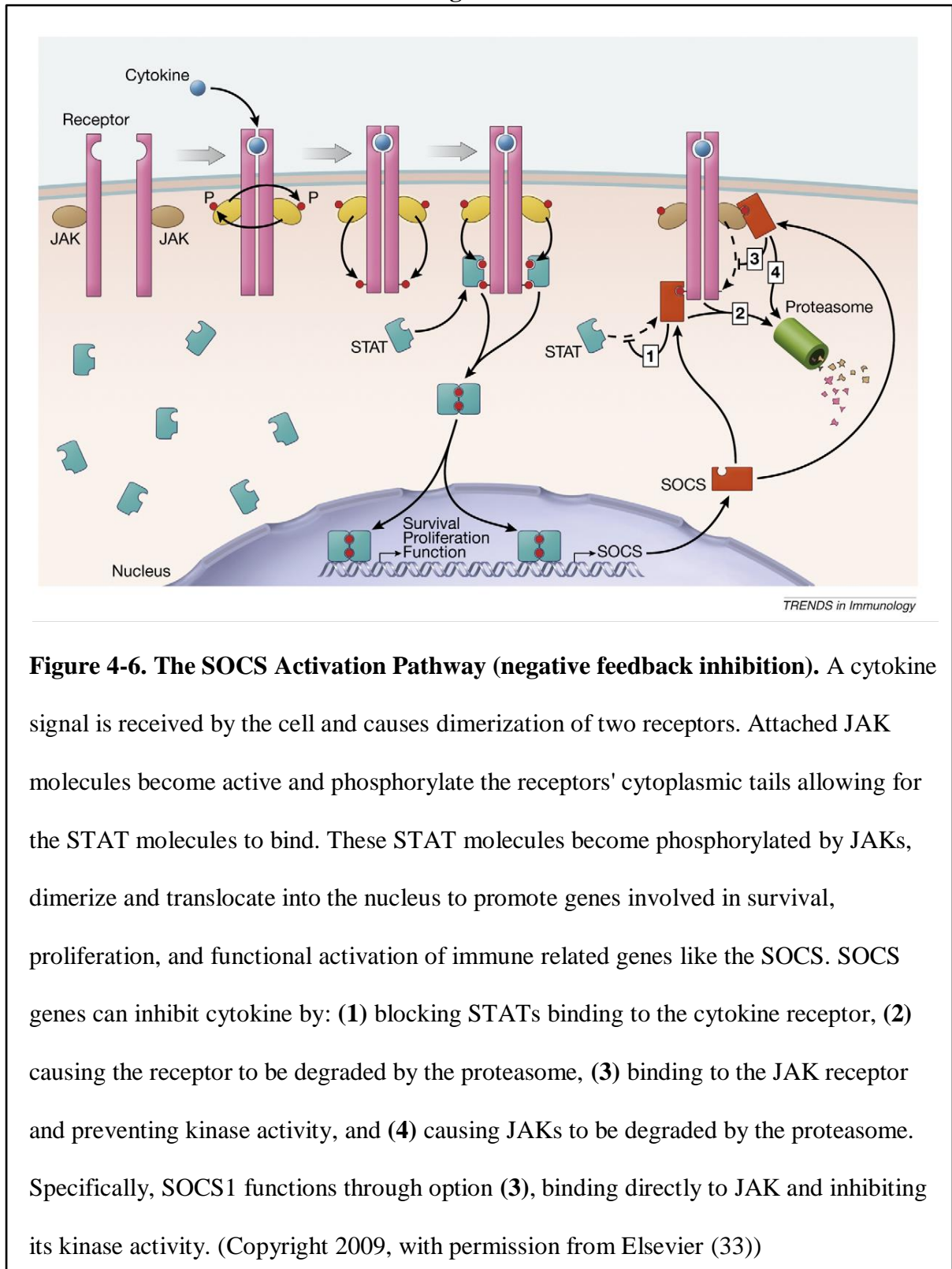
**Figure 4-5**

**Figure 4-5. Co-localization of Spinster and *Ema*.** The *drosophila* CLEC16A orthologue *ema* (Ema::GFP fusion (green)) was found to co-localize with the late endosomal protein spinster (magenta) in a instar larval garland cells (Image use permitted by Rockefeller University Press Noncommercial third-party reuse (32)).

### *SOCS1*

The last important gene in the region is *SOCS1*, a small gene located ~80 KB downstream from *CLEC16A*. *SOCS1*, a negative feedback regulator of cytokine signaling and is <1 KB. It is part of the family of suppressors of cytokine signaling (SOCS) family that is made up of eight proteins SOCS1-7 and CIS. All SOCS family proteins have a Src homology 2 (SH2) domain, C-terminal motif referred to as the SOCS box and a variable N-terminal region. When cytokines are received by a cell, they induce transcription of genes for survival, proliferation, and functional activation; however, they also induce *SOCS* genes. *SOCS* genes function to dampen/turn off the signaling in a negative feedback loop; SOCS1 specifically binds to JAK and inhibits its kinase activity (Figure 4-6) (33,34). When *SOCS1* is completely knocked out in mice, neonatal lethality occurs (35-37) and when over expressed results in increased apoptosis of peripheral T-cells (38). *SOCS1* is clearly an essential gene for proper cytokine regulation and is important in T-cell development (33,34). In an interesting "diabetes" connection, *SOCS1*, *SOCS3* and *SOCS6* were found to block insulin signaling, causing insulin resistance, however this finding maybe be more relevant to type 2 diabetes (39,40).



**Figure 4-6**

## Project Purpose

Several GWAS studies (3-5) identified the chromosome 16p13.13 region as associated with type 1 diabetes and these reported associations were replicated in several follow up studies (16,18,41-44). While these replication studies clearly confirm a role for the 16p13.13 region in conferring risk for type 1 diabetes, they do not clearly delineate which specific genes or SNPs are responsible for this effect. There is, however, complementary functional data suggesting a possible model for how SNPs in this region might act. Most of the variants most strongly associated with type 1 diabetes in this region occur within introns of the *CLEC16A* gene. It has been reported that intronic regions of *CLEC16A* can physically interact with the putative promoter region of *DEXI* and thereby modulate *DEXI* expression in monocytes and in Epstein Barr Virus (EBV) transformed lymphoblastoid cell lines (45-47). As previously mentioned, the intronic region of *DEXI* then works as a distal enhancer for *CIITA* (23,24). These functional findings provide a potential mechanism for variation at intronic sites in *CLEC16A* affecting neighboring genes expression. However they do not clearly resolve the causative variants or whether this mechanism is the actual mode of action of these variants with regards to modulating risk of type 1 diabetes. In the current study, in order to refine the type 1 diabetes association at 16p13.13 and identify the relevant risk variants and genes, a 455 KB segment of the region was re-sequenced in 128 type 1 diabetes patients and 64 controls. Both the newly identified SNPs from the sequencing and the prior reported SNPs from HapMap (48) and 1000 Genomes (49) were tested for an association with type 1 diabetes and with mRNA transcript levels for genes located in the

region. A group of highly correlated SNPs were found to be both significantly associated with type 1 diabetes and with the expression of DEXI.

## **Research Design and Methods**

### **Subjects**

This study was reviewed and approved by the Institutional Review Board at the University of Virginia. DNA from anonymous type 1 diabetes cases and controls were obtained from Virginia Mason Medical Center and Puget Sound Blood Center (Seattle, WA, USA), respectively. DNA from affected sib-pair families (ASP) was obtained from the Type 1 Diabetes Consortium (T1DGC) (50) and Human Biological Data Interchange Repository (HBDI) (51).

### **Sequencing of Region to Identify SNPs**

We targeted a 455 kb region (10,943,936 - 11,399,037 (Hg19)) on 16p13.13 for deep sequencing, which encompassed the four genes (*CIITA*, *DEXI*, *CLEC16A*, and *SOCST1*) and all SNPs previously reported to be significantly associated with type 1 diabetes. The region was tiled with ~10 kb PCR fragments with ~0.5 kb overlaps. Amplifications were carried out in 48 pools that each contained DNA from 4 individuals (128 type 1 diabetes patients and 64 controls). Amplified PCR fragments were pooled in equimolar amounts to create DNA libraries. DNA libraries were prepared for sequencing using Illumina's Paired-End Sample Preparation Kit (Illumina Inc.). Sequencing was performed on an Illumina Genome AnalyzerIIx (Illumina Inc) using 63 base paired-end reads. Sequencing reads were assembled using Burrows-Wheeler Aligner (BWA) tool (52). SAMtools (53) was used for conversion, indexing, and aligning of the data to the reference genome (Hg18 ver. 36) (25,54) as well as for identifying SNPs.

## SNP Filtering and Selection

Putative SNPs identified from sequencing were filtered to limit false positives caused by strand bias by including only minor alleles with a frequency rate  $\geq 0.1\%$ , for which the proportion of forward reads was not statistically different from the proportion of reverse reads. The two proportions were only declared not statistically different if the absolute value of the t-statistic for the proportions test was less than one. The SNP list was then parsed into two groups: annotated SNPs and novel unreported SNPs (defined as lacking a dbSNP ID).

SNPs annotated in dbSNP were excluded from analysis if they had a minor allele frequency (MAF)  $< 5\%$  (except exonic SNPs) or had  $< 50$  mapped reads. The annotated SNP list was further filtered to exclude SNPs that had previously been genotyped by the HapMap project (48), leaving a total of 355 annotated SNPs.

Novel variants, were filtered to exclude those with a MAF  $< 5\%$  (for exonic SNPs exclusion was for MAF  $< 1\%$ ) or the combination of  $< 50$  aligning reads and  $> 20\%$  non-aligning reads, which left a total of 93 novel SNPs.

## Genotyping

We designed two custom GoldenGate genotyping assays (Illumina, CA). The first panel consisted of 123 SNPs which were selected as tagging SNPs ( $r^2 = 0.9$ ) to capture information on the 505 reported SNPs in the CEU HapMap population using Haploview 4.2 (18). The second panel consisted of SNPs identified through sequencing. The compiled SNP lists were verified for design-ability using Illumina's GoldenGate Panel

Design software (www.illumina.com). Genotyping was performed on a total of 413 SNPs on 1850 HBDI and 1750 TIDGC samples. Data were examined in GenomeStudio (Illumina Inc.) and exported in PLINK (55) format.

The following quality control steps were applied to the SNP genotyping data. Signal cloud graphs were manually reviewed for all of the SNPs and those with poor clustering were excluded. SNPs with >5 Mendelian transmission errors were identified and excluded from further analyses. SNPs genotyped in duplicate were examined for concordance using PLINK; those with  $\geq 5\%$  discordance were excluded. SNPs with significant deviations from Hardy-Weinberg equilibrium, >5% missingness or <1% MAF.

For sample quality control in GenomeStudio, DNA samples with a genotype call rate of <90% were removed from further analyses. Subjects with >5 Mendelian transmission errors were identified with PLINK (55) and excluded from the analyses. All remaining Mendelian transmission errors were identified and problematic genotypes removed with PEDCHECK (56).

### **ImmunoChip Dataset**

Genotypes from 1,734 SNPs in the 16p13.13 region obtained from 10,796 individuals were extracted from the ImmunoChip dataset (data release Feb 2012) (57). Quality control included the exclusion of SNPs with significant deviations from Hardy-Weinberg equilibrium or >5% missingness. PEDCHECK was used to identify Mendelian

errors and remove conflicting genotypes. Finally, the data was filtered to remove all duplicate SNPs and SNPs with <1% MAF.

### **Imputation**

After quality control, the data set that was derived from targeted sequencing included 319 SNPs with genotypes from 749 families containing 3370 individuals. The ImmunoChip dataset included 1040 SNPs with genotypes from 2689 families containing 10,746 individuals. There were 252 SNPs and 1581 individuals in common between the two datasets.

The two datasets were merged and missing genotypes were filled in by imputation. A reference haplotype population (same individuals from both panels) was generated using MACH (58) and data from 481 parental Caucasian founders whose ethnicity was confirmed using KING (59). Imputation was performed using MACH and the reference population. SNPs with  $r^2 \leq 0.5$  (values generated by MACH) were removed. The same quality control steps as applied to the originally generated genotype data were applied to the imputed data.

### **Association Testing**

The final merged datasets was tested for association with type 1 diabetes using the TDT option in PLINK and results were visualized using LocusZoom (60). Test for residual evidence of disease association after conditioning on the SNP with the most significant association with type 1 diabetes were carried out using UNPHASED (61).

Statistical significance for association was determined by the conservative Bonferroni correction based on the number of SNPs tested(  $P \leq [0.05 / (\text{Number of SNPs tested})]$  (62,63)

## **RT-PCR**

RNA was isolated from the 33 HapMap B lymphoblastoid cell lines (Coriell Institute for Medical Research, Camden NJ) using TRIzol (Invitrogen Inc.) followed by a DNase treatment (Applied Biosystems). RNA was reverse transcribed using random primers (Invitrogen Inc.) and Superscript II (Invitrogen Inc.) following manufacturer's protocol. The cDNA samples and standards were then plated in triplicate using a Beckman FX-Robotic workstation and air dried (24-48 hours). PCR primers were designed using cDNA transcripts from the NCBI database and Primer3 (64) (Table 4-1). Amplifications were carried out on an Applied Biosystems Inc. 7900 [95°C for 10 minutes, (95°C for 15 seconds, (60°-61°C) for 30 seconds, 72°C for 30 seconds) for 35-40 cycles, followed by a dissociation curve]. The data was examined using ABI's SDS 2.3 software, and CT (threshold cycle) and baseline adjusted for each PCR run.



