A FUNCTIONAL ANALYSIS OF THE ASSOCIATION OF TYPE 1 DIABETES WITH THE CYTOTOXIC T-LYMPHOCYTE ANTIGEN-*4* (*CTLA4*) GENE

Suzana M. Anjos

Faculty of Medicine, Department of Human Genetics

McGill University

February 2005

A dissertation submitted to the Faculty of Graduate Studies and Research in fulfillment of the requirements of the degree of Doctor of Philosophy

Your file Votre référence
ISBN: 978-0-494-21614-9
Our file Notre référence
ISBN: 978-0-494-21614-9

"All life is an experiment
The more experiments you make the better"
R.W. Emerson

ABSTRACT

Type 1 Diabetes (T1D) is characterized by the autoimmune destruction of the insulin producing beta cells of the pancreas. Linkage with Type 1 Diabetes has been reported on chromosome 2q31-33. Within this 23 cM interval, T1D association with *CTLA4* polymorphisms has been localized to a linkage disequilibrium (LD) block covering the D2S72-*CTLA4*-D2S116 interval. Within this interval, T1D is associated with a haplotype in the *CTLA4* gene, containing several potentially functional polymorphisms. Due to the extent of LD in the region, genetics studies alone are not sufficient to identify the causative polymorphism(s). To this end, we undertook a systematic analysis of each of these polymorphisms and their functional relevance in the pathogenesis of Diabetes.

We investigated the role of the only coding polymorphism in the *CTLA4* gene, a Threonine to Alanine substitution in the N-terminal signal sequence. The Ala17 (+49G) allele is associated with increased risk to T1D. Since the major function of the signal sequence is to direct trafficking of the pre-protein to the Endoplasmic Reticulum (ER), we hypothesized that a substitution could alter these events. The Ala17 is inefficiently *N*-glycosylated in the ER and its expression is lower than the Thr17 at the cell surface in *COS1* cells.

Using the family-based transmission disequilibrium test, (TDT) we report an association of the -318C-allele at the proximal promoter polymorphism and T1D. Conditional TDT analysis dissected the contributions of the linked –318C>T and +6230G>A polymorphisms, as independent contributors to the genetic effect. We report

transcriptional effects of the promoter alleles *in vitro*, and *in vivo* with the –318T allele inducing stronger transcriptional activity. This effect may be mediated by alternative transcription factor binding at the Single Nucleotide Polymorphism (SNP).

We confirmed association of the 3'+6230G>A SNP with T1D. In evaluating its contribution to the expression of soluble and full-length *CTLA4* isoforms in an allele-dependent manner, we found no correlation between alleles at +6230 and levels of either isoform. The 3'SNP does not alter *in cis* the expression of the downstream *ICOS* gene.

We propose that susceptibility at the *CTLA4* locus is mediated by a haplotype exerting both transcriptional and post-translational effects.

RÉSUMÉ

Le diabète juvénile est caractérisé par la destruction de nature auto-immune des cellules beta produisant l'insuline dans le pancréas. La susceptibilité génétique localisée à de nombreux loci est un facteur déterminant dans le développement de la maladie. Le locus *IDDM12* est localisé dans la région 2q33 qui contient plusieurs gènes impliqués dans la régulation immunitaire, incluant *CD28*, *CTLA4* et *ICOS*. Parmi ceux-ci, la région comprenant *CTLA4* a été associée avec une susceptibilité au diabète. Cependant, celle-ci contient plusieurs polymorphismes, tous en *linkage disequilibrium*, dont la contribution individuelle ne peut pas être disséquée par les seules méthodes génétiques. C'est pourquoi, nous avons entrepris une étude systématique de chaque polymorphisme dans ce pic d'association, en définissant la base fonctionnelle d'une telle association.

Le seule polymorphisme codant se situe dans le signal peptide dont la fonction est de diriger le trafique de protéines membranaires ou sécrétées traduites sur le complexe ribosomal vers le reticulum endoplasmique (ER). Un changement d'acide aminé de thréonine à alanine pourrait ainsi affecter la fonction du signal peptide. Nous avons montré que l'allèle prédisposant, l'alanine, compromet l'efficacité de glycosylation au niveau du ER et que celui-ci se rend à la surface des cellules COS1 en quantité réduite par rapport à l'allèle protecteur, la thréonine. Ce mécanisme pourrait expliquer l'association avec le diabète.

Au niveau des régions non-codantes, on trouve une forte association avec les allèles au promoteur (-318C>T), qui se trouvent liées au polymorphisme dans le 3' (+6230G>A), et le diabète dans notre collection de familles diabétiques. Nous avons montré que ces deux polymorphismes agissent indépendamment et que le -318C>T altère

le niveau de transcription de *CTLA4 in vitro* et *in vivo*. De plus, nous avons investigué le rôle du polymorphisme au 3' dans l'expression des deux isoformes de *CTLA4*, la forme soluble et membranaire sans trouver une corrélation entre le génotype au 3' et le niveau d'expression de ces isoformes. Nous n'avons pas vu de relation entre ce polymorphisme et la régulation *en cis* du gêne en amont, *ICOS*.

Nous proposons ainsi un modèle où la susceptibilité au diabète à *CTLA4* est contenue dans un haplotype régulant l'expression au niveau transcriptionnel et à la post-traduction.

FOREWORD

The McGill University guidelines and regulations regarding the fulfillment of a Doctor of Philosophy Degree (PhD) stipulate that it must display original scholarship, be expressed in good literary style, and must be a distinct contribution in the pursuit of knowledge. The dissertation itself can consist of a collection of original research articles of a cohesive nature, which together report on a single program of research and includes a comprehensive literature review. This is a manuscript-based dissertation intended to conform to these regulations.

CONTRIBUTIONS OF AUTHORS

The contributions of authors are listed here and at the beginning of each chapter for the convenience of the reader.

CHAPTER 3: Suzana Anjos, Audrey. Nguyen, Houria Ounissi-Benkalha, Marie-Catherine Tessier and Constantin Polychronakos.
"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**, 277(48): 46478-46486. 2002

Suzana Anjos: Experimental design and strategy, including all plasmid cloning, protein translation assays, cell-based assays and confocal microscopy, data collection, analysis and interpretation, hypothesis development, manuscript organization and preparation.
Audrey Nguyen: Technical expertise in cell-free translation systems, interpretation of immunofluorescence data.
Houria Ounissi-Benkalha: Technical advice and strategy in matters of Molecular Biology.
Marie-Catherine.Tessier: Site-directed mutagenesis.
Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

CHAPTER 4: Suzana M. Anjos, Marie-Catherine Tessier and Constantin Polychronakos.
"Association of the cytotoxic T-lymphocyte associated antigen-4 with Type 1 Diabetes: evidence for independent effects of two polymorphisms on the same haplotype block". **The Journal of Clinical Endocrinology and Metabolism**, 89(12): 6257-65. 2004

Suzana Anjos: Experimental design and strategy, hypothesis development, gene cloning and reporter assay implementation, genotyping, allelic imbalance studies, data interpretation and analysis, manuscript preparation.

Marie-Catherine Tessier: Technical assistance in subcloning promoter reporter constructs, high-throughput genotyping design and implementation, data collection.
Constantin Polychronakos: Research Director, hypothesis development, statistics expertise, manuscript editing and preparation.

CHAPTER 5: Suzana M. Anjos, Wei Shao, Luc Marchand and Constantin Polychronakos. "Allelic effects on gene regulation at the autoimmunity-predisposing *CTLA4* locus: a re-evaluation of the 3' +6230G>A polymorphism". Genes and Immunity, accepted for publication, 2005

Suzana M. Anjos: Collection and extraction of RNA samples, experimental design and strategy, hypothesis development, supervision and training of summer student, manuscript preparation.
Wei Shao: Experimental data acquisition in the allelic imbalance assay.
Luc Marchand: Allelic imbalance assay design and experimental data acquisition with respect to the *ICOS* gene data.
Constantin Polychronakos: Research Director, hypothesis development, manuscript editing and preparation.

APPENDIX I: Suzana M. Anjos, Constantin Polychronakos
Evaluating the effect of the diabetes-associated *CTLA4* signal peptide polymorphism on its focal localization in T cells.

Suzana M. Anjos: Hypothesis development, experimental data, confocal microscopy, image analysis and manuscript preparation
Constantin Polychronakos: Research Director, hypothesis development, manuscript editing and preparation.

APPENDIX II: **Suzana M. Anjos**, Constantin Polychronakos

The dinucleotide $(AT)^n$ repeat in the 3'UTR of the diabetes associated gene, *CTLA4* does not alter the stability of the messenger RNA. *Manuscript in preparation for submission.*

Suzana Anjos: Hypothesis development, Experimental data, and manuscript preparation
Constantin Polychronakos: Research Director, Hypothesis development and manuscript editing.

ACKNOWLEDGEMENTS

"Not everything that counts can be counted,
and not everything that can be counted counts."
Albert Einstein

To Dr. Polychronakos who gave me both an opportunity to study in his laboratory and a passionate subject to evolve with. As my Supervisor, Dr. Polychronakos demonstrated his humanity and understanding countless times, and his patience and trust propelled me forward. I am most thankful for the freedom he granted me with respect to this project, for his trust in my abilities, his patience through my many mistakes, and his enthusiasm and encouragement for the discoveries we made.

I would like to thank the members of my supervisory committee, Dr. Jacquetta Trasler, Dr. Feige Kaplan and Dr. Andrew Karaplis for their suggestions and advice with regards to science and beyond. Dr. Karaplis' suggestions in the earlier parts of this project proved to be invaluable and I am thankful to him for his generosity in sharing his expertise and lab. I would like to thank Dr. Trasler for excellent advice in matters of microscopy and for lending an ear in trying times. Dr. Kaplan's practical approach and insightful questions, as well as friendly advice were very important in this process and for that I am thankful.

To the members of the Polychronakos lab, for your support, for the lessons learned and the experience of a lifetime.

To Dr. Houria Ounissi-Benkalha who taught me to be skeptical, rigorous and critical and that perfection is an attainable goal in experimental science. Your

encouragement and thoughtful discussions made my foray into molecular biology seamless. Thank you Houria.

To Jennifer McCann for the countless hours of friendship, support and encouragement, for showing me to be fearless and tenacious and for invaluable life lessons.

To Tamara Kelly for keeping me sane in hours of despair, for being a good friend throughout this process and for encouragement and support in the writing of this Thesis and at all times.

To Laura Benner and Fran Langton our Human Genetics Gate-Keepers for their expediency, friendliness and just taking care of stuff without a fuss.

I thank my parents, Maria and Victor Anjos, for the transmission of the genetic material that made this degree possible. I am grateful for their constant, unwavering encouragement in the pursuit of a degree they did not understand ("When are you going to finish and get a job?"). Their awe at the smallest of my achievements, their interminable curiosity about the world around them, and their strength of character and work ethic make them great examples to follow. Thanks mom and dad.

To my brother, whose curiosity made him my best audience. Thank you for always asking questions and for challenging my answers. You are the smartest and wittiest person I know.

To the Martinez family for unconditional support, generosity and their enduring spirit.

Finally, I would like to thank my husband, Martin Martinez without whom this would not have been possible. No other person endured the daily rendering of lab

occurrences with the understanding and enthusiasm he displayed. Thank you for being my audience countless times in my "DNA is..." talks and for creative direction in matters of "Scientific Art". Thank you for encouraging and supporting me, always.

TABLE OF CONTENTS

Abstract	iii
Résumé	v
Foreword	vii
Contributions of Authors	viii
Acknowledgements	xi
Table of Contents	xiv
List of Figures	xx
List of Tables	xxi
List of Abbreviations	xxii
1.0 CHAPTER ONE	1
1.1 Statement of Purpose and Objective of Research	1
2.0 CHAPTER TWO: Comprehensive Literature Review	5
2.1 Diabetes Mellitus	5
2.2 Type 2 Diabetes (T2D)	6
2.2.1 Maturity onset diabetes in the young (MODY)	7
2.3 Epidemiology of Type 1 Diabetes	7
2.3.1 Age, gender and time	7
2.3.2 Rates and Geography	8
2.4.1 Pathogenesis of Type 1 Diabetes	9
2.4.1 Tolerance mechanisms: Central Tolerance	9
2.4.2 Peripheral Tolerance and Immunoregulatory mechanisms	10
2.4.2.1 Regulatory T cells	11

2.5 Evidence for Diabetes as an autoimmune disorder 12

2.5.1 Autoantibodies in T1D 12

2.6 Mechanisms of beta cell destruction 14

2.6.1 Cell-mediated immunity 14

2.6.2 The non-obese diabetic (NOD) mouse 16

2.7 Etiology of Type 1 Diabetes 17

2.7.1 Environmental Risk Factors 17

2.7.1.1 Viral Infections 18

2.8 Genetics of Complex Diseases: Type 1 Diabetes 19

2.8.1 Approaches and tools in identifying genes in complex diseases 19

2.8.1.1 Genome-wide scans and linkage analysis 19

2.8.1.2 Fine-mapping linkage analysis and positional cloning 20

2.8.1.3 Genetic Association studies 21

2.8.1.4 The Transmission Disequilibrium Test (TDT) 22

2.8.2 Allelic variations in protein function and/or gene expression 21

2.8.2.3 Detecting allele-specific expression differences: *In vitro* 26

2.8.2.4 Detecting allele-specific expression: *In vivo* 26

2.9 Whole genome and fine linkage analysis in Type 1 Diabetes 29

2.10 *IDDM1*: HLA-encoded susceptibility to T1D 29

2.10.1 Class II *HLA* genes and susceptibility to T1D 30

2.10.2 Other *HLA* genes 33

2.10.3 Mechanism underlying *HLA* encoded susceptibility 34

2.10.4 Synteny with the NOD mouse: the MHC locus 36

2.11 IDDM loci outside the HLA 36

2.12 *IDDM2*: the *INS*-VNTR 37

2.12.1 INS-VNTR and function 38

2.12.1.1 Insulin VNTR and thymic insulin expression 40

2.14 *IDDM12*: The *CTLA4* gene 41

2.14.1 Linkage and association of the cytotoxic T lymphocyte antigen-4 (*CTLA4*) T1D: the early studies 41

2.14.2 Association of *CTLA4* in the presence of *HLA* risk alleles 41

2.14.3 Fine Mapping, linkage disequilibrium, and association studies at *CTLA4* 43

2.14.4 NOD mouse genetics: the *Ctla4* gene 45

2.15 *CTLA4* in other autoimmune diseases 49

2.15.1 The autoimmune thyroid disorders (AITD) 49

2.15.1.1 Graves' disease (GD) 50

2.15.1.2 Hashimoto Thyroiditis (HT) 50

2.15.2 Systemic Lupus Erythematosus (SLE) 50

2.16 CTLA-4 function 52

2.16.1 T cell activation 52

2.16.2 CTLA-4 as a negative regulator of T cell function 53

2.16.3 CTLA-4's mode of action 54

2.16.4 The CTLA-4 Knockout mouse 56

2.16.5 Human examples of CTLA-4 deficiency: Chediak-Higashi syndrome 56

2.16.6 CTLA-4's cell-surface expression and intracellular cycling pattern 57

2.16.7 CTLA-4 in the autoimmune process 58

2.17 Allelic variation at the *CTLA4* gene 62

2.17.1 Functional studies at the signal peptide 62

2.17.2 The role of the Signal Peptide 64

2.17.2.1 Signal Peptide mutations causing disease 65

2.17.2 Regulatory variation at the *CTLA4* gene 66

2.17.2.1 The *CTLA4* promoter 66

2.17.2.2 The *CTLA4* 3'UTR: The dinucleotide repeat 67

2.17.2.3 The alternatively spliced soluble CTLA-4 69

2.18 **Project Rationale** 72

2.18.1 Objective 1. The Signal Peptide polymorphism 72

2.18.1.1 Approach 73

2.18.2 Objective 2: Regulatory variation at the *CTLA4* gene: 5'end 74

2.18.2.1 Approach 75

2.18.3 Objective 3: Regulatory variation at the *CTLA4* gene: 3'end 77

2.18.3.1 Approach 77

3.0 CHAPTER THREE: The Signal Peptide 78

3.1 Contributions of Authors 78

3.2 Abstract 79

3.3 Introduction 80

3.4 Materials and Methods 87

3.5 Results 92

3.6 Discussion 112

3.7 Acknowledgements 116

4.0 CHAPTER FOUR: Regulatory Variation at *CTLA4*: the 5'end 117

4.1 Contributions of Authors 117

4.2 Abstract 118

4.3 Introduction 119

4.4 Materials and Methods 123

4.5 Results 132

4.6 Discussion 144

4.7 Footnotes 149

5.0 CHAPTER FIVE: Regulatory Variation at *CTLA4*: the 3'end 150

5.1 Contributions of Authors 150

5.2 Abstract 151

5.3 Introduction 152

5.4 Materials and Methods 155

5.5 Results 163

5.6 Discussion 170

5.7 Acknowledgements 172

6.0 CHAPTER SIX: Discussion 173

6.1 The *CTLA4* signal peptide 173

6.1.1 Proposed Molecular Model 174

6.2. Cell-surface targeting of CTLA-4 175

6.3 Future Studies 176

6.4. Regulatory Variation at *CTLA4* 176

6.4.1The *CTLA4* promoter 177

6.4.2 Future Studies 178

6.4.3 The *CTLA4* 3'UTR dinucleotide repeat 179

6.4.4 The alternatively spliced soluble *CTLA4* isoform 180

6.4.5 Future Directions 181

6.5 Evaluating the hypothesis using transgenic mouse models 182

6.6 Implications for Type 1 Diabetes (and autoimmune disease) 183

6.7 Summary 186

6.7.1 Main Contributions 186

A1. APPENDIX ONE: The Signal Peptide: Intracellular localization in T cells 189

A1.1 Contributions of authors: 189

A1.2 Introduction 190

A1.3 Materials and Methods 192

A1.4 Results and Discussion 194

A2. APPENDIX TWO: Regulatory Variation at *CTLA4*: The 3' $(AT)^n$ repeat 200

A2.1 Contributions of Authors 200

A2.2 Introduction 201

A2.3 Materials and Methods 202

A2.4 Results 204

A2.5 Discussion 210

A3: APPENDIX THREE: COPYRIGHTS 211

REFERENCES 212

LIST OF FIGURES

Chapter 2:

Figure 1: Detecting allelic expression differences *in vivo.* 28
Figure 2: Simplified map of the *HLA* locus. 35
Figure 3: The association map of *CTLA4* gene region and Graves' disease. 46
Figure 4: Simplified version of the T cell activation and the co-stimulatory network 55
Figure 5: CTLA-4 cycling. 59
Figure 6: Type I diabetes associated *CTLA4* haplotype and potential functional role of disease-associated polymorphisms 63

Chapter 3:

Figure 1. Multiple sequence alignment of CTLA-4 signal peptides 85
Figure 2. CTLAAla17 is inefficiently processed in the Endoplasmic Reticulum 94
Figure 3: Signal peptide cleavage of the major products of CTLAAla17 and CTLAThr17 is unaffected. 96
Figure 4: Deglycosylation abolishes allelic differences in processing efficiency 99
Figure 5: Differential processing of the two alleles depends on glycosylation 101
Figure 6: The partially glycosylated CTLAAla17intermediate 104
Figure 7: CTLAAla17 and CTLAThr17 in COS1 cells 109

Chapter 4:

Figure 8: Summary of genetic and functional significance of SNPs within the *CTLA4* gene 122
Figure 9: Allelic imbalance in the RNA is quantified by sequencing 130
Figure 10: Higher *CTLA4* expression from the promoter –318 T allele in vitro. 140
Figure 11: Allelic imbalance at the *CTLA4* promoter locus 143

Chapter 5

Figure 12: Schematic representation of primers used in this study located in the *CTLA4* region 158
Figure 13: Validation of the SNuPE methodology 162
Figure 13: Relative allelic abundance by *CTLA4* haplotype 165
Figure 14: Effect of +6230G>A on *ICOS* expression 169

Appendix I

Figure 1: CTLA-4 signal peptide allelic proteins do not co-localize in T cells 198
Figure 2: Quantification of cell-surface expression of CTLA-4 198

Appendix II

Figure 3: Allelic steady-state mRNA does not depend on length of the 3'UTR $(AT)^n$ repeat 208
Figure 4: The length of the *CTLA4* 3'UTR repeat does not affect mRNA stability 209

LIST OF TABLES

Chapter 2:

Table 1: Suspected loci involved in T1D susceptibility.....31
Table 2: Class II *HLA* haplotypes and their association with Type 1 Diabetes.....32
Table 3: Summary of *Insulin* expression *in vivo* in pancreas and thymus.....39
Table 4: Genetic association studies of *CTLA4* polymorphisms with Type 1 Diabetes.....47
Table 5: Family-based association studies of *CTLA4* with Type 1 Diabetes.....48

Chapter 4:

Table 6: Primers and probes used for fluorescence polarization (FP).....125
Table 7: Transmission disequilibrium of SNPs within the *CTLA4* promoter and of +6230G>A in the 3' end.....133
Table 8: Haplotype tag SNPs (htSNPs) were genotyped in 498 T1D families.....136
Table 9 :Two SNP haplotype analysis in 498 T1D families.....137
Table 10: Testing independent effects of the 5' and 3' ends of *CTLA4* by the conditional TDT test.....137

Chapter 5

Table 11: Summary of difference in allelic variation of *CTLA4* haplotypes.....163

Appendix II

Table 1: Regulatory variation at the 3'UTR.....206
Table 2: Frequency of *CTLA4* haplotypes (n=154, chromosomes).....206

ABBREVIATIONS

AITD: Autoimmune Thyroid Disease

AH: Autoimmune Hypothyroidism

APC: Antigen Presenting Cell

ARE: AU-rich element

ASPs: Affected Sibling Pairs

BAC: Bacterial Artificial Chromosome

CHIP: Chromatin Immunoprecipitation

CMV: Cytomegalovirus

CTLA4: Cytotoxic T-lymphocyte antigen-4

DC: Dendritic Cell

EMSA: Electrophoretic Mobility Shift Assay

ER: Endoplasmic Reticulum

ERK: Extra-cellular-signal-regulated kinase

EST: Expressed Sequence Tag

FBAT: Family-based association test

FP: Fluorescence Polarization

GD: Graves' disease

HBAT: haplotype-Based association test

HLA: Human Leukocyte Antigen

HMG: High-Mobility Group

HtSNPs: Haplotype Tag SNPs

ICOS: Inducible Co-stimulator

IFN-γ: Interferon-gamma

Ig: Immunoglobulin

IL-2: Interleukin-2

INS: Insulin

LD: Linkage Disequilibrium

LEF1: Lymphocyte Enhancer Factor-1

MHC: Multi-HistoCompatibility

MS: Multiple Sclerosis

NOD: Non-Obese Diabetic

PBMCs': Peripheral Blood Mononuclear cells

PCR: Polymerase Chain Reaction

PFA: Paraformaldehyde

PHA: Phytohemagglutinin

PMA: Phorbol Myristate Acetate

RT: Reverse Transcriptase

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEM: Standard Error of the Mean

SLE: Systemic Lupus Erythematosus

SNP: Single nucleotide polymorphism

SNuPe: Single Nucleotide Primer Extension

SP: Signal Peptide

SPP: Signal peptide peptidase

SRP: Signal recognition particle

T1D: Type 1 Diabetes

TCF1: T-cell Factor

TCR: T-cell receptor

TDT: Transmission Disequilibrium Test

TNF-α: Tumor Necrosis factor-alpha

TNF: Tumor Necrosis Factor

Treg's: Regulatory T cells

VNTR: Variable Number of Tandem Repeats

1.0 CHAPTER ONE

1.1 Statement of Purpose and Objective of Research

Type 1 Diabetes (T1D) is characterized by the autoimmune mediated destruction of the insulin producing beta cells of the pancreas; it is primarily a result of the failure of immune tolerance though it manifests itself as a metabolic disease of insulin insufficiency.

Its etiology is defined by the complex interaction of networks of genes and their environment, the nature of which we are only beginning to unravel. In doing so, it has become apparent that several regions or loci are overrepresented in affected individuals and are associated with increased risk. In fact, a number of such regions have been reported, though only a minority has been confirmed by multiple independent studies.

Of these, the *HLA* on the 6p21 chromosomal fragment, contributes the highest risk for developing T1D with certain alleles at the *HLA* explaining up to 50% of the inheritable risk while the region upstream of the Insulin gene, the variable number of tandem repeats, *VNTR* contributes 10% of the susceptibility to T1D (Bell et al. 1984; Julier et al. 1991; Bain et al. 1992; Bennett et al. 1995). More recently, a region on chromosome 2q33 has been associated with T1D (Nistico et al. 1996; Marron et al. 1997; Marron et al. 2000; Ueda et al. 2003). This region encompasses several genes involved in immune regulation including the co-stimulator *CD28*, its negative counterpart, the Cytotoxic T lymphocyte antigen-4, *CTLA4* and the Inducible Co-Stimulator, *ICOS*.

Any one of these genes could conceivably contribute to pathogenesis if the delicate balance between tolerance and immunity is tipped.

In this region, a haplotype within a linkage disequilibrium (LD) block spanning the *CTLA4* gene and its 5' and 3' flanking regions contributes the highest risk, while the other genes in the block have been excluded from the association (Marron et al. 2000; Ueda et al. 2003). Within this block there are several markers in linkage disequilibrium any of which could influence gene expression levels, and/or protein function, including a –318C>T single nucleotide polymorphism (SNP), a non-synonymous SNP in the signal peptide, +49>G resulting in a Threonine to Alanine substitution, a dinucleotide repeat in the 3'untranslated region (UTR) and a SNP in the 3'flanking region outside the known polyadenylation signal of *CTLA4*, +6230G>A. Although the highest association was recently reported at the 3'flanking region at the +6230G>A SNP in a large collection of T1D families, not all of the SNPs in the block were tested in this study, the overall effect was modest in T1D and the 5'end of the *CTLA4* gene could not be excluded (Ueda et al. 2003). Thus, any one of these polymorphisms could alter gene function by mediating allele specific differences in transcript abundance and /or protein function. Due to the high degree of LD in this region, genetics alone will not suffice in narrowing down the causative allele; systematic and detailed functional analysis is required.

The aim of these original studies to follow is to narrow down the causative polymorphism(s) in the *CTLA4* gene region and to determine their mechanism of action through functional studies.

We began by addressing the role of the only coding polymorphism in the *CTLA4* gene, a Threonine to Alanine substitution resulting from a single base change (+49A>G)

in the signal peptide of CTLA-4. The Alanine (G) allele is strongly associated with disease risk and its location in the signal peptide led us to hypothesize that it may alter the early protein trafficking events of the CTLA-4 glycoprotein, events that take place in the endoplasmic reticulum (ER) prior to cleavage of the signal peptide. We showed that this signal peptide amino acid change, altered the extent of *N*-glycosylation of the CTLA-4 protein and resulted in reduced CTLA-4 (Ala17) at the cell surface (Chapter 3, (Anjos et al. 2002)). We found similar differences in T cells transfected with the different colored allelic proteins, (Appendix I).

In our own family-based genetic association and linkage study (using the transmission disequilibrium test, TDT), we found that several polymorphisms within regulatory regions around *CTLA4*, all in tight LD with one another, were associated with T1D risk including the +6230G>A at the 3'end (Anjos et al. 2004). Particularly striking, was the association of the proximal promoter polymorphism, at –318 C, located in highly conserved minimal region for *CTLA4* transcription, and T1D in our dataset. Conditional TDT analysis successfully dissected the contributions of the tightly linked –318C>T and +6230G>A polymorphisms, as independent contributors to the genetic effect (Anjos et al. 2004). We showed that the *CTLA4* –318 C allele directs lower transcriptional activity *in vitro* using an allele specific luciferase reporter system in Jurkat T cells, findings supported by *in vivo* quantification of relative amounts of allele-specific transcripts in PBMCs derived from individuals heterozygous at the –318C>T promoter SNP (Chapter 4, (Anjos et al. 2004)). We postulate that this effect is mediated by altered transcription factor binding to a response-element for two related regulators of lymphocyte development, *LEF1* and *TCF1*.

Though we found a genetic effect at the +6230G>A SNP, this association was modest compared to that at the –318C>T SNP. Nonetheless, previous work by Ueda et. al., (Ueda et al. 2003) prompted us to evaluate the contribution of this SNP to the expression of the soluble and full-length isoforms of *CTLA4*. We report that there was no significant difference between levels of soluble or full-length (+6230G>A) allele specific transcripts in PBMCs of heterozygous individuals, and we confirm our previous findings of allele-specific promoter differences, in soluble and full-length isoforms (Anjos S.M. et al. 2005). In doing so, we have also indirectly shown that the same promoter drives the transcription of both soluble and full-length transcripts (Anjos S.M. et al. 2005). We also showed that the +6230G>A does not modulate the expression of the downstream *ICOS* gene (Chapter 5 (Anjos S.M. et al. 2005)).