**Table 4-1**

Primer Sequences used for the expression quantitative trait loci (eQTL) analysis of the genes and isoforms in LD block 16p13.13

NCBI Ref Seq	Gene	Forward Primer	Reverse Primer	Size (bp)
BQ221200	<i>CIITA III</i> *	ggggaagctgagggcacg	gaagctccaggtagccaccttcta	182
BM147267	<i>CIITA IV</i> *	gcggccccagagctgg	gaagctccaggtagccaccttcta	125
NM_015226.2	<i>CLEC16A Iso I</i>	cgcataagatgcagagaat	ggtggcgactgtgaggac	257
BC112897.1	<i>CLEC16A Iso II</i>	ctctgtggagcctgtgtgag	tgtttaatcttgcgctgtgc	102
AK126771.1	<i>CLEC16A Iso III</i>	acagcccagagtcagca	actcatggggcatctgaat	157
NM_014015.3	<i>DEXI</i> †	agtagggctcgcactgctgttc	acctctctgttggcacgaag	192
NM_003745.1	<i>SOCS</i>	gcccttctgtaggatggta	ggaggaagaggaggaagggtt	104
NM_002046.3	<i>GAPDH</i>	cagccgagccacatcgc	catgggtggaatcatattgaaca	186
Novel Transcript	<i>CIITA V</i>	acctgcaacaacaggattcac	caccaacacctcagacttcac	105
Novel Transcript	<i>CIITA VI</i>	agctcgctgccagccttc	ggattctcgctgtccctgt	159

\*CIITA III and CIITA IV primers were originally designed by Hornell 2003 (65)

†There are two intron-less *DEXI* pseudogenes (retro-*DEXI*) located on chromosome 15 at coordinates (Hg19) (29,033,933-29,035,470) and (23,157,046-23,158,548) whose genomic DNA matches almost exactly the cDNA for *DEXI*. The forward primer has one nucleotide difference at the 3' end (**bold**) that differ from the pseudogenes.

### Expression Quantitative Trait Loci (eQTL) Analysis

Amplification efficiencies were calculated per transcript and compared to that for the control *GAPDH* gene. Median CT values were determined from the raw triplicate CT values. The  $\Delta\Delta\text{CT}$  values were calculated correcting for primer efficiencies and were quantile normalized. The quantile normalized  $\Delta\Delta\text{CT}$  were merged with genotyping data on the 33 HapMap cell lines to create a PLINK PED file. An eQTL analysis (linear model with quantitative trait) was then performed in PLINK using the normalized expression data. The eQTL results from the custom and ImmunoChip panels were merged together. The dataset was filtered to remove SNPs with a MAF <5% as well as all duplicate SNPs. The final SNP dataset was visualized using LocusZoom. To correct for limited sample size, instead of using stringent Bonferroni corrected  $P$ -values (63), a program was designed to calculate the total number of independent tests. The program examined the correlations ( $r^2$ ) for all 915 SNPs by examining each SNPs genotyping pattern, determining how many SNPs had the same pattern, versus how many unique patterns and represented independent tests. The number of independent tests performed was  $N = 366$  with a corrected  $P$ -value of  $0.05 / 366 = 1.37 \times 10^{-4}$ .

## **Results**

### **Re-sequencing of the 16p13.13 Region Identifies Novel SNPs**

A region of 455 kb from 16p13.13 was amplified in 192 individuals (128 T1D patients and 64 control individuals) and sequenced to an average depth of coverage of 35X. When aligned to the human reference sequence, there were two gaps at coordinates (Hg19) (11,028,710-11,083,336) and (11,340,104-11,350,943) from which no sequence was obtained that passed quality control. When compared to the human reference genome, 962 SNPs were identified, of which 93 had no prior annotation. A genotyping panel spanning the region was designed containing both tagging SNPs selected from HapMap, and novel SNPs identified from our re-sequencing (60 of the 93 SNPs were designable and included in the panel).

### **Fine Mapping Identifies Statistically Significant Associations with Type 1 Diabetes**

#### **Custom Panel**

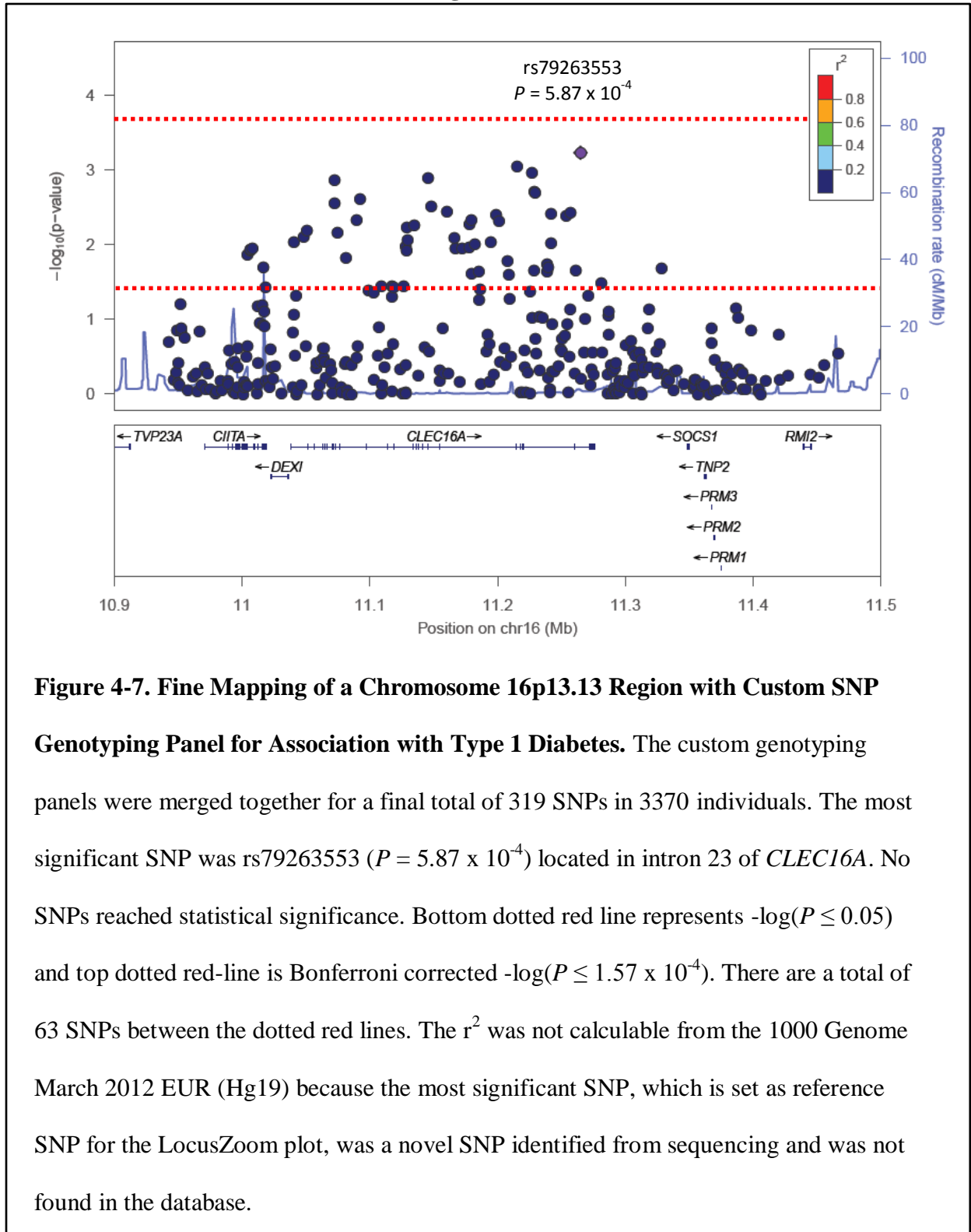
A total of 413 SNPs were genotyped on 3600 individuals from the HBDI and T1DGC. After data cleaning and merging, the custom panel contained 319 SNPs (MAF >1%) genotyped on 3370 individuals from 749 families (Table 4-2 and Figure 4-7). The data was analyzed by the TDT test using PLINK

**Table 4-2**

Functional Annotation of SNPs on Custom Panel

<b>Type of SNP</b>	<b>Combined</b>
Exonic-coding synonymous	6
Exonic-coding missense	4
Near 5' UTR	4
3' UTR	9
Intron	191
Intergenic	105
<b>Total</b>	<b>319</b>

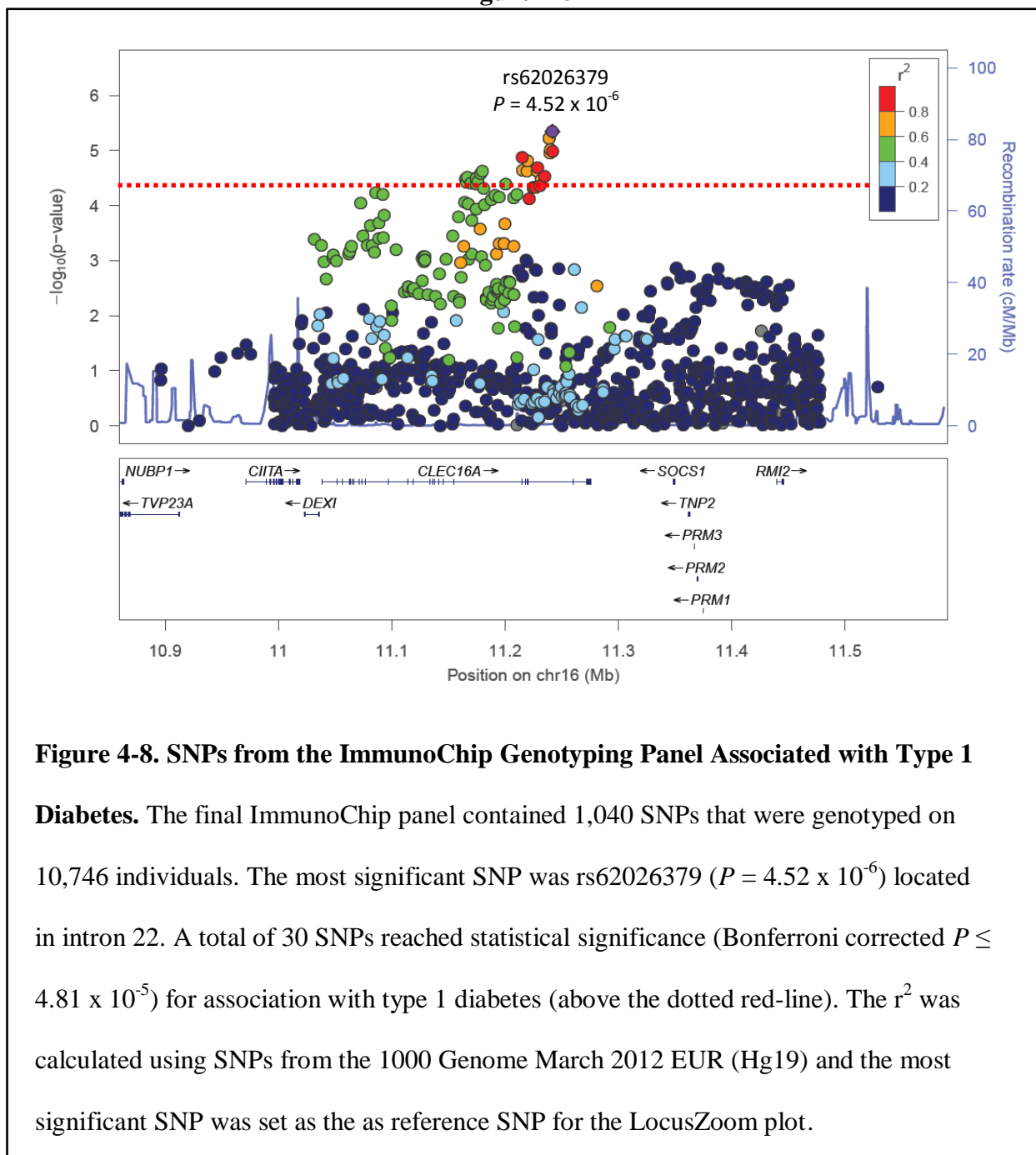
Function was determined from UCSC Genome Table Browser (25)

**Figure 4-7**

### ImmunoChip

In order to increase the statistical power of the study, the data were merged with the ImmunoChip affected sib-pair dataset using imputation to fill in the missing genotypes. The ImmunoChip dataset originally contained 1,734 SNPs genotyped on 10,796 individuals. After quality control steps described in the methods section, the final numbers were filtered down to 1,040 SNPs ( $MAF > 1\%$ ) and 10,746 individuals (Figure 4-8). In the ImmunoChip a total of 30 SNPs were identified to have alleles significantly associated with disease. The ImmunoChip dataset was then merged with the custom genotyping panels.

Figure 4-8



**Figure 4-8. SNPs from the ImmunoChip Genotyping Panel Associated with Type 1 Diabetes.** The final ImmunoChip panel contained 1,040 SNPs that were genotyped on 10,746 individuals. The most significant SNP was rs62026379 ( $P = 4.52 \times 10^{-6}$ ) located in intron 22. A total of 30 SNPs reached statistical significance (Bonferroni corrected  $P \leq 4.81 \times 10^{-5}$ ) for association with type 1 diabetes (above the dotted red-line). The  $r^2$  was calculated using SNPs from the 1000 Genome March 2012 EUR (Hg19) and the most significant SNP was set as the as reference SNP for the LocusZoom plot.