We also addressed the effect of the dinucleotide $(AT)^n$ repeat in the 3'untranslated region of *CTLA4* on the stability of the messenger RNA (mRNA) and report that the two most common alleles of the microsatellite (88 bp and 106 bp) do not affect mRNA stability (Appendix II).

2.0 CHAPTER TWO: Comprehensive Literature review

Diabetes is characterized clinically as a disruption in glucose homeostasis. It is a major health concern in developed countries, taking up a sizable fraction of health budgets . While 90% of the diabetes cases fall under the Type 2 diabetes (T2D) classification, and its rates have been rising alongside a concomitant pandemic of obesity, Type 1 Diabetes (T1D), which affects children and adolescents has also increased in the last 30 years for reasons which are not entirely clear (Williams and Pickup 2001).

2.1 Diabetes Mellitus

Under normal circumstances, the entry of glucose into the insulin producing beta cells of the pancreas, triggers the release of insulin. The release of insulin into the tissue network where it binds to insulin receptors, initiates a cascade of biochemical events which result in the uptake of glucose into cells with subsequent metabolism into energy or storage as glucagon, reviewed in (Notkins 2002). Defects anywhere along the beta-cell peripheral tissue axis can result in hyperglycemia. Clinically, the two major forms of diabetes can be distinguished as Type 1 Diabetes (also referred to as juvenile diabetes or insulin dependent diabetes mellitus, IDDM) and Type 2 diabetes (Notkins 2002). Although Type 1 Diabetes is characterized by a complete deficiency in the production of insulin, Type 2 Diabetes' etiology is characterized by decreased sensitivity to insulin coupled with an inability of the beta cell to compensate for the increased insulin requirements. Though the pathogenesis of T2D remains a source of debate and intense research, in T1D it is accepted that the pathogenesis involves the autoimmune destruction

of beta cells of the pancreas, a process which appears to be hardwired into the genetic makeup of the individual that results in disease when triggered by environmental factors.

2.2 Type 2 Diabetes (T2D)

Loosely, defined as "adult onset" diabetes, there are an increasing number of Type 2 Diabetics diagnosed in young adults while also becoming increasingly common in children who present with obesity. The major risk factors are predetermined by one's genetic makeup, though lifestyle factors likely trigger and/or exacerbate disease onset and progression. Thus, T2D can be defined as a polygenic disorder influenced by gene-gene interaction and gene-environment interactions. Since it is clinically heterogeneous the identification of genes involved is challenging. There are disproportionally high rates of T2D in African Americans and Latino populations and native American populations and the high incidence of diabetes in certain populations and among first-degree relatives of T2D as well as the high concordance rate in identical twins, provides strong evidence that genetic factors underlie susceptibility to the common form of diabetes. So far several loci have been found to be associated with T2D including a common polymorphism (Pro12 to-Ala) in PPARγ contributing a modest but significant diabetes risk associated with the common proline variant (85% frequency) (Altshuler et al. 2000). Its modest effect translates into a large population-attributable risk-influencing as much as 25% of Type 2 Diabetes in the general population because it is so common (Altshuler et al. 2000). A small proportion of diabetics get the disease because of mutations in a single gene falling into the distinct category of Maturity Onset Diabetes in the Young, known as MODY.

2.2.1 Maturity onset diabetes in the young (MODY)

MODY is an uncommon form of diabetes, resulting from mutations in a single gene, inherited in an autosomal dominant manner. As in T2D, it does not require insulin treatment at least in the early stages and as a result MODY is often misdiagnosed as T2D. It develops in young adulthood, before the age of 25 with often-missed mild hyperglycemia. In the presence of hyperglycemia it can be misdiagnosed as T1D in the young, and insulin treatment is prescribed. Through genetic analyses several types of MODY have been found and are characterized by the mutations in the genes they incur. While MODY2 (the most common form of MODY) is caused by a mutation in the Glukokinase gene, the enzyme responsible for sensing circulating levels of glucose, MODY1 and MODY3-6 are caused by mutations in genes encoding transcription factors, *HNF4A*, *TCF1*, *TCF2*, and *IPF1* (Fajans et al. 2001) involved in the crucial cascade of transcriptional control at the level of the insulin gene and the glucose transporter. Though MODY2 and MODY3 are the most common forms of monogenic diabetes, the diabetes caused by mutations in single genes remain nonetheless, rare.

2.3 Epidemiology of Type 1 Diabetes

2.3.1 Age, gender and time

The prevalence of T1D in the United States is approximately 2 per 1000 for children under age 20 (Dorman and Bunker 2000). The incidence of T1D in Caucasians is higher than other childhood chronic diseases, such as Cystic Fibrosis and juvenile arthritis for instance. Although the onset of diabetes can occur at any time it is usually diagnosed during childhood and adolescence (hence Juvenile Diabetes) with a peak

around the time of puberty (a pattern which has been reported throughout the world and is common to most populations) (Dorman and Bunker 2000).

The gender distribution in T1D is similar for males and females, although in low-risk populations such as the Japanese, there appears to be a distortion towards increased incidence for the female. In contrast, in regions where overall incidence is high, an excess risk for males has been observed. Notably, there is also a seasonal variation in risk, with the highest reported incidences in the winter months in both Northern and Southern populations and a concomitant lower incidence in late spring and summer (Dorman and Bunker 2000). There are also recent temporal trends in incidence reported. A significant increase in T1D has been reported by most population registries in Northern and Central Europe (Dorman and Bunker 2000) as well as Asian and Western Pacific countries.

2.3.2 Rates and Geography

There is a striking difference in the incidence of T1D relative to geographic and ethnic variation. There are pockets of exceptionally high T1D incidence in the world, including the island of Sardinia and Finland (Dorman and Bunker 2000). Whereas the incidence of T1D (age-adjusted) is 0.1-3/100 000 per year in Asian countries, including Japan, China and Korea, it is the highest in Sardinia and Finland at >36/100 000 per year (Dorman and Bunker 2000). Interestingly, Cuban, Chilean, and Mexican populations also have extremely low incidences of T1D. In most other Caucasian populations in Europe and the Americas, the incidence rates are moderate (~10-20/100 000 per year)(Dorman and Bunker 2000). The reasons for these differences along ethnic lines are the focus of intense research, but are likely to reflect differences in environmental factors as well as genetic factors.

2.4 Pathogenesis of Type 1 Diabetes

It is increasingly apparent that failures in tolerance mechanisms manifest themselves as autoimmune disease. Self-tolerance is established primarily in the thymus (central tolerance) and maintained in the periphery (peripheral tolerance). These mechanisms have evolved to provide a comprehensive and complex web of controls to ensure that lymphocytes do not become auto-aggressive the failure of which results in devastating consequences for the individual. The following sections will review what is known about central and peripheral tolerance in the development of T1D.

2.4.1 Tolerance mechanisms: Central Tolerance

Central tolerance is established in the thymus. It broadly involves clonal deletion and inactivation of self-reactive lymphocytes, and concerns immature T cells as they differentiate through the thymus (Mathis and Benoist 2004).

Most immature T-lymphocytes are not released onto the periphery because they are eliminated in the thymus through apoptosis. Initially, deletion occurs when thymocytes come into contact with stromal cells acting as Antigen Presenting Cells (APCs) in the thymic cortex and the T-cell Receptor (TCR) undergoes rearrangement after interacting with the antigen presented by the various MHC molecules (reviewed in (Marrack et al. 1993)). If TCR rearrangement does not occur, the lymphocyte is deleted by apoptosis (Marrack et al. 1993).

Further differentiation occurs through positive selection whereby thymocytes differentiate into either cytotoxic ($CD8^+$) or helper ($CD4^+$) T-lymphocytes. In addition, T-lymphocytes whose TCR reacts too strongly with the self-peptide presented are negatively selected.

The discovery of tissue-specific expression of transcripts in the thymus restricted in particular to the thymic medullary epithelial cells (MECs) reignited the interest in the role of central tolerance mechanisms in organ-specific autoimmunity (Derbinski et al. 2001). It is believed that the promiscuous thymic expression of proteins characteristic of peripheral organs, might enhance tolerance induction to those organs reviewed by (Kyewski et al. 2002). The discovery of insulin transcripts in the thymus whose levels of expression correlated with susceptibility or protection in human T1D (Pugliese et al. 1997; Vafiadis et al. 1997) is an example of an organ-specific autoimmune disease whose pathogenesis may involve genetically regulated clonal deletion of insulin auto-antigens. It has been suggested that higher levels of insulin expressed in the thymus mediated *in cis* by the variable number of tandem repeats (VNTR), may enhance the negative selection of auto-reactive thymocytes (Pugliese et al. 1997; Vafiadis et al. 1997). This has been more directly demonstrated in a mouse model engineered to express graded levels of thymic insulin expression (but normal pancreatic insulin levels) with mice expressing low levels of thymic insulin presenting with detectable peripheral reactivity to insulin (Chentoufi and Polychronakos 2002).

2.4.2 Peripheral Tolerance and Immunoregulatory mechanisms

Peripheral tolerance relates to the circulating mature T or B cells after they have exited the primary lymphoid organs and are freely circulating through the blood, lymph, and secondary lymphoid organs. Antigens of concern would then be those restricted to the tissues and not those in the thymus (Mathis and Benoist 2004). Clonal deletion and anergy are mechanisms of tolerance typically used in the thymus but also in the periphery though a variety of other selection mechanisms are used including, clonal ignorance,

deviation suppression and helplessness (Mathis and Benoist 2004). Peripheral mechanisms of tolerance are crucial because central tolerance is not perfect and at any given time, self-reactive T cells can be detected in individuals (Liblau et al. 1991; Sun et al. 1991; Link et al. 1992). Recently, the activities of tolerance mechanisms in the periphery, particularly those of the Regulatory T Cells (Tregs) have been the focus of intense research in the context of their involvement in autoimmunity and will be reviewed in the next section.

2.4.2.1 Regulatory T cells

Increasingly, evidence has been mounting in favor of a subset of T cells particularly those with the $CD4^+CD25^+$ phenotype as the crucial keepers of immunological self-tolerance and the negative regulation of immune responses. The vast majority of Tregs are naturally differentiated in the thymus as a functionally distinct and mature subpopulation of T cells and migrate to the periphery where they are functional (Fehervari and Sakaguchi 2004). It is thought that the generation of cells with this phenotype is controlled in development, (reviewed in Sakaguchi 2004), and congenital deficiency in this cell population in humans results in severe damage to the mechanisms of self-tolerance and immunoregulation, and leads to severe autoimmune disease, immunopathology and allergy (Gambineri et al. 2003). The single most important feature of $CD4^+CD25^+$ Tregs is their immuno-suppressive capabilities in a variety of immune cells (Fehervari and Sakaguchi 2004). Though the mechanism by which it dampens the immune response remains unknown, the constitutive expression of the IL-2 receptor (CD25, or IL-2R) may act as a sink for IL-2 thereby depriving autoreactive T-cells of an essential growth factor (de la Rosa et al. 2004) Alternatively, a model relying on a more

antagonistic form of suppression involving Tregs constitutively expressing the CTLA-4 receptor, itself a potent negative regulator of T cell function, has received much attention (Grohmann et al. 2002; Fallarino et al. 2003; Munn et al. 2004).

Although key aspects of Treg suppressive function and mechanisms are still unknown, and so far, a reliable, unambiguous surface marker for naturally occurring $CD4^+$ Tregs remains undiscovered, efforts are ongoing in the potential clinical applications. Particularly, enhancing Treg function in autoimmunity, organ transplantation and allergy or dampening it in the context of cancer and infectious diseases are aspects being considered (Fehervari and Sakaguchi 2004).

2.5 Evidence for Diabetes as an autoimmune disorder

Initially, the evidence for autoimmunity in patients with T1D came from studies in which sera from patients newly diagnosed with T1D reacted with pancreatic islets eventually leading to identification of auto-antigens with which the islet cell auto-antibodies (ICA) reacted.

2.5.1 Autoantibodies in T1D

2.5.1.1. GAD65

The first autoantibody to be detected was an isoform of glutamic acid decarboxylase, GAD65 encoded by a gene on chromosome 10p11 (Baekkeskov et al. 1990). It is expressed in neuroendocrine cells including the beta cells of the pancreas is located within neuron-like vesicles (Baekkeskov et al. 1990). Though discovered years

ago its function within the beta cell remains unknown, but 60-80% of newly diagnosed patients have autoantibodies to GAD65 (Notkins 2002)

2.5.1.2. IA-2

A second autoantigen, IA-2, is a member of the transmembrane protein-tyrosine phosphatase family encoded by a gene on chromosome 2q35 (Lan et al. 1996). IA-2 is a transmembrane protein found in secretory vesicles of both endocrine and neuronal cells (Solimena et al. 1996) and its function has remained a mystery though recent experiments in animals suggest that it may play a role in insulin secretion (Saeki et al. 2002). IA-2 is found in nearly the same proportion of newly diagnosed patients with T1D as GAD-65 (60-70%) (Notkins 2002).

2.5.1.3. Insulin

The third autoantigen associated with autoimmunity is an obvious one, insulin. Autoantibodies to insulin are among the first indicators of Type 1 Diabetes, appearing well before onset of disease during the prediabetic state called insulitis in 30 to 50% of children (Atkinson and Eisenbarth 2001).

2.5.2 Autoantibodies as tools in disease prevention

Though IA-2, GAD65 and insulin were at first used for diagnosis and clinical classification of Type 1 diabetics it was soon realized that they appear long before onset of disease and were good predictors of progression to clinical disease and had a cumulative effect. In fact, presence of one autoantibody predicts the likelihood of onset of disease within 5 years at 10%, 50% in the presence of two autoantibodies and 60-80% in the presence of three autoantibodies (Maclaren et al. 2003). Given this, it was possible

to define sub-clinical phenotypes and patients with the highest indicators of risk could be earmarked for preventative therapeutic strategies long before their beta cell mass is destroyed.

Beyond their clinical usefulness in screening, antibodies may provide clues as to the cause and the primary target of the autoimmune process. It is unlikely that autoimmunity is driven by a single autoantigen given that there is a multiplicity of autoimmune manifestations in patients with autoimmune disease. This is the case in both murine models of autoimmune disease as in human models. In the NOD for instance, diabetes is present along with thyroiditis, sialadenitis, autoimmune hemolytic anemia and antinuclear antibodies, reviewed in (Bach 2003) while human autoimmune diabetes often exhibits extrapancreatic manifestations, including Graves' disease. There is a well established association between T1D and autoimmune thyroid disease (AITD) with up to 20% of T1D patients carrying detectable thyroid antibodies, and 50% of these progressing to clinical AITD (Levin and Tomer 2003), making the case strong for a common immuno-regulatory mechanism in autoimmunity (Bach 2003).

2.6 Mechanisms of beta cell destruction

The destruction of the beta cell is a cell mediated immune process and this is central to the cause of autoimmune diabetes.

2.6.1 Cell-mediated immunity

It is well established that the principle effectors of the immune system are B and T lymphocytes. T-lymphocytes can be divided into T-helper (T_H) cells, which express $CD4^+$ surface antigens, and cytotoxic T-cells, which express $CD8^+$ surface antigens. The

$CD4^+$ T_H precursor cells can be further divided into two populations, T_H1 and T_H2 cells which are distinguished by the cytokines they secrete, T_H1 cells secrete interleukin-2 (IL-2), interferon-y (IFN-y) and tumor necrosis factor alpha (TNF-α) which regulate the cell-mediated response and (can) induce tissue damage. T_H2 cells secrete interleukins, IL-4, IL-5, IL-6 and IL-10 and are activated to assist B cells for specific immunoglobulin (Ig) production.

In Type 1 Diabetes, the direct cell-to-cell interaction between antigen-specific $CD8^+$ cytotoxic T-lymphocytes and autoantigens on beta cells results in beta cell killing via a combination of perforin-granzyme and Fas activation mediated cytolysis, (reviewed in Polychronakos 2001). In contrast, antigen-specific $CD4^+$ helper T lymphocytes do not recognize autoantigens on beta cells because beta cells do not express MHC class II molecules. Rather, they act by recognizing autoantigens that have been picked up and processed by APCs expressing class II molecules (Grogg et al. 1992). This indirect mechanism results in the release of a variety of effector molecules, such as INF-γ, TNF-α and nitric oxide, and results in bystander killing.

Based on animal models it is now generally accepted that multiple effector molecules and pathways are involved in the destruction of the beta cell. Much of the evidence supporting this notion comes from animal studies (since pancreatic biopsies are not routinely performed in humans) and the following sections address the current state of knowledge of animal models of Type 1 Diabetes and the (beta) cell-mediated immune response.

2.6.2 The non-obese diabetic (NOD) mouse

Developed in the late 1970s in Japan, the NOD mouse is the most commonly used animal model of Type 1 Diabetes. In the NOD mouse, disease develops spontaneously in 60-80% of females and 20-30% in males (Kikutani and Makino 1992; Bach 1994). The phenotype shares several clinical features with the human. Notably, as in the human, there is a period of insulitis that begins at about 4 weeks of age with relatively few immune cell infiltrates detected in histological studies before that. Diabetes onset occurs between the 15th and 25th week of life and though males exhibit similar levels of insulitis as the females they fail to develop diabetes at the same rate, (reviewed by Mathis and Benoist 2004). In addition, the NOD mouse is also particularly prone to developing other autoimmune diseases, such as autoimmune thyroiditis (Salomon et al. 2001) and autoimmune sialadentis (Hu et al. 1992). The pathogenesis of autoimmune destruction of the beta cells in the NOD mouse and in the human, is T cell mediated, and involves islet infiltration by macrophages and B cells (reviewed in Anderson and Bluestone 2004). As in the human form of diabetes, diabetes in the NOD mouse is also a polygenic disorder, with the single largest genetic contributor being the major histocompatibiity locus (MHC) (McDevitt et al. 1996), syntenic with the human leukocyte antigen (HLA) locus in humans. The individual contribution of loci in NOD syntenic with the human loci will be briefly discussed in later sections

2.6.3 T-cell mediated autoimmune diabetes in the NOD mouse

Initially, evidence for the involvement of autoreactive T cells in the NOD mouse came from immunohistological analyses of the pancreas revealing that the islet infiltrate are T cells (Wicker et al. 1986; Bendelac et al. 1987; Bach 1994). A large body of

evidence has established that T-cells are responsible for disease initiation and progression in the NOD mouse, including the ability of CD4+ T cells to transfer disease (Wicker et al. 1986; Bendelac et al. 1987) and the blocking of CD4+ by anti-CD4 mAbs therapy (Shizuru et al. 1988) preventing diabetes onset in the NOD. Although CD4+T cells are more actively involved in disease pathogenesis, CD8+ T-cells promote disease particularly early on and may play a role in effector function (reviewed by Anderson and Bluestone 2004).

Much effort has been devoted to identifying the antigen(s) involved in the pathogenesis of T1D in the NOD mouse and these include insulin, GAD, insulinoma associated protein-2 (IA-2) and heat shock protein 60, all produced in the pancreatic islets, with insulin being exclusively expressed in the islets (Lieberman and DiLorenzo 2003). However, epitope-spreading, whereby an initial immune response elicits other antigen responses in the area, prevents the identification of the antigens responsible for triggering disease. It is probable that various processes, including antigen mimicry, nonspecific inflammation, and defective tolerogenic processes, may combine to promote responses against the beta cell (Anderson and Bluestone 2004).

2.7 Etiology of Type 1 Diabetes

2.7.1 Environmental Risk Factors

Though there is a clear relationship between the degree of genetic identity with a proband and genetic susceptibility to T1D, evidence from monozygotic twin (who share almost 100% of their genes) studies pegs the concordance rate of T1D at 30-50% followed by first and second degree relatives (Onengut-Gumuscu and Concannon 2002). This indicates that genetic susceptibility alone does not explain the entire risk. There is

accumulating evidence that environmental factors can trigger autoimmunity in individuals with an underlying genetic predisposition.

2.7.1.1 Viral Infections

That infectious agents can induce experimental autoimmune diseases in several experimental conditions is well known but much of the evidence in humans is from epidemiological studies, (reviewed in Bach 2002) . Epidemiological evidence has shown a steady rise in autoimmune diseases in developed countries over the last three decades including multiple sclerosis (Rosati et al. 1988; Poser et al. 1989), Type 1 Diabetes (Green et al. 1992; 2000) among others with a concomitant decrease in infectious diseases due to the use of antibiotics, vaccination, improved hygiene and socio-economic conditions (Bach 2002). Whether this relationship is causal in humans remains to be determined, but the best evidence of a causal relationship between the two has been derived from experimental animal models. For instance, breeding the NOD mouse in an isolated and sterile environment can double the incidence of diabetes (Like et al. 1991) and that infection with mycobacteria (Martins and Aguas 1999), lymphocytic choriomeningitis virus (Oldstone 1990), murine hepatitis virus (Wilberz et al. 1991) among other pathogens actually prevents diabetes in the NOD. Though the mechanisms underlying the infection-autoimmunity relationship remain unknown, these may involve the balance between T_H1 and T_H2 cytokines, their interaction with regulatory T cells as well as other mechanisms, reviewed by (Bach 2002).

2.8 Genetics of complex diseases: Type 1 Diabetes

2.8.1 Approaches and tools in identifying genes in complex diseases

Two approaches have been used in identifying disease loci in Type 1 Diabetes: genome-wide linkage studies and association studies based on a candidate gene approach. The former has been successful in identifying the highest genetic risk locus, the HLA gene region and the latter in identifying variants within the insulin (*INS*) and *CTLA4* gene that confer diabetes susceptibility. The following is a brief review of these approaches in the search for genes in complex diseases like T1D.

2.8.1.1 Genome-wide scans and linkage analysis

Whole genome scans involve genotyping evenly spaced markers, spanning across the entire genome in nuclear families where there are affected sibling pairs (ASPs) or other affected relative pairs, reviewed in (Onengut-Gumuscu and Concannon 2002). Evidence for linkage in a specific chromosomal region is determined based on the extent of allele sharing summed across all ASPs genotyped (Onengut-Gumuscu and Concannon 2002). If an allele does not segregate with the disease locus, the inheritability of that allele between pairs of siblings would be 50%; thus deviations from the expected 50% would suggest that the alleles are in linkage with the disease locus (Hirschhorn 2003). The success of these studies depends largely on the sample size and it has been argued that it has limited ability to detect loci whose effects are modest, as is the case in most polygenic diseases. One definite advantage is that the entire genome can be analyzed (Onengut-Gumuscu and Concannon 2002) without prior assumptions (or bias) regarding the loci involved.

The statistical significance of the findings in these studies is typically reported in the form of a Logarithm of the Odds (LOD) score, or as a *P*-value. The LOD score is defined as: the likelihood of observing the data at the marker(s) assuming a particular genetic model with a disease allele linked to the marker and the likelihood of observing the data under the null hypothesis of no linkage (Hirschhorn 2003). Frequently, in complex diseases the genetic model underlying the association is unknown, so non-parametric methods of data analysis are used, which means that no assumptions are made about the dominance, frequency or magnitude of the effect of the disease allele (Hirschhorn 2003).

A useful measure of genetic contribution to disease risk is λ, the ratio of the risk to an individual with a defined relation to an affected individual to the population risk for a given disorder (Onengut-Gumuscu and Concannon 2002). Higher λ values indicate increased familial contributions to risk. For instance, in T1D, there is significant familial clustering with an average prevalence risk in siblings of 6% compared to 0.4% in the general population. Thus, the λ_s, or familial clustering, calculated from the increased T1D risk in siblings over population incidence (6.0/0.4) results in a λ_s value of 15 (Onengut-Gumuscu and Concannon 2002).

Once a linkage peak has been identified using anonymous markers, the goal is to pinpoint the gene responsible for the linkage. Although this can be done by positional cloning, there are relatively few successful examples of genome scans in complex diseases followed by positional cloning which led to the identification of the gene causing the association.

2.8.1.2 Fine-mapping linkage analysis and positional cloning

Once a linkage peak has been obtained for a particular region by a genome scan and subsequently confirmed in other scans, a typical approach to narrow the linkage peak (where markers are interspersed at ~10cM and the regions showing evidence of linkage can contain many genes) involves genotyping polymorphic markers at a high density within that linkage peak for association studies with the disease phenotype. Association studies are then performed to compare allele frequencies between affected and unaffected individuals. This requires testing either the disease-causing variant itself, or a marker in close proximity with it, based on a phenomenon known as Linkage Disequilibrium (LD). Linkage Disequilibrium describes the non-random correlation between alleles at a pair of physically linked polymorphic sites (Carlson et al. 2004). A given allele at one site coexists on the same chromosome with a given allele at the other site more often than expected by chance alone and, thus, non-functional polymorphisms (markers) in the vicinity can be used to obtain information about allele frequency at the functional site. Because LD occurs over distances in the genome, that typically range from a few kilobases (kb) to 100 kb (haplotype blocks), association studies require a markedly higher SNP density.

2.8.1.3 Genetic Association studies

The tool of choice for a geneticist wishing to narrow the causative marker to disease correlation is the association study. The basis of genetic association studies lies in comparing allele frequencies in unrelated patients and ethnically matched unrelated controls (a case-control study) under the assumption that if the allele tested is in linkage disequilibrium with the disease allele, or *is* the disease-allele, it will be appear more

frequently in patients than controls (Tomer and Davies 2003). A caveat in the case-control study is that spurious associations can be found if the patients and controls are not properly matched (i.e. they differ in attributes unrelated to the presence or absence of the disease) and this is known as population stratification (Spielman et al. 1993). The most obvious (though by no means the only) source of stratification is ethnic origin. Both disease risk and allele frequencies at many markers differ between populations. Therefore, if the case and control sets are not perfectly matched in the representation of two or more ethnic groups, any polymorphism with different allele frequencies between populations will be spuriously associated with the disease (association in the absence of linkage). Association studies are more sensitive in detecting *modest* gene effects (Risch and Merikangas 1996).

2.8.1.4 The Transmission Disequilibrium Test (TDT)

Population stratification is not trivial, and can lead to false associations which can be avoided by a family-based test, the Transmission Disequilibrium Test, developed by Speilman et. al., (Spielman et al. 1993). Since it evaluates the transmission of alleles from informative (heterozygous) parents to affected children, it uses the parents' untransmitted allele as an internal control thereby avoiding the difficult matching of patients and controls. Thus, the TDT evaluates the significance of the distortion (by a χ^2 or binomial test) between the expected equal transmission of alleles from heterozygous parents at a defined SNP (or any other marker) to their affected child and the observed transmission of alleles. In this way it functions both as a test of association between marker and disease and a test of linkage, since it also demonstrates that the marker used is in close proximity to the disease locus (in LD). Given that whole genome linkage scans are

unlikely to detect loci with modest effects, association studies are the tool of choice when a gene (or gene region) is suspected in disease susceptibility. At the present time whole genome high-density association studies are not feasible due to limitations in current genotyping technologies (too many SNPs required to account for human variability). Awaiting the results of the HAPmap consortium project (2003), which aims to restrict the number of required SNPs by using the SNP(s) which best characterize the LD block, referred to as "haplotype tag SNPs" (Johnson et al. 2001), and for technologies with the required high throughput and manageable cost, researchers have resorted to the candidate gene approach in association studies.

Association studies performed on genes with an *a priori* hypothesis based on biological considerations (*i.e.* the *INS* gene in T1D) or information obtained from extensive research in the NOD mouse use the candidate gene approach. This approach aims to use the case-control or the family-based TDT to test "tag SNPs", (or the smallest number of SNPs possible) for association with disease. The aim is to discover a DNA variant (typically a SNP) which can explain the disease effect, through effects on protein function if the SNP is in a protein-coding region, or through regulatory effects (altering expression levels rather than sequence of protein products) modulated by 5'and/or 3'SNPs.

2.8.2 Allelic variations in protein function and/or gene expression

Underlying the genetic association and linkage studies is the notion that human variation in disease susceptibility and protection can be explained or accounted for by the variation in the genes or regions surrounding genes encoded by DNA. In addition, phenotypic variation between individuals is crucial to evolutionary adaptations of natural

and artificial selection. These notions are central to modern human genetics and to this Thesis, and a section devoted to it follows.

2.8.2.1 Coding sequence polymorphisms

At the present time, most of the DNA variation in the form of single nucleotide polymorphisms (SNPs) associated with complex diseases has been found in the coding exons of genes (Glazier et al. 2002) though this likely reflects a bias because regulatory regions of genes are less well defined and characterized. Typically after sufficient evidence has been amassed and one gene has been narrowed down in susceptibility to a complex disease, functional tests where swapping one genotypic variant for another in a transgenic system or in cell lines would provide supportive evidence of its effect in disease. Though not all positions in a protein are equally important for function, conserved sequences are more likely than not to harbor important functional tasks, and substitutions at these residues could conceivably alter protein function. Whereas most Mendelian diseases are caused by coding mutations which significantly disrupt protein function, and tend to be localized in highly conserved regions, coding SNPs associated with complex traits are involved in moderate effects on protein function and are unlikely to result in its complete disruption (Thomas and Kejariwal 2004).

2.8.2.2 Regulatory variation in complex diseases

Whereas coding regions are relatively well characterized and it is easy to envision how changes at the amino acid level could lead to altered protein function, the role of regulatory variation in modulating disease susceptibility is more difficult to unravel, if only because the gene regulation infrastructure is still poorly understood. However, it has

been estimated that it is found in 6% of mouse genes (4 out of 69 genes tested in three tissues derived from adult mouse strains, showed regulatory variation and is likely an underestimate of the true variation) (Cowles et al. 2002) and in 18-20% of genes in human lymphoblastoid cell lines (Cheung et al. 2003; Pastinen et al. 2004), though there is considerable variability in the level of variation reported (Pastinen et al. 2004). There are also studies of such gene expression variations in budding yeast, *Drosophilia* and between populations of fish and primates (Knight 2004). It appears that a majority of regulatory variation is controlled *in trans*, while a minority is regulated *in cis* (Morley et al. 2004) and shows familial aggregation (Yan et al. 2002; Cheung et al. 2003). Regulatory regions typically encompass the adjacent regions around genes, and include the promoter region as well as the 5' and 3'untranslated regions (UTR) and can extend over long distances. Allele-specific differences in expression have been reported to influence mRNA stability, mRNA processing efficiency or mRNA isoform expression (Pastinen and Hudson 2004). The following section will summarize what is known about regulatory variation at promoter regions and beyond, with an emphasis on the tools used to identify these allelic differences.

Regulatory variation at promoters

It was recently estimated that 34% of promoter polymorphisms significantly altered reporter gene expression by more than 1.5 fold in at least one of three transfected human cell lines in a screening exercise of the proximal promoters of 170 genes tested (Hoogendoorn et al. 2003). In addition, systematic assays of promoter haplotypes by reporter gene analysis on chromosomes 21 (Buckland et al. 2004) and 22q11

(Hoogendoorn et al. 2004) show functional variation among 18% and 20% of polymorphisms tested respectively

2.8.2.3 Detecting allele-specific expression differences: *In vitro*

Determining allelic expression differences at the promoter is more commonly assayed by reporter gene systems (Knight 2004; Pastinen and Hudson 2004) where cells are transiently transfected with allele specific promoter constructs. A caveat in this method is that the results are highly context and cell line dependent and typically only include 1kb or less of genomic context (Knight 2004). However, a recent analysis of human polymorphisms assayed by transient reporter assays in physiologically relevant cell lines revealed that most functional polymorphisms were in proximal promoter regions, with just over one quarter located beyond 1Kb (Knight 2004) but these are likely due to detection bias, as promoter regions, too are poorly defined. Promoter allelic differences can be prioritized for experimental verification by computational based approaches used to predict regulatory regions, combined with interspecies comparisons to identify evolutionary conservation, since conserved regulatory sequences are more likely to be functional (non-coding regions are more degenerate than exonic regions) (Elgar 1996; Hardison et al. 1997) though these approaches need not be restricted to promoter regions.

2.8.2.4 Detecting allele-specific expression: *In vivo*

It is possible to directly assess the relative abundance of allelic transcripts in their normal chromosomal context by quantifying marker alleles in cells or tissues (Figure 1). This is possible when an individual is heterozygous for a polymorphism located in exonic sequences. The underlying concept is that both copies of autosomal genes are expressed

in a codominant manner or in equal proportions such that the quantification of relative abundance of transcripts in RNA would not be expected to deviate from an equimolar ratio if both copies are equally expressed; if it does, then one allele is preferentially expressed over the other (Pastinen et al. 2004). This same approach has been used extensively in detecting imprinted genes (Reik and Walter 2001) but in the case of "allelic imbalance" there is no parent-of-origin influence on gene expression (Pastinen et al. 2004). When *cis*-acting variation is present, the mechanism via which it modifies gene expression can involve altered transcriptional activity (mainly at the promoter) through differential binding of transcription factors or enhancers, or mRNA stability especially if the causative allele is in the 3'UTR region. It may also alter polyadenylation signals or result in alternative splicing. Although there are numerous reports of allelic imbalance studies in the literature, deciphering the mechanisms underlying these differences has proven to be more difficult.

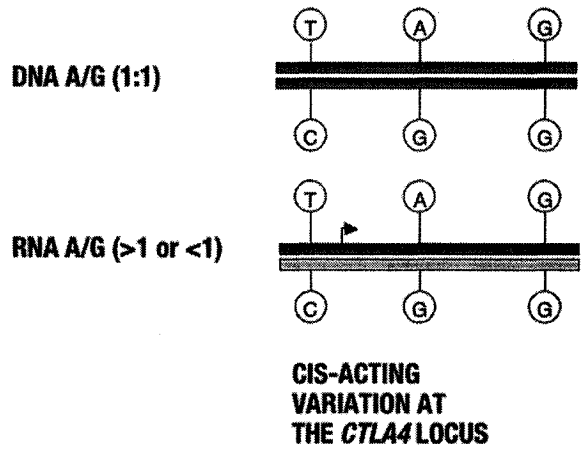


Figure 1: Detecting allelic expression differences *in vivo*.

Using the transcribed signal peptide polymorphism +49A>G in the *CTLA4* gene as a marker, for instance, one can detect allelic differences in expression by comparing relative amounts of transcript derived from each allele in the PCR derived from RNA and DNA from a heterozygous individual at the marker of interest.

2.9 Whole genome and fine linkage analysis in Type 1 Diabetes

At least 5 genome-wide linkage scans have been performed for Type 1 Diabetes with the latest failing to find significant evidence of linkage beyond the Human Leukocyte Antigen, *HLA* (Cox et al. 2001). These have been followed by many studies on portions of the genome for which there was suggestive evidence of linkage. No region outside of the *HLA* achieved genome-wide significance in all genome scans but there are about 18 regions (Table 1) which showed some evidence of linkage, reaching the suggestive level, sometimes, but not in all cases (Cox et al. 2001)

The following sections will briefly review the data on the loci with the highest evidence for linkage with Type 1 Diabetes particularly, the *HLA*, *INS* and *CTLA4* with what is known about the etiological mutations to follow.

2.10 *IDDM1*: HLA-encoded susceptibility to T1D

The HLA region exhibits high genetic complexity with multiple genes in linkage disequilibrium. class I genes, which include the *HLA* genes, A, B and C; class II genes including the *HLA-DR*, *DP* and *DQ* genes; and class III genes including *TNFα*, heat shock protein 70 and several members of the complement family are all contained within this cluster, (reviewed in Tomer and Davies 2003) (Figure 2). The statistically highest association within the HLA cluster is obtained with the class II genes, particularly with alleles of the *HLA-DQA1*, *DQB1*, and *DRB1*, loci all involved in the genetic predisposition to T1D in human diabetes.

2.10.1 Class II *HLA* genes and susceptibility to T1D

Earlier associations with T1D were detected mainly in DR alleles, particularly with the DR3 and DR4 alleles which bore the highest risk (Todd et al. 1987; Morel et al. 1988). Currently, the high risk alleles are defined in the context of their haplotype which includes various alleles at DR and DQ (Table 2) and whose individual contribution to T1D risk is very difficult to ascertain due to the strong LD in the region. It is known that several combinations of alleles are associated with disease, for instance DQ2/DQ8 alleles are associated with increased susceptibility to T1D (30% of patients are heterozygous for DQ2/DQ8 alleles) (Thomson et al. 1988) while a particular DQ6 molecule appears to confer dominant protection and is inversely associated with disease even in the presence of T1D autoantibodies *and* in the presence of high-risk *HLA* alleles (DQ6 alleles are reported in less than 1% of patients, compared to 15-20% in the general population) (Pugliese et al. 1995). Close to 95% of patients with Type 1 Diabetes have either the DRB1*0301,DQB1*0201 or the DRB1*04, DQB1*0302 haplotype and while the heterozygous genotype DRB1*0301,DQB1*0201/DRB1*04, DQB1*0302 is present in only 2% of the general population it is found in nearly 40% of Type 1 Diabetics in many, but not all populations (van der Auwera et al. 1995; Rewers et al. 1996).

Table 1: Suspected loci involved in T1D susceptibility

Generated from the latest genome-wide scan generated by Cox et. al.(Cox et al. 2001). Adapted from (Pugliese and Eisenbarth 2000).

Locus	Chromosome	Candidate Genes	Markers	LOD
IDDM1	6p21.3	*HLA DR/DQ*	-	65.8
IDDM2	11p15.5	*INSULIN-VNTR*	-	4.38
	16q22-24		D16S3098	3.93
PTPN22	1p13	*LYP=PTPN22*	SNP (R620W)	
IDDM3	15q26	-	D15S107	-
IDDM4	11q13.3	*MDU1, ZFM1, RT6, ICE, LRP5, FADD, CD3*	FGF3, D11S1917	
IDDM5	6q25	*MnSOD*	ESR, a046Xa9	1.96
IDDM6	18q12-q21	JK (Kidd), ZNF236	D18S487, D18S64	
IDDM7	2q31-33	*NEUROD*	D2S152, D251391	2.62
IDDM8	6q25-27	-	D6S281, D6S264 D6S446	1.81
IDDM9	3q21-25	-	D3S1303, D10S193	-
IDDM10	10p11-q11		D10S565	2.8
IDDM11	14q24.3-q31	*ENSA, SEL-1L*	D14S67	-
IDDM12	2q33	*CTLA4*	SNP=T17A (49A>G), 3'UTR: $(AT)_n$, +6230G>A*	
IDDM13	2q34	*IGFBP2, IGFBP5, NEUROD, HOXD8*	D2S137, D2S164, D2S1471	
IDDM15	6q21	-	D6S283, D6S434, D6S1580	-
IDDM16	IGH	-	-	-
IDDM17			D10S1750, D10S1773	1.56
IDDM18	5q31.1-33.1	IL-12βp40		
	1q42			2.2
	16p11-13			1.74
	17q25			1.81
	19q11			1.80

Table 2: Class II *HLA* haplotypes and their association with Type 1 Diabetes.

Adapted from Melanitou et. al.,(Melanitou et al. 2003).

DRB1	DQA1	DQB1	Diabetes Risk
*0301(DR3)	*0501	*0201(DQ2)	High Risk
*0401(DR4)	*0301	*0302(DQ8)	High Risk
*0405(DR4)	*0301	*0302(DQ8)	High Risk
*0801(DR8)	*0401	*0402(DQ4)	Moderate Risk
*1601(DR2)	*0102	*0502(DQ5)	Moderate Risk
*0403(DR4)	*0301	*0302(DQ8)	Neutral or weakly protective
*1501(DR2)	*0102	*0602(DQ6)	Strongly protective
*0401(DR6)	*0101	*0503(DQ5)	Strongly protective
*0701(DR7)	*0201	*0303(DQ9)	Strongly protective
*0701(DR7)	*0201	*0201(DQ2)	Protective

2.10.2 Other *HLA* genes

Though most of the evidence for association with T1D comes from the class II alleles at the HLA, other genes within the Class I cluster, in particular have been reported to be associated with Type 1 Diabetes. In the first such study, parents with sibling pairs affected with T1D who were homozygous at the class II DR3 haplotype but heterozygous at the class I B locus, showed significant deviation from the expected 50% allele sharing between siblings under the null hypothesis of no HLA variation outside the defined DR3 haplotype (Robinson et al. 1993). This implicated other HLA loci in diabetes susceptibility. Several studies since have also suggested that class I genes may influence susceptibility in the presence of other clinical aspects, such as age of onset (Demaine et al. 1995; Fujisawa et al. 1995; Pitkaniemi et al. 2004), and the rate of beta cell destruction (Nakanishi et al. 1997; Panagiotopoulos et al. 2003). Evidence for a second *IDDM* locus in the HLA class I region, telomeric to HLA-F was provided (Lie et al. 1999). In this study, the parents were chosen to be homozygous for the HLA class II DR-DQ haplotypes so as to eliminate possible confounding effects due to linkage disequilibrium. Other studies have since provided evidence to support the existence of a second locus outside the HLA class II region in T1D susceptibility (Lie et al. 1999; Lie et al. 2002; Johansson et al. 2003; Johansson et al. 2003). Evidence for yet another critical region for diabetes susceptibility was provided by (Hanifi Moghaddam et al. 1998) located between the *TNF* and *Hsp70* genes, a region of approximately 200kb in the class III HLA locus and other studies have also provided evidence in support of a class III association with T1D independent of the class II region, (reviewed by Pugliese and Eisenbarth 2000).

2.10.3 Mechanism underlying *HLA* encoded susceptibility

The strong linkage disequilibrium across the HLA locus makes it challenging to distinguish the contributions of individual alleles to diabetes susceptibility or protection. It is thought that HLA-DQ β-chain (DQB1) alleles lacking an aspartic acid at position 57, near a region on the MHC molecule involved in MHC-peptide binding interactions, are strongly associated with increased susceptibility to T1D (Hoover and Marta 1997; Sanjeevi et al. 1997). Disrupting the peptide binding interactions may result in inefficient binding and presentation of islet-antigen derived peptides. A study of the homologous I-Ag7 molecule in the (NOD) mouse found that it was generally a poor peptide binder (Carrasco-Marin et al. 1996). However, the presence of the aspartic acid at position 57 has also been reported in Japanese patients (Awata et al. 1992), and in studies of Caucasian patients (Todd et al. 1989; Mijovic et al. 1991; Ronningen et al. 1991). A presence of an arginine at position 52 of the HLA-DQ β-chain (DQA1) is also associated with increased risk to diabetes (Khalil et al. 1990), but the mechanism by which these polymorphisms affect Type 1 Diabetes susceptibility, though believed to involve differences in efficiency of presentation of peptides by the various alleles, largely remains unknown.

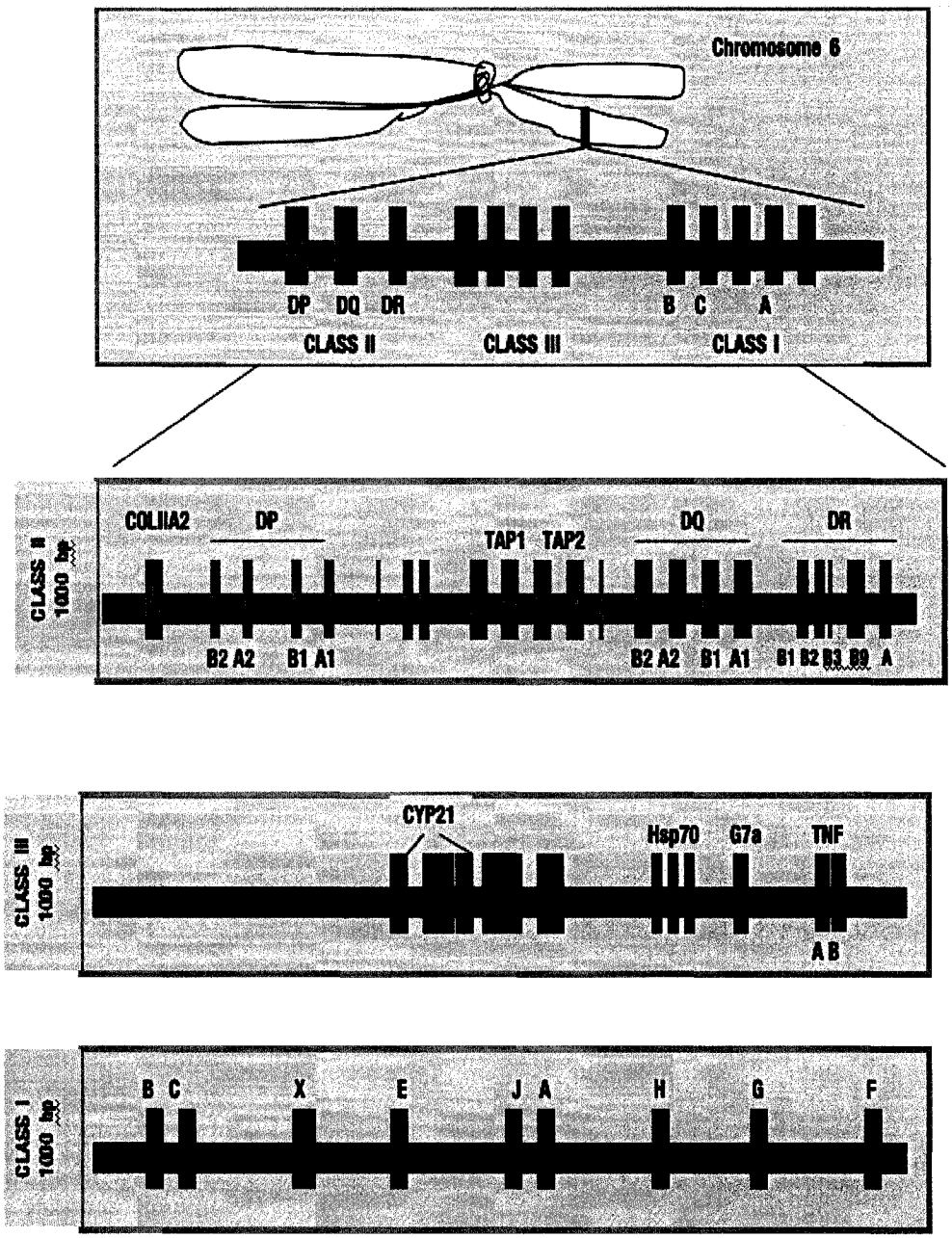


Figure 2: Simplified map of the *HLA* locus.

The major genes within each class indicated.

Adapted from (Williams and Pickup 2001)

2.10.4 Synteny with the NOD mouse: the MHC locus

The NOD mouse harbors a unique major histocompatibility complex (MHC) class II molecule, with a haplotype in particular, the $H2^{g7}$ essential for disease susceptibility (Tisch and McDevitt 1996). Homozygosity at this locus may be necessary for disease development but the exact mechanism by which the MHC susceptible alleles act is not clearly defined though it is believed to function by skewing thymic selection processes, modulating regulatory populations or skewing T_H1 versus T_H2 effector responses, (reviewed in Anderson and Bluestone 2004).

2.11 *IDDM* loci outside the HLA

HLA-encoded susceptibility may account for less than 50% of the inheritable disease risk: the 25% of siblings that share no *HLA* haplotype with the T1D-affected proband still have a disease risk seven-fold higher than the general population (Risch 1987). This means that there is a substantial role for non-HLA encoded susceptibility and that T1D behaves as a polygenic disease. Though multiple genome-wide scans have been performed and numerous genomic regions implicated in diabetes risk, these regions are seldom replicable across the studies (the exception being the *HLA* itself). It is now recognized that none of these studies had the power to detect the small genetic effects that make up the rest of the susceptibility (Cox et al. 2001). The second locus identified (*IDDM2*), which included the candidate Insulin gene, though widely accepted as a substantial genetic contributor to T1D has only modest effects in linkage studies, with a

reported Odds Ratio (OR) of 3 and a calculated sibling risk ratio (λ_s) of 1.12 (Pociot and McDermott 2002). In fact, there has been a lack of significant evidence for linkage at the *IDDM2* region (possible due to a lack of power) and its acceptance as a susceptibility locus is based on highly significant evidence from multiple association.

2.12 *IDDM2*: the *INS*-VNTR

The insulin gene region was one of the first candidate loci to be examined and early reports of association with T1D susceptibility (Bell et al. 1984) were repeatedly confirmed (Julier 1991, Bennett 1995). The association of this region with T1D can be attributed with a large degree of probability to a region located upstream of the insulin gene, a series of variable number of tandem repeats (VNTR) unique to humans and primates (Julier et al. 1991; Bain et al. 1992). The minimal region encompassing the *INS*-VNTR, the insulin gene itself as well as the promoter region of the Insulin like Growth Factor-2 *(IGF2)* and the 3' flanking region of thyrosine hydroxylase (*TH*) genes, extends approximately 4.1kb and harbors several other polymorphisms in LD with the VNTR (all outside coding regions). While any one of these could have allelic effects on gene expression, multiple studies have successfully excluded neighbouring *INS*-VNTR polymorphisms as well as the *IGF2* gene (Owerbach and Gabbay 1993; Bennett et al. 1995; Undlien et al. 1995). However, since these analyses have been restricted to haplotypes within the 4.1kb region and not beyond it, the possibility remains (albeit small) that the *INS*-VNTR may actually be in LD with the causative variant outside the 4.1kb region (Doria et al. 1996). There are three classes of VNTR alleles, defined by the number of repeats they contain, with Class I alleles being the shortest and Class III the longest (the intermediate sized Class II alleles are infrequent in Caucasian populations,

occasionally seen in Africans). The Class I alleles are found reproducibly associated with diabetes susceptibility, when homozygous (denoted as I/I) while the longer Class III alleles are associated with dominant protection (Bennett and Todd 1996).

2.12.1 *INS*-VNTR and function

The *INS* VNTR was postulated early on to exert its effect through transcriptional regulation (*in cis*) since it does not encode protein and is directly upstream of the insulin gene promoter though a possible regulatory effect of the VNTR on *IGF2* was an alternative possibility.

The effect of the various VNTR alleles on insulin expression was investigated using *in vitro* and *in vivo* approaches, by our lab and others, in the pancreas and subsequently in the thymus. The Class I VNTR alleles correlated with higher pancreatic insulin expression *in vivo* and *in vitro*, and lower insulin expression in the thymus (*in vivo*, as there are no thymic cell culture lines), while the Class III alleles, associated with a strong protective effect in T1D, modulated a lower pancreatic insulin expression, but a higher thymic insulin level (Table 3) (Pugliese et al. 1997; Vafiadis et al. 1997). When *IGF2* was tested as an alternative possible target for the effect of the insulin VNTR, no effect on expression in tissues relevant to T1D was found (Vafiadis et al. 1998; Vafiadis et al. 2001) although it may have a small effect on placental *IGF2* expression (Paquette et al. 1998).

Table 3: Summary of *Insulin* expression *in vivo* in pancreas and thymus

Homozygosity for the longer alleles (class III/III) is associated with a 3-to-5-fold reduction in Type I Diabetes susceptibility compared with the Class I/I genotype.

Tissue	Relative expression Class I vs. Class III VNTR alleles	Fold Increase	References
Pancreas (adult)	I > III	1.25	(Bennett et al. 1995; Bennett et al. 1996)
Pancreas (fetal 7-20wks)	I > III	1.15	(Vafiadis et al. 1996)
Thymus (fetal 7-20wks)	III > I	~2.7±0.7	(Vafiadis et al. 1997)
Thymus (fetal/post-natal)	III > I	~ 2.4 ± 0.4	(Pugliese et al. 1997)

2.12.1.1 Insulin VNTR and thymic insulin expression

Feedback metabolic mechanisms regulating insulin secretion were likely to correct any differences in insulin transcription in the pancreas. However, in the thymus, altered expression of insulin provided an attractive hypothesis leading us to postulate that higher insulin thymic transcription modulated by the Class III alleles enhanced the negative selection of insulin-autoreactive T-lymphocytes in the thymus, reviewed in (Anjos and Polychronakos 2004). We have recently confirmed this prediction in a rodent model in which there is graded thymic insulin levels with no change in pancreatic insulin levels (Chentoufi and Polychronakos 2002). A mechanism accounting for the altered levels of insulin expression in the thymus via the VNTR remains unknown. A possible hypothesis is that the *INS*-VNTR alleles affect chromatin unraveling and/or RNA polymerase loading, but this remains untested. The recent isolation and creation of the rare thymic stroma cells that produce insulin will allow to address some of these questions.

Another region which has been implicated in Type 1 Diabetes susceptibility is that on the long arm of chromosome two at 2q31-33 known to encompass several genes involved in immune regulation and will be the subject of the following sections as well as the remainder of the thesis.

2.14 *IDDM12*: The *CTLA4* gene

2.14.1 Linkage and association of the cytotoxic T lymphocyte antigen-4 (*CTLA4*) T1D: the early studies

Within a 23 cM interval at 2q31-35 in humans, syntenic with the NOD mouse locus *Idd5* on chromosome 1, early reports proposed at least three Type I Diabetes loci (*IDDM7*, *IDDM12* and *IDDM13*) (Copeman et al. 1995; Morahan et al. 1996; Nistico et al. 1996). Within this 23cM interval, T1D association with *CTLA4* polymorphisms was reported (Nistico et al. 1996; Marron et al. 1997) and localized within a linkage disequilibrium block covering the D2S72-*CTLA4*-D2S116 interval (Marron et al. 2000). The *CTLA4* gene was the preferred candidate because of its role in T cell function and the bearing of potentially functional polymorphisms, though *CD28* and *ICOS* are also perfectly suitable functional candidates.

Earlier association studies focused on two markers within the *CTLA4* gene itself, a single base substitution in the signal peptide (+49A>G), resulting in a Threonine to Alanine substitution, as well as an $(AT)^n$ microsatellite repeat in the 3'untranslated region (UTR), but stronger effects were since found at two other SNPs, the –318C>T in the proximal promoter region and at +6230G>A in the 3' flanking region. The first evidence of association was initially reported in Italian families and in combined sets of Sardinian, UK and USA populations (n=818) (Nistico et al. 1996). The TDT test showed there was increased transmission of the G allele at the signal peptide of *CTLA4* to affected siblings. More evidence of association was subsequently reported in a multi-ethnic collection of families with one or more affected offspring which included families from China, Korea, Mexican-Americans, Spain and France (Marron et al. 1997). In these studies the TDT

revealed significant transmission distortion for alleles at both the $(AT)^n$ repeat and the +49A>G polymorphism (again, the G allele and the longer alleles at the microsatellite (106 bp) were overtransmitted). The overall evidence for linkage remained highly significant when combining data sets from both studies where increased transmission distortion for the G allele was seen (Marron et al. 1997). The strongest evidence for association was obtained in Italian, French and Spanish populations, Mexican-Americans and Koreans ((Marron et al. 1997). At the dinucleotide repeat, Marron et. al., found significant transmission distortion from parents heterozygous for the 88bp and 106bp alleles but not for any of the other alleles (long, >106bp or intermediate, between 88bp and 106bp) (Marron et al. 1997), a negative result of uncertain significance, because of lack of statistical power due to the rarity of these alleles. In a follow-up study by the same group, similar results were obtained for the 88bp allele (P=0.001) in a combined data set of US, European and Mexican American T1D families (178 with a single affected child, 350 with more than one affected child) (Marron et al. 2000).

Evidence for association of *CTLA4* polymorphisms with Type 1 Diabetes in German families was also reported in the first analysis of haplotypes. Increased transmission of the G allele at the *CTLA4* signal peptide in combination with the longer alleles at the microsatellite marker was found, and a –318 C>T promoter dimorphism seemed to have a small protective effect when transmitted *in cis* with the +49A and the shorter $(AT)^n$ alleles (Braun et al. 1998; Donner et al. 1998).

2.14.2 Association of *CTLA4* in the presence of *HLA* risk alleles

HLA genotype does not appear to affect *CTLA4*'s s modest but significant effect on T1D. In a study of the Belgian Diabetes Registry, the common +49G allele at *CTLA4*

was associated with T1D regardless of status at high risk haplotypes *HLA-DR3* or *DR4* alleles; HLA-DQA1*0301-DQB1*0302 and/or DQA1*0501-DQB1*0201 (and *INS-VNTR* status) (Van der Auwera et al. 1997). In separate study, *CTLA4* haplotypes were analyzed with *HLA* high risk *DQ8* and *DQ2* alleles in a German population and similarly, did not confer a protective effect (Braun et al. 1998; Donner et al. 1998). More recently, in a Japanese study where high risk *HLA* alleles unique to the Japanese were taken into account when evaluating the association of the +49G allele, they found no correlation between *HLA* and *CTLA4* genotypes in T1D (Kikuoka et al. 2001). Similar findings were reported in a Turkish population (Genc et al. 2004).

2.14.3 Fine Mapping, linkage disequilibrium, and association studies at *CTLA4*

Attempts at identifying the functional variant in the linkage interval at 2q33 were first undertaken by Marron et. al., (Marron et al. 2000) who showed that the markers linked and associated with T1D were contained within a phagemid artificial chromosome clone of 100 kb including the *CTLA4* gene but not *CD28* nor *ICOS* suggesting that the T1D gene in this region is either *CTLA4* or another gene in close proximity to it (Marron et al. 2000). In this study, there was significant transmission distortion of three markers within the *CTLA4* gene; the +49A>G, an intronic C>T SNP and the $(AT)^n$ repeat in the combined data sets from the US, Mediterranean-European and Mexican-American populations ($P=0.01$) but none with *CD28* markers (Marron et al. 2000). CD28 was also excluded in a study by Ban et. al., (Ban et al. 2003) and more recently by (Ueda et al. 2003).