### Final Merged Dataset

The merged dataset, after imputation, contained genotypes for 939 SNPs on 12,485 individuals (4,057 ASP). A breakdown of the type of SNPs can be found in Table 4-3. The combined panel contained 26 SNPs from the Custom Genotyping panel, 661 SNPs from the ImmunoChip and 252 SNPs that were genotyped on both panels (Figure 4-9). The family based TDT (66) was used to test for association of alleles at these SNPs with type 1 diabetes (Figure 4-10). A total of forty eight statistically significant SNPs were identified (Bonferroni corrected  $P \leq 5.32 \times 10^{-5}$ ) (Table 4-4). A magnified plot of the forty-eight statistically significant SNPs can be seen in Figure 4-11. The strongest evidence for association was obtained at rs34306440 located in intron 20 of *CLEC16A* (OR = 0.838 (0.776-0.904),  $P = 5.74 \times 10^{-6}$ ).

**Table 4-3**

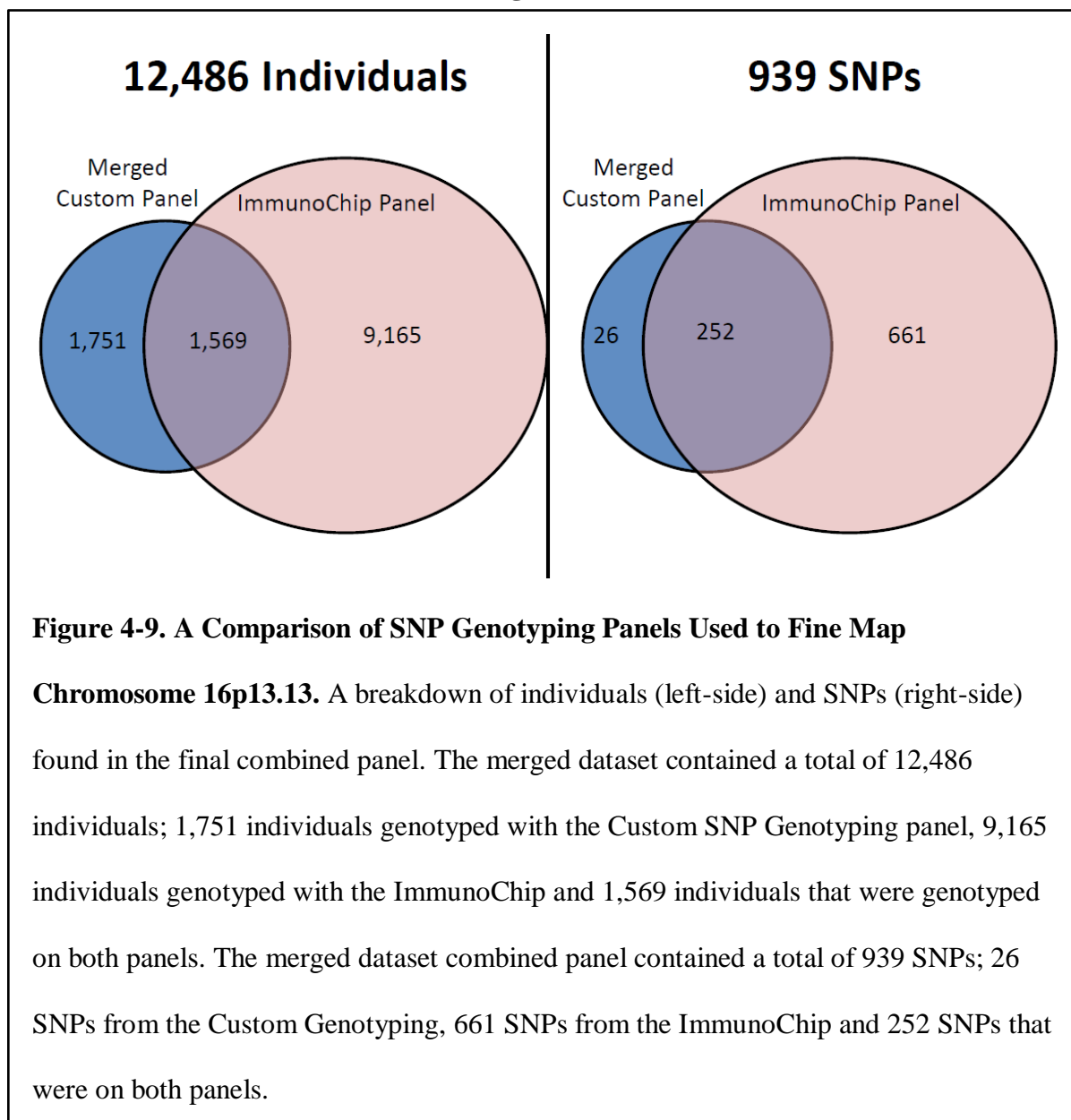
Functional Annotation of SNPs on Final Panel (939 SNPs)

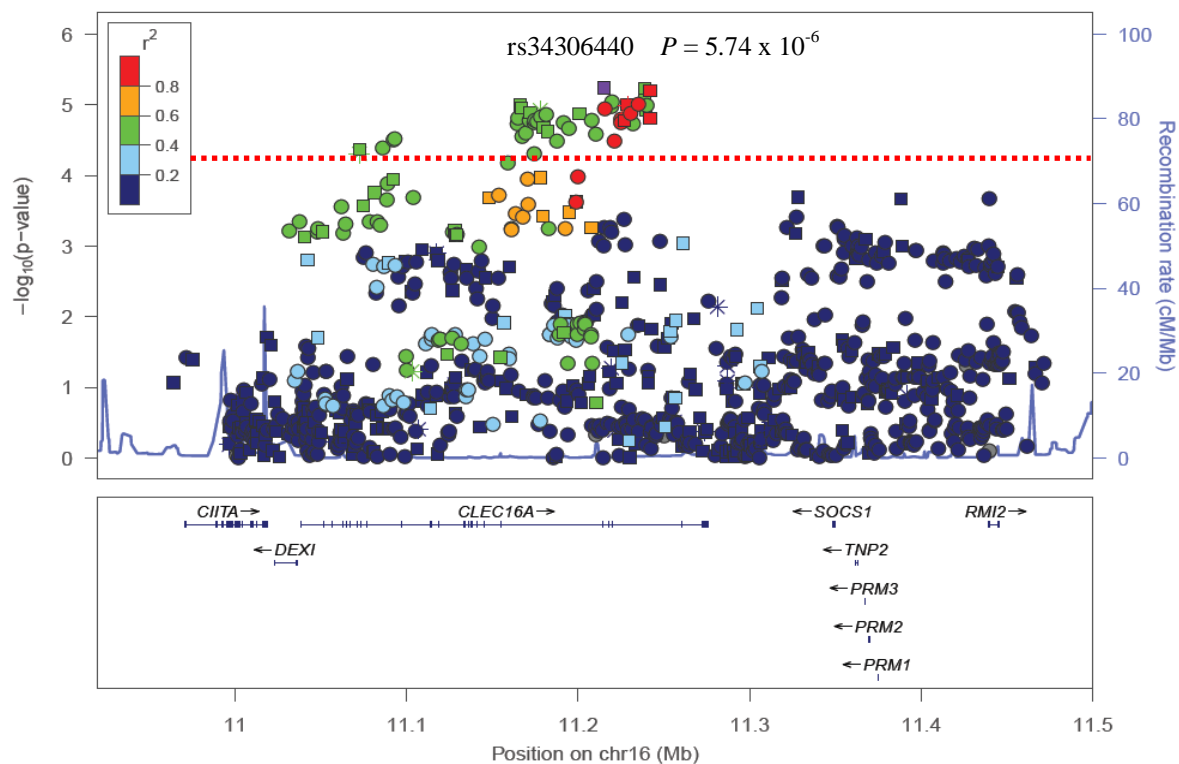
Type of SNP	Count
Exonic-coding synonymous	9
Exonic-coding missense	3
Near 5' UTR	31
3' UTR	15
Intron	555
Intergenic	326
<b>Total</b>	<b>939</b>

Function was determined from UCSC Genome Table Browser (25)

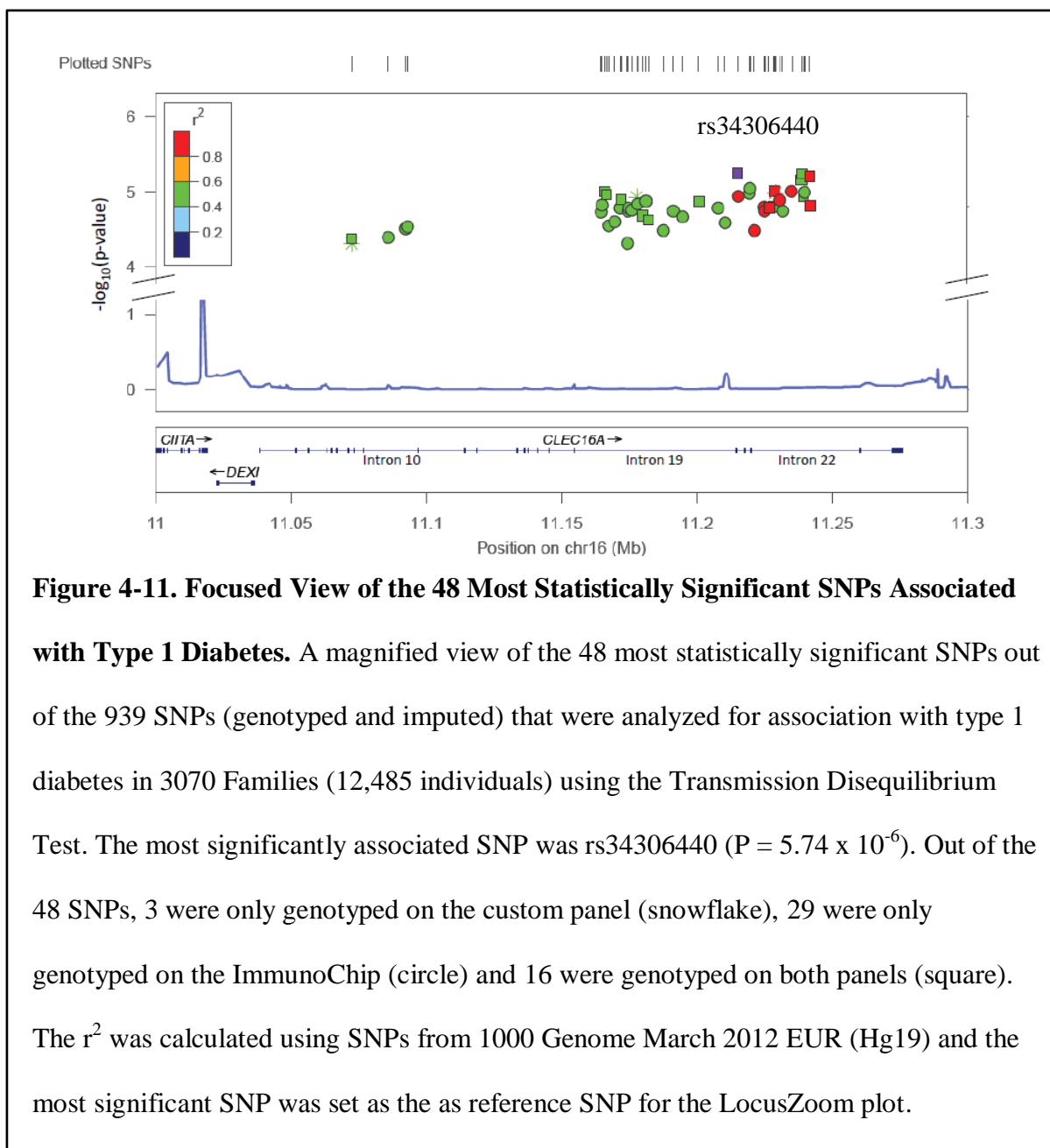


Figure 4-9



**Figure 4-10****Figure 4-10. SNPs Associated with Type 1 Diabetes from the Final Genotyping**

**Panel A** total of 939 SNPs (genotyped and imputed) were analyzed for association with type 1 diabetes in 3070 Families (12,485 individuals) using the Transmission Disequilibrium Test. The most significantly associated SNP was rs34306440 ( $P = 5.74 \times 10^{-6}$ ). Out of the 939 SNPs 26 were only genotyped on the custom panel (snowflake), 661 were only genotyped on the ImmunoChip (circle) and 252 were genotyped on both panels (square). SNPs above the dotted red-line is considered significant ( $P \leq 0.05$ ) after Bonferroni correction. The  $r^2$  was calculated using SNPs from the 1000 Genome March 2012 EUR (Hg19) and the most significant SNP was set as the reference SNP for the LocusZoom plot.

**Figure 4-11**

**Table 4-4**

SNPs in *CLEC16A* most strongly associated with type 1 diabetes (total 48) from TDT analysis sorted in order of base position.

SNP Name	Position (Hg19)	MAF*	Minor Allele	Major Allele	Transmitted Minor Allele Count	Untransmitted Minor Allele Count	TDT <i>P</i> -value†	Odds Ratio†	Location ‡
rs12708713	11072459	0.167	T	C	1045	1239	4.92E-05	0.843 (0.777-0.916)	Intron 8
rs12708714	11072525	0.168	C	T	1048	1244	4.24E-05	0.842 (0.776-0.915)	Intron 8
rs7403919	11085968	0.341	C	T	1691	1938	4.13E-05	0.873 (0.818-0.931)	Intron 10
rs8062322	11092319	0.289	A	C	1529	1768	3.15E-05	0.865 (0.808-0.926)	Intron 10
rs9926367	11093179	0.288	C	T	1520	1759	3.00E-05	0.864 (0.807-0.926)	Intron 10
rs7200940	11164567	0.307	G	C	1556	1804	1.88E-05	0.863 (0.806-0.923)	Intron 19
rs11860603	11165017	0.305	C	T	1546	1796	1.53E-05	0.861 (0.804-0.921)	Intron 19
rs9934969	11166002	0.303	A	G	1540	1795	1.01E-05	0.858 (0.802-0.918)	Intron 19
rs11865121§	11166688	0.304	A	C	1544	1798	1.11E-05	0.859 (0.802-0.919)	Intron 19
rs7198621	11167458	0.304	G	C	1538	1779	2.86E-05	0.865 (0.808-0.926)	Intron 19
rs725613	11169683	0.336	C	A	1625	1874	2.56E-05	0.867 (0.811-0.927)	Intron 19
rs12925642	11171602	0.304	G	A	1548	1797	1.67E-05	0.861 (0.805-0.922)	Intron 19
rs2058531	11172109	0.305	G	A	1554	1807	1.28E-05	0.860 (0.804-0.920)	Intron 19
rs9652601	11174365	0.306	A	G	1557	1792	4.89E-05	0.869 (0.812-0.93)	Intron 19
rs9652582	11174564	0.300	A	G	1536	1783	1.81E-05	0.862 (0.805-0.922)	Intron 19
rs2041670	11174652	0.300	T	C	1537	1785	1.69E-05	0.861 (0.804-0.922)	Intron 19
rs11648679	11175984	0.335	G	A	1623	1877	1.76E-05	0.865 (0.809-0.924)	Intron 19
rs12708715	11177824	0.301	T	C	1542	1795	1.19E-05	0.859 (0.803-0.92)	Intron 19
rs9929994	11178245	0.336	G	A	1617	1873	1.47E-05	0.863 (0.808-0.923)	Intron 19
rs12708716	11179873	0.333	G	A	1615	1866	2.10E-05	0.866 (0.810-0.925)	Intron 19
rs9888908	11181244	0.300	A	C	1539	1790	1.36E-05	0.860 (0.803-0.92)	Intron 19
rs7203793	11182134	0.335	G	C	1626	1876	2.39E-05	0.867 (0.811-0.926)	Intron 19
rs12924729¶	11187783	0.305	A	G	1565	1806	3.31E-05	0.867 (0.810-0.927)	Intron 19
rs12928537	11191400	0.303	A	G	1550	1798	1.82E-05	0.862 (0.806-0.923)	Intron 19
rs12927355	11194771	0.302	T	C	1556	1802	2.18E-05	0.864 (0.807-0.924)	Intron 19
rs741172	11200798	0.300	T	C	1551	1803	1.35E-05	0.860 (0.804-0.921)	Intron 19
rs9746695	11207894	0.302	C	T	1550	1799	1.69E-05	0.862 (0.805-0.922)	Intron 19
rs12935413	11210447	0.333	A	G	1617	1865	2.64E-05	0.867 (0.811-0.927)	Intron 19
rs34306440#	11215035	0.209	G	A	1203	1436	5.74E-06	0.838 (0.776-0.904)	Intron 20
rs35032408	11215424	0.208	G	T	1204	1429	1.16E-05	0.843 (0.780-0.910)	Intron 20
rs8064154	11219419	0.338	G	A	1610	1870	1.05E-05	0.861 (0.806-0.920)	Intron 21
rs12924112	11219720	0.336	G	T	1589	1849	9.24E-06	0.859 (0.804-0.919)	Intron 21
rs34540843	11221287	0.242	G	A	1332	1555	3.32E-05	0.857 (0.796-0.922)	Intron 22
rs36045143	11224966	0.242	G	A	1332	1564	1.62E-05	0.852 (0.792-0.916)	Intron 22
rs2241099	11225064	0.241	G	C	1325	1555	1.82E-05	0.852 (0.792-0.917)	Intron 22

SNP Names	Position (Hg19)	MAF*	Minor Allele	Major Allele	Transmitted Minor Allele Count	Untransmitted Minor Allele Count	TDT P-value†	Odds Ratio†	Location ‡
rs12919828	11226779	0.243	G	A	1335	1567	1.66E-05	0.852 (0.792-0.917)	Intron 22
rs35732840	11228355	0.336	G	T	1606	1860	1.60E-05	0.863 (0.808-0.923)	Intron 22
rs62026376	11228712	0.243	T	C	1331	1569	9.89E-06	0.848 (0.789-0.913)	Intron 22
rs62026377	11229128	0.243	T	G	1327	1564	1.04E-05	0.849 (0.789-0.913)	Intron 22
rs7203459	11230703	0.242	C	T	1324	1558	1.31E-05	0.850 (0.790-0.914)	Intron 22
rs2867880	11231857	0.334	A	G	1604	1856	1.84E-05	0.864 (0.808-0.924)	Intron 22
rs12919732	11235123	0.244	G	A	1321	1558	1.00E-05	0.848 (0.788-0.912)	Intron 22
rs2903692	11238783	0.330	A	G	1581	1844	6.99E-06	0.857 (0.802-0.917)	Intron 22
rs4322688	11238991	0.332	G	C	1590	1856	5.86E-06	0.857 (0.801-0.916)	Intron 22
rs12930373	11239599	0.331	A	G	1590	1847	1.17E-05	0.861 (0.805-0.921)	Intron 22
rs12917893	11239978	0.331	T	A	1584	1842	1.04E-05	0.860 (0.804-0.92)	Intron 22
rs62026379	11241806	0.241	A	G	1314	1556	6.27E-06	0.845 (0.785-0.909)	Intron 22
rs17673553	11241906	0.242	G	A	1326	1558	1.56E-05	0.851 (0.791-0.916)	Intron 22

\*MAF was calculated from the founders (4,655 individuals).

†TDT P-values and Odds Ratio were calculated from the 3,070 multiplex families (12,485 individuals).

‡ Intronic regions were based on Ensembl (CLEC16A - 001 ENST00000409790) (67).

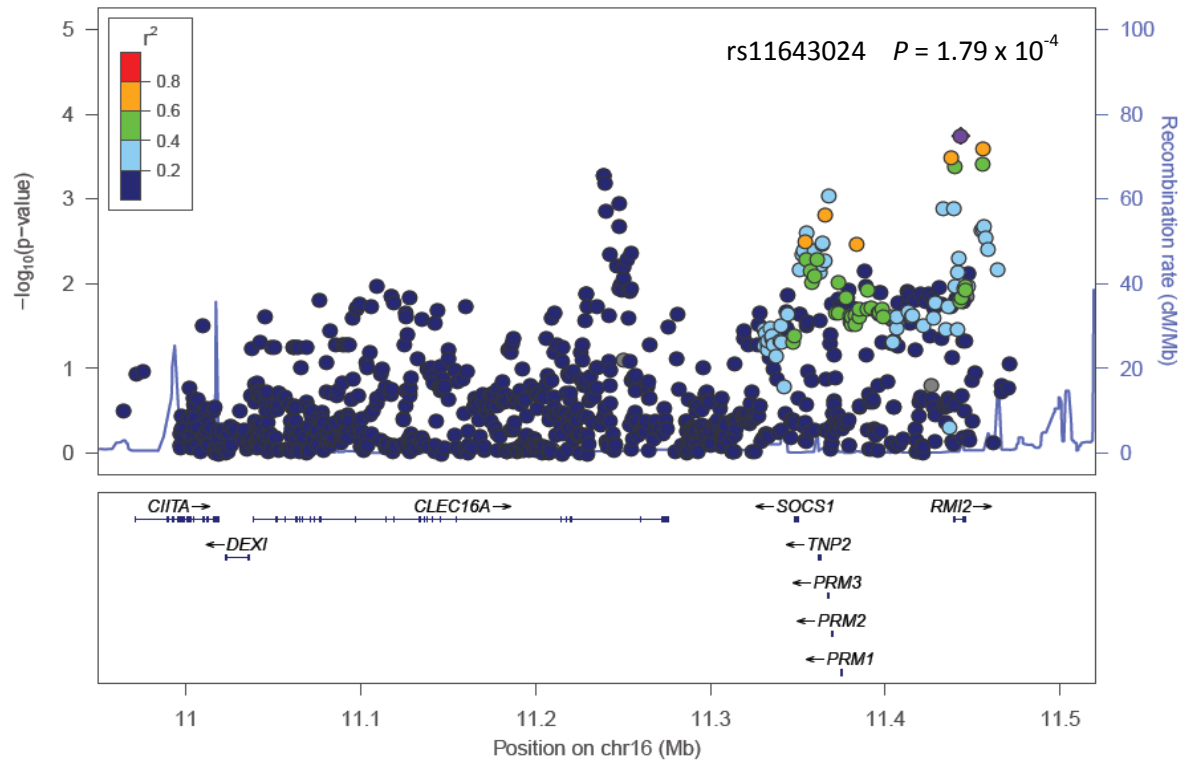
§GWAS reported SNP in association with multiple sclerosis (7). || GWAS reported SNPs in association with type 1 diabetes (2,3).

¶GWAS reported SNP previously reported in association with primary biliary cirrhosis (10).GWAS reported SNPs were limited to the dbGAP NHGRI GWAS catalog.

#Most significant SNP associated with type 1 diabetes. MAF, minor allele frequency.

### Conditioning on Top SNP

In order to determine whether any other SNPs were independently associated with type 1 diabetes in this region, the data was reanalyzed conditioning on the most significantly associated SNP using Unphased (61). An independent, but not statistically significant association was found downstream of *CLEC16A* at SNP rs11643024 ( $P = 1.79 \times 10^{-4}$ ) (Figure 4-12). After identifying the most significant SNPs associated with type 1 diabetes, we investigated whether any of these SNPs showed correlation with mRNA expression level changes for any of the genes in the region. Prior to assessing expression levels, we catalogued alternative splice forms of genes in the region.

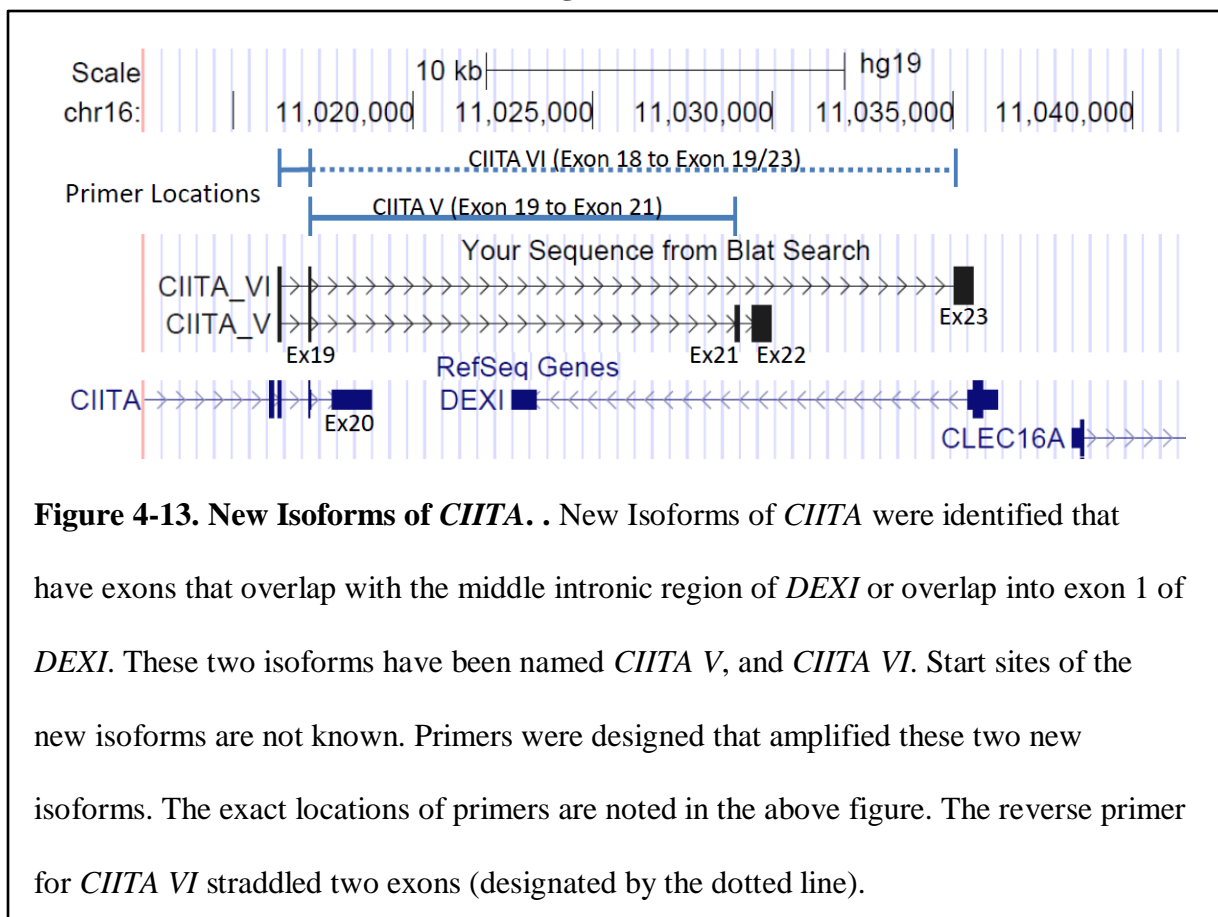
**Figure 4-12****Figure 4-12. Conditioning on the SNP Most Significantly Associated with Type 1**

**Diabetes to Reveal Independently Associated SNPs.** The Unphased program was used to analyze the genotyping data conditioned on the SNP most significantly associated with type 1 diabetes. An independent association was identified downstream with SNP rs11632024 ( $P = 1.79 \times 10^{-4}$ ). No SNPs reached statistical significance. The  $r^2$  was calculated using 1000 Genome March 2012 EUR (Hg19) and the most significant SNP as reference.