Association studies of *CTLA4* markers with T1D have been carried out extensively in the last few years (Table 4), and show overwhelming evidence of association at the signal peptide and at microsatellite markers (Vaidya and Pearce 2004). One study reported significant association of the promoter –318C>T polymorphism in the Chinese; though contradictory reports also exist for all three markers (reviewed by (Vaidya and Pearce 2004)). Recently, a meta-analysis of reported genetic associations with *CTLA4* (at the +49A>G), showed the number of statistically significant follow-up studies for *CTLA4* was above that expected by chance alone, and was not due to population stratification and/or admixture, nor could it be accounted for by genetic heterogeneity (Lohmueller et al. 2003). These studies did not pin-point the functional variant and the association of *CTLA4* polymorphisms with T1D may be due to LD between the markers tested and the causative polymorphism, though this should be limited to a region of a few hundred kilobases based on the defined LD block in the region (Johnson et al. 2001).

Recently association studies with tag SNPs, identified by resequencing a 300kb region including *CTLA4*, *ICOS* and *CD28,* further defined this LD block to contain the *CTLA4* gene region, the 5' flanking *ICOS* region but not *CD28* (Ueda et al. 2003). The studies were carried out in three autoimmune endocrinopathies: Graves' disease (GD), autoimmune Hashimoto Thyroiditis (HT) and T1D. They genotyped a total of 108 SNPs and the *CTLA4* $(AT)^n$ microsatellite polymorphism in nearly 400 patients with GD and controls for association studies. By constructing linkage disequilibrium blocks and employing various statistical models they mapped the susceptibility locus in GD to three main peaks within this LD block (Figure 3). These included: the 5'(UTR) *CTLA-4* region

and 5'flanking *ICOS* region, with the strongest linkage peak in the 3'UTR of *CTLA-4* (Ueda et al. 2003). The *ICOS* and *CD28* gene regions were completely excluded from this association and the authors found no other genes in the region. The region with the highest susceptibility peak in GD harbored four SNPs, designated CT60, JO31, JO30 and JO27_1 and while the CT60 SNP was the most highly associated marker, the authors were unable to further dissect the susceptibility locus with respect to the four SNPs at the association peak. In Graves' disease the prevalence of the G-allele at CT60 was 63% in patients and 53% in healthy controls (Odds risk (OR), 1.51), and was also associated with Autoimmune Hypothyroidism (AH). (OR, 1.45).

In T1D, the authors used the TDT in 3600 Type 1 Diabetic families with one or more affected children. They genotyped 9 SNPs including the +49A>G, and the CT60, and found a modest effect of the 3'SNP on disease susceptibility (OR, 1.15) but could not rule out the 5'end of *CTLA4* (Ueda et al. 2003).

2.14.4 NOD mouse genetics: the *ctla4* gene

Chromosome 1 of the mouse harbors the fifth susceptibility locus (termed insulin dependent diabetes-5 or *Idd5*) in the NOD mouse, which contains several genes involved in immune response including *Ctla4* and *Cd28,* reviewed in (Anderson and Bluestone 2004). *Cd28* was recently excluded as the causal gene in *Idd5* and *Ctla4* is currently thought to be the causal gene in the region (Anderson and Bluestone 2004).

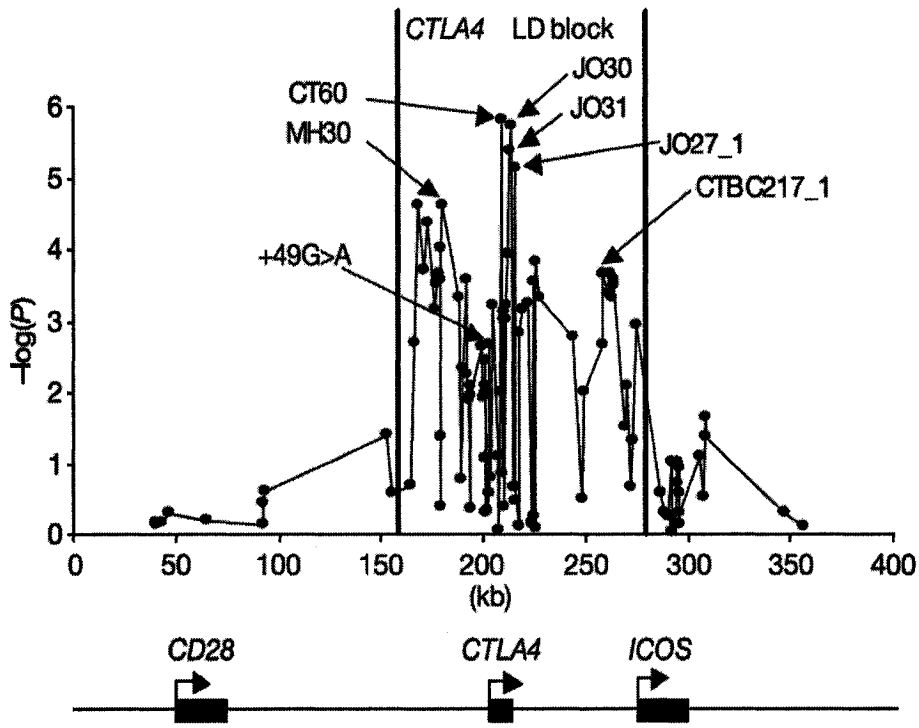


Figure 3: The association map of *CTLA4* gene region and Graves' disease.

The log *P* values (y axis) of the differences between cases of Graves' disease (n=384) and controls (n=652) were plotted against physical distance (x axis). Vertical lines indicate the two boundaries of the LD blocks. Lines drawn between points are for the sake of clarity as the association with disease and markers is not linear. Figure reproduced from (Ueda et al. 2003) with permission, copyright Nature Publishing Group 2005.

Table 4: Genetic association studies of *CTLA4* polymorphisms with Type 1 Diabetes.

Patients/ Controls	Population	Marker	Odds Ratio	P value	Reference
606/502	Caucasian (Swedish)	$(AT)_n$106bp	1.84	0.002	(Lowe et al. 2000)
160/200	Japanese	$(AT)_n$88bp	0.54	0.0012	(Ihara et al. 2001)
118/195	Japanese	$(AT)_n$	NS	NS	(Ban et al. 2003)
751/502	Caucasian (Swedish)	$(AT)_n$106bp	1.3	0.0001	(Graham et al. 2002)
293/325	Caucasian (German)	+49A>G	1.4 Gallele	<0.004	(Donner et al. 1997)
244/274	Caucasian (US)	+49A>G	NS	NS	(Marron et al. 1997)
89/57	Caucasian (Spanish)	+49A>G	NS	NS	
97/112	Korean	+49A>G	NS	NS	
180/379	Chinese	+49A>G	NS	NS	
525/530	Caucasian (Belgium)	+49A>G	1.5 Gallele	<0.005	(Van der Auwera et al. 1997)
173/425	Japanese	+49A>G	NS	NS*	(Awata et al. 1998)
112/100	Caucasian (French)	+49A>G	1.8 Gallele	0.002	(Djilali-Saiah et al. 1998)
192/136	Caucasian (Polish)	+49A>G	1.7 Gallele	0.002	(Krokowski et al. 1998)
117/141	Japanese	+49A>G	NS	NS**	(Hayashi et al. 1999)
110/200	Japanese	+49A>G	NS	NS	(Yanagawa et al. 1999)
111/445	Japanese	+49A>G	NS	NS***	(Abe et al. 1999)
253/91	Chinese (Taiwan)	+49A>G	1.7 Gallele	0.0051	(Lee et al. 2000)
125/200	Japanese	+49A>G	1.5 Gallele	0.018	(Kikuoka et al. 2001)
160/200	Japanese	+49A>G	1.8 Gallele	0.0002	(Ihara et al. 2001)
144/307	UK	+49A>G	NS	NS	(McCormack et al. 2001)
74/107	Japanese	+49A>G	1.3 Gallele	0.01	(Takara et al. 2000)
90/94	Filipino	+49A>G	1.9 Gallele	0.003	(Klitz et al. 2002)
62/84	Alsacian	+49A>G	5.6 Gallele	<0.05	(Ongagna et al. 2002)
134/273	Caucasian (French)	+49A>G		NS	(Fajardy et al. 2002)
305/289	Caucasian (Czech)	+49A>G	NS	NS	(Cinek et al. 2002)
97/60	Japanese	+49A>G	NS	NS**	(Mochizuki et al. 2003)
347/260	Chinese (Taiwan)	-318C>T	1.9 Callele	0.0026	(Lee et al. 2001)
160/200	Japanese	-318C>T	NS	NS	(Ihara et al. 2001)

NS, not significant.

*Association found in a subgroup of patients with insulin depletion and who required insulin within 1 month of diagnosis ($P = 0.012$).

**Significant association found in a subgroup of patients with GAD autoantibodies ($P<0.05$).

***Significant association found in a subgroup of patients with islet-specific cell antibody 512 antibodies ($P=0.004$).

Table 5: Family-based association studies of *CTLA4* with Type 1 Diabetes

Families n	Population	Marker -318C>T (*P*-value)	Marker 49A>G (*P*-value)	Marker: $(AT)_n$ (*P*-value)	Marker +6230G>A (*P*-value)	Reference
301	US		G (NS)	106		(Marron et al. 1997)
44	French		G (NS)	106		
18	Spanish		G (NS	106		
97	Mexican American		G (0.002)	106 (0.001)		
31	Chinese		G (NS)	106 (NS)		
41	Korean		G (0.03)	106 (NS)		
231	**All excluding US**		**G ($2x10^{-5}$)**	**106 (0.002)**		
109	German	C (NS)	G (<0.05)	84 (<0.05) 102 (NS) 118 (0.04)		(Donner et al. 1998)
254	Danish		G (NS)			(Larsen et al. 1999)
39	Spanish		G (NS)			
187	Italian		G (0.004)			(Nistico et al. 1996)
44	Spanish		G (0.004)			
284	UK		G (NS)			
180	US		G (NS)			
123	Sardinian		G (NS)			
818	**Combined Total**		**G (0.002)**			
297	UK		G (0.02)			(McCormack et al. 2001)
3 671	UK, US Norwegian, Irish, Romanian and Finnish		G (0.013)		G (0.00006)	(Ueda et al. 2003)

2.15. *CTLA4* in other autoimmune diseases

Since individual autoimmune diseases share certain features in common, while also being widely diverse, these might be taken as clues when considering what events take place in the autoimmune process. While it is unlikely that the same genes are all involved across different diseases, one gene locus, the *HLA*, has been consistently implicated as a major disease contributor. Could another gene, involved in negative regulation of T cell function, also be a common genetic susceptibility gene? There have also been numerous genetic studies implicating *CTLA4* with other autoimmune diseases, particularly the autoimmune thyroid diseases (AITD) (Yanagawa et al. 1995; Awata et al. 1998; Braun et al. 1998; Huang et al. 1998; Fukazawa et al. 1999; Heward et al. 1999; Ligers et al. 1999; Ahmed et al. 2001; Hadj Kacem et al. 2001; Lee et al. 2001; Hudson et al. 2002; Kouki et al. 2002; Wang et al. 2002; Wang et al. 2002; Yung et al. 2002), raising the interesting possibility of common genetic factors in T1D and AITD. The following sections address the current state of knowledge with respect to *CTLA4* and several autoimmune disorders.

2.15.1 The autoimmune thyroid disorders (AITD)

Among the most common human autoimmune disorders, autoimmune thyroid disease affects up to 5% of the general population and includes two related disorders already mentioned above, Graves' disease (GD) and Hashimoto Thyroiditis (HT) which are thought to share common pathogenic mechanisms though they have very different

phenotypes, reviewed in (Kristiansen et al. 2000). In these two disorders, the *HLA* locus seems to contribute only modest effects on susceptibility.

2.15.1.1 Graves' disease (GD)

Individuals with Graves' disease present with antibodies directed against the thyroid follicular cell membrane thyrotropin receptor thereby stimulating thyroid function and leading to the classical Graves' disease phenotype of hyperthyroidism. Graves' disease is the most common form of hyperthyroidism with an estimated incidence/year of 20/100 000 in Caucasian populations, reviewed in (Tomer and Davies 2003). The *CTLA*4 locus has been repeatedly associated with Graves' disease and the pattern of association has been discussed above, also reviewed in detail by (Simmonds and Gough 2004).

2.15.1.2 Hashimoto Thyroiditis (HT)

The most common form of autoimmune thyroid disease is Hashimoto Thyroiditis (incidence in Caucasians between 0.5%-7.7% (Tomer and Davies 2003)) and the disease phenotype is hypothyroidism. Gradual destruction of the thyroid gland is mediated by infiltrating lymphocytes. Association of *CTLA4* polymorphisms with HT in case-control studies have been extensively reported (Sale et al. 1997; Awata et al. 1998; Akamizu et al. 2000; Takara et al. 2000; Ban et al. 2003; Mochizuki et al. 2003; Ueda et al. 2003).

2.15.2 Systemic Lupus Erythematosus (SLE)

Long considered the prototype of human autoimmune diseases, SLE is a generalized, multi-organ autoimmune disease. It has been suggested that the tightly regulated process of apoptosis of T and B cells in healthy individuals is impaired in lupus patients resulting in the persistence of autoreactive B cells which leads to a dangerous

levels of autoantibodies with a concomitant breakdown of peripheral tolerance, (reviewed in Sekigawa et al. 2004). The prevalence of SLE in the general population is 0.2-0.3% and SLE can manifest itself at any age, it typically affec()ts those in the second to fifth decade of life and is significantly more common in females, (reviewed in Sekigawa et al. 2004). Though association of common *CTLA4* polymorphisms, particularly the +49A>G and the $(AT)^n$ microsatellite polymorphism have been investigated in SLE, conflicting results abound (Kristiansen et al. 2000). In a recent meta-analysis which grouped the results from 7 case-control SLE studies grouping 821 patients and 1329 controls, the authors found that the signal peptide polymorphism contributed significantly to disease risk with an estimated OR for the GG genotype of 1.39 (Barreto et al. 2004).

Some evidence for the association of *CTLA4* polymorphisms with other autoimmune diseases such as Rheumatoid Arthritis, Multiple Sclerosis, Celiac disease, Addisons' disease and Myasthenia Gravis etc, have also been reported and remains to be corroborated as contradictory reports also exist (Kristiansen et al. 2000).

From the perspective of function, *CTLA4* is an excellent candidate gene, fulfilling a crucial role in immune regulation, the disruption of which could lead to autoimmunity. The following sections address its function in the normal immune system, as well as its role in the autoimmune process.

2.16. CTLA-4 function

There are two known isoforms of CTLA-*4* in humans: The full-length CTLA-4 is a transmembrane glycoprotein expressed transiently on the surface of activated $CD4^+$ and $CD8^+$ T-cell populations, although it has also been detected on B-cells and is the main isoform found in adult thymocytes. The soluble CTLA-4 isoform is generated by alternative splicing of the transmembrane domain and is expressed mainly in resting T-cells (Magistrelli et al. 1999; Oaks et al. 2000). Its function is at present poorly characterized. Since CTLA-4's function is to dampen T cell activation, the next few sections address the critical steps leading to T cell activation.

2.16.1. T cell activation

The ability to discriminate between self and non-self can be considered as the most fundamentally important aspect of immune regulation. This translates into an ability to recognize and discriminate between foreign infectious invaders and normal host tissues. This highly selective process is characterized by a complex set of T cell regulatory mechanisms designed to maintain the fidelity of the immune responses, and includes the requirement of two distinct signals for effective activation of antigen-specific T cells. The first signal is delivered via the binding of the antigenic peptide presented on the APC to the T cell receptor (TCR) and is known as Signal 1, while a costimulatory signal delivered via CD28's interaction with its ligand B7 expressed on the surface of the APC (Signal 2) is a requirement for cell division and expansion of effectors and regulators of the immune system, (reviewed in Salomon and Bluestone 2001). The

mechanism by which CD28/B7 interactions regulate immunity involves cell expansion as CD28 ligation stimulates transcription of IL-2 critical in T cell growth, and the anti-apoptotic agent Bcl-X_L (Salomon and Bluestone 2001) (Figure 4). In addition to costimulation, CD28/B7 interactions regulate T_H1 and T_H2 differentiation, cell migration and homeostasis of the already discussed regulatory cells, $CD4^+CD25^+$Tregs (Salomon and Bluestone 2001).

Built into the delicate balance of immune homeostasis is a mechanism to counter costimulatory induction of T cells. This is offered by the interaction of CTLA-4 with the B7 ligand shared in common with CD28.

2.16.2 CTLA-4 as a negative regulator of T cell function

For several years, CD28 and CTLA-4 were thought to have redundant roles as they share extensive sequence similarity and bind the same B7 ligand. However, research has shown that, CTLA-4 expression, dependent on CD28/B7 engagement inhibits IL-2 production, cell cycle progression, and anti-CD3-induced cyclins and CTLA-4 mAbs which block CTLA-4 signal *in vitro* enhance proliferation of T cells in mice (Walunas et al. 1996; Tivol et al. 1997; Walunas and Bluestone 1998). Thus, the CD28/B7/CTLA-4 costimulation and inhibition pathway has emerged as a powerful regulator of peripheral T cell function. The full-length CTLA-4 receptor binds the ligand B7 expressed on the surface of APC's shared in common with its counter-receptor CD28. Its expression on the cell surface is up-regulated only 2-3 days following T-cell activation and is a mere fraction of the level of the constitutively expressed co-stimulator, CD28 (Figure 4). Interestingly, the strength of binding of CTLA-4 to B7 is at least one order of magnitude

stronger than CD28/B7 interactions (Ellis et al. 1996), which has led some to propose that CTLA-4's role is to competitively block CD28/B7 binding. To date, however, the exact mode of action of CTLA-4's inhibition is not completely understood. The next paragraph summarizes the latest data on the subject.

2.16.3 CTLA-4's mode of action

CTLA-4's mode of action may involve the displacement of the costimulatory binding, inhibition of cell cycle progression (Krummel and Allison 1996), skewing of T cell differentiation towards a T_H2 phenotype (Alegre et al. 1998; Walunas and Bluestone 1998) and/or regulation of signal transduction in the $CD4^+CD25^+$ Tregs subset (known to constitutively express CTLA-4) (Salomon et al. 2000) and possibly others, reviewed in (Salomon and Bluestone 2001). In fact these may all be interrelated as Jeffrey Bluestone, a leading researcher on the subject believes. Overall, the major effect of CTLA-4 is to alter the threshold of T cell activation by modifying the early events in TCR signaling, by modulating T cell receptor signaling, via TCRζ chain phosphorylation (Lee et al. 1998). This may be accomplished via the intracytoplasmic tail of CTLA-4 (100% conserved across species) which has PI-3 kinase and SHP-2 phosphatase activity (Bradshaw et al. 1997; Schneider et al. 2001) and its interaction with the TCRζ chain. Downstream biochemical signaling may also be modulated by CTLA-4, for instance through ERK

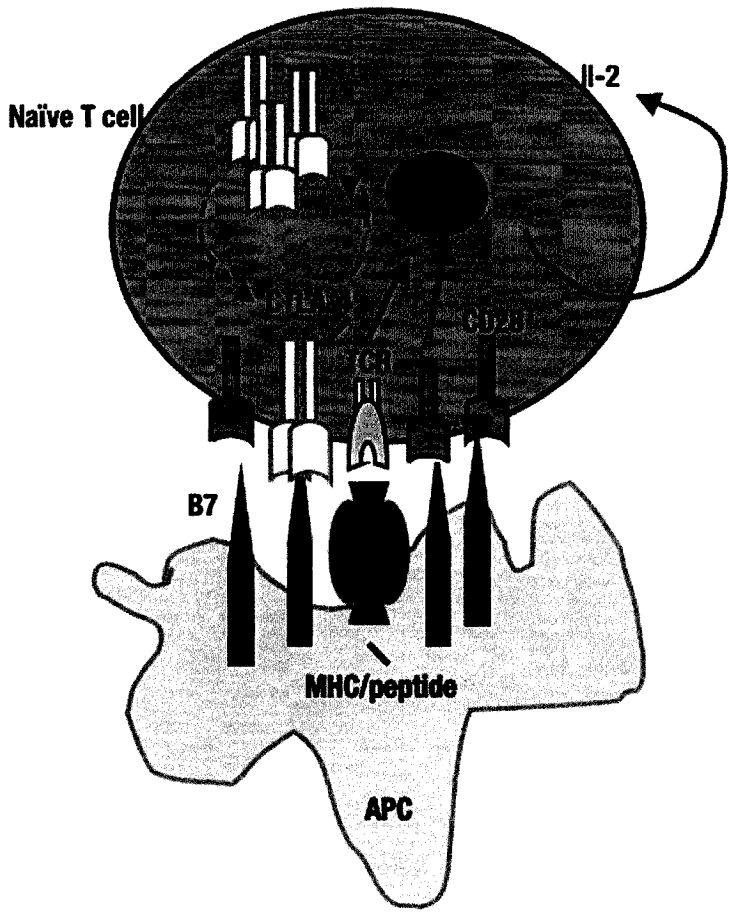


Figure 4: Simplified version of the T cell activation and the co-stimulatory network

Antigen presentation through an MHC-Class II molecule on the APC and the T cell receptor (TCR) without co-stimulation results in T cell anergy. The interaction between the constitutively expressed CD28 molecule on the T cell and the B7-1/2 ligands expressed on the APC provides the costimulatory signal necessary for augmenting and sustaining a T cell response while CTLA-4 expression is upregulated in response to activation and downregulates T cell function by binding its ligand B7-1/2. The mechanism by which CTLA-4 arrests proliferation of T cells is not known but is thought to act via competitive displacement of CD28/B7 interactions, since its avidity for the B7 ligand is at least 100 times higher than CD28. Alternatively, it may also antagonize CD28-mediated signaling and/or TCR-mediated signaling. These mechanisms are not mutually exclusive.

activation (Calvo et al. 1997). Whether the effects of CTLA-4 engagement are directed at inhibition of costimulatory signals, regulation of proximal TCR signaling, or downstream effector pathways of T cell activation they result in altered T cell differentiation and downregulation of immune responses.

No other example of CTLA-4's regulatory role tells the story better than the CTLA4 knockout mouse.

2.16.4 The CTLA-4 Knockout mouse

CTLA-4 deficient mice develop massive lymphoproliferative disease with multiorgan T cell infiltrates and die at 2-3 weeks post-natally (Tivol et al. 1995; Waterhouse et al. 1995). The lymphoproliferative disorder observed in these mice is mediated by loss of the negative regulatory role of CTLA-4 in T cell activation (Chen 2004). Evidence that CTLA-4 actively suppresses CD28 costimulation is provided when introduction of CD28 deficiency eliminates the fatal lymphoproliferative symptoms (Chen 2004).

2.16.5 Human examples of CTLA-4 deficiency: Chediak-Higashi syndrome

Interestingly, there are several human examples of induced or acquired CTLA-4 deficiency. These further illustrate the importance of this molecule in the regulation of peripheral tolerance and deserved a mention. 85-90% of patients with mutations in the lysosomal trafficking regulator (*LYST*) present with fatal uncontrollable T cell activation (clinical management of this aspect of the disorder is particularly challenging and most succumb at 3.1 years of age); this rare autosomal recessive disorder is known as the

Chediak-Higashi syndrome (Barrat et al. 1999). At least the lymphoproliferative aspect of this syndrome is thought to be mediated by disruptions in CTLA-4 cell surface expression with some patients totally lacking CTLA-4 expression (Barrat et al. 1999).

2.16.6 CTLA-4's cell-surface expression and intracellular cycling pattern

Since subtle regulation of CTLA-4 membrane expression is required to orient immune response to either T cell activation or tolerance induction, CTLA-4 expression is tightly regulated by compartmentalization in intracellular vesicles, i.e., in the endocytic compartments and secretory granules (Marengere et al. 1996; Perkins et al. 1996; Chuang et al. 1997) (Figure 5). It cycles in and out of the intracellular stores to the cell surface in a clathrin-dependent manner (association with the coated pit adaptor protein AP-2) (Chuang, Alegre et al. 1997), a process thought to be regulated by the tyrosine phosphorylation of the CTLA-4 cytoplasmic tail (Marengere, Waterhouse et. al. 1996, Shiratori et al. 1997) and requires a polarized export to the site of TCR contact (Linsley et al. 1996) (Figure 5). The site of TCR contact with an APC is referred to as the immunological synapse, a concept which evolved from the collective work on T cell activation and cytoskeletal reorganization, essential to TCR signal transduction, (reviewed in Bromley et al. 2001). Thus, the synapse is essential for the generation and maintenance of signals that determine T cell fate upon antigenic stimulation. Upon T cell activation, reorientation of the microtubule organizing center (MTOC) to a position facing the site of APC contact takes place, concomitant with localized actin polarization and accumulation of proteins at the T-cell-APC interface (Kupfer and Singer 1989). Both CTLA-4 and CD28 concentrate at the immunological synapse, but while CD28 goes from a uniform cell-surface distribution to a concentrated expression at the synapse, CTLA-4,

residing in intracellular vesicles orients the vesicles containing it upon TCR engagement towards the APC/TCR contact site (Egen and Allison 2002). Interestingly, the quantity of CTLA-4 which concentrates at the synapse is correlated with the strength of the TCR signal, suggesting that CTLA-4 might preferentially restrict T cell responses to stronger TCR signals (Egen and Allison 2002). B7-1/2 binding to CTLA-4, is required for this polarization to occur and CTLA-4 preferentially binds B7-1 (Pentcheva-Hoang et al. 2004).

2.16.7 CTLA-4 in the autoimmune process

CTLA-4 plays an important role in the initiation of autoimmune response as well as in the maintenance and termination of an ongoing autoimmune response in experimental animal models (Karandikar et al. 1996; Perrin et al. 1996; Luhder et al. 1998; Karandikar et al. 2000; Luhder et al. 2000). In the murine model for Multiple Sclerosis (MS), experimental autoimmune encephalitis (EAE), which like MS is a demyelinating disease of the central nervous system that is mediated by CD4+ T helper cells (Th1), disease can be induced by immunization with various myelin proteins (or epitopes) such as MBP, PLP or MOG. It is increasingly clear that co-stimulatory pathways, involving CD28/B7, mediate disease. Blockade of CTLA-4/B7 interactions in vivo by administration of monoclonal anti-CTLA-4 mAbs, exacerbates autoimmune encephalitis disease (Karandikar et al. 1996; Perrin et al. 1996), while CTLA-4/B7 blocking in recipient mice for primed EAE-specific T cells also worsened the disease (Karandikar et al. 1996).

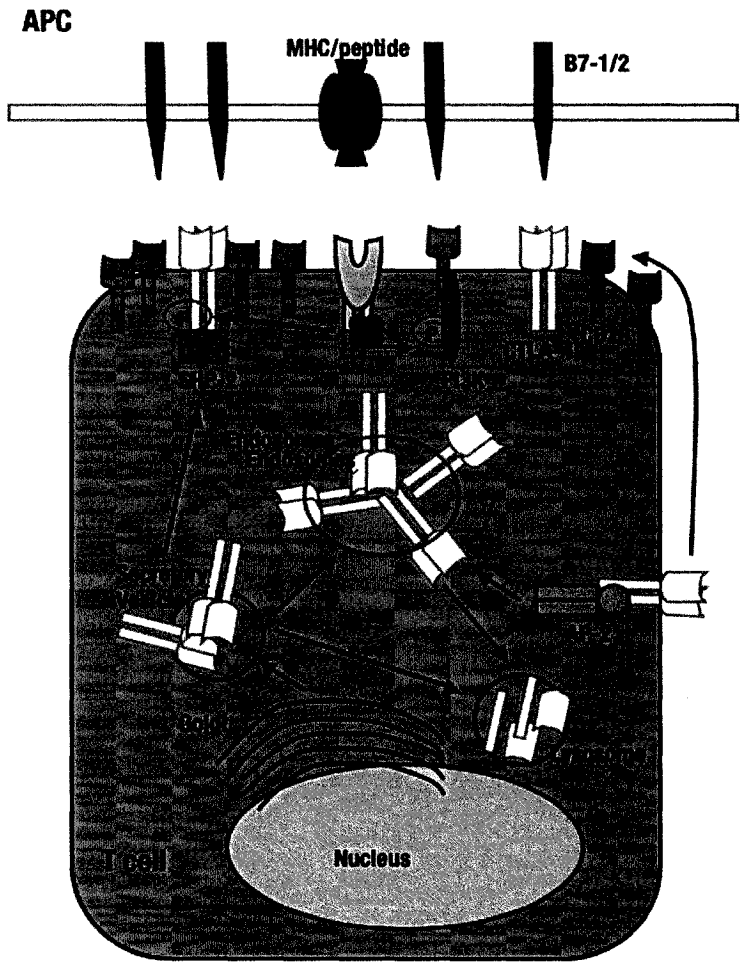

APC
MHC/peptide
B7-1/2
Nucleus

Figure 5: CTLA-4 cycling.