### Identification of Novel Splicing Isoforms of *CIITA* Created by Alternative Splicing

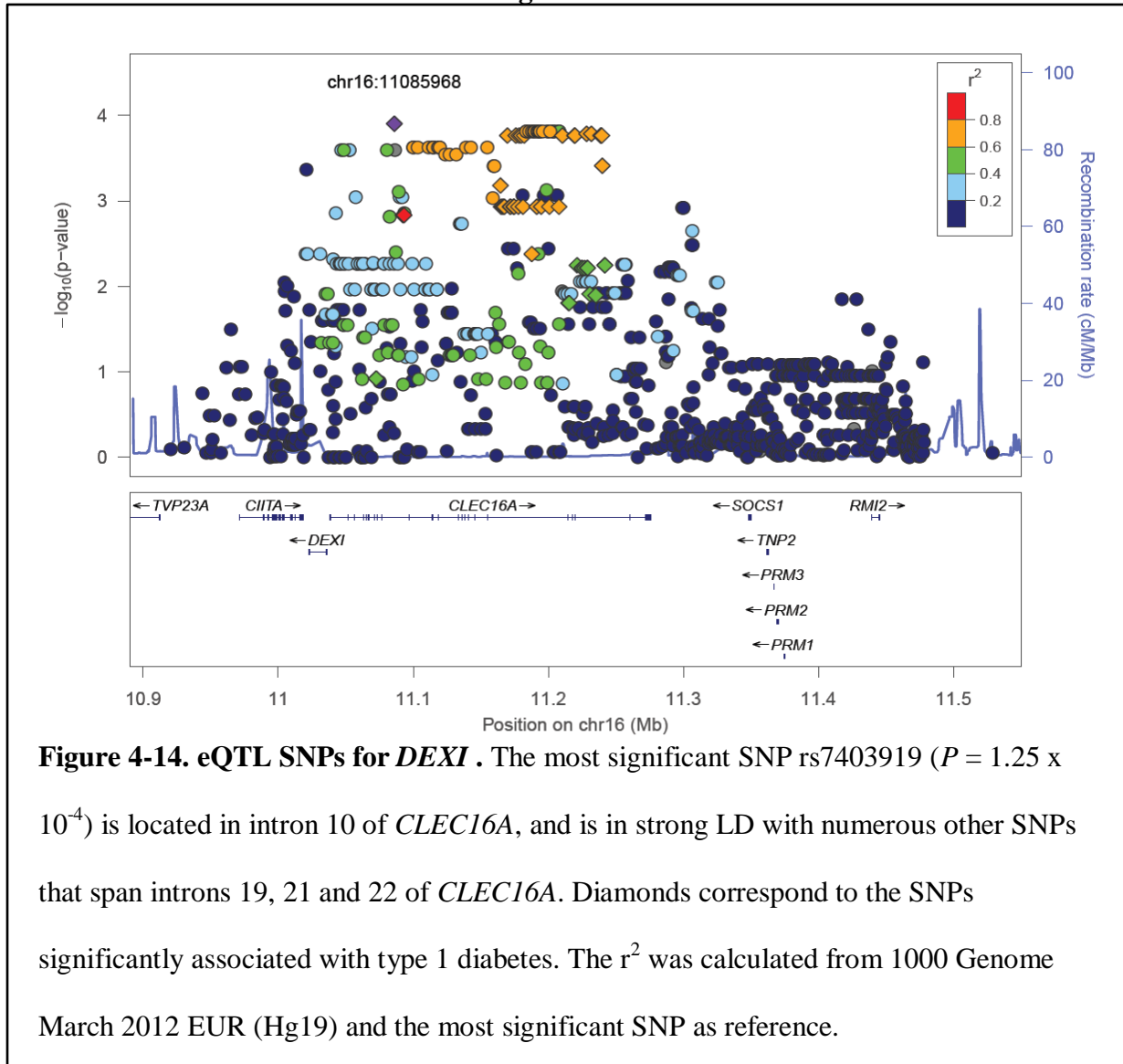
Examination of RNA-Seq data from the Human BodyMap 2.0 project (Illumina Inc.) (data available at (<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/novel>) ) at coordinates (Hg19 chr16:11,010,386,-11,039,093) revealed evidence of infrequent transcripts that included sequences from both *CIITA* and *DEXI*, which are neighboring genes that are transcribed in opposite directions (Figure 4-13). This observation raised the possibility that these genes could exert some regulatory role on each other via anti-sense inhibition. The transcripts were validated and characterized by amplification and Sanger sequencing. Two types of novel isoforms of *CIITA* were identified that overlap with the *DEXI* transcript. One isoform that ends in the intronic region of *DEXI*, referred to here as *CIITA V* and another isoform that overlaps with exon 1 of *DEXI*, referred to as *CIITA VI* (Fig. 4-13). The novel isoforms of *CIITA* have the same stop codon as other *CIITA* isoforms, but differ in the 3'UTR. Real time PCR primers were designed that amplified the two isoforms found and are noted in Figure 4-13.



**Figure 4-13**

### SNPs in *CLEC16A* are Associated with Variation in Expression Level of *DEXI*

Given that one of the significant associations with type 1 diabetes were observed at SNPs located within a coding region, and previous reports suggested that sequences in the 16p13.13 region could act at a distance (*cis*) to regulate the expression of genes in the region, an expression quantitative trait locus analysis (eQTL) was performed using the expression levels of *CIITA*, *DEXI*, *CLEC16A* and *SOCS1* in B-cell lines as phenotypes. Primers that specifically amplified each of the common reported alternative spliced transcripts for the target genes were designed and used to assess transcript levels by quantitative RT-PCR. The full panel of SNPs ( $N = 915$ ) from the region were then tested for association with quantitative variation in the levels of each of the transcripts. The only significant associations were observed for *DEXI* transcripts and a number of SNPs spanning the *CLEC16A* gene, many of which had alleles that were modestly correlated ( $r^2 > 0.6$ ) (Figure 4-14). The most statistically significant association was observed at rs7403919 ( $P = 1.24 \times 10^{-4}$ ). A second independent biological replicate focusing on *DEXI* expression yielded similar results (rs7403919,  $P = 5.45 \times 10^{-5}$ ). A full breakdown of the forty-eight type disease associated SNPs and their correlation ( $r^2$ ) with SNP rs7403919 can be found Table 4-5. All other SNPs tested as eQTL for other genes in the region (*CIITA Isoforms III-VI*, *CLEC16 Isoforms 1-3*, and *SOCS*) did not show statistically significant association with expression levels. The most significant SNP for each gene eQTL test and corresponding  $P$ -value can be found in Table 4-6. Note: *CIITA I* was not measured because it is myeloid dependent (19).

**Figure 4-14**

**Table 4-5**

SNPs Significantly Associated with Type 1 Diabetes with their Corresponding eQTL Results for *DEXI*

SNP	Position (Hg19)	Minor Allele	$\beta^*$	STAT*	P-value*	$r^2$ rs7403919†	$r^2$ rs34306440†	Location‡
rs7403919	11085968	C	0.9029	4.404	1.25E-04	1.000	0.533	Intron 10
rs725613	11169683	C	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs11648679	11175984	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs9929994	11178245	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs12708716	11179873	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs7203793	11182134	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs12935413	11210447	A	0.8884	4.289	1.72E-04	0.940	0.501	Intron 19
rs8064154	11219419	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 21
rs12924112	11219720	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 21
rs4322688	11238991	G	0.8884	4.289	1.72E-04	0.940	0.501	Intron 22
rs12930373	11239599	A	0.8884	4.289	1.72E-04	0.940	0.501	Intron 22
rs2903692	11238783	A	0.8884	4.289	1.72E-04	0.940	0.529	Intron 22
rs35732840	11228355	G	0.8934	4.306	1.64E-04	0.875	0.467	Intron 22
rs2867880	11231857	A	0.8934	4.306	1.64E-04	0.875	0.467	Intron 22
rs12917893	11239978	T	0.8518	3.994	3.88E-04	0.875	0.467	Intron 22
rs7200940	11164567	G	0.7586	3.799	6.61E-04	0.844	0.645	Intron 19
rs8062322	11092319	A	0.7974	3.502	1.47E-03	0.834	0.578	Intron 10
rs9926367	11093179	C	0.7974	3.502	1.47E-03	0.834	0.578	Intron 10
rs11860603	11165017	C	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs9934969	11166002	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs11865121	11166688	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs7198621	11167458	G	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs12925642	11171602	G	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs2058531	11172109	G	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs9652601	11174365	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs9652582	11174564	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs2041670	11174652	T	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs12708715	11177824	T	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs9888908	11181244	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs12928537	11191400	A	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs12927355	11194771	T	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs741172	11200798	T	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs9746695	11207894	C	0.7596	3.586	1.17E-03	0.790	0.622	Intron 19
rs12924729	11187783	A	0.6429	3.1	4.19E-03	0.737	0.511	Intron 19

SNP	Position (Hg19)	Minor Allele	$\beta^*$	STAT*	P-value*	$r^2$ rs7403919†	$r^2$ rs34306440†	Location‡
rs34540843	11221287	G	0.9382	2.98	5.67E-03	0.600	0.680	Intron 22
rs62026379	11241806	A	0.9382	2.98	5.67E-03	0.600	0.680	Intron 22
rs17673553	11241906	G	0.9382	2.98	5.67E-03	0.600	0.680	Intron 22
rs34306440§	11215035	G	0.8521	2.561	1.57E-02	0.533	1.000	Intron 20
rs35032408	11215424	G	0.8521	2.561	1.57E-02	0.533	1.000	Intron 20
rs36045143	11224966	G	0.9253	2.95	6.10E-03	0.500	0.600	Intron 22
rs2241099	11225064	G	0.9253	2.95	6.10E-03	0.500	0.600	Intron 22
rs62026376	11228712	T	0.9253	2.95	6.10E-03	0.500	0.600	Intron 22
rs62026377	11229128	T	0.9253	2.95	6.10E-03	0.500	0.600	Intron 22
rs12919732	11235123	G	0.8494	2.649	1.28E-02	0.500	0.600	Intron 22
rs12919828	11226779	G	0.9253	2.95	6.10E-03	0.485	0.592	Intron 22
rs7203459	11230703	C	0.8583	2.663	1.23E-02	0.412	0.529	Intron 22
rs12708713	11072459	T	0.5901	1.6	1.20E-01	0.394	0.638	Intron 8
rs12708714	11072525	C	0.5901	1.6	1.20E-01	0.394	0.638	Intron 8

Samples are sorted in order of  $r^2$  relative to rs740319 (most significant SNP that is associated with *DEXI* expression levels).

\*Statistical analysis performed using expression and genotyping data.

† $r^2$  values were calculated using genotyping data for 33 HapMap samples and the designated SNP as reference. ‡Intronic regions were based on Ensembl (CLEC16A - 001 ENST00000409790) (67).

§Most Significant SNP associated with type 1 diabetes.