Protein trafficking and localization during the process of T cell activation is thought to play a major role in the ability of a protein to regulate the T cell response. CTLA-4 is not detectable in naïve T cells but is rapidly upregulated upon TCR ligation with the MHC/peptide. Surface expression of CTLA-4 expression itself is tightly regulated by a 100% conserved tyrosine-based intracellular localization signal in the cytoplasmic tail. This motif results in the rapid endocytosis of CTLA-4 from the cell surface to endosomal compartments, and targets at least some CTLA-4 to the lysosomes for degradation. Once at the cell surface CTLA-4's' intracellular motif may be phosphorylated by protein tyrosine kinases (PTKs, shown as red spheres) associated with the TCR. This results in the stabilization of surface CTLA-4 and prevents association with the clathrin-coated pit adaptor protein AP-2. CTLA-4 is also known to migrate to the site of TCR engagement creating upon T cell activation as the arrow indicates. Even subtle defects in the trafficking of CTLA-4 are likely to disturb this tightly regulated intra-cellular cycling process and could result in decreased and/or altered T cell responses. Figure adapted from (Chambers et al. 2001)

Moreover, in CD28-deficient mice the EAE disease effects are milder (Girvin et al. 2000) and blocking the entire CD28/CTLA-4/B7 system by systemic administration of an immunoglobulin-CTLA-4 fusion protein (CTLA-4-Ig) resulted in suppression and amelioration of disease (Khoury et al. 1995). In the NOD mouse, similar effects have been observed in mice that express TCR transgenes specific for an islet antigen but CTLA-4 (anti-CTLA-4 mAbs) administration worsened disease only when administered early on; before the onset of insulitis, and only in a particular strain of NOD (BDC2.5) which are populated by both transgenic and endogenous T cells (Luhder et al. 1998). This suggests that CTLA-4 inhibition may also be affecting the Treg subset (Luhder et al. 1998). But treatment of NOD mice with CTLA-4-Ig early on prevented disease and had no effect on the severity of insulitis (Luhder et al. 1998). Thus various combinations of recombinant human or recombinant mouse proteins with human and mouse Ig's have been constructed with varying results in the murine models with the timing of intervention being crucial to the effect.

In humans, the direct role of CTLA-4 mediated pathogenesis as seen in experimental mouse models is more difficult to assess for obvious reasons. Mainly, the bulk of evidence supporting the role of CTLA-4 in autoimmune disease comes from genetic studies that correlate risk with a particular genotype described above. Much more difficult is the task of correlating altered function and/or expression with any particular genotype. Many of the studies addressing this have focused on the role of the only coding SNP in the *CTLA4* gene, the +49A>G SNP, while others have addressed the role of regulatory SNPs associated with disease on variability of expression located in the promoter and the 3'UTR (Figure 6). These are summarized below.

2.17 Allelic variation at the *CTLA4* gene

2.17.1 Functional studies at the signal peptide

The +49G allele (the Ala-17) of the signal peptide has been consistently over-transmitted from heterozygous parents to affected offspring by the transmission disequilibrium test (TDT) in multiple ethnic groups and is the most studied SNP in the CTLA4 gene. Located in the signal peptide, this single base change results in Threonine to Alanine substitution that could alter the function of the CTLA-4 protein. Given the tight regulation of cell-surface/intracellular CTLA-4 disruption of CTLA-4 trafficking or cell-surface expression could alter immune function; this has been addressed in some studies, either by directly measuring cell surface CTLA-4 surface expression, or by indirect assessment of CTLA-4 function through proliferation and activation assays.

The first such study measured the proliferation ability of T-cells in individuals homozygous for the predisposing +49G allele at the signal peptide coding SNP and found increased proliferation associated with GG genotype (P=0.047 for combined homozygous individuals with Graves' disease (n=45), Hashimoto Thyroiditis (n=18) and normal controls (n=43) (Kouki et al. 2000). The increase in proliferation was shown to be CTLA-4 dependent as blockade of CTLA/B7 signaling with an anti-CTLA-4 mAb resulted in a reversal of the pattern of proliferation when correlated with genotype. Under these conditions, the authors report reduced proliferation of T cells in GG homozygous patients, vs. AA homozygotes (P=0.019) (Kouki et al. 2000). In a follow-up study by Maurer et. al., they found a similar pattern of proliferation in normal individuals homozygous for the G-allele (n=7) or A-allele (n=29) under suboptimal PBMC stimulation conditions (Maurer et al. 2002).

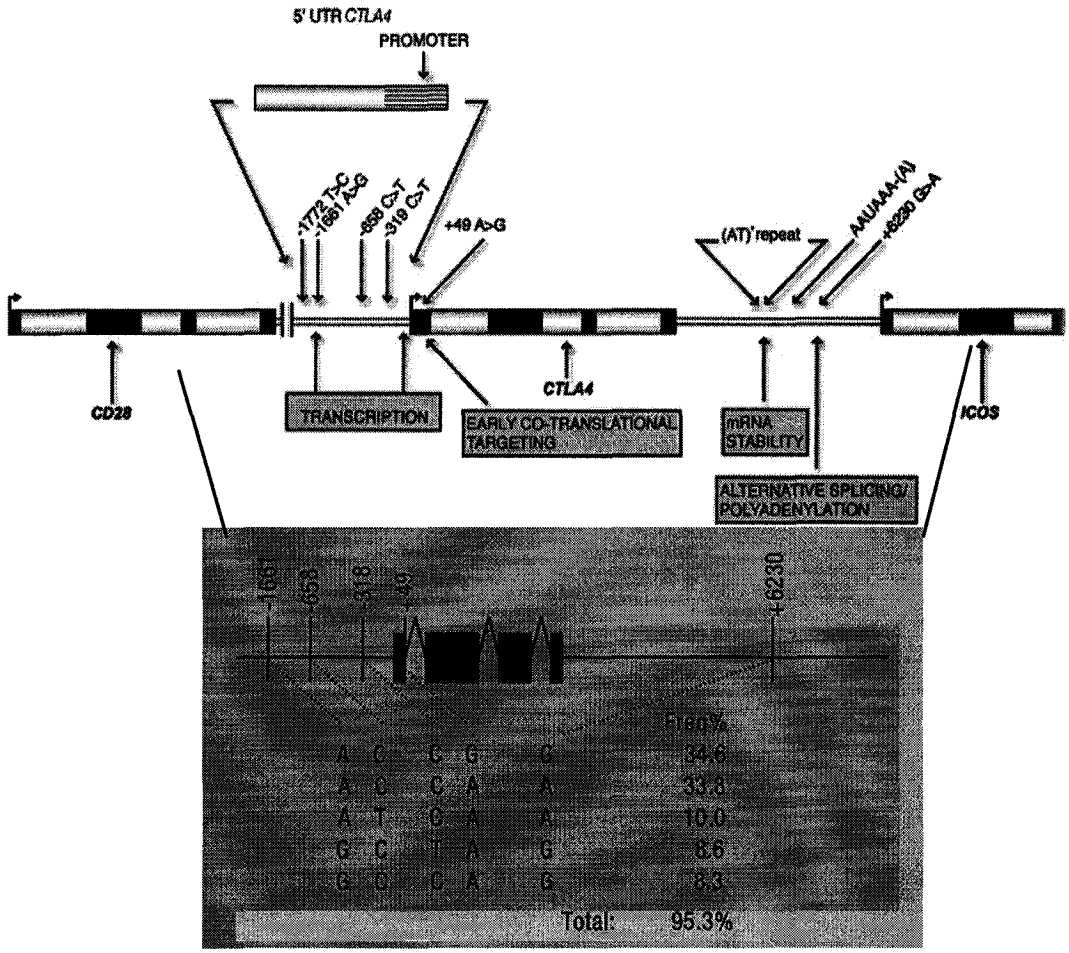


Figure 6: Type I diabetes associated *CTLA4* haplotype and potential functional role of disease-associated polymorphisms.

The *CTLA4* gene and neighbouring 5' and 3' regions are all contained within a haplotype block, where 5 haplotypes describe >95% of the variability within the region. SNPs and their possible functional significance are indicated, as is the $(AT)^n$ repeat and the predicted polyadenylation site for the major *CTLA4* transcript. SNPs in the 5' flanking region and in the promoter region are indicated as is the minimal promoter region (335bp, striped bar); only the –318C>T SNP is located within this minimal region conserved across rodents and humans. Not all *CTLA4* SNPs are indicated and the figure is not drawn to scale. Figure reproduced adapted from (Anjos and Polychronakos 2004) with written permission from Elsevier. *Copyright 2005.*

Moreover, they also analyzed the surface expression of CTLA-4 after primary mixed leukocyte reaction (MLR) against allogeneic blood-derived dendritic cells in the two genotype groups and report statistically significant differences (Maurer et al. 2002). Intracellular staining pattern of CTLA-4 in cells from individuals from the two groups revealed differential staining pattern (Maurer et al. 2002). In addition, they also showed that this allele negatively affects the downregulation of T-cell activation in response to IL-2 (Maurer et al. 2002). A single contradictory *in vitro* study reported no difference between cell surface CTLA-4 in Jurkat T cells transfected with +49A or +49G containing constructs though following activation there was an increase in mRNA expression derived from A49 clones vs. G49 clones, though it did not reach statistical significance (Xu et al. 2002). They also report no difference in levels of IL-2 in response to T cell activation between the +49A and +49G clones (Xu et al. 2002). None of these studies address the molecular mechanism if any, underlying the functional endpoints.

An integral part of this thesis deals with the effect of the signal peptide amino acid change on the function of the signal peptide. The next section summarizes some important concepts on this subject.

2.17.2 The role of the Signal Peptide

Signal sequences, or peptides, play a crucial role in targeting and membrane insertion of secretory and membrane proteins to the Endoplasmic Reticulum (ER), reviewed in (Martoglio and Dobberstein 1998). Signal sequences have a tripartite structure with a hydrophobic core, important for targeting and membrane insertion, flanked on its C-terminal side by a polar region containing (often, but not always) small uncharged residues which form the recognition site for cleavage of the peptide and on its

N-terminus by a polar region, with a net positive charge (von Heijne 1985). Although almost all signal sequences share this structure, they otherwise share no particular sequence identity (von Heijne 1985) and their length is also highly variable.

Once co-translational membrane insertion has taken place the signal peptide is cleaved by the Signal Peptide Peptidase (SPP), and the cleaved signal peptide is liberated into the ER membrane, while the protein is further translocated into the ER lumen or integrated into the membrane (Martoglio 2003). Recently it has become evident that signal sequences have a more diverse role and may also affect the orientation of membrane insertion (Spiess 1995), discriminate between different targeting pathways (Ng et al. 1996; Weiner et al. 1998) and the recent finding that signal peptides fragments are liberated from the ER membrane upon intramembrane proteolysis (Weihofen and Martoglio 2003) supports the hypothesis that certain signal peptide fragments may have functions beyond protein targeting and may be bioactive. A post-targeting role for a signal peptide has been described for the HLA-E epitopes generated from the signal sequence of the MHC class I molecules (Braud et al. 1997) and for the signal sequence of the Hepatitis C Virus (HCV) (McLauchlan et al. 2002).

2.17.2.1 Signal Peptide mutations causing disease

It is conceivable that an amino acid substitution in the signal peptide may alter its function in any of the events listed above, particularly, if mutations occur at the SPP cleavage recognition site, or in the hydrophobic core. Thus far, naturally occurring mutations in human signal sequences with a direct correlation with defective secretion and pathology have been reported for familial forms of Diabetes Insipidus (Ito et al. 1993; McLeod et al. 1993; Repaske et al. 1997; Beuret et al. 1999; Siggaard et al. 1999),

familial hypoparathyroidism (Arnold et al. 1990; Karaplis et al. 1995; Sunthornthepvarakul et al. 1999), and in other rare inheritable syndromes (Parrott et al. 1992; Racchi et al. 1993; Lagerstrom-Fermer et al. 1995; Witt et al. 1999; Kato et al. 2000; Lanza et al. 2002; Fingerhut et al. 2004).

Altered function in relation with other polymorphisms in the *CTLA4* non-coding region was also addressed in some studies, with mixed results. These polymorphisms lie within the boundaries of the regulatory regions, are untranslated and may alter expression levels of mRNA (Figure 6).

2.17.2 Regulatory variation at the *CTLA4* gene

2.17.2.1 The *CTLA4* promoter

The promoter SNP at –318 has been the least used in genetic association and linkage probably because it has a very low heterozygosity rate (16%), due to the low frequency of the –318T allele (0.9%). In Type 1 Diabetes, one single case-control study found a significant association with this SNP, in a large cohort of Chinese background (p=0.0026, OR=1.9) (Lee et al. 2001) while a smaller case-control study in the Japanese failed to find an effect (Ihara et al. 2001). Smaller family-based studies have also failed to find an effect (Donner et al. 1998) likely due to lack of power.

The first study to address the role of the promoter alleles in expression was a case-control study by Ligers et. al., in which CTLA-4 cell-surface levels were quantified by flow cytometry in multiple sclerosis patients versus matched-controls. When the data were analyzed by promoter and signal peptide (–318C>T;49A>G) haplotype,

significantly higher surface expression in both patients and controls for the –318T/C;+49A/A genotype was found and not for any other genotype (Ligers et al. 2001). They also compared *CTLA4* expression levels in lymphocytes from 44 heterozygous individuals at the promoter (T/C) vs. 186 homozygotes (C/C) and found an effect in Myasthenia Gravis patients (n=4) but not in MS patients or normal controls (Ligers et al. 2001). This inconsistent result is not surprising, as comparisons between individuals are subject to factors such as inter-subject variation due to genetic background as well as past and present immune experiences and may easily miss subtle differences expected in multigenic disorders.

More recently, a report by Wang et. al., found higher promoter activity associated with the rare –318T allele of the promoter in a luciferase reporter assay containing 329bp of upstream *CTLA4* promoter sequence (Wang et al. 2002). This SNP is well within the minimal region required for inducible *CTLA4* expression as shown in the mouse (Perkins et al. 1996) and could very well harbor important regulatory elements capable of modulating expression in an allele-dependent manner. In addition, linkage disequilibrium in the area extends well beyond the minimal promoter region, and may harbor other SNPs which could explain altered expression. These possibilities remain to be explored.

There are other regulatory variants located in the 3'UTR that may modulate expression differences. The next section summarizes the known data.

2.17.2.2 The *CTLA4* 3'UTR: The dinucleotide repeat

Another obvious candidate for functional effect is the $(AT)^n$ repeat at the 3' UTR, a region typically involved in maintaining mRNA stability. Regulation of mRNA decay is an important mechanism by which the level of gene expression is controlled. Certain

AU-rich elements (AREs) commonly associated with mRNA degradation process, particularly the AUUUA pentamer and the nonamer UUAUUUA(U/A)(U/A), are not present in the *CTLA-4* 3' UTR. Since these are not always required for a functional ARE, as is the case of the *zif278, c-jun, c-myc* genes (Chen and Shyu 1995) it is conceivable that the $(AT)^n$ repeat might play this role for *CTLA4*. There are two common alleles, referred to as 88bp and 106bp allele, which together account for two thirds of all Caucasian chromosomes and are in tight linkage disequilibrium with the A and G allele of the signal peptide, respectively (Holopainen and Partanen 2001).

In attempts to determine if the alleles at the microsatellite alter mRNA stability a study by Wang et. al., reported a ~50% increase in steady-state *CTLA-4* mRNA level for the 86 bp allele, compared to longer alleles in heterozygous individuals (Wang et al. 2002). They also show a faster RNA disappearance curve after actinomycin D treatment in one individual homozygous for a long $(AT)^n$ repeat compared to one homozygous for a shorter one (Wang et al. 2002). Another study correlated the levels of circulating IL-2 receptor alpha chain (IL-2Rα) and telomerase activity, both parameters of T cell activation status, with genotype at the *CTLA4* dinucleotide expansion (Huang et al. 2000). They found a positive relationship between longer alleles at the $(AT)^n$ repeat and telomerase activity in stimulated PBMCs' derived from homozygous myasthenia gravis patients. They reported similar findings in relation to circulating levels of IL-2Rα (Huang et al. 2000).

At the present time, it remains unclear as to whether the common alleles for the *CTLA-4* $(AT)^n$ microsatellite differentially mediate mRNA stability. Rigorous allelic

imbalance studies in heterozygous individuals for the more common alleles are needed to elucidate the effect, if any.

2.17.2.3 The alternatively spliced soluble CTLA-4

The soluble isoform was initially identified by Magistrelli et. al., who detected the smaller isoform in mRNA derived from human PBMCs by RT-PCR. It lacked the transmembrane domain as a result of the splicing of exon 3, its expression was downregulated by T cell activation and it was detectable in healthy human sera (Magistrelli et al. 1999). These findings were corroborated by a subsequent study where the newly identified isoform was sequenced and found to be 49 amino acids shorter than the full length, but still contained a cytoplasmic tail (as a result of splicing) albeit with an altered reading frame and arose by alternative splicing (Oaks et al. 2000). It was also detectable only in resting human lymphocytes and not in the thymus or lymph nodes (rats) (Oaks et al. 2000). Similar soluble isoforms were also identified in the mouse and rat (Oaks et al. 2000).

At present, the function of this soluble isoform is not well understood. It has an intact B7-binding motif, the MYPPPY amino acid sequence, 100% conserved in CTLA-4/CD28 proteins. A recombinant form of sCTLA4 was able to retain B7 binding properties (Oaks et al. 2000) which could lead to negative signaling.

2.17.2.4 The soluble form in autoimmune diseases

There has been much excitement recently over a study in which a newly identified SNP, +6230G>A, in the 3' flanking region of *CTLA-4* was found to be highly associated with Graves' disease. In Type 1 Diabetes the predominance of this SNP was weaker

(Ueda et al. 2003). It is unclear how this SNP may modulate expression of *CTLA4* though the authors proposed an alternative splicing or polyadenylation mechanism as a possible explanation for the genetic effect. They investigated the level of expressed soluble vs. full length *CTLA4* in heterozygous individuals using the transcribed signal peptide SNP as a marker. In three individuals heterozygous for the +49A>G SNP the authors report higher soluble *CTLA4* mRNA from the disease *protective* haplotype (+49A;+6230A) than from the *predisposing* haplotype (+49G;+6230G) in unstimulated $CD4^+$ T-cells (Ueda et al. 2003). The observation was limited to the soluble isoform since they did not find allelic differences at the full-length *CTLA4* isoform. They propose that this observed effect is mediated by the +6230G>A SNP (Ueda et al. 2003).

Two studies recently compared serum soluble *CTLA-4* levels in individuals with systemic lupus erythematosus (n=40) (Liu et al. 2003) and autoimmune thyroid disease (Graves' disease n=17 and Hashimoto Thyroiditis, n=4) (Oaks and Hallett 2000), reporting higher levels of soluble CTLA-4 were found in patients vs. controls.

Whether this SNP is a functional determinant in the differential expression of *CTLA4* isoforms remains speculative. It is yet unclear whether the +6230G polymorphism is part of any *CTLA4* transcripts, as it is outside the known polyadenylation site of *CTLA4* (Ling et al. 1999; Ling et al. 2001) (Figure 6), and if it does, if it is part of *both* the full-length and soluble *CTLA-4* transcripts. Ueda et. al., propose that the +6230G>A SNP determines the efficiency of splicing and production of soluble *CTLA-4* although a mechanism was not proposed (Ueda et al. 2003). There are several examples in which a cell uses splicing factors to influence the decision of when and how to polyadenylate an mRNA precursor (Zhao et al. 1999); it remains plausible

that this SNP may regulate splicing efficiency depending on the cell context which in turn may affect the choice of polyadenylation site. Different transcripts may have different stabilities or translation abilities as is the case of the eukaryotic initiation factor-2 (*eIF-2α*) (Miyamoto et al. 1996) which would lead to differential expression of different isoforms. This remains to be investigated.

2.18 Project Rationale

The aim of the research conducted in this thesis is to determine the mechanism underlying the association of the *CTLA4* gene with T1D. Prior evidence from whole genome-wide linkage analysis and fine linkage mapping in addition to a large number of case-control and family-based association studies, made the case for *CTLA4* as a candidate gene in T1D susceptibility a strong one. In addition, its role as a negative regulator in the tightly woven network of the T cell response system made the genetic findings more exciting. Based on these earlier studies, we believed that allelic variation at *CTLA4* underlies this genetic association. However, the extent of LD in the region encompassing the highest association required detailed functional analysis to untangle the contributions from SNPs genetically linked to disease. The question asked in this thesis is which of the SNPs in the *CTLA4* gene region explains this association and through which mechanism.

We proceeded to answer this question in three distinct but interrelated steps. The first step addressed the only coding SNP in the entire region, while the two other objectives dealt with the regulatory regions at the 5' and 3' ends.

2.18.1 Objective 1. The Signal Peptide polymorphism

Establishing a working hypothesis: background

The signal peptide, located at the *N*-terminus of proteins destined for membrane expression or secretion, functions early on in the co-translational process and is cleaved before the polypeptide leaves the Endoplasmic Reticulum (ER). Thus, an amino acid

substitution there could only alter the function of the mature protein by effects on the early stages of protein transport or processing.

Though, Maurer et.al., reported different allelic localization patterns in T cells from individuals homozygous for each signal peptide allele by confocal microscopy (Maurer et al. 2002) the molecular mechanism underlying this was not investigated, while yet another study reported no difference in cell surface expression nor cytokine levels between allelic variants at the signal peptide (Xu et al. 2002). Given this evidence, we established a working hypothesis which formed the basis for the work undertaken in the next chapter. We hypothesized that the Thr17Ala amino acid substitution in the signal peptide of CTLA-4 could alter the early events of protein trafficking and/or post-translational processing which take place in the Endoplasmic Reticulum resulting in altered levels of cell surface CTLA-4.

2.18.1.1 Approach

We employed a combination of *in silico* proteomics tools, cell-free *in vitro* systems and *in vitro* cell-based assays to test this hypothesis.

We used an *in vitro* system developed for the study of pre-ER and ER protein processing events based on the work of Nobel winner Günter Blobel and used extensively in these types of studies. The cell-free rabbit reticulosyte system with ER containing pancreatic membranes allows for the detailed study of events leading up to insertion in the ER membrane, during co-translational membrane integration and post-membrane integration. In addition, post-translational events which take place in the ER lumen, such as protein folding, high-mannose addition (*N*-linked Glycosylation) exclusive to the ER,

phosphorylation and acetylation as well as polypeptide assembly and quality control mechanisms are all integral to this system.

In designing our cell-based experiments several points were considered. Unlike Mendelian inheritable disorders due to signal peptide mutations where the effects are dramatic, we expected rather subtle differences in targeting between alleles at the signal peptide. So as not to obscure these differences the assay design involved a co-transfection system in *COS1* cells of the two allelic forms of the full-length CTLA-4 gene fused to a different color fluorescent protein of the Green Fluorescent protein (GFP) family. In this manner, we devised a co-localization experiment with the full-length CTLA-4 proteins, and *C*-terminal truncated CTLA-4 constructs that constitutively cycle to the cell-surface were used in quantification of cell-surface levels. *COS1* cells were selected for their availability, ease of transfection and large morphology. In using a co-transfection assay with the two alleles in one single cell, we eliminated differences which arise as a result of transfection efficiency.

The results of this study were peer-reviewed and have been published.

2.18.2 Objective 2: Regulatory variation at the *CTLA4* gene

Establishing a working hypothesis: *cis*-acting variation at the *CTLA4* promoter

Turning our attention towards the regulatory regions at the *CTLA4* gene was based on both intuition and prior evidence. At the promoter, there were relatively few genetic studies done using the most proximal promoter SNP located –318 bp from the transcriptional start site, undoubtedly due to the low heterozygosity rate at this SNP. Of the two genetic case control studies performed in T1D, one showed a positive association

(Ihara et al. 2001) while the other showed no association at all (Lee et al. 2002). Though a significantly larger TDT-based study in T1D was highly suggestive of an association at a newly identified SNP in the 3'flanking region of *CTLA4*, the overall effect on T1D was modest and the 5'end of *CTLA4* could not be excluded.

Functional evidence in support of an allele-dependent functional effect at the promoter was provided in the form of a reporter-based transcriptional assay reporting significant differences in transcription of promoters bearing –318T alleles vs. –318C alleles by Wang et. al., (Wang et al. 2002). *In vivo* allelic expression differences in Multiple Sclerosis patients were not detected in individuals homozygous for each allele at the promoter, but were found in a small number of Myasthenia Gravis patients (Ligers et al. 2001).

Based on this data we constructed an *a priori* hypothesis that the –318C>T SNP located in the minimal region required for *CTLA4* transcription alters the expression of *CTLA4* through modulation of transcriptional effects.

2.18.2.1 Approach

We employed a combination of genetic and functional tools to dissect the contribution of the promoter SNPs, and of the newly identified 3'SNP, +6230G>A. By selecting SNPs across the *CTLA4* gene, previously designated as haplotype "tag" SNPs (Johnson et al. 2001), we effectively accounted for the extent of variability at the *CTLA4* locus (without genotyping all 108 SNPs). Our approach involved genotyping the "tag" SNPs in a collection of 500 T1D families and performing the TDT on individual SNPs and whole haplotypes. In parallel, we devised *in vitro* and *in vivo* allelic variation assays described in detail previously. Briefly, our *in vitro* system included a significantly larger

genomic promoter context than (Wang et al. 2002) encompassing the –318C>T SNP, and used the highly sensitive Dual-Luciferase reporter system which allows for simultaneous integrated quantification of transcription efficiency of a promoter via Luciferase and correction for transfection efficiency with the Renilla gene under the control of the *SV40* promoter. We chose the human transformed Jurkat T cell line, because it contains no endogeneous *CTLA4* mRNA (Lindsten et al. 1993) or protein but has the transcriptional mechanisms required for transcribing the CTLA-4 promoter. *In vivo*, we used the highly sensitive allelic imbalance assay in heterozygous individuals at the –318C>T SNP, using the +49A>G SNP as a marker (because it is transcribed). In comparing expression of the two alleles in the same RNA sample from cells of the same individual, we avoid inter-individual differences that can obscure such effects when comparisons are made across genotypes as in (Ligers et al. 2001)

Allele-specific quantification by cDNA sequencing

To quantify allele specific differences *in vivo* we employed a sequencing-based approach previously validated by Qiu et. al. 2003. It assumes that the peak height at a particular base is proportional to the amount present in the sample. Comparisons are made between PCR fragments that include the marker SNP, derived from genomic DNA and from matched reverse transcribed RNA from individuals who are heterozygous at the marker SNP. There is very little variation between peak heights at any single base across multiple replicates, or between individuals. The results of these experiments have been peer-reviewed and published.

2.18.3 Objective 3: Regulatory variation at the *CTLA4* gene: 3'end

Re-evaluating the *CTLA4* 3'SNP: +6230G>A

This SNP emerged in 2001, in a study by (Johnson et al. 2001) which identified it in the context of determining the extent of LD in the region and identifying haplotype "tag" SNPs, which the +6230G>A isn't. Subsequently, in the largest genetic association and linkage study to date, in T1D and autoimmune thyroid disease, it was identified as the SNP most likely to be responsible for the association in Graves' disease and to a lesser extent in T1D (Ueda et al. 2003). In our own TDT data, a haplotype with the –318T;+49A;+6230G alleles, shows highly significant under-transmission while haplotypes harboring the –318C;+49G;+6230G alleles are over-transmitted. Clearly, further investigation of cis-acting variation requires evaluation in a haplotype-sensitive manner and in a larger collection of individuals. Both the soluble and full-length isoforms also need to be evaluated in light of the evidence put forth by (Ueda et al. 2003).

2.18.3.1 Approach

We used an *in vivo* approach in assessing *cis*-acting variation as described previously by comparing allelic expression differences in heterozygous individuals, in conjunction with bioinformatics tools. Due to new technology available to us, we chose a slightly different detection system for "allelic imbalance" based on the single nucleotide primer extension reaction, SNuPe. Briefly, a primer (or probe) is designed complimentary to the template sequence so that its last nucleotide is directly adjacent to the target SNP. After annealing to the template, a primer extension reaction is performed in which the

nucleotide corresponding to that immediately adjacent to the annealed probe is fluorescently labeled and detection is done. The data from this study has been peer-reviewed and accepted for publication.

3.0 CHAPTER THREE

A Common, Autoimmunity Predisposing Signal Peptide Variant of the Cytotoxic T-Lymphocyte Antigen-4 Results in Inefficient Glycosylation of the Susceptibility Allele

Suzana Anjos, Audrey Nguyen, Houria Ounissi-Benkalha, Marie-Catherine Tessier and Constantin Polychronakos

Journal of Biological Chemistry, 277 (48): 46478-46489. 2002

3.1 Contributions of Authors:

Suzana Anjos. Experimental design and strategy, including all plasmid cloning, protein translation assays, cell-based assays and confocal microscopy, data collection, analysis and interpretation, hypothesis development, manuscript organization and preparation.

Audrey Nguyen: Technical expertise in cell-free translation systems, interpretation of immunofluorescence data.

Houria Ounissi-Benkalha: Technical advice and strategy in matters of Molecular Biology.

Marie-Catherine.Tessier: Site-directed mutagenesis.

Constantin Polychronakos: Research Director, hypothesis development, manuscript preparation and editor, corresponding author.

The association of *CTLA4* with T1D has been narrowed down to a haplotype block encompassing one coding and several regulatory polymorphisms in tight linkage disequilibrium, any one of which could be the causative functional polymorphism. In the pursuit of the causative polymorphism, we began by addressing the function of the only coding SNP, the +49A>G, resulting in a Threonine to Alanine subsitution in the signal peptide of CTLA-4. Because of its location in the signal peptide, whose function is to target and anchor growing polypeptides on the ribosomal complex to the ER, we hypothesized that a Thr17Ala substitution may alter the early events of protein transport.

3.2 Abstract

A common Thr17Ala polymorphism in the signal peptide of the cytotoxic T-lymphocyte antigen 4 (CTLA-4), a T-cell receptor that negatively regulates immune responses, is associated with risk for autoimmune disease. Since the polymorphism is absent from the mature protein, we hypothesized that its biological effect must involve early stages of protein processing, prior to signal peptide cleavage. Constructs representing the two alleles were compared by *in vitro* translation, in the presence of endoplasmic reticulum (ER) membranes. We studied glycosylation by Endoglycosidase-H digestion and glycosylation mutant constructs, cleavage of peptide with inhibitors, and membrane integration by ultracentrifugation and Proteinase-K sensitivity. A major cleaved and glycosylated product was seen for both alleles of the protein but a band representing incomplete glycosylation was markedly more abundant in the predisposing Ala allele (32.7% ± 1.0 *versus* 10.6% ± 1.2 for Thr, $P<10^{-9}$). In addition, differential intracellular/surface partitioning was studied with co-transfection of the alleles fused to distinct fluorescent proteins in *COS1* cells. By quantitative confocal microscopy we found a higher ratio of cell-surface/total CTLAThr17 versus CTLAAla17 (P=0.01).

Our findings corroborate observations, in other proteins, that the signal peptide can determine the efficiency of post-translational modifications other than cleavage and suggest inefficient processing of the autoimmunity-predisposing Ala allele as the explanation for the genetic effect.

3.3 Introduction

The cytotoxic T-Lymphocyte antigen-4 (CTLA-4), a disulfide-linked homodimer expressed on the cell-surface of activated T cells is responsible for the attenuation of immune response by binding to ligands (B7-1 and B7-2) expressed on the surface of antigen presenting cells (Walunas et al. 1994; Krummel and Allison 1996; Lee et al. 1998; Walunas and Bluestone 1998). Recent reports have implicated *CTLA4* in the modulation of autoimmune responses (Karandikar et al. 1996; Perrin et al. 1996; Chambers et al. 1997; Hurwitz et al. 1997; Luhder et al. 1998; Bachmann et al. 1999; Karandikar et al. 2000) and in the maintenance of peripheral tolerance (Perez et al. 1997; Walunas and Bluestone 1998; Eagar et al. 2002). The *CTLA4* knockout mouse exhibits a severe lymphoproliferative disorder, autoimmune disease and early lethality, demonstrating the importance of CTLA-4 in the modulation of T-cell response (Tivol et al. 1995; Waterhouse et al. 1995). In Humans, patients with the Chediak-Higashi syndrome present with symptoms not unlike those of the CTLA-/- mouse, due to a defect in the CTLA-4 cycling pathway caused by mutations in the lysosomal trafficking regulator gene (*LYST*) (Barrat et al. 1999). More subtle reductions in expression or function of *CTLA4* may determine susceptibility to common autoimmune diseases.

Indeed, common *CTLA4* polymorphisms have been found to confer susceptibility to Type 1 Diabetes (Nistico et al. 1996; Todd and Farrall 1996; Donner et al. 1997; Marron et al. 1997; Van der Auwera et al. 1997; Awata et al. 1998; Marron et al. 2000), autoimmune thyroid disease (Yanagawa et al. 1995; Donner et al. 1997; Donner et al. 1997; Kotsa et al. 1997; Yanagawa et al. 1997; Barbesino et al. 1998; Heward et al. 1999;

Vaidya et al. 1999; Tomer et al. 2001; Kouki et al. 2002; Nithiyananthan et al. 2002), and several other autoimmune disorders (Djilali-Saiah et al. 1998; Kemp et al. 1998; Giscombe et al. 2002; King et al. 2002; Vaidya et al. 2002). By the transmission disequilibrium test, association with diabetes has been narrowed down to a haplotype encompassing *CTLA4* but not adjacent genes (Marron et al. 2000; Kouki et al. 2002). The haplotype consists of three *CTLA4* polymorphisms in tight linkage disequilibrium (LD) with each other which includes a –318C>T transition in the promoter, an +49A>G (Thr17Ala) signal peptide amino acid substitution and a microsatellite $(AT)^n$ repeat in the 3'UTR (Marron et al. 2000; Holopainen and Partanen 2001; Kouki et al. 2002). Due to the tight LD, contribution to diabetes susceptibility cannot be genetically dissected and functional studies are required to define the etiological variant(s).

We decided to focus on the non-synonymous signal peptide polymorphism as the most likely candidate. Two recent reports present evidence that T-lymphocytes from subjects homozygous for the diabetes-predisposing G (Ala) allele of the CTLA-4 signal peptide showed enhanced proliferation and cytokine production after in vitro stimulation compared to cells from homozygotes for the protective A (Thr17) allele (Kouki et al. 2000; Maurer et al. 2002). Although Muerer et al. (Maurer et al. 2002) presented some non-quantitative evidence of defective CTLA-4 targeting to the cell surface by confocal microscopy, neither study addressed the molecular mechanism for the differential behavior of T-lymphocytes from homozygotes for each genotype.

Since the signal peptide is co-translationally cleaved in the endoplasmic reticulum (ER) and is not a part of the mature protein, we hypothesized that the Thr17Ala

polymorphism determines differential targeting to the cell surface by altering early intracellular trafficking of CTLA-4. Signal peptides function in directing ribosome-bound nascent polypeptides to the (ER) membrane where they assure the translocation of growing polypeptide chains into the ER lumen. In conformity with most signal sequences, the CTLA-4 signal peptide has three distinct regions (Figure 1): a predicted hydrophobic sequence of twelve amino acids flanked by two helix-breaking prolines, a small polar C-terminus (c-) region encompassing the recognition site for signal peptide cleavage and a long N-terminus (n-) region of twenty amino acids that includes the Thr to Ala subsitution at position 17. As shown in Figure 1A, the Ala allele introduces a hydrophobic amino acid in a highly conserved position, occupied by a serine or threonine in 24 of 25 other species found in a BLAST search we performed (four species are shown for illustration purposes in Figure 1A). This change somewhat alters hydrophobicity and alpha-helix propensity, two properties known to be important in signal peptide function (Figure. 1B). Since Ala homozygotes are very frequent in the general population, the functional consequences of this substitution must be subtle, as expected of common alleles predisposing to complex disorders. These allelic effects may involve differential binding to the signal recognition particle (SRP), anchoring of the signal peptide to the ER membrane, entry of the growing polypeptide into the lumen, signal peptide cleavage and, possibly, association with chaperones and other quality-control elements of the ER lumen that might affect folding and glycosylation (Li et al. 1996).

The work reported here was aimed at testing the hypothesis that the (Thr17Ala) substitution in the signal peptide of CTLA-4 alters the early ER trafficking and/or processing of CTLA-4 and leads to its differential expression on the cell surface. To test

our hypothesis we used a cell-free *in vitro* translation system suited for examining early ER transport events. To also look for decreased expression at the cell surface, the hypothesized ultimate consequence of defective early processing, we examined allelic differences in intracellular vs. cell-surface CTLA-4 levels in a dual-transfection system with fusion of each allele to a different fluorescent protein and simultaneous quantification by confocal microscopy. The results suggest defective ER processing of a significant portion of the CTLAAla17 molecules resulting in an aberrantly glycosylated product and decreased cell-surface expression.

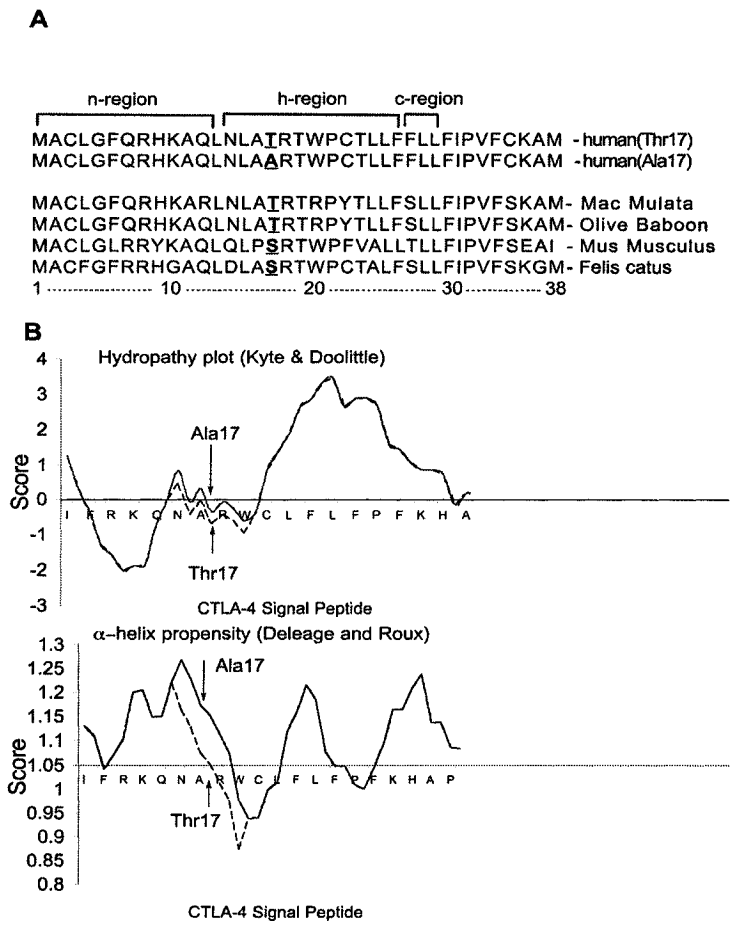


Figure 1. Multiple sequence alignment of CTLA-4 signal peptides and prediction of signal peptide function

A, Alignment of CTLA-4 signal peptides shows a conserved Thr17 or Ser17 (underlined and bold) in 24 out of 25 species in a BLAST search (Karplus et al. 1998). The predicted signal peptide cleavage site is indicated. ***B***, Prediction of factors which influence signal peptide function (Matoba and Ogrydziak 1998) was done using algorithms available at ProtScale (http://us.expasy.org/cgi-bin/protscale.pl). Hydrophobicity was calculated by

the Kyte and Doolittle method (Kyte and Doolittle 1982) and alpha helix propensity by Deleage and Roux (Deleage et al. 1987). We found that the Threonine to Alanine change resulted in increased hydrophobicity and in a higher propensity to form α-helices in the area directly adjacent to the change.

3.4 Materials and Methods

Construction of DNA plasmids

Full-length *CTLA4* was amplified by RT-PCR from cDNA of a heterozygous individual for the signal peptide polymorphism at codon 17. For the fluorescent fusion proteins, a forward primer containing the linker *Nhe*I (underlined) ATAGCTAGCATGG CTTGCCTTTGGATTTCAG and an antisense primer containing the linker *Age*I (underlined) with two additional bases (lowercase) CACACCGGTgc ATTGA TGGGAATAAAATAAGGC ensured an *in-frame*, seamless fusion protein. PCR was performed with a high-fidelity Polymerase/Taq mixture. A-3'overhangs were added and the PCR product was cloned into pCR2.1 (Invitrogen, Carlsbad, CA). Clone genotyping was done by PCR and digestion with *Fnu4HI* (New England Biolabs, Beverly, MA). Full-length CTLAThr17 and CTLAAla17 digested with *Nhe*I + *Age*I purified from plasmid was fused to restricted *Nhe*I + *Age*I ECFP and EYFP (Clontech, Palo Alto, CA). For *in vitro* transcription, full-length *CTLA4* with the termination codon was amplified from plasmid with the mutagenic antisense primer CACACCGGTTcaATTGATGGGAA TAAAATAAGGC (the stop codon is in lowercase and the *Age*I linker is underlined). A-3' overhangs were added to the PCR product and ligated to pCR2.1 T-vector and clones selected for orientation downstream of the T7 promoter.

Truncation of amino acids Tyr201-Asn223 from the CTLA-4 COOH terminal results in a mutant constitutively directed to the cell surface with no requirement for the machinery that normally directs this translocation in activated T-cells (Leung et al. 1995).

The corresponding construct was prepared by PCR amplification from cDNA of a heterozygous individual using the antisense primer ATAACCGGTgaCCCCTGTTGTA AGAGGGCTTC (*Age*I underlined) and T/A cloned into pCR2.1. COOH-truncated CTLAThr17 and CTLAAla17 will be denoted herein as CTLAGlyThr17 and CTLAGlyAla17 in reference to Gly200, the last amino acid in the truncated constructs (numbering of all constructs assumes an *uncleaved* signal peptide of 37 amino acids). The truncated DNA plasmids were then subcloned *in-frame* into ECFP and EYFP plasmids, using the *Nhe*I + *Age*I sites. All resulting DNA constructs were sequenced.

N-Glycosylation mutants

The two predicted *N*-linked glycosylation sites in the CTLA-4 sequence were deleted by site-directed mutagenesis by the mismatched-primer method (Kunkel 1985) using the QuikChange Multi site-directed mutagenesis kit (Stratagene, LaJolla, CA). Asparagine-113 and Asparagine-145 were converted to Aspartate residues in CTLAAla17 and CTLAThr17. Both sites were mutated in each allele yielding CTLAAla17(Asn113/145Asp) and CTLAThr17(Asn113/145Asp) or one at a time, yielding CTLAAla17(Asn113Asp), CTLAAla17(Asn113Asp) and CTLAThr17 (Asn145Asp), and CTLAAla17(Asn145Asp). All mutant constructs were verified by sequencing.

Cell culture, transfection and antibodies

COS1 cells were grown in Dulbecco's modified Eagle medium (DMEM) (Gibco, Frederick, MD) supplemented with 10% fetal bovine serum (FBS), 100 units/ml

penicillin and 100 units/ml streptomycin, at 37°C and 5% CO_2. Transient transfections were carried out using Fugene 6 (Boeringher Mannhein, Indianapolis, IN) in 6-well plates on glass cover slips according to the manufacturer's instructions. Cells were processed 24-48 hours after transfection.

In vitro transcription, translation and Endoglycosidase H digestions

Full-length cDNA encoding CTLAAla17 and CTLAThr17 (or their glycosylation mutants) in pCR2.1 was linearized by *SpeI* and *in vitro* transcribed using T7 RNA polymerase (Promega, Madison, WI). The resulting mRNA was standardized to 1mg/ml and equal amounts were added to a premixed cocktail containing Rabbit Reticulosyte Lysate with [^{35}S]-Methionine (>1000Ci/mmol) (Amersham, Arlington Hts, IL), in the presence or absence of canine pancreatic microsomal membranes according to the manufacturer's instructions (Promega). Translation products were denatured by boiling in 2x-SDS loading buffer and 10-15 μl aliquots were run on 15% SDS-PAGE, followed by autoradiography at -70°C. Endoglycosidase H (EndoH) (New England Biolabs) digestions were performed on in vitro translated protein. Briefly, proteins were treated with 75mM dithiothreitol and 10% SDS and boiled for 5 min. The resulting supernatant was incubated with EndoH, for the indicated times at 37°C. Products were analyzed by SDS-PAGE and visualized by autoradiography as described above.

Proteinase K digestions, ultracentrifugation and signal peptidase inhibition

To distinguish *in vitro* translated *CTLA4* molecules integrated in the ER membrane from those in the cytosolic phase, we used proteinase K digestion. Translation

products were treated with 100 mM $CaCl_2$, and incubated for 1 hour with 20 μg/ml Proteinase K (Sigma) with or without Triton X-100 (1%) at 0°C. The reaction was terminated with phenylmethylsulfonyl fluoride (PMSF) (2mM final concentration). As an additional measure of cytosolic/ER partition the translation products were separated by ultracentrifugation. Briefly, reactions were diluted 100-fold with Na_2CO_3 (pH 11.5), and centrifuged at 100,000 rpm for 30 min at 4°C. The pellet was resuspended in 1x SDS sample buffer and denatured by boiling for 3-4 min. The supernatant was concentrated by column filtration (Microcon, Bedford, MA) to 1/10 of its original volume. The sample was prepared for SDS-PAGE by denaturation in 2X SDS sample buffer. All products were resolved by SDS-PAGE and visualized by autoradiography. Finally, inhibition of the signal peptidase was achieved by incubation of the translation mix with 5mM (final concentration) *N*-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone (Sigma) in 10%Me_2SO for 1 hour at 30°C.

Image Processing and Quantification and Confocal Microscopy

Transfected cells were washed 3 times in PBS and fixed with ice-cold 50:50 acetone/methanol for 3 min or fixed with freshly prepared 4% paraformaldehyde for 30 min. The fixed cells were mounted on standard microscope slides with conventional mounting media. Samples were analyzed by confocal microscopy (LSM 510, Zeiss Axiophot, Germany). Simultaneous double acquisitions were performed using the 458nm and 514nm laser lines to excite Cyan Fluorescent Protein (CFP) and Yellow Fluorescent Protein (YFP) respectively using 63X oil immersion Neofluar objectives. The fluorescence was selected with double fluorescence dichroic mirror and a band pass

filter of 480-520nm for CFP and a longpass filter of 560nm for YFP. The images were processed and analyzed using NIH Image freeware available from the National Institutes of Health (NIH) website (http://rsb.info.nih.gov/nih-image/Default.html). An outline of the cell-surface was drawn with the cursor and the mean density was calculated for each color acquisition simultaneously. The same was done for the intracellular fluorescence (both procedures are illustrated in Figure. 7C).

For densitometry measurements taken from autoradiographs background density was subtracted from each band and the same area was selected for all measurements within one experiment.

3.4 Results

CTLAAla17 is inefficiently processed in the ER

When otherwise identical full-length CTLAAla17 and CTLAThr17 cDNA constructs were *in vitro* transcribed and translated in the absence of microsomal membranes, [^{35}S]-methionine labeled protein products from both allelic forms migrated with an apparent molecular weight of 26 kDa. This is consistent with the calculated molecular mass of monomeric, uncleaved, unglycosylated CTLA-4 (Figure. 2A, *lanes 1 and 4*). Upon addition of microsomal membranes translation products of 29 kDa were apparent for both alleles, corresponding to the size change expected from high-mannose glycosylation in two positions and signal peptide cleavage. In addition, the CTLAAla17 reaction (Figure. 2A, *lanes 2 and 3*) contained an intermediate band migrating with an apparent molecular weight of 25 kDa. In multiple experiments, this band was absent or only faintly visible in the CTLAThr17 lane. Quantitatively, the CTLAAla17 25 kDa product represented 32.7%±1.0 of total processed CTLAAla17 versus 10.6%±1.2 for the CTLAThr17 allele (Figure. 2B) ($P<1x10^{-9}$, n=11 independent experiments). This ratio was independent of the amount of microsomal membranes in the reaction, within a range from 0.8 eq to 2.8 eq membranes added to the reticulosyte lysate. One observation worth noting was that translation efficiency was substantially increased in the presence of microsomal membranes suggesting that co-translational CTLA-4 processing increases translation efficiency. Having established that a portion of CTLAAla17 molecules is differentially processed in the ER, we next examined the nature of this allele-specific intermediate.

The 25 kDa CTLA-4Ala17 fraction, close in size with the band seen in the absence of microsomes could represent an uncleaved and unglycosylated form which has not undergone any processing in the ER. Alternatively, it could represent correctly translocated but incorrectly cleaved and/or glycosylated product.

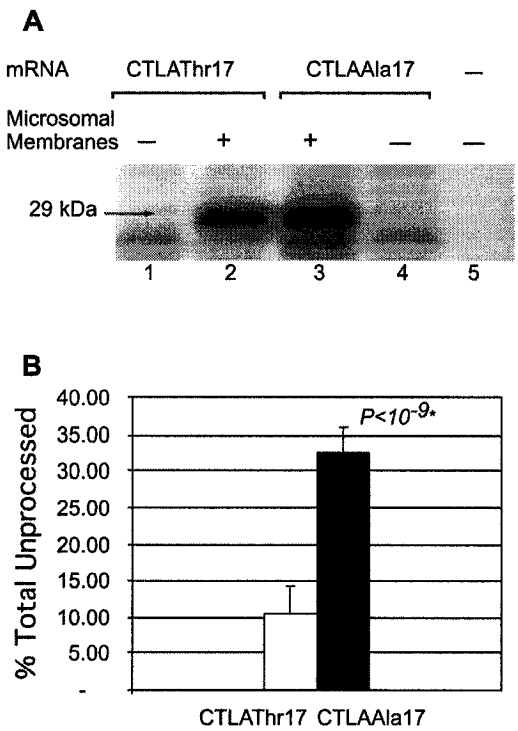


Figure 2. CTLAAla17 is inefficiently processed in the Endoplasmic Reticulum

A, CTLAAla17 and CTLAThr17 constructs were *in vitro* transcribed and translated under conditions outlined in Experimental Procedures. Where indicated, ER-containing microsomal membranes were added. In the presence of membranes an intermediate is apparent in the CTLAAla17 reaction at ~25 kDa, virtually absent in the *CTLA4*Thr17 reaction. This suggests incomplete ER processing. A control reaction that contains no mRNA is shown in *lane 5*. ***B***, We quantified the % total unprocessed by measuring densities of the 29 kDa band (considered processed) and the 25 kDa (considered unprocessed) within each lane for CTLAAla17 and CTLAThr17. All densities were measured with NIH Image on scanned autoradiographs. Background was subtracted for each band and represented are the mean and standard error of 11 independent experiments.

Signal peptide is cleaved in the major products of both CTLAAla17 and CTLAThr17

According to SignalP, a signal peptidase recognition site prediction algorithm (Nielsen et al. 1997) (www.cbs.dtu.dk/services/SignalP/index.html) both alleles had the same cleavage site at Ala37-Met38 (Figure 1A), resulting in a 37 amino acid cleaved signal peptide. Cleavage by the signal peptidase was experimentally confirmed by using a signal peptidase inhibitor (N-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone). For both allelic forms, inhibition of cleavage gave rise to a full-length intermediate of ~33 kDa resulting from a gain of 4 kDa by the 29 kD intermediate, corresponding to a gain of 37 amino acid signal peptide (Figure 3). Thus the major bands in the translated products of both CTLAAla17 and CTLAThr17 represent cleaved intermediates. The cleavage status of the additional CTLAAla17 25 kDa band was less clear because of band overlap (Figure 3), as inhibition was not complete at the maximum effective concentration of inhibitor (higher concentrations inhibited translation). Thus the major bands for both alleles must represent fully glycosylated and fully cleaved product.

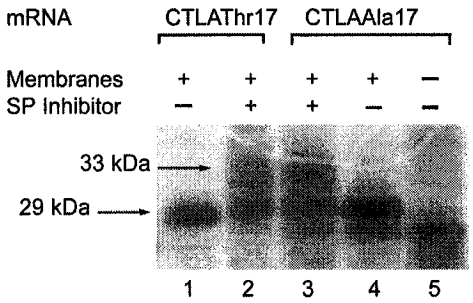


Figure 3: Signal peptide cleavage of the major products of CTLAAla17 and CTLAThr17 is unaffected.

We assessed signal peptide cleavage by performing the translation reactions for each allele in the presence of the signal peptidase inhibitor, N-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethyl ketone. Inhibition of cleavage results in a 4 kDa increase in molecular weight (band shift to 33 kDa) as expected for the gain of the 37 amino acid signal peptide. This indicates that the major translation products of both alleles are cleaved.

CTLA-4 has two *N*-linked glycosylation sites

CTLA-4, a cell-surface receptor, has two predicted *N-linked* glycosylation sites located in the extracellular domain, at Asn113 and Asn145, thought to be important for structural integrity (Metzler et al. 1997). Although not all instances of the consensus sequon Asn-X-Thr/Ser (X=any amino acid) are necessarily glycosylated (Gavel and von Heijne 1990; Kasturi et al. 1995), the observed size of the major band clearly indicated full glycosylation of most of the molecules for either allele. This was further confirmed with partial digestion of the *in vitro* translation products with EndoH that, in glycoproteins with several glycosylation sites, produces a ladder of partially digested molecules differing by only one *N-linked* chain. Two additional major bands with an estimated molecular weight of 25 kDa and 22 kDa appeared after 1-2 hours of digestion (Figure. 4A, *lanes 2-5*) representing cleavage of, respectively, one or two N-linked glycosylation moieties from the 29 kDa form that was still visible. Further digestion for up to 16 hours never produced smaller products. The 25 kDa band was indistinguishable in size from the aberrant intermediate seen with CTLAAla17. This allowed us to conclude firmly that, in addition to being cleaved, the 29 kDa product was glycosylated at both sites and, tentatively, that the aberrant CTLAAla17 product represents cleaved, monoglycosylated protein.

This was further confirmed with complete Endo-H deglycosylation. As expected, prolonged digestion for 16 hours with EndoH resulted in a major band of 22 kDa for both alleles, with no extra band for CTLAAla17 (Figure. 4B). Thus complete deglycosylation abolished any difference between allelic forms of the protein, allowing us to conclude

that the extra band seen after microsomal processing of nascent CTLAAla17 represents cleaved but aberrantly glycosylated CTLA-4.

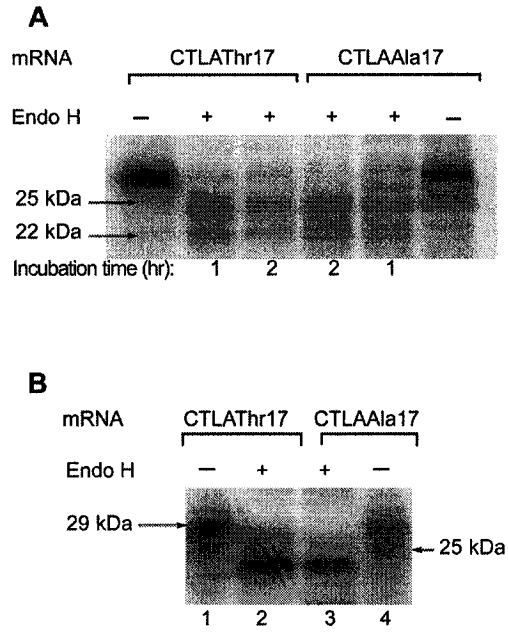


Figure 4: Deglycosylation abolishes allelic differences in processing efficiency

A, CTLAThr17 and CTLAAla17 were *in vitro* translated in the presence of microsomal membranes and subjected to partial EndoH digestions for 1 or 2 hours at 37°C. A ladder of intermediates was obtained representing the deglycosylated proteins. The smaller digestion product obtained in *lanes 2-5* corresponds to a completely deglycosylated protein and migrates at 22 kDa. The 25 kDa intermediate observed in *lanes 2-5* represents a monoglycosylated protein and migrates with the same apparent molecular weight as the intermediate observed in the CTLAAla17 reaction (*lane 6*). ***B***, *In vitro* translated CTLAThr17 and CTLAAla17, in the presence of microsomal membranes, were subjected to overnight digestion with EndoH resulting in a major digestion product at 22 kDa (*lanes 2-3*) abrogating the difference between the two alleles. This confirms that the 25 kDa precursor observed following microsomal processing of the CTLAAla17 represents a cleaved but improperly glycosylated CTLA-4.

Differential processing of the two alleles depends on glycosylation:

The deglycosylation results regarding the nature of the unprocessed intermediate observed in CTLAAla17 were further corroborated by site-specific mutagenesis designed to abrogate glycosylation by removing each glycosylation site separately and both together in each allele. Asn113 and Asn145, the predicted *N-linked* glycosylation sites were replaced with Asp residues at the same positions (Figure. 5). The resulting constructs were *in vitro* transcribed and translated, as described above.