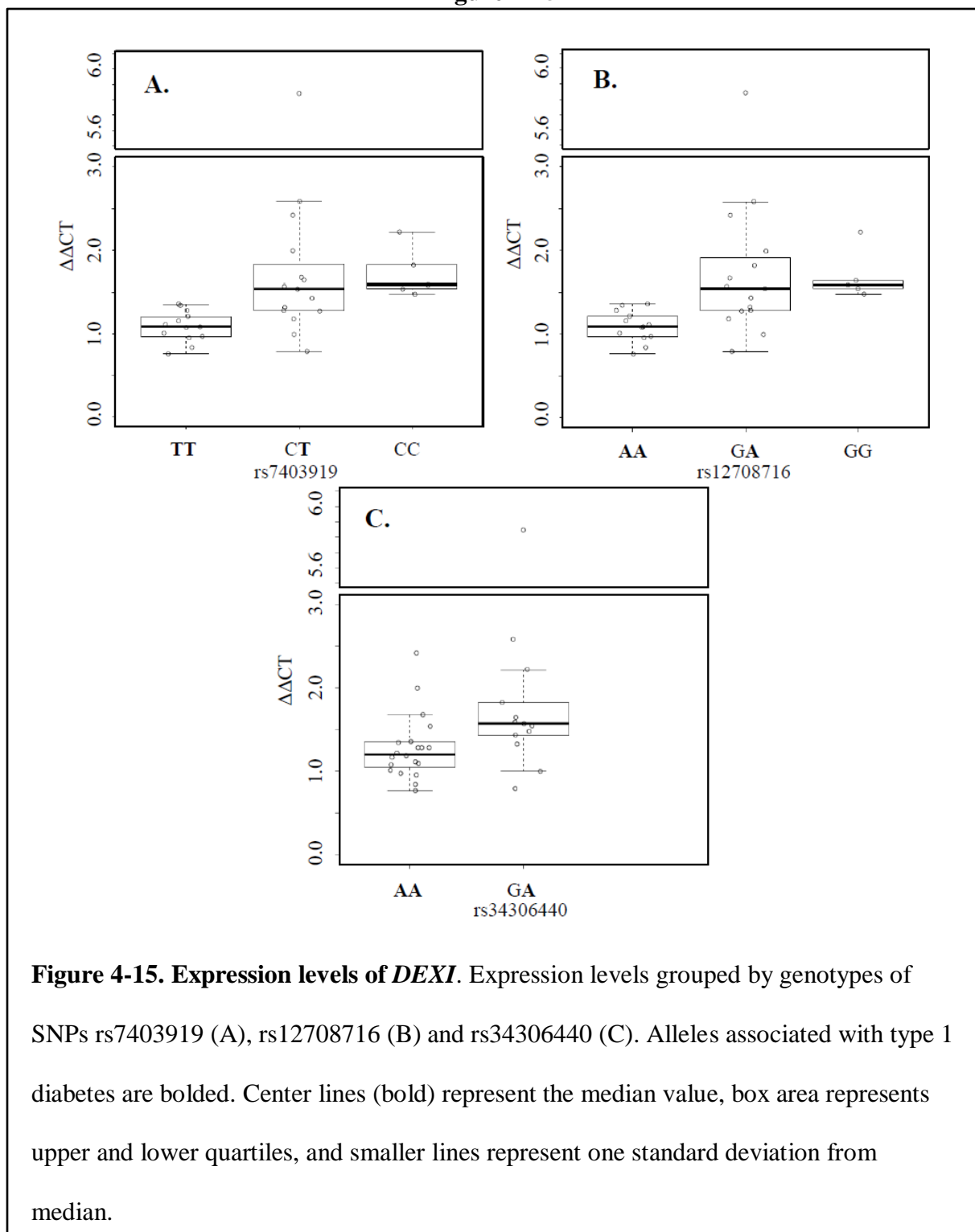
$\beta$ , linear regression coefficient; STAT, coefficient t-statistic;  $r^2$ , correlation.

**Table 4-6**Top eQTL SNP *P*-values for each gene tested

<b>Gene</b>	<b>Most Sig. SNP</b>	<b>BP (Hg19)</b>	<b><i>P</i>-value</b>
<i>CIITA III</i>	rs7199828	11006262	$2.14 \times 10^{-3}$
<i>CIITA IV</i>	rs4381620	11477757	$1.26 \times 10^{-2}$
<i>CLEC16A Iso I</i>	rs431918	11365416	$5.36 \times 10^{-3}$
<i>CLEC16A Iso II</i>	rs243329	11352313	$3.33 \times 10^{-2}$
<i>CLEC16A Iso III</i>	rs11074964	11455169	$7.64 \times 10^{-4}$
<i>DEXI</i> *	rs7403919	11085968	$1.25 \times 10^{-4}$
<i>SOCS1</i>	rs77309215	11261183	$3.20 \times 10^{-3}$
<i>CIITA V</i>	rs34033173	10982933	$1.41 \times 10^{-2}$
<i>CIITA VI</i>	rs4129933	11528893	$2.68 \times 10^{-2}$

\*Only SNP *P*-value that reached statistical significance was found in the eQTL study for *DEXI* located in intron 10 of *CLEC16A*. (Bonferroni corrected  $P \leq 1.37 \times 10^{-4}$ )

We specifically examined the expression levels of *DEXI* in relation to three SNPs: rs34306440 (the SNP most strongly associated with type 1 diabetes in the region), rs7403919 (the SNP most strongly associated with variation in *DEXI* expression in the region) and rs12708716 (most often cited in type 1 diabetes literature) (Figure 4-15). For all three SNPs the risk allele always corresponded to the major allele. The homozygotes for the risk alleles had lower expression of *DEXI*, whereas the heterozygotes and homozygotes for the alternative allele displayed higher levels of expression.

**Figure 4-15**

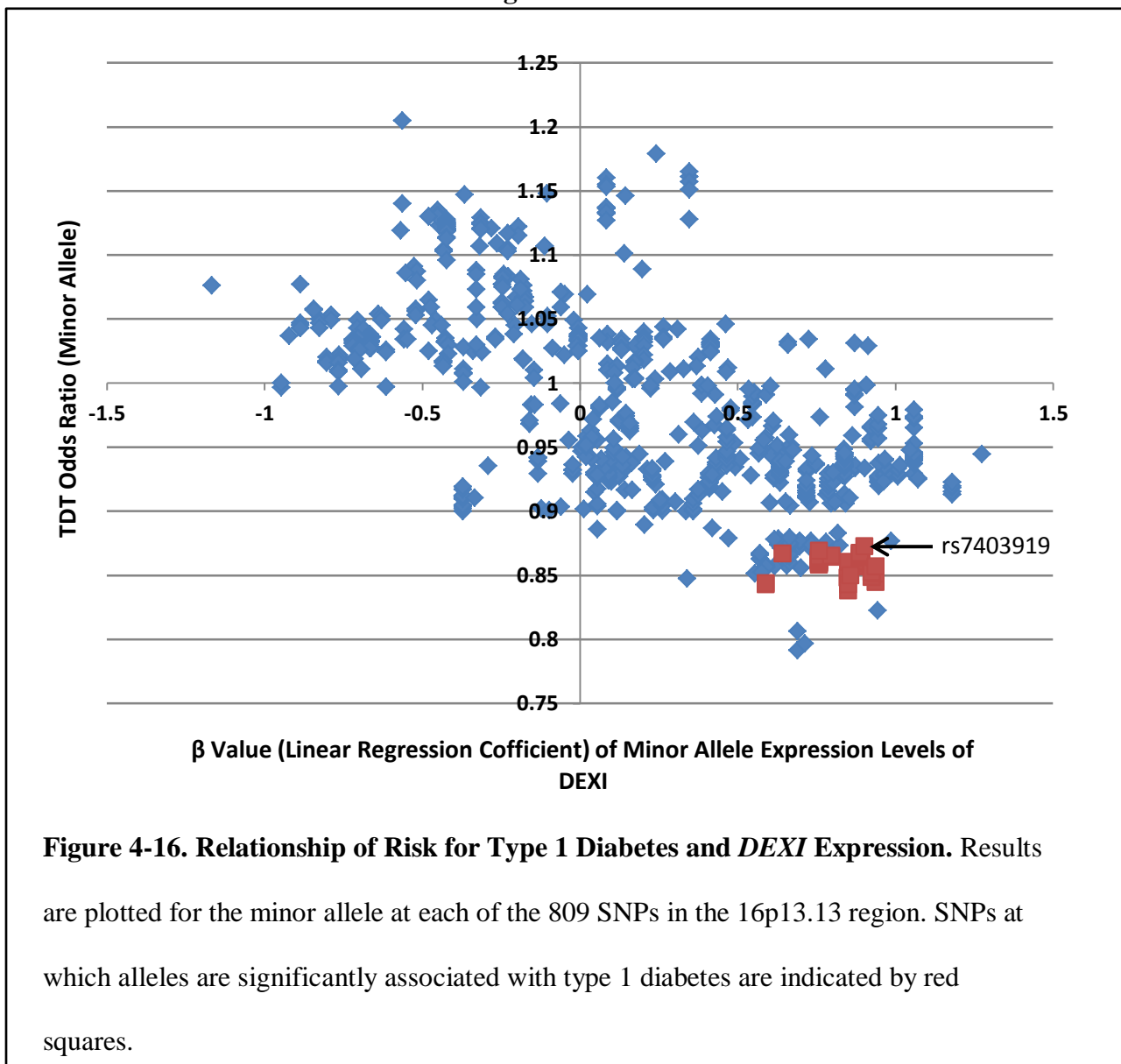
**Figure 4-15. Expression levels of *DEXI*.** Expression levels grouped by genotypes of SNPs rs7403919 (A), rs12708716 (B) and rs34306440 (C). Alleles associated with type 1 diabetes are bolded. Center lines (bold) represent the median value, box area represents upper and lower quartiles, and smaller lines represent one standard deviation from median.



**Relationship of SNPs Associated with Type 1 Diabetes and/or DEXI Expression**

In order to more broadly characterize the relationship between disease risk and *DEXI* expression we plotted the odds ratio (OR) for type 1 diabetes versus the linear regression coefficient ( $\beta$ ) representing *DEXI* expression level for the minor alleles of all 809 SNPs for which these data were available (Figure 4-16). Disease risk was negatively correlated with *DEXI* transcript levels, Pearson correlation ( $R = -0.66$ ). All SNPs for which statistically significant associations with type 1 diabetes were observed clustered in the lower right quadrant of the graph corresponding to the low disease risk and increased *DEXI* expression.

Figure 4-16



## **Discussion**

Chromosomal region 16p13.13 is one of a large number of regions where there is significant, replicated evidence of allelic association with type 1 diabetes (2-5), as well as with other autoimmune disorders. As these different disorders have disparate target organs and pathologies, such findings of association imply that the loci in question have basic autoimmunity-predisposing effects.

Repeated findings of association for multiple autoimmune disorders in a single chromosomal region like 16p13.13 increases both confidence in the initial findings and the importance of the locus in understanding the underlying mechanism whereby autoimmunity develops. However, functional studies that might elucidate the mechanism depend upon the ability to specify the relevant gene in the region, if not the specific causative genetic variants. In the current study, we addressed this need for the 16p13.13 region by resequencing the region in a large collection of type 1 diabetes cases and unaffected controls to identify all common variants, genotyping these variants in a collection of multiplex type 1 diabetes families, and then separately testing for association of alleles at these SNPs with disease state and transcript levels for each of the genes in the region.

There were 46 novel variants from our re-sequencing that were included among the 939 SNPs that we genotyped in 3,070 multiplex families with type 1 diabetes. Forty-eight of these SNPs displayed a statistically significant association with type 1 diabetes. These included several SNPs that provided the most significant findings in this region in other autoimmune diseases such as rs11865121 in MS (7) and rs12924729 in PBC (10).

In all cases, it was the major allele that conferred risk for type 1 diabetes. Alleles at these 48 SNPs were not all highly correlated (pairwise  $r^2$  range from 0.4 to 0.9), nor were the SNPs physically clustered. All of these 48 SNPs were located in introns that spanned much of the *CLEC16A* gene (introns 8-22).

The broad physical distribution of disease associated SNPs and their modest correlation raise the possibility that they either act through multiple mechanisms or the SNPs could affect multiple genes. Given their intronic locations and prior reports of cis-oriented transcriptional effects in the region, we tested for association between our entire panel of SNPs and transcripts levels for each of the documented alternatively spliced forms of the four genes in the region, *CIITA*, *DEXI*, *CLEC16A* and *SOCS1*. Only for *DEXI* were statistically significant results obtained. Similar to the disease association results, SNPs associated with *DEXI* transcript levels appear to be distributed across much of the *CLEC16A* gene. A comparison of *DEXI* expression levels in cell lines grouped by genotype at one of three disease-associated SNPs suggested that risk for disease was associated with reduced expression of *DEXI*. To explore this hypothesis more globally, we examined the relationship between the odds ratio for the minor allele at each SNP in the region and the corresponding linear regression coefficient for *DEXI* expression for the same allele. *DEXI* expression was negatively correlated with disease risk ( $R = -0.66$ ).

Our results suggest that genetic variation over a broad region spanning much of the *CLEC16A* gene influences the expression of a neighboring gene, *DEXI* and that reduced *DEXI* expression is correlated with increased risk for type 1 diabetes. This model is consistent with a prior report that sequences in intron 19 of *CLEC16A* can interact with

the promoter region of *DEXI* (45). Our results suggest that the region of interaction may be large or that distal sequences may act indirectly to affect the stability or structure of the DNA loops formed.

Our results and those from prior studies focus attention on *DEXI* as the gene most likely to affect risk for type 1 diabetes in the 16p13.13 region. Little is known about the function of *DEXI* and its predicted 95 amino acid protein product. The protein is well-conserved, with orthologues readily detectable in multiple species. There is 95% identity between the human and mouse orthologues and 84.5% between human and zebrafish.

There are no identifiable domains in *DEXI* that might suggest a possible function for the protein. However, Miyaki *et al* showed that siRNA mediated knockdown of *DEXI* reduced the incidence of apoptosis in cells treated with the chemotherapy agent camptothecin (68). A role in a basic cellular function such as apoptosis would be consistent with the involvement of *DEXI* in multiple autoimmune diseases with different target tissues.

**Acknowledgements for Research Work**

The author acknowledges Dr. Stephen Rich for providing the T1DGC ImmunoChip data. The author acknowledges the support of staff of the Genome Sciences Laboratory Emily Farber, Daniel Gallo, Jamie Artale and Benjamin Artale in the Center for Public Health Genomics at the University of Virginia for training and maintenance of the genotyping equipment used in the current study and for generation of the ImmunoChip data. This work was supported by grants from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (R01DK46635 and DP3DK085678) and Juvenile Diabetes Research Foundation International (JDRF) (33-2008-409) to PC. This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418.

### **Work Cited**

1. Concannon P, Rich SS, Nepom GT: Genetics of Type 1A Diabetes. *The New England Journal of Medicine* 2009;360:1646-1654
2. Barrett JC, Clayton DG, Concannon P, *et al.*: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics* 2009;41:703-707
3. Hakonarson H, Grant SF, Bradfield JP, *et al.*: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007;448:591-594
4. Todd JA, Walker NM, Cooper JD, *et al.*: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics* 2007;39:857-864
5. Consortium WTCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-678
6. Consortium IMMSG: Risk Alleles for Multiple Sclerosis Identified by a Genomwide Study. *The New England Journal of Medicine* 2007;357:851-862
7. De Jager PL, Jia X, Wang J, *et al.*: Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nature genetics* 2009;41:776-782
8. Dubois PC, Trynka G, Franke L, *et al.*: Multiple common variants for celiac disease influencing immune gene expression. *Nature genetics* 2010;42:295-302
9. Gateva V, Sandling JK, Hom G, *et al.*: A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature genetics* 2009;41:1228-1233

10. Mells GF, Floyd JA, Morley KI, *et al.*: Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nature genetics* 2011;43:329-332
11. Nischwitz S, Cepok S, Kroner A, *et al.*: More CLEC16A gene variants associated with multiple sclerosis. *Acta neurologica Scandinavica* 2011;123:400-406
12. Zhang Z, Cheng Y, Zhou X, *et al.*: Polymorphisms at 16p13 are associated with systemic lupus erythematosus in the Chinese population. *Journal of medical genetics* 2011;48:69-72
13. Marquez A, Varade J, Robledo G, *et al.*: Specific association of a CLEC16A/KIAA0350 polymorphism with NOD2/CARD15(-) Crohn's disease patients. *European journal of human genetics : EJHG* 2009;17:1304-1308
14. McGovern DP, Gardet A, Torkvist L, *et al.*: Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nature genetics* 2010;42:332-337
15. Skinningsrud B, Husebye ES, Pearce SH, *et al.*: Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. *The Journal of clinical endocrinology and metabolism* 2008;93:3310-3317
16. Cooper JD, Smyth DJ, Smiles AM, *et al.*: Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nature genetics* 2008;40:1399-1401
17. Patsopoulos NA, Bayer Pharma MSGWG, Steering Committees of Studies Evaluating I-b, *et al.*: Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of neurology* 2011;70:897-912



18. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265
19. LeibundGut-Landmann S, Waldburger JM, Krawczyk M, *et al.*: Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. *European journal of immunology* 2004;34:1513-1525
20. Reith W, Mach B: The bare lymphocyte syndrome and the regulation of MHC expression. *Annual review of immunology* 2001;19:331-373
21. Swanberg M, Lidman O, Padyukov L, *et al.*: MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nature genetics* 2005;37:486-494
22. Holling TM, van Eggermond MC, Jager MJ, van den Elsen PJ: Epigenetic silencing of MHC2TA transcription in cancer. *Biochemical pharmacology* 2006;72:1570-1576
23. Ni Z, Abou El Hassan M, Xu Z, Yu T, Bremner R: The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. *Nature immunology* 2008;9:785-793
24. Reith W, Boss JM: New dimensions of CIITA. *Nature immunology* 2008;9:713-714
25. Kent WJ, Sugnet CW, Furey TS, *et al.*: The Human Genome Browser at UCSC. *Genome Research* 2002;12:996-1006
26. Edgar AJ, Birks EJ: Cloning of Dexamethasone-Induced Transcript. *Am J Respir Cell Mol Biol* 2001;25:119-124
27. Stelzl U, Worm U, Lalowski M, *et al.*: A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005;122:957-968

28. Kladney RD, Bulla GA, Guo L, *et al.*: GP73, a Novel Golgi-Localized Protein Upregulated by Viral Infection. *Gene* 2000;249:53-65
29. Lewis SA, Tian G, Cowan NJ: The alpha and beta Tubulin Folding Pathways. *Cell Biology* 1997;7:479-484
30. Abbas AK, Lichtman AH, Pillai S: *Cellular and Molecular Immunology*. China, Saunders Elsevier, 2010
31. Wu C, Orozco C, Boyer J, *et al.*: BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology* 2009;10:R130
32. Kim S, Wairkar YP, Daniels RW, DiAntonio A: The novel endosomal membrane protein Ema interacts with the class C Vps-HOPS complex to promote endosomal maturation. *The Journal of cell biology* 2010;188:717-734
33. Palmer DC, Restifo NP: Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. *Trends in immunology* 2009;30:592-602
34. Illangumaran S, Rottapel R: Regulation of cytokine receptor signaling by SOCS1. *Immunological Reviews* 2003;192:196-211
35. Naka T, Matsumoto T, Narazaki M, *et al.*: Accelerated apoptosis of lymphocytes by augmented induction of Bax in SSI-1 (STAT-induced STAT inhibitor-1) deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 1998;95:15577-15582
36. Starr R, Metcalfe K, Elefanty AG, *et al.*: Liver degeneration and lymphoid deficiencies in mice lacking suppressor of cytokine signaling-1. *Proceedings of the National Academy of Sciences of the United States of America* 1998;95:14395-14399

37. Marine JC, Topham DJ, McKay C, *et al.*: SOCS1 deficiency causes a lymphocyte-dependent perinatal lethality. *Cell* 1999;98:609-616
38. Seki Y, Yang J, Okamoto M, *et al.*: IL-7/STAT5 cytokine signaling pathway is essential but insufficient for maintenance of naive CD4 T-cell survival in peripheral lymphoid organs. *Journal of Immunology* 2007;178:262-270
39. Mooney RA, Senn J, Cameron S, *et al.*: Suppressors of cytokine signaling-1 and -6 associate with and inhibit the insulin receptor. A potential mechanism for cytokine-mediated insulin resistance. *The Journal of biological chemistry* 2001;276:25889-25893
40. Rui L, Yuan M, Frantz D, Shoelson S, White MF: SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *The Journal of biological chemistry* 2002;277:42394-42398
41. Awata T, Kawasaki E, Tanaka S, *et al.*: Association of type 1 diabetes with two Loci on 12q13 and 16p13 and the influence coexisting thyroid autoimmunity in Japanese. *The Journal of clinical endocrinology and metabolism* 2009;94:231-235
42. Cooper JD, Walker NM, Smyth DJ, *et al.*: Follow-up of 1715 SNPs from the Wellcome Trust Case Control Consortium genome-wide association study in type I diabetes families. *Genes and immunity* 2009;10 Suppl 1:S85-94
43. Reddy MV, Wang H, Liu S, *et al.*: Association between type 1 diabetes and GWAS SNPs in the southeast US Caucasian population. *Genes and immunity* 2011;12:208-212
44. Wu X, Zhu X, Wang X, *et al.*: Intron polymorphism in the KIAA0350 gene is reproducibly associated with susceptibility to type 1 diabetes (T1D) in the Han Chinese population. *Clinical endocrinology* 2009;71:46-49

45. Davison LJ, Wallace C, Cooper JD, *et al.*: Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Human molecular genetics* 2012;21:322-333
46. Nica AC, Montgomery SBD, A. S., Stranger BE, *et al.*: Candidate Causal Regulatory Effects by Intergration of Expression QTLs with Complex Trait Genetic Associations. *PLOS Genetics* 2010;6:11
47. Zeller T, Wild P, Szymczak S, *et al.*: Genetics and Beyond - The Transcriptome of Human Monocytes and Disease Susceptibility. *PLOS One* 2010;5:15
48. International HapMap C: A haplotype map of the human genome. *Nature* 2005;437:1299-1320
49. Genomes Project C: A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073
50. Type 1 Diabetes Genetics Consortium (T1DGC) [article online], Available from <https://www.t1dgc.org>. 2013
51. Lernmark A, Ducat L, Eisenbarth G, *et al.*: Family Cell Lines Available for Research. *American journal of human genetics* 1990;47:1028-1030
52. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-1760
53. Li H, Handsaker B, Wysoker A, *et al.*: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078-2079
54. Consortium IHGS: Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921

55. Purcell S, Neale B, Todd-Brown K, *et al.*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;81:559-575
56. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American journal of human genetics* 1998;63:259-266
57. Cortes A, Brown MA: Promise and pitfalls of the Immunochip. *Arthritis Research & Therapy* 2011;13:1-3
58. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* 2010;34:816-834
59. Manichaikul A, Mychaleckyj JC, Rich SS, *et al.*: Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;26:2867-2873
60. Pruim RJ, Welch RP, Sanna S, *et al.*: LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336-2337
61. Dudbridge F: Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Human heredity* 2008;66:87-98
62. Klein RJ, Zeiss C, Chew EY, *et al.*: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385-389
63. McIntyre LM, Martin ER, Simonsen KL, Kaplan NL: Circumventing multiple testing: a multilocus monte carlo approach to testing for association. *Genetic epidemiology* 2000;19:18-29

- 64.** Koressaar T, Remm M: Enhancements and modifications of primer design program Primer3. *Bioinformatics* 2007;23:1289-1291
- 65.** Hornell TMC, Beresford GW, Bushey A, Boss JM, Mellins ED: Regulation of the Class II MHC Pathway in Primary Human Monocytes by Granulocyte-Macrophage Colony-Stimulating Factor. *Journal of Immunology* 2003;171:2374-2383
- 66.** Spielman RS, McGinnis RE, Ewens WJ: Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM). *American journal of human genetics* 1993;52:506-516
- 67.** Flicek P, Ahmed I, Amode MR, *et al.*: Ensembl 2013. *Nucleic acids research* 2013;41:D48-55
- 68.** Miyaki Y, Suzuki K, Koizumi K, *et al.*: Identification of a potent epigenetic biomarker for resistance to camptothecin and poor outcome to irinotecan-based chemotherapy in colon cancer. *International journal of oncology* 2012;40:217-226

## **Chapter 5**

### **Future Directions**

## **Introduction**

Clearly from this study we have unanswered questions about chromosomal region 16p13.13. Based on our data future research should be focused on understanding *DEXI*'s function and how changes in its expression confer risk for type 1 diabetes. Even though in our experiments we encountered novel isoforms of *CIITA* that could influence the region, none of our data supported these isoforms involvement. This finding may be due to lack of statistical power, wrong cell type tested or cytokine stimulation is needed, however due the lack of evidence for involvement the novel isoforms will not be examined further. Since chromosomal region 16p13.13 is one region out of many associated with type 1 diabetes, a much broader discussion of future studies with a more global approach will conclude the dissertation.

## **Investigating *DEXI***

Understanding the function of the previously uncharacterized protein, *DEXI*, is the most essential part of understanding how this region confers risk for type 1 diabetes and other autoimmune diseases. As mentioned previously, *DEXI* is a 14 KB gene that is transcribed in the opposite direction to *CIITA* and *CLEC16A*. *DEXI* is composed of two exons (only 1 coding) and codes for a protein 95 amino acids long (1).

*DEXI* has been implicated in two possible immune response related pathways. In yeast two-hybrid screens, *DEXI* was found to interact with the Golgi membrane protein 1 (*GOLM1*) and tubulin-specific chaperone D protein (*TBCD*) (2). *GOLM1*, also known as *GP73*, has been found to be up-regulated in a viral infection (3) and *TBCD* is a protein



involved in helping beta-tubulin fold correctly (4). In another study siRNA mediated knockdown of *DEXI*, reduced the incidence of apoptosis in cells treated with the chemotherapy agent camptothecin (5). This indicates that *DEXI* is possibly involved in either an anti-viral response or apoptosis or a combination of both.

Interestingly, there are two intronless retro-*DEXI* genes located at 15q11.2 and 15q13.1 in a chromosomal region known for duplications and imprinting which have been implicated in causing Prader-Willi/Angelman syndrome (6,7). Davison *et al* showed that the 15q11.2 version is minimally expressed in cells, especially lymphoblastoid cells (8) and version 15q13.1 straddles across two genes. Therefore the duplication of *DEXI* raises evolutionary questions and also complicates future experiments that need to be carefully controlled for the presence of these pseudogenes.

*DEXI* has no identifiable domains that could suggest function. The major focus for future research should be investigating expression levels of *DEXI* in other tissues and specifically investigating individuals with the risk alleles. Based on our data it appears expression levels of *DEXI* are the key mode of conferring risk for disease. However further investigation into the protein *DEXI* is needed in order to completely understand how it functions in a cell and causes apoptosis.

Using the previously utilized primer pair the expression levels of *DEXI* can be investigated in other tissues and cell types. It is important when performing this assay that activated T and B cells be investigated as well as resting cells because the levels of expression can easily change in inactivated/activated immune cells. *DEXI* was originally identified as being upregulated in the lungs of emphysema patients (9), and this

upregulation was suggested to be likely due to the exposure to cigarette smoke (8), hinting at an immune response to toxic chemicals (apoptosis??). This also implies that investigating DEXI expression specifically in type 1 diabetic patients is essential for defining the specific immunological response. However, the change in mRNA level does not always correspond to measureable difference in protein levels, so investigating the protein levels is important.

Measuring DEXI's protein levels and determining what proteins it interacts with are essential. In order to begin such studies, high affinity antibodies that stain DEXI and other possible interacting proteins need to be designed or purchased. Currently, commercial antibodies for DEXI are available from Sigma Aldrich and other sources. As previously mentioned, in a yeast two-hybrid screen DEXI was found to interact with GOLM1 and TBCD (2). GOLM1 might be a good candidate for further study because of its involvement in virus response. Commercial antibodies are available for this protein from Sino Biological Inc. The TBCD protein seems like an unlikely candidate for involvement in an autoimmune disease although an unknown biochemical relationship could exist.