All glycosylation mutants translated in the absence of microsomal membranes, behaved as the wild-type alleles and a 26 kDa translation product was detected. In the presence of ER-containing membranes, the double glycosylation mutant migrated at 22 kDa, as expected of an unglycosylated CTLA-4 with a cleaved signal peptide. No extra band corresponding to differentially processed CTLAAla17 band was detectable (Figure. 5, *lanes 10 and 11*), clearly indicating that the allelic difference in processing requires the presence of glycosylation sites. The uniform 25 kDa electrophoretic mobility of the single-site mutants CTLAThr17 (Asn113Asp) and CTLAAla17 (Asn113Asp), CTLAThr17 (Asn145Asp) and CTLAAla17 (Asn145Asp) (Figure 5, *lanes 2,3 and 6, 7*) is consistent with loss of the signal peptide (4 kDa) and gain of a single high-mannose glycosylation site (2-3 kDa). Again there was no difference between alleles, indicating that differential processing occurs only in the presence of both glycosylation sites. These results also confirm with certainty that both CTLA-4 signal peptide alleles are ER-translocation competent and can be cleaved, since a completely deglycosylated product migrates at 22 kDa, 4 kDa less than the monomeric CTLA-4 product translated in the absence of membranes.

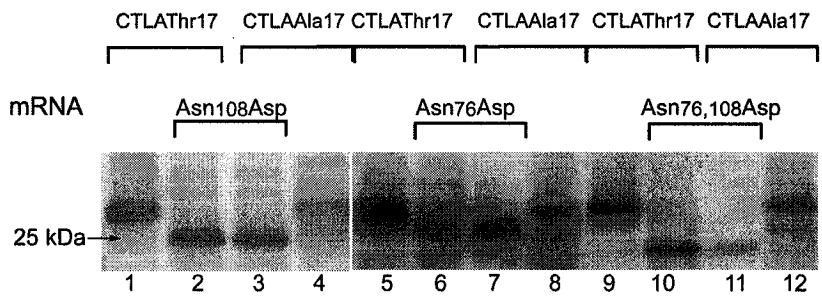


Figure 5: Differential processing of the two alleles depends on glycosylation

CTLA-4 mutants with abolished *N*-linked glycosylation sites were *in vitro* transcribed and translated in the presence of microsomal membranes. In *lanes 2 and 3*, CTLAThr17 (Asn145Asp) and CTLAAla17 (Asn145Asp) respectively, translated a 25 kDa product, the expected size for a monoglycosylated CTLA-4; these monoglycosylated mutants translate an intermediate equal in size as the partially glycosylated form in CTLAAla17 (wild-type, *lane 4)*. Similarly, in lanes *6* and *7,* mutants at the second glycosylated site, Asn113, translate a 25 kDa monoglycosylated CTLA-4, the same size as the CTLAAla17 intermediate (WT-Ala in *lane 8* for comparison). *Lanes 10* and *11* correspond to double mutants, CTLAThr17 (Asn113/145Asp) and CTLAAla17 (Asn113/145Asp), respectively where a 22 kDa translation product is obtained corresponding to a cleaved but unglycosylated CTLA-4. Differences between the two alleles are therefore abolished when one or the other site is mutated; they appear to be dependent on the presence of both glycosylation sites.

The partial glycosylation intermediate is not integrated in the ER membrane.

We next addressed the question of whether the cleaved but aberrantly glycosylated CTLAAla17 intermediate, representing almost one third of all molecules of the autoimmunity-predisposing allele in the ER, can be correctly targeted to the cell surface. Misprocessed proteins are retained in the ER bound to chaperones such as calnexin and calreticulin and eventually translocated back to the cytoplasmic phase where they undergo ubiquitin-driven, proteasome-dependent degradation (Ellgaard et al. 1999). Partition of the aberrantly glycosylated CTLAAla17 between cytosolic phase and ER membrane was evaluated by two independent assays. Protection from Proteinase K digestion and resistance to extraction by alkaline high salt is both indicators of integration into the microsomal membranes. Since CTLA-4 is a Type I transmembrane glycoprotein, its 36 amino acid COOH-tail is exposed in the cytoplasm and will be digested, resulting in a 4 kDa loss.

Indeed, Proteinase K digestion resulted in the expected 4 kDa reduction of the 29 kDa major bands seen with both alleles (Figure. 6A, *lanes 2 and 5*). A band corresponding to a 4 kDa reduction in the size of the 25 kDa intermediate seen with CTLAAla17 is also seen (Figure 6A, *lane 5*) but this form is considerably less intense relative to the upper band, indicating partial sensitivity to Proteinase K digestion. By densitometric quantification, prior to digestion it constitutes 34.5%$\pm$0.6 of total CTLAAla17 product, but only 18.5%$\pm$0.8 after digestion ($P=0.0001$, n=3 independent experiments) (Figure.6B, in reference to *lanes 5 and 6* in Figure. 6A). The corresponding percentages for CTLAThr17 are 10.8%$\pm$1.0 and 9.9%$\pm$1.8 respectively (NS, n=3) (Figure 6B, in reference to Figure. 6A, *lanes.1-2*). Thus, molecules located outside the ER can

account for the allelic difference in the abundance of the incompletely processed form almost entirely.

Strikingly similar results were obtained upon treatment of translation reactions with high Na_2CO_3 which releases soluble and peripheral membrane proteins while transmembrane proteins remain inserted in the ER lipid bilayer, followed by separation of pellet (P) and supernatant (S) fractions by ultracentrifugation. The major 29 kDa bands were largely recovered in the pellet (Figure. 6C, *lanes 5-6*) for CTLAThr17 and CTLAAla17. This supports the results found in the protease sensitivity assay and confirms that CTLAAla17 and CTLAThr17 are both translocation competent and integral ER membrane proteins. It was apparent, however, that significantly less incompletely glycosylated CTLAAla17 intermediate was recovered in the pellet. By densitometric quantification it was found to be 36.2%±1.0 of the total prior to separation by ultracentrifugation and 24.9%±0.4 in the pellet (*P=0.0078*, n=3) (Figure. 6D). The corresponding values obtained for the CTLAThr17 intermediate were 8.9%±2.7 prior to ultracentrifugation and 10.0%±1.2 recovered in the pellet (NS, n=3). This remarkable concordance between two independent methods demonstrates that roughly half of the aberrantly processed CTLAAla17 is in the cytoplasmic phase. Since it has undergone signal-peptide cleavage and glycosylation, this fraction does not represent failure of translocation to the ER but rather retrotranslocation back to the cytoplasmic phase for proteasomal degradation.

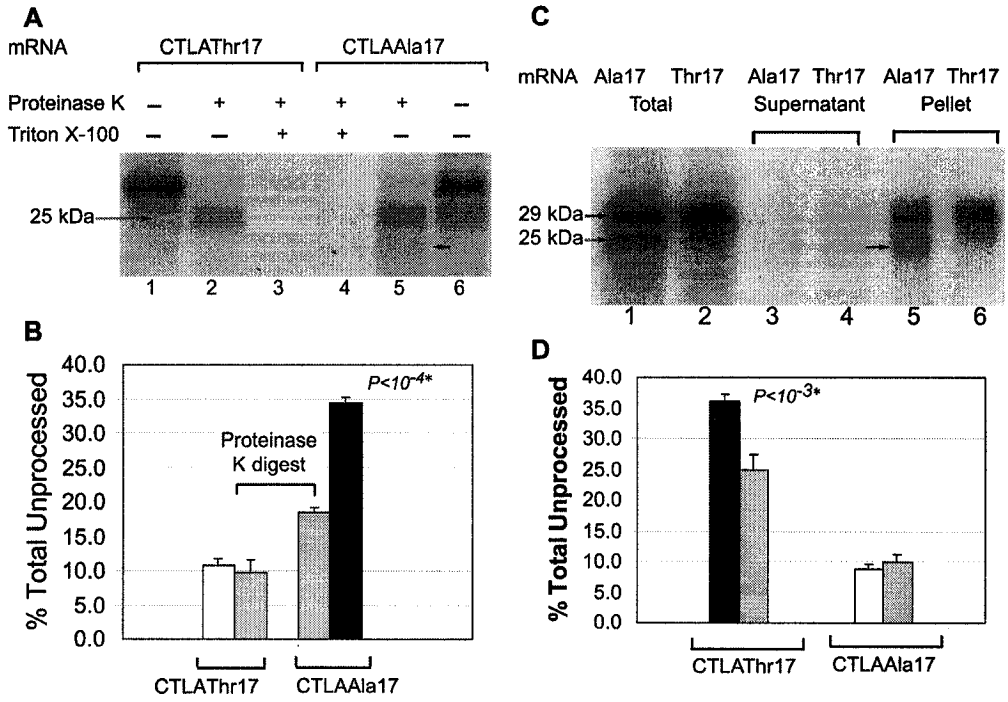


Figure 6: The partially glycosylated CTLAAla17 intermediate is located outside the ER membrane

A. We subjected *in vitro* translated CTLAThr17 and CTLAAla17 in the presence of microsomal membranes, to Proteinase K in the presence of Triton X-100 (where indicated). We observed a ~4 kDa reduction in the size for both alleles of their major 25 kDa band, as a result of digestion of the 36 aa cytoplasmic tail (*lanes 2 and 5*). We detected a minor product in the CTLAAla17 digest (*lane 5*, arrow) migrating at ~22 kDa which represents the incompletely glycosylated precursor minus the COOH-tail. Since we recovered considerable less intermediate upon digestion this suggests it is outside the ER membrane. ***B***, Quantification of relative densities of the major and minor bands before and after Proteinase K digestion. The bar graph shown represents three independent experiments evaluated by a one-tailed Student's T-test. ***C.*** As an alternative assay of ER membrane integration, translation products were extracted with Na_2CO_3 (pH 11.5) and ultracentrifuged. ER-integrated proteins are found in the membrane pellet, while cytosolic molecules stay in the supernatant. ***D***. Quantification of the proportion of incompletely glycosylated intermediate recovered in the pellet, as a fraction of total.

Checkered bars represent percentage in the pellet. Statistics as in B, n=3 experiments. This independently confirms that a substantial portion of the incompletely glycosylated fraction in the Ala allele retrotranslocates to the cytoplasmic phase.

CTLAAla17 and CTLAThr17 do not co-localize in co-transfected *COS*1 cells

To determine if the differential processing observed in the cell-free system would translate into cell-surface expression differences, we devised a dual-transfection system in COS1 cells with fusion proteins of CFP and YFP downstream of otherwise identical CTLAAla17 and CTLAThr17. In these experiments reciprocal constructs were always tested to exclude possible effects due to the different properties of the fluorescent proteins. In addition, we directly demonstrated that co-transfections with the same allele tagged with each of the two FPs (e.g. CFP-CTLAThr17 and YFP-CTLAThr17) showed that the same allele did not behave differently as a result of the different fluorescent protein fusion and was targeted and expressed in the same compartments (Figure. 7A, *panels a-c*). Similar results were observed for CFP-CTLAAla17 and YFP-CTLAAla17 co-transfections (Figure 7A, *panels d-f*)

CTLA-4 is known to localize to endosomal compartments in T-cells as well as in non-T-cell systems (Leung et al. 1995; Linsley et al. 1996). To demonstrate that the fusions of CTLA-4 resulted in a properly targeted protein, each allele was transfected independently and the cells were stained with an antibody to the transferrin receptor followed by detection by a *Cy-5* conjugate. The transferrin receptor has been previously shown to co-localize with CTLA-4 in post-Golgi compartments as was the case for our fusion proteins (data not shown).

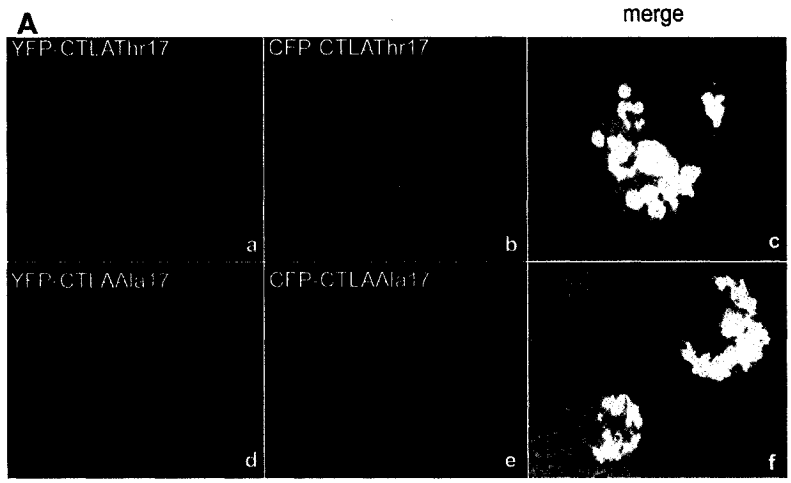
When the two alleles, now labeled with different fluorescent proteins (CFP-CTLAThr17 and YFP-CTLAAla17) were introduced into COS1 cells clear differences in targeting were seen, with distinct regions where only one allele could be found (Figure. 7B, *panels a-c, see arrows*). Similar observations were made for reciprocal transfections

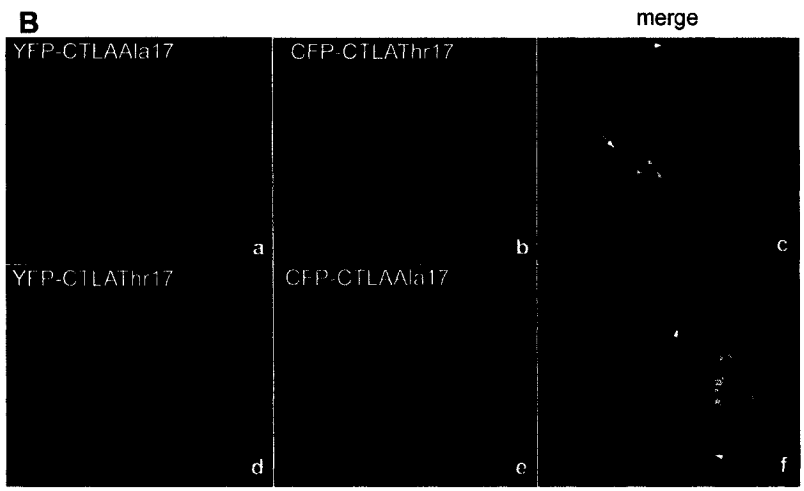
of YFP-CTLAThr17 and CFP-CTLAAla17 (Figure. 7B, *panels d-f*). These results are consistent with our *in vitro* observations of cytoplasmic retrotranslocation of CTLAAla17. The expected resulting difference in targeting to the cell surface was then investigated.

Cell-surface levels of CTLAAla17 in *COS1* cells are lower than CTLAThr17

Using the same dual-transfection system described above, CTLA-4 constructs with signal peptide allelic variants were tested for quantitative differences in expression at the cell-surface of *COS1* cells. Surface expression is specific for activated T-lymphocytes, but truncating the last 22 amino acids of the carboxy-tail allows partial cell-surface expression in a non-T cell system (Leung et al. 1995). By co-expressing fluorescent fusion proteins of the COOH-truncated CTLA-4 alleles in one single cell, it was possible to simultaneously measure fluorescence intensity at the cell-surface for both alleles and compare it to intracellular fluorescence (Figure. 7C, *panels a-f*). Ratios of cell-surface/intracellular mean fluorescence were quantified in single cells for both CTLA-4 alleles and compared with a paired, 2-tailed student t-test. Mean cell-surface/intracellular fluorescence intensity ratios are 9.4 ± 2.2 for YFPGlyThr17 and 1.5 ± 0.8 for CFPGlyAla17 ($P=5 \times 10^{-4}$, n=9 cells, example illustrated in Figure. 7C). The average ratios of cell-surface/intracellular fluorescence intensity for YFPGlyAla17 and CFPGlyThr17 co-transfections were 6.0 ± 2.4 and 10.0 ± 2.8 respectively ($P=0.01$) (Figure. 7C). The ratio varied somewhat from cell to cell, accounting for the relatively large standard error of the mean but, since the comparisons were paired within the same cell, the difference was highly significant statistically. The results presented here were

calculated from at least 9 different cells obtained in at least two independent transfections for each allele. Cells were selected for expressing the two colors at roughly equal intensities, and their selection was finalized prior to any knowledge of the quantitative results. By choosing to measure ratios within the same cell for both alleles correction for transfection efficiency differences is inherent as well as correction for any differences in fluorophore efficiency. Again, in this case same-allele co-transfections were performed and similar calculations done but no statistical significance was found ($P>0.05$). This allowed us to conclude that, for any given level of expression, there is significantly less CTLAGlyAla17 expressed at the cell-surface of COS1 cells than CTLAGlyThr17.


A
merge
YFP-CTLAThr17
CFP-CTLAThr17
a
b
c
YFP-CTLAAla17
CFP-CTLAAla17
d
e
f


B
merge
YFP-CTLAAla17
CFP-CTLAThr17
a
b
c
YFP-CTLAThr17
CFP-CTLAAla17
d
e
f

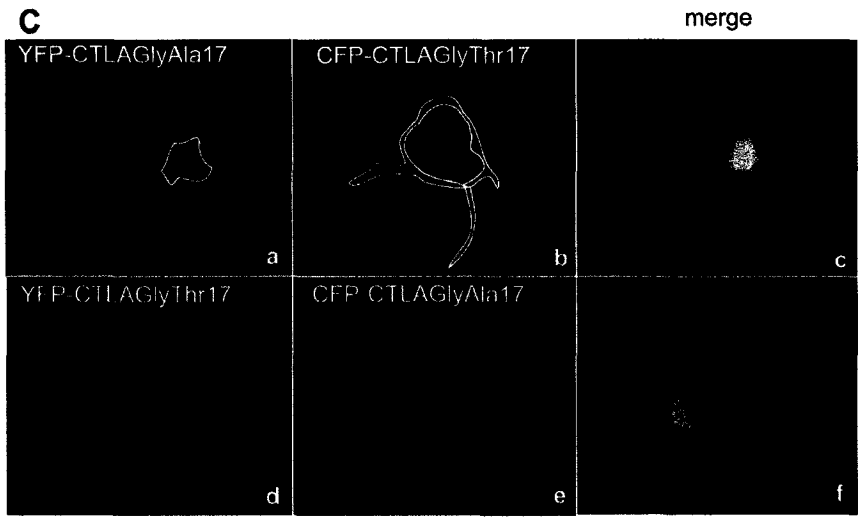


Figure 7: CTLAAla17 and CTLAThr17 do not co-localize in transfected cells and there is significantly less Ala17 expressed at the cell-surface

We transfected *COS1* cells with carboxyterminal-truncated CTLA-4 fused to cyan or yellow fluorescent protein (respectively CFP, green and YFP, red), capable of some constitutive translocation to the cell surface. ***A***, The same allele fused to different color fluorescent proteins was co-transfected into COS1 cells: YFP-CTLAThr17 + CFP-CTLAThr17 (*panels a-c*) and YFP-CTLAAla17 + CFP-CTLAAla17 (*panels d-f*). We processed the cells 24-48 hrs following transfection and analyzed them by confocal microscopy at 63X magnification. All transfections were repeated independently at least 5 times. Complete co-localization of the different colored fusion proteins validates the approach despite inherent differences in the fluorophore intensities. ***B***, *Panels a and b* represent the co-transfection of YFP-CTLAAla17 + CFP-CTLAThr17 into *COS1* cells. The two alleles do not co-localize in COS1 cells (panel c, see arrows, note areas where no red is found and areas where no green is found). *Panels d-f* are a representative of the reciprocal transfection YFP-CTLAThr17 + CFP-CTLAAla17. ***C***, We performed quantitative fluorescence analysis in COS1 cells co-transfected with YFP-CTLAAla17 + CFP-CTLAThr17 (and the reciprocal transfection), exclusively on cells where the cell-surface was clearly visible. Measurements were collected by outlining the cell-surface and the intracellular fluorescence with a cursor (see *panels a-f*) and measuring pixel densities.

surface was clearly visible. Measurements were collected by outlining the cell-surface and the intracellular fluorescence with a cursor (see *panels a-f*) and measuring pixel densities. All raw data was processed in NIH Image and the cells shown are a representative of at least 2 independent transfections and at least 9 different cells.

3.5 Discussion

Our data demonstrate that a common amino acid polymorphism in a signal peptide can have measurable consequences on the efficiency of processing the protein and ultimately on its expression level at the cell surface. In the case of CTLA-4, a molecule involved in the inhibitory regulation of the immune response, this mechanism offers an attractive explanation for the higher frequency of diabetes and other autoimmune diseases in individuals homozygous for the Ala allele (Nistico et al. 1996; Donner et al. 1997; Donner et al. 1997; Kotsa et al. 1997; Marron et al. 1997; Awata et al. 1998; Heward et al. 1999). Signal peptide mutations completely abrogating proper targeting of a protein have been described in Mendelian disorders (Ito et al. 1993; Racchi et al. 1993; Karaplis et al. 1995; Beuret et al. 1999) but, to our knowledge, this is the first demonstration of a subtler effect in the context of susceptibility to a common complex disease. However subtle, such effects are extremely important to define towards composite molecular prediction and better understanding of mechanism leading to effective prevention/treatment.

To summarize our findings: by a widely used model of *in vitro* reconstitution of translation and ER processing, we have shown that both alleles of the signal peptide are capable of translocation to the ER and both are completely and correctly cleaved. The difference appears to lie in the fact that up to one-third of CTLAAla17 molecules are glycosylated on only one of the two possible sites.

The absence of an effect on translocation or cleavage was not surprising, as the polymorphism does not alter any of the known consensus elements required for

translocation or the signal peptidase cleavage site. Somewhat less expected was the effect on glycosylation, as this modification is not ordinarily thought of as depending on the signal peptide. However, evidence is arising that in the time interval between entry of the amino terminal of the nascent polypeptide into the ER and signal peptidase cleavage, the signal peptide may participate in the process of protein folding and alter its interactions with ER chaperone proteins and modifying enzymes (Li et al. 1996; Chen et al. 2001). Thus, the highly inefficient processing of the HIV glycoprotein 120 was shown to be due strictly to its signal peptide, as the exact same mature-protein sequence is processed efficiently with a control signal peptide (Li et al. 1996). The native HIV gp120 signal sequence is translocation-competent, but affects processing by prolonging gp120 association with ER chaperones and results in incorrect folding which causes ER retention (Li et al. 1996).

Moreover, it is now clear that although the oligosaccharyltransferase complex is tethered to the ER membrane, transfer of N-linked high-mannose moieties to proteins is not synchronous with translocation to ER, but may follow folding of the protein and is profoundly influenced by it (Holst et al. 1996). It is known that not all consensus Asn-X-Ser/Thr *N-linked* glycosylation sites are equally glycosylated or glycosylated at all. A proline in position X (or immediately after Ser/Thr) eliminates glycosylation (Gavel and von Heijne 1990; Shakin-Eshleman et al. 1996), while Trp, Asp, Glu, and Leu decrease glycosylation efficiency (Shakin-Eshleman et al. 1996). By introducing glycosylation sequons into different positions of the *saccharomyces* carboxypeptidase Y, Holst et al. showed dependence of glycosylation on protein folding and the position of the sequon in the context of the folded protein (Holst et al. 1996), results that led them to conclude that

glycosylation does not necessarily precede folding, and can be affected by it. In addition, deletion of downstream sequences was shown to abrogate glycosylation of the hepatitis C virus E1 protein (Dubuisson et al. 2000). Taken together, these observations suggest a mechanism whereby a signal peptide variant may affect the glycosylation efficiency through altered chaperone association and folding.

We performed the single mutant experiments to define which of the two sites remained unglycosylated in the partially processed CTLAAla17 fraction. Assuming that each mutation did not otherwise change CTLA-4 processing, we had expected that elimination of the specific site that remained unglycosylated in a fraction of CTLAAla17 would result in a single monoglycosylated band for both alleles, while elimination of the site fully glycosylated in both alleles would give a monoglycosylated band in both alleles, plus an additional minor unglycosylated band in CTLAAla17. Instead we found a single monoglycosylated band in both alleles, regardless of which site was mutated. The simplest explanation for this is that abolition of the consistently glycosylated site leads to alterations in protein folding and/or interactions with ER chaperones that abolish the glycosylation inefficiency seen with CTLAAla17. Alternatively, eliminating one of two glycosylation sites through mutagenesis may relieve competition between them, although a mechanism for such competition is not obvious and no previous paradigm exists.

Our cell-surface targeting studies used a truncated CTLA-4 mutant that bypasses the activated-lymphocyte specific regulatory mechanism of translocation from Golgi to the cell surface. We believe that this does not detract from the validity of our conclusions, as passage from Golgi to the cell surface is a post-ER event, occurring after cleavage of the signal peptide and its efficiency should not differ between the Ala17 and

Thr17. Given the allelic differences in early processing we found in the *in vitro* system of isolated ER, the obvious explanation is that less of the Ala17 makes it to the Golgi, and therefore less is translocated to the cell surface.

Throughout this discussion we have assumed that doubly glycosylated CTLA-4, the major band in both alleles, represents correct ER processing leading to functional expression at the cell surface; it is practically the only form seen with the Thr17 allele, homozygotes for which are common in the general population and healthy. Extrapolating from our *in vitro* assay, we propose that Ala homozygotes express one-third less CTLA-4 on the surface of their T-cells than Thr homozygotes, which might tip the balance in favor of immune response and predispose them to autoimmune disease. One might speculate that this allele's survival in evolution despite this disadvantage is due to the counterbalancing effect of better defense against infectious diseases.

3.6 Acknowledgements

We would like to thank Dr. Andrew Karaplis for helpful discussions and technical support as well as Dr. Dominique Nouël for expert help with confocal microscopy. The Juvenile Diabetes Research Foundation funded this study and Suzana Anjos was supported by a Thyroid Foundation of Canada/Canadian Institute of Health Research doctoral research award.

4.0 CHAPTER FOUR

Association of the *CTLA4* gene with Type 1 Diabetes: evidence for independent effects of two polymorphisms on the same haplotype block

Suzana M. Anjos, Marie-Catherine Tessier, Constantin Polychronakos
Journal of Clinical Endocrinology and Metabolism, 89(12): 6257-65.2004

4.1 Contributions of Authors

Suzana Anjos: Experimental design and strategy, hypothesis development, gene cloning and reporter assay implementation, genotyping, allelic imbalance studies, data interpretation and analysis, manuscript preparation.

Marie-Catherine Tessier: Technical assistance in subcloning promoter reporter constructs, high-throughput genotyping design and implementation, data collection.

Constantin Polychronakos: Research Director, hypothesis development, statistical analysis, manuscript preparation and editing.

Having addressed the role of the only coding polymorphism associated with T1D susceptibility, we next addressed the nature of regulatory variation at the *CTLA4* gene and the association of several SNPs located at the promoter and 3'flanking region of *CTLA4* with T1D in our family database. Based on sequence conservation and predictions of altered transcription factor binding sites we formulated a working hypothesis that the most proximal promoter polymorphism –318C>T was most likely to alter expression levels of *CTLA4* in an allele-dependent manner. The following work tests this hypothesis and the association of these SNPs with T1D.

4.2 Abstract

A recent study mapped the known association of Type 1 Diabetes with the *CTLA4* gene to a polymorphism at the 3'end (+6230G>A) but could not rule out additional contribution from the 5' end of the gene. To examine this possibility, we analyzed four polymorphisms at the 5' flanking region for effects independent of +6230G>A. We confirm, by the transmission disequilibrium test (TDT) in 498 family trios, over-transmission of the susceptibility allele (G) at +6230 (217/168, P=0.013). Of the four promoter polymorphisms, one (-318C>T) showed over-transmission of the C allele (97/58, P=0.0017). Since the undertransmitted T at the promoter is in linkage disequilibrium with the overtransmitted G at +6230G>A,the effect observed at the promoter cannot be accounted for by the latter. We confirm this by showing that parents heterozygous at the promoter but homozygous at +6230 over-transmit the C promoter allele even more significantly (53/24, $P=9x10^{-4}$). *In vitro*, the T promoter allele directs higher luciferase expression in Jurkat cells by 42% ($P=0.006$), a difference also found in lymphocyte mRNA from eight individuals heterozygous at the promoter but homozygous at +6230 ($P=1.3 \times 10^{-4}$). Thus the +6230G>A cannot be the sole functional variant. Either the two polymorphisms define a haplotype carrying the (yet unexamined) functional variant or the -318C>T contributes to the genetic association independently, a possibility suggested by the functional evidence we present.

4.3 Introduction

Type 1 Diabetes (T1D) is due to the autoimmune destruction of the insulin-producing pancreatic beta cells. Its etiology involves an important element of genetic susceptibility that behaves as a complex trait. Of the several putative loci identified by linkage analysis, only three have been multiply confirmed and narrowed down to specific genes by the transmission disequilibrium test (TDT). In addition to *IDDM1,* which maps to the HLA region (Horn et al. 1988; Thomson et al. 1988; Todd et al. 1988) and *IDDM2* (Julier et al. 1991; Bennett et al. 1995) mapping to the insulin gene, significant association with T1D has been found in a haplotype encompassing the cytotoxic T-lymphocyte antigen-4 gene (*CTLA4*) (Marron et al. 2000; Ueda et al. 2003). This haplotype maps to a broad linkage peak that overlaps the loci previously designated as *IDDM12* (Marron et al. 1997), *IDDM7* (Nistico et al. 1996) and *IDDM13* (Van der Auwera et al. 1997) on chromosome 2q31 (Owerbach et al. 1997). This association has been extensively confirmed in both case-control and transmission-disequilibrium studies (Nistico et al. 1996; Marron et al. 1997; Owerbach et al. 1997; Van der Auwera et al. 1997; Awata et al. 1998; Badenhoop et al. 1999; Lowe et al. 2000; Klitz et al. 2002; Ma et al. 2002; Yung et al. 2002). Moreover, association of *CTLA4* polymorphisms with other autoimmune endocrinopathies, especially Graves' disease, has been repeatedly confirmed (Kotsa et al. 1997; Heward et al. 1999; Vaidya et al. 1999; Takara et al. 2000; Hadj Kacem et al. 2001; Nithiyananthan et al. 2002), though a lack of association has also been reported (Yanagawa et al. 1999; Ban et al. 2001).

CTLA-4 is a critical T-cell surface receptor, the lack of which results in

uncontrolled T-cell mediated lymphoproliferative disease in the *CTLA4*$^{-/-}$ mouse (Tivol et al. 1995; Waterhouse et al. 1995). CTLA-4 down regulates autoimmune response in experimental animal models (Perrin et al. 1996; Chambers et al. 1997; Luhder et al. 1998; Karandikar et al. 2000; Luhder et al. 2000) and blockade of CTLA-4 signaling in human cancer patients elicits autoimmune manifestations (Phan et al. 2003); its importance in immune regulation makes it a very strong functional candidate for association with autoimmune disease.

A recent study (Ueda et al. 2003) defined the extent of the associated haplotype based on a case-control study of 840 Graves' disease patients and on TDT in 3,600 T1D families as well as 210 Graves' disease families. It confirmed the previous finding (Marron et al. 2000) that the association is confined to *CTLA4* and does not extend to neighboring genes. For both diseases, the strongest association was found at +6230G>A, a newly described single-nucleotide polymorphism in the 3' flanking region, downstream of the previously known polyadenylation site. This polymorphism may be associated with altered levels of steady-state mRNA of a soluble *CTLA4* isoform (*sCTLA4*) through a mechanism that remains unclear but could involve allele-dependent differential alternative splicing and/or polyadenylation of *sCTLA4*. In Graves' disease all of the association could be attributed to +6230G>A, while the more rigorous TDT test on 3,600 T1D families failed to narrow the causative polymorphism exclusively to the 3'end and effects from the 5'end could not be excluded (Ueda et al. 2003). The aim of this study is to dissect the contribution of the 5' and the 3'-ends of *CTLA4* to T1D susceptibility with a combination of functional and genetic studies.

The T1D-associated *CTLA4* haplotype contains several polymorphisms in tight

linkage disequilibrium (LD), any one of which or a combination thereof could determine the functional effect (Figure 8). Of these, the only non-synonymous polymorphism is a +49A>G base substitution causing a Thr17Ala change in the signal peptide. We have recently shown that the predisposing Ala17 allele is incompletely glycosylated in the endoplasmic reticulum leading to retrograde transport of a portion of the molecules to the cytoplasm for degradation (Anjos et al. 2002). This ultimately results in less mature CTLA-4 (Ala17) at the cell-surface which may explain, in part, the reduced inhibitory function of CTLA-4 reported in individuals with the +49G allele (Kouki et al. 2000; Maurer et al. 2002). There are also several promoter polymorphisms in the LD block that could contribute to the T1D association through transcriptional effects on expression, independently of and in addition to the effect of the 3'flanking +6230G>A on splicing or RNA stability. In this study we have found that the most proximal promoter polymorphism -318C>T is highly associated with T1D and that this association is independent of the effect of the 3' SNP +6230G>A. We also show through *in vitro* and *in vivo* studies, an allele dependent effect of the promoter on *CTLA4* transcription levels.

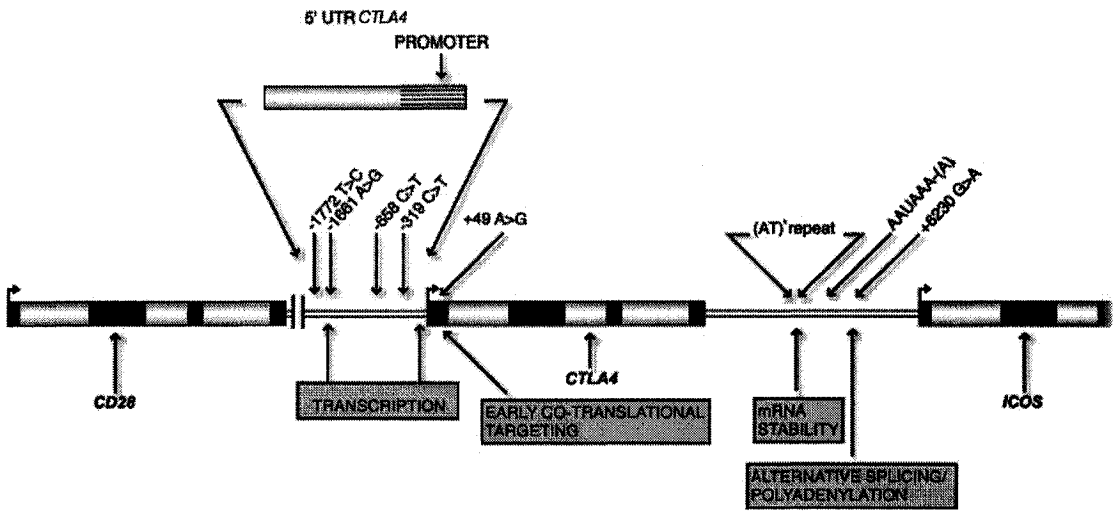


Figure 8*:* Summary of genetic and functional significance of SNPs within the *CTLA4* gene.

The –318 and +49 SNPs lie within a highly conserved region (overall sequence homology between mouse, rat and human, 79%). Evolutionary sequence conservation is the best *a priori* criterion for selecting regulatory sequences to evaluate for functional importance (Elgar 1996; Hardison et al. 1997). The shaded area in the 5'end is the highly conserved promoter region (~400 bp) which includes the –318 SNP but not the –658 SNP. Potential functional roles are shown in boxes and some are addressed in this study.

4.4 Materials and Methods

Subjects

Genomic DNA was obtained from 498 family trios, most of who attended the diabetic clinic at the Montreal Children's Hospital. Ethnic backgrounds were mostly mixed European, the largest single group being Quebec French-Canadian. All participants were diagnosed under the age of 18 and required insulin treatment continuously from the time of diagnosis. Informed consent was obtained for all individuals involved in this study, approved by the Research Ethics Boards of the hospitals where recruitment took place.

Genotyping

The +49A>G signal peptide polymorphism was genotyped by restriction digest and/or Taqman. For the manual genotyping: The forward primer, 5'ATGGCTTGCCTTGGATTTCA-3' and reverse primer, 5'-CTTTGCAG AAGACAGGGATG-3' were used to amplify a 110 bp fragment, followed by digestion by *Tse*I (New England Biolabs). The products were resolved on 10% PAGE. For genotyping using Taqman, the primers used for amplification were: forward: 5'-ATGGCTTGCCTTGGATTT-3', reverse: 5'-GCAGAAGACAGGG ATGAAGA-3'. The forward probe used was: 5'-TET-CCAGGTCCTGGTAGCC AGGTTC9-TAMRA-3' and the reverse: 5'-FAM-AGGTCCTGGCAGCCAGG TTC9-TAMRA-3'.The –1722(T>C), -1661(A>G), -658(C>T) and -318(C>T) SNPs in the *CTLA4* promoter and the +6230(G>A) SNP in the 3'-end were genotyped using single-

nucleotide primer extension with ddNTPs labeled with different fluorochromes corresponding to each allele. Primer and probe sequences for fluorescence polarization (FP) are listed in Table I. Unincorporated PCR primers and dNTPs were removed according to the PE AcycloPrime PCR Clean–Up protocol and primer extension was performed according to the PE AcycloPrime-FP protocol in a GeneAmp PCR system 9700 (MJ research, Waltham, Massachusetts). Final detection of the SNP by FP used the Criterion System with Analyst HT, (Molecular Devices, Sunnyvale, CA). The genotyping call rate was ≥97% for all SNPs and a Mendelian error rate of 0.001 was detected.

Statistics and data analysis

Functional data were compared with the 2-tailed, paired Student t-test. The paired test was used when comparing DNA and RNA from the same individual, or *in vitro* expression of each allelic constructs in each of the 5 experimental replications, to account for variation among experiments that affected both alleles equally.

The binomial distribution with a $P=0.05$ under the null hypothesis was used to evaluate the TDT data. The TDT is based on the distortion from the expected 50% transmission of alleles at a marker locus from heterozygous parents to affected children when the marker is both linked and associated with the disease (Spielman et al. 1993). Because of strong prior evidence, the +6230G/A polymorphism was evaluated at an unadjusted alpha level of 0.05. For the multiple testing of the four promoter SNPs, the Bonferroni-adjusted a level for each SNP was calculated as $1-(1-a)^{1/n}$, where a is the overall significance threshold (0.05). Thus, individual results were considered significant if $P \leq 0.0127$.

Table 1: Primers and probes used for fluorescence polarization (FP).

All sequences are in the 5'→3' direction.

SNP	Sense primer	Anti-sense primer	T_m for PCR (°C)	Probe (forward)	Probe (reverse)	Tm (°C) Ext.
-1722T>C (rs733618)	AGCCCTTTC TGACTTCCA CA	AAGCGCCAACA AGCAATAAC	62	ACTCTATCATGATCATGGG TTTAGCTG	ACACAGCAGTGGCAGG GACAG	55
-1661A>G (rs4553808)	AGCCCTTTC TGACTTCCA CA	AAGCGCCAACA AGCAATAAC	62.5	GCAGGAACATTTGTTTTTC ACTTTTT	CAGACTGGGCAACAGAG GTTTTT	55
-658C>T rs# n.a.	TCCTTCTGC AAAACCAGA GG	AAATCCATTTAG CATTTGGTTAAG A	60	AAACCAGAGGCAGCTTCTT TTC	AATCACAAGAAATAAAC TGAAAATAGGC	55
-318C>T (rs5742909)	GGGATTTAG GAGGACCC TTG	AGCCGTGGGTTT AGCTGTTA	58	AAGTCTCCACTTGTTATCC AGATCCT	TGAAACTGAAGCTTCAT GTTCACTTT-3'	55
+6230G>A (rs3087243)	TCATGAGTC AGCTTTGCA CC	CTGAGAAAGCA GGCGGTAG	59	GATTTCTTCACCACTATTT GGGATATAAC	AAGGACTGTTATGTCTG TGTTAACCCA 2.GGACTGCTATGTCTGT GTTAACCCA	55

All primers and probes were designed using primer3 software available online at: http://www-genome.wi.mit.edu/ genome_software/oth- er/primer3.html (Quandt et al. 1995).

Genotype odds ratios were computed using pseudocontrols generated from the untransmitted parental alleles assuming Hardy-Weinberg equilibrium. For haplotype analysis including haplotype TDT, the Family-Based Association Test (FBAT) software was used (http://www.biostat.harvard.edu/~fbat/fbat.htm).

Promoter constructs

Genomic DNA from a heterozygote -318C>T was used to amplify a 662 bp fragment of the *CTLA4* promoter. The forward primer used was 5'-GCTGTCAAG GGACCATTAGAAGGA-3' and the reverse primer, 5'-GGCTTTATGGGAGCGG TGTTG–3'. PCR product was T-cloned into pCR2.1 (Invitrogen). Clones representing each allele were screened by PCR and *Mse*I (NEB) restriction digestion as previously described (Heward et al. 1998). Clones were sequenced and ligated into pGL3-Basic at *Kpn*I and *Xba*I sites (NEB). The constructs are denoted herein as pCTLAT^{-318} and pCTLAC^{-318}.

Transfections and Luciferase reporter assays

Jurkat T cells were maintained in RPMI 1640 supplemented with 10% FBS, 1 mM Sodium Pyruvate, 1 mM MEM and 100 units/ml penicillin/streptomycin at 37 °C, 5% CO_2 (cell-culture reagents from Invitrogen). 5 X 10^5 cells were transfected in triplicates with 2 μg of *CTLA4* promoter constructs and 10 ng of the transfection control renilla luciferase, pRL-CMV (Promega) using Lipofectamine 2000 as per manufacturer's instructions. All transfections were performed with the same plasmid DNA maxiprep preparation. The cells were activated with 1 μg/ml PHA and 50 ng/ml PMA (Sigma) 4

hours following transfection for 48 hours.. Although PHA and PMA are physiologically irrelevant stimuli they were found to be specifically required for constitutive expression of Renilla-Luciferase under the control of the CMV promoter in the Jurkat cell line(Schifferli 1996). Luciferase assaying was performed using Promega's Dual Luciferase Reporter Assay following manufacturer's instructions. Assay was performed on 96-well plates and detection by a multi-sample, plate reading luminometer MicroLumat Plus (EG&Berthold, Germany).

DNA and RNA samples

Genomic DNA was extracted from whole blood by standard phenol-chloroform methods. RNA from whole blood was extracted using QIAamp RNA Blood Mini kit (Qiagen, Germany). PBMCs were obtained from whole blood by standard Ficoll gradient and 5-8x10^6 cells were cultured in RPMI 1640 supplemented with 5% heat-inactivated FBS, 1 mM sodium pyruvate, 2 μM β-mercaptoethanol and activated with 5 μg/ml Concanavalin (Sigma) and 10 nM PMA (Sigma) overnight. RNA was isolated from PBMCs using the Rneazy Mini Kit (Qiagen, Germany) with on column Dnase treatment.

PCR, RT-PCR and quantification of allelic differences by dideoxy sequencing

Reverse transcription was performed under standard conditions using random primers and 250-1,000 ng of total RNA template. Primers for exon 1 of *CTLA4* (includes the +49A>G polymorphism) in DNA were: forward, 5'ATGGCTTGCCTTG GATTTCA-3' and reverse, 5'-CTTT GCAGAAGACAGG GATG-3'. For RT-PCR an antisense primer, 5'-GCTGGGCCACGT GCATTGCT-3' spanning exons 1 and 2 was

used. These primers do not distinguish between the soluble and full-length splicing isoforms. PCR amplicons were purified by the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and were sequenced.

Allele-specific quantification by sequencing was done by comparing ratios of peak heights of each allele in one single reaction in matching RNA and DNA samples from heterozygous individuals amplified, respectively, by RT-PCR or PCR (Figure 9A). The peak height at a particular base is proportional to the amount present in the sample. This methodology has been validated elsewhere (Qiu et al. 2003) but we tested the principle by mixing known quantities of PCR derived from genomic DNA of individuals homozygous for each allele and sequencing the mixtures. We used the +49A>G SNP as a proxy for the –318 and +6230 SNPs due to the strong LD between all three loci (D'~0.8); hence a T allele at –318 is in LD with an A at +49 and a G at +6230, a haplotype that was also found by Johnson et. al. (Johnson et al. 2001). All haplotypes of individuals used in the allelic imbalance studies were confirmed with genotypes of family members. The proportions of +49A/A:+49G/G mixed were: 90:10, 70:30, 50:50, 30:70 and 10:90 (Figure 9B). The observed ratios of A49 over G49 were plotted against the expected and the correlation between observed and expected was strikingly good (r^2 =0.95 and 0.98 for A and G, respectively), thereby validating the method for allelic-specific quantification (Figure 2A). Ratios of peak height corresponding to each allele in DNA samples ranged between 0.8-1.2. In the corresponding RNA samples, an allelic imbalance was said to exist when this ratio deviated from the 0.8-1.2 range. All RNA and DNA peak heights are reported as a ratio of (peak height at allele A)/(average peak height of 3 "A" bases upstream and 3 "A" bases downstream of +49A) over (peak height at allele G)/(average

peak height of 3 "G" bases upstream and 3 "G" bases downstream of +49G). We refer to this ratio as a standardized ratio.

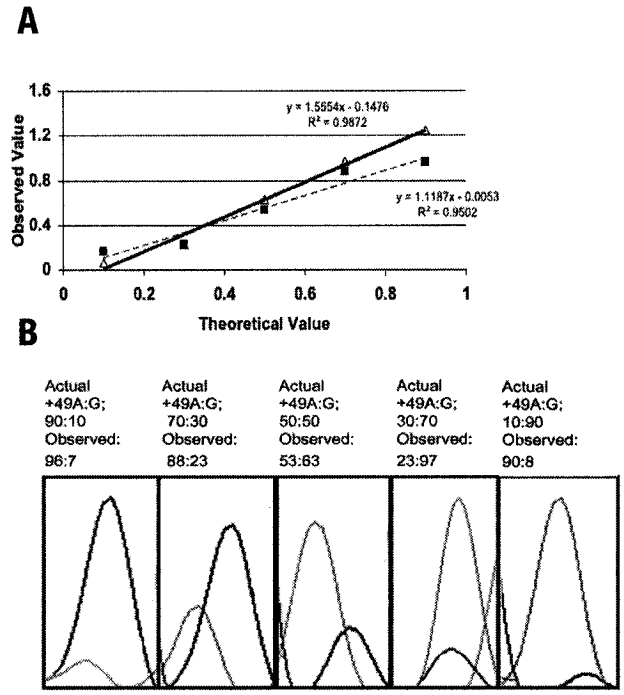


Figure 9: Allelic imbalance in the RNA is quantified by sequencing

To quantify differential expression of the two alleles, RT-PCR product is sequenced and the ratio of the peak heights of the two allelic bases is used to estimate relative abundance in steady-state mRNA.

A. The method was validated by sequencing mixtures, in defined proportions, of PCR product derived from homozygotes for each +49 allele. The observed relative amounts of each allele are plotted against the known amount. (The dashed line is the +49A allele and the black line the +49G allele). The observed values in the sequencing results correlate strikingly well with those expected from the mixture.

B. The peak height at the SNP is proportional to the amount sample(Qiu et al. 2003) (black chromatogram, +49A and gray, +49G). Each value at the SNP base peak is corrected by dividing it by the average peak height of adjacent identical bases defined throughout the text as standardized. This accounts for variations within individual sequencing reactions independent of the relative amount of allele.

Although peak heights for the same base vary within the same sequence, the nucleotide peak in any given position in relation to its same-base neighbors is remarkably stable across samples run in the same sequencing reaction.

4.5 Results

Genetic studies: The -318 promoter SNP is associated with T1D independently of the 3' +6230 SNP.

The recent study by Ueda et. al. (Ueda et al. 2003) finds a strong association between a 3'flanking SNP, +6230G>A and Graves' disease that cannot be explained by another SNP in the haplotype thus strongly implicating this SNP as the causative variant in *CTLA4*. However, in the same study, the association in T1D is much weaker and an effect at the 5'end cannot be ruled out (Ueda et al. 2003). To separate these effects on genetic grounds and narrow down the potential functional SNPs, we genotyped five SNPs contained within the *CTLA4* haplotype block, previously defined as haplotype tag SNP's (htSNPs) (Johnson et al. 2001). htSNPs can be used to probe the haplotype diversity of an LD block without genotyping all polymorphisms within a block. In addition, we genotyped the +6230G>A 3' SNP in 498 diabetic family trios (Figure 8). We confirm association of +6230G>A with Type 1 Diabetes in our dataset. We find significant over-transmission of the +6230G allele by the transmission disquilibrium test (TDT) (217 /168, *P=0.013)* (Table 2). However, the transmission distortion is stronger in the promoter where the C allele is significantly overtransmitted (97/58, *P=0.0017*) (Table 2). The Mendelian error rate for the -318C>T polymorphism is 1/1100 and parental genotypes for both SNPs are in Hardy-Weinberg equilibrium. The odds ratios are: 1.29 (0.9-1.73, 95% CI) for the +6230 (AA *vs.* AG+GG) and 1.67 (1.19-2.39, 95% CI) for -318 (CC vs. CT+TT). The other three SNPs examined in the *CTLA4* 5' flanking region were not associated with T1D (Table 2) and the +49A>G is associated with T1D in our

dataset as reported in multiple studies (Nistico et al. 1996; Marron et al. 1997; Owerbach et al. 1997; Awata et al. 1998) likely through its LD with the +6230G>A, as proposed by Ueda et. al., (Ueda et al. 2003). To explore the full haplotype diversity at *CTLA4*, we genotyped the previously reported htSNPs (Johnson et al. 2001) required to define the five haplotypes that represent 97% of European chromosomes. On haplotype-based TDT analysis, only one haplotype is significantly over-transmitted after correction for multiple tests and it contains C at –318C and G at +6230G (*P=0.0021*). None of the 5-SNP haplotypes containing the +6230A alleles approach significance although they are frequent and account for almost 43% of chromosomes (Table 3). Two-SNP TDT analysis on the promoter (-318C>T) and 3'SNP (+6230G>A) reveals significant over-transmission of the –318C;+6230G haplotype (*P=0.00028*) (Table 4) and although the –318C;+6230A haplotype is very frequent its under-transmission does not reach significance. The only haplotype significantly under-transmitted (even when corrected for multiple tests) is the–318T; +6230G haplotype (*P*=0.012), despite the considerably smaller number.

Examination of the haplotype frequencies suggests that the +6230G>A (which accounts for all other previously reported *CTLA4*-T1D associations (Ueda et al. 2003) cannot account for the transmission distortion at -318, as the over-transmitted G allele at +6230 is in near-perfect Linkage Disequilibrium with the under-transmitted -318T allele (Johnson et al. 2001) (and D'>0.9, confirmed in our dataset). One possibility is that they function independently. We examined this possibility by performing the conditional TDT analysis, whereby transmissions from parents heterozygous at one locus but homozygous for the other are counted. We perform the test on parents who are homozygous at +6230,

but heterozygous at the promoter locus (-318), hence; -318T; +6230G *vs.* -318C;+6230G and the very few -318T; +6230A *vs.* -318C;+6230A parents, thereby eliminating any effect the 3'flanking SNP may have on the probability of transmission. In this group, the C-allele remained significantly over-transmitted (53/27, $P=9x10^{-4}$) (Table 5). Similarly, when TDT is performed on parents heterozygous at +6230 but homozygous at -319, the G allele at +6230 is still significantly over-transmitted (145/99, $p=0.003$) (Table 5). In fact, removal of the -318T effect strengthens the association at +6230 ($P=0.013$ with the promoter *vs.* $P=0.003$ despite the smaller number).

The simplest explanation for this observation is that neither of the two SNPs is responsible for the functional effect, but that the C and G alleles mark a haplotype that contains a yet unexamined functional polymorphism. Alternatively, both polymorphisms function independently of each other. To search for a possible mechanism underlying a contribution of the –318C>T promoter SNP under the second hypothesis, we examined its effects on *CTLA4* transcription. The following sections address these issues.

Table 2: Transmission disequilibrium of SNPs within the *CTLA4* promoter and of +6230G>A in the 3' end.

SNP & Allele	T	UT	Total	P	Odds ratio	Allele frequencies
-1722 C (T>C)	50	66	116	0.13 (n.s)		C:0.13 T:0.87
-1661 A (A>G)	97	92	189	0.71 (n.s)		A: 0.83 G: 0.17
-658 C (C>T)	68	85	153	0.17 (n.s)		C: 0.9 T: 0.1
-318T (C>T)	58	97	155	0.0017**	1.70 (CI:1.19-2.39)	C: 0.91 T: 0.09
+49 G (A>G)	194	244	433	0.024	1.26	A: 0.62 G: 0.38
+6230 G (G>A)	217	168	385	0.013**	1.29 (CI:0.9-1.73)	A: 0.44 G: 0.57

** *P*-value calculated by the 2-tailed binomial distribution test, significant at the *0.05* level under the null hypothesis. T=transmitted, UT=untransmitted

Table 3: Haplotype tag SNPs (htSNPs) were genotyped in 498 T1D families

Association with T1D was tested using haplo-FBAT software. The LD block containing the entire *CTLA4* gene can be represented by 5 common haplotypes (frequency >5%) which account for >97% of all chromosomes observed.

Common haplotypes* -1661	-658	-318	+49	+6230	Frequency (%)	Z score**	*p*
A	C	C	G	G	34.6	3.069	*0.0021*
A	C	C	A	A	33.8	-1.141	*0.253 (n.s)*
A	T	C	A	A	10.0	-0.354	*0.723 (n.s.)*
G	C	T	A	G	8.6	-2.187	*0.029*
G	C	C	A	G	8.3	2.230	*0.026*

*The –1722 htSNP was not included in this analysis because it is redundant with the +49A>G htSNP as reported by (Johnson et al. 2001) (confirmed in our own dataset, such that a T allele at –1722 appears exclusively on +49G chromosome and both D'and r^2 values ~1). We chose the +49 SNP due to its higher heterozygosity rate.

**Z-score is positive when the haplotype is overtransmitted and negative when undertransmitted

Table 4 :Two SNP haplotype analysis in 498 T1D families

Haplotypes -318	+6230	Frequency (%)	Z-Score	*P*
C	G	47.2	3.636	***0.00028***
C	A	43.5	-2.002	*0.05*
T	G	9.0	-2.204	***0.012***

**Z-score is positive when the haplotype is overtransmitted and negative when undertransmitted. Haplotypes containing the G at +6230 and the T at –318 have been bolded.

Table 5: Testing independent effects of the 5' and 3' ends of *CTLA4* by the conditional TDT test.

SNP & allele	Transmitted	Untransmitted	Total	P
-318T	24	53	77	0.0009*
+6230 G	145	99	244	0.0032**

P-value calculated by a 2-tailed binomial distribution test.
*TDT done on heterozygous parents at –318 but homozygous at +6230.
**TDT performed on +6230G>A heterozygous parents who were homozygous at -318.

Functional Studies: The proximal *CTLA4* promoter region is highly conserved and the -318C allele is part of the lymphoid enhancer factor-1 (*LEF1*) transcription consensus site

The SNPs upstream of –318 were not associated with T1D in our population (Table 2), but it is still possible that an unknown SNP in LD with -318C>T might be responsible for the independent genetic effect. Unlike coding sequences, the bounds of promoter-enhancer regions are not clearly defined. In genes such as *CTLA4*, with highly conserved structure and function (Ling et al. 1999), evolutionary sequence conservation is a good indicator of functional importance (Elgar 1996; Hardison et al. 1997). Comparative sequence analysis between mouse, rat and human, showed that the -318C>T SNP lies within a 400bp region of high conservation (overall similarity, 79%), directly upstream of the +1 transcriptional start site (Figure 8). This sequence completely encompasses the 335 bp promoter shown by Lindsten *et al.* to be sufficient for regulated in vitro expression of *CTLA4* (Lindsten et al. 1993). None of the other 5' flanking SNPs mapped to a highly conserved sequence.

In addition, we analyzed whether any of the promoter SNPs altered transcription factor (TF) binding sites by submitting 20-30 bases flanking each SNP to MatInspector (http://www.genomatix.de/cgi-bin/matinspector_prof/mat_fam.pl (Quandt et al. 1995)). Two potential TF elements were detected although only one mapped to a phylogenetically conserved region, the lymphocyte enhancer factor-1 (*LEF1*) that binds the consensus sequence $^{-315}$ttCAAAG^{-323} (-318C bolded, the core sequence is underlined and the TF consensus sequence is in uppercase). The –318T allele did not correspond to the binding element of any known transcription factors. Although SNP –1661 also

modified an *MEF2* response element, it lies outside the 5' conserved region and is not associated with T1D. Based on this data, we chose to focus our efforts on the promoter SNP most likely to play a functional role in regulation, the -318C>T SNP.

Functional studies: Transcriptional effects of the -318T allele at the *CTLA4* promoter

We investigated the effect of the C-to-T base change in the promoter of *CTLA4* by testing its ability to drive the expression of the luciferase reporter gene. Constructs containing 662 bp of the *CTLA4* promoter region including the SNP at -318bp were expressed in the Jurkat T cell line. Expression of luciferase was inducible in Jurkat cells by the addition of PHA and PMA and standardized against the transfection control plasmid Renilla, pRL. The results are depicted in Figure 3, and are expressed as relative Luciferase units (luciferase/pRL*100). There was no detectable difference between pCTLA-T^{-318} and pCTLA-C^{-318} in Jurkat cells when not activated. Upon activation by PHA and PMA, luciferase expression increased by 3 and 4 fold for pCTLA-C^{-318} and pCTLA-T^{-318} respectively (Figure