Using the high affinity antibodies or epitope tag, protein complex immunoprecipitation (co-IP) experiments can be performed to confirm DEXI's expression levels and its interactions with GOLM1. Once a successful co-IP has been performed, the protein complex can be subjected to mass spectrometry. Mass spectrometry will allow for the validating of known interacting proteins and identification of any unknown interactions. It will also allow for identification of any post translational

modifications that occur. Identification of unknown interacting proteins will also help decipher the exact pathway that *DEXI* interacts in that up regulates cell apoptosis.

Determining the location and movement of the *DEXI* protein is the final phase to better understand how it triggers cell apoptosis. The protein will be tagged with a fluorescent tag like GFP and expressed in a cell. Using immunofluorescent microscopy the protein's exact location can be visualized. It can be speculated based on the interaction with *GOLM1* that *DEXI* could interact with/around the Golgi apparatus.

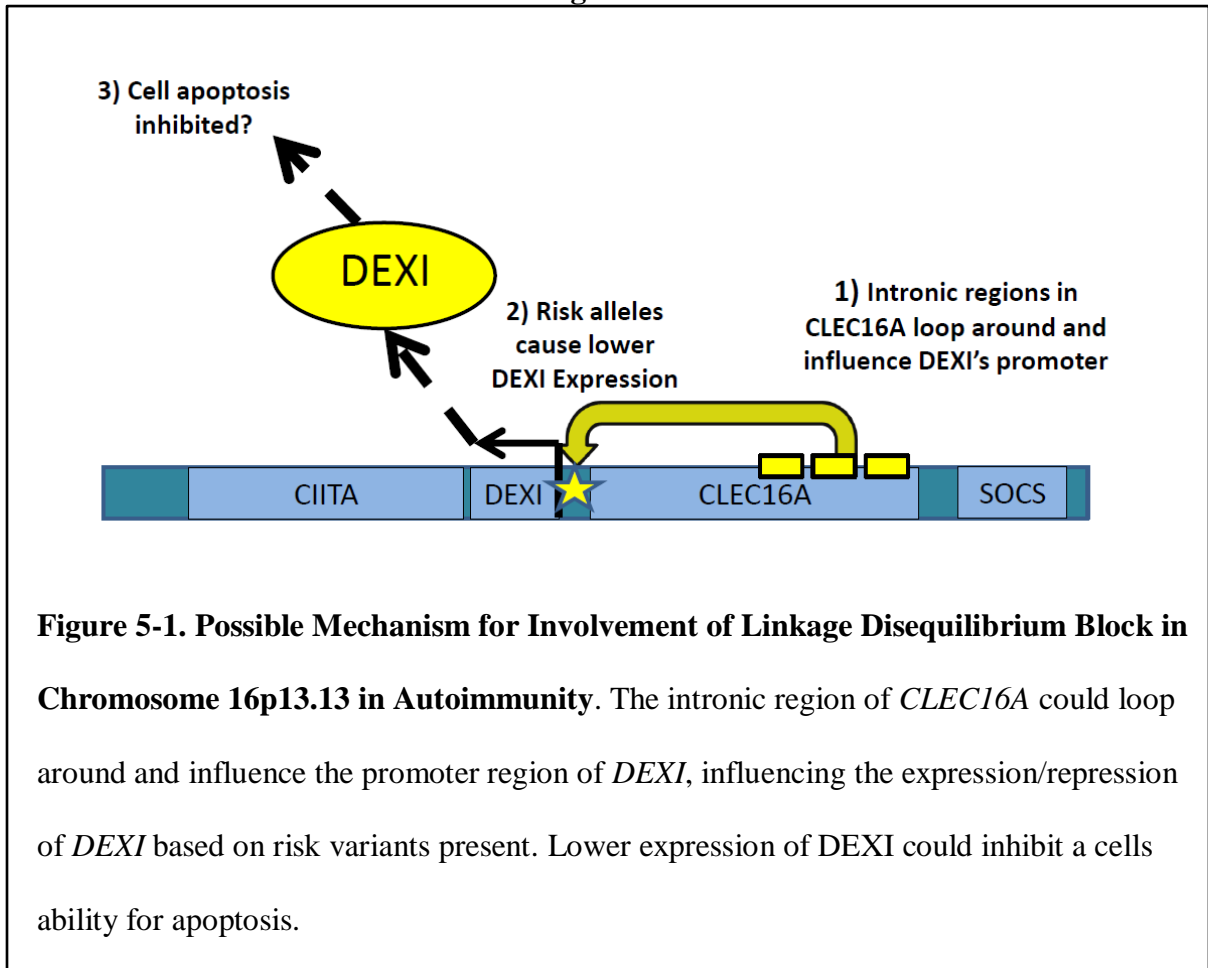
Deciphering the expression levels of *DEXI* in various tissues and individuals with type 1 diabetes is key to confirming how the region confers risk. Further investigations into the protein are needed to validate the change in expression seen in mRNA. Finally, validating/determining the proteins that *DEXI* interacts with and its location in the cell is important to decipher how it possibly causes apoptosis in a cell.

### **Possible Mechanisms Conferring Risk**

Based on our findings the *CLEC16A* intronic region influences the expression of neighboring gene *DEXI*. Alleles associated with risk for type 1 diabetes correspond to lower expression of *DEXI*. This model is consistent with a previous finding where intron 19 of *CLEC16A* was found to interact with the promoter region of *DEXI* (8). However, our data suggests that this interaction may span multiple introns or the distal sequences in neighboring introns, which surround intron 19, influences the stability or structure of the loop. Now exactly how lower expression of *DEXI* confers risk is more theoretical and currently based on limited findings. As mentioned previously it was recently found that

DEXI interacts with GOLM1 (2), a golgi membrane protein upregulated in a virus infection. Also it was recently found when DEXI is knocked out by siRNA it inhibits the cell's ability for apoptosis (5). Based on all these findings a possible model can be hypothesized where the intronic regions of *CLEC16A* loop around and influence *DEXI*'s expression which inhibits a cells ability for apoptosis (Figure 5-1).

Exactly how this increases a individuals risk for type 1 diabetes will be explored further, but is all speculative. Assuming type 1 diabetes is triggered by a virus, lower *DEXI* expression could confer risk by inhibiting cell death during a viral infection in the pancreatic beta cells. By inhibiting apoptosis of cells a longer infection could occur, resulting in a more aggressive immune response later. A more aggressive immune response could result in a greater chance for autoimmunity to develop in the pancreas. This model is supported by the fact that DEXI interacts with GOLM1, a protein upregulated by a virus. A second model that is not viral based can also be hypothesised based on the inhibition of apoptosis. During the development of B and T cells they undergo a process of negative selection where cells that tightly bind to self antigens are removed. If any of these self reactive cells are inhibited from apoptosis, possibly caused by lower DEXI expression, they could allow for the escape of autoreactive T and B cells into the body and favor the development of autoimmunity. Clearly, these two models are very speculative and further investigations into the exact function of DEXI are needed to identify the exact mechanism of action.

**Figure 5-1**

### **Future Directions of Type 1 Diabetes Research**

Clearly more time and resources are needed for deciphering the genetics behind type 1 diabetes. With the advent of GWAS the list of possible candidate regions has grown drastically and on average GWAS reported loci contain over 8.5 protein coding genes per region (10). Each of these regions require further fine mapping to identify the causative risk variant/variants. However, as seen in Chapter 4, fine mapping the region is very time consuming and may not truly identify the causative risk variant because of strong LD. Complicating issues further, as seen with chromosomal region 16p13.13, the risk variant identified by GWAS may confer risk for the disease by influencing a neighboring gene.

Future genetic studies of type 1 diabetes and other autoimmune diseases should utilize large expression microarrays in combination with GWAS panels to perform comprehensive eQTL analysis. This combination of investigation will ensure resources are not wasted pursuing genes that are associated with risk, yet have no direct influence on disease as seen with *CLEC16A*. This will narrow down the functional regions that confer risk for disease. Also, utilizing GWASs and expression data from other autoimmune disorders will help prioritize follow-up genes to study. The exact phenotypic change could ultimately be different as seen with *IL2RA* in type 1 diabetes and multiple sclerosis.

After thorough investigations of several regions identified in GWASs by fine mapping, genotyping and functional studies, SNPs were identified that clearly increase or decrease an individual's risk for disease. These risk-associated SNPs will be entered in a

database for type 1 diabetes and other autoimmune related diseases. This will add to the screening of individuals for risk with disease in a tiered risk system (11). It can be hypothesized that individuals that have the high risk MHC class II haplotype DR3/DR4, can have further increased risk if they have the HLA-A\*02 allele (12) and an even greater risk if they have the shortest VNTR (class I) of the insulin gene (13).

Functional SNPs from GWAS constitute the next tier for example if the individual has risk variant/variants from *CLEC16A* that lowers *DEXI* expression this further increases their risk. However, this additive predicative model is indeed speculative at the time and per current findings the predictive power for type 1 diabetes risk after accounting for HLA is small (14,15). This limitation may be overcome by utilizing larger datasets that look at the complex interactions of SNPs (15). After all regions identified by GWAS have been fine mapped and risk variants identified, the predictive power may show a modest improvement. Currently, the established tool for predicting type 1 diabetes is identifying individuals with the high risk haplotype and monitoring them for development of autoantibodies (16,17). As previously mentioned in Chapter 1, an individual's risk for developing type 1 diabetes increases as the number of autoantibodies increases rather than the titre of autoantibodies (18-20). Utilizing the tiered screening system in combination with testing for antibodies may help improve the predictive model and future screening.

This tiered system would allow for more precise and personally adapted treatment/prevention for an individual's risk for developing type 1 diabetes and other autoimmune diseases. Utilizing this screening method might also help to ultimately

identify the environmental trigger/triggers of type 1 diabetes and sub types of the disease. Currently, type 1 diabetics are diagnosed based on the fact they no longer produce insulin, but they could be sub-grouped based on age of onset and severity. In the future, each individual may be assessed for combinations of factors which ultimately predict their risk for developing type 1 diabetes and other autoimmune diseases and allow for medical intervention/prevention.



### **Work Cited**

1. Kent WJ, Sugnet CW, Furey TS, *et al.*: The Human Genome Browser at UCSC.  
*Genome Research* 2002;12:996-1006
2. Stelzl U, Worm U, Lalowski M, *et al.*: A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005;122:957-968
3. Kladney RD, Bulla GA, Guo L, *et al.*: GP73, a Novel Golgi-Localized Protein Upregulated by Viral Infection. *Gene* 2000;249:53-65
4. Lewis SA, Tian G, Cowan NJ: The alpha and beta Tubulin Folding Pathways. *Cell Biology* 1997;7:479-484
5. Miyaki Y, Suzuki K, Koizumi K, *et al.*: Identification of a potent epigenetic biomarker for resistance to camptothecin and poor outcome to irinotecan-based chemotherapy in colon cancer. *International journal of oncology* 2012;40:217-226
6. Kelly M, Edgar AJ, Wevrick R: Analysis of DEXI/Dexi refines the organization of the mouse 7C and human 15q11->q13 imprinting clusters. *Animal Cytogenetics and Comparative Mapping* 2001;92:149-152
7. Christian SL, Bhatt NK, Martin SA, *et al.*: Integrated YAC contig map of the Prader-Willi/Angelman region on chromosome 15q11-q13 with average STS spacing of 35?kb. *Genome Research* 1998;8:146-157
8. Davison LJ, Wallace C, Cooper JD, *et al.*: Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Human molecular genetics* 2012;21:322-333
9. Edgar AJ, Birks EJ: Cloning of Dexamethasone-Induced Transcript. *Am J Respir Cell Mol Biol* 2001;25:119-124

10. Human Type 1 Diabetes Loci [article online], 2013. Available from <http://www.t1dbase.org/page/Regions/display/species/human/disease/1/type/assoc>. Accessed 7-04-2013
11. Concannon P, Rich SS, Nepom GT: Genetics of Type 1A Diabetes. *The New England Journal of Medicine* 2009;360:1646-1654
12. Robles DT, Eisenbarth GS, Wang T, *et al.*: Millennium award recipient contribution. Identification of children with early onset and high incidence of anti-islet autoantibodies. *Clinical immunology* 2002;102:217-224
13. Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nature genetics* 1995;9:293-298
14. Clayton DG: Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLOS Genetics* 2009;5:1-6
15. Wei Z, Wang K, Qu HQ, *et al.*: From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet* 2009;5:e1000678
16. Kimpimake T, Kupila A, Hamalainen AM, *et al.*: The first signs of B-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the finnish type 1 diabetes prediction and prevention study. *journal of Clinical Endocrinology & Metabolism* 2001;86:4782-4788
17. Barker JM, Barriga KJ, Yu L, *et al.*: Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *The Journal of clinical endocrinology and metabolism* 2004;89:3896-3902

- 18.** Ziegler AG, Hummel M, Schenker M, Bonifacio E: Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes. *Diabetes* 1999;48:460-468
- 19.** Orban T, Sosenko JM, Cuthbertson D, *et al.*: Pancreatic islet autoantibodies as predictors of type 1 diabetes in the Diabetes Prevention Trial-Type 1. *Diabetes Care* 2009;32:2269-2274
- 20.** Gardner SG, Gale EA, Williams AJK, *et al.*: Progression to diabetes in relatives with islet autoantibodies. *Diabetes Care* 1999;22:2049-2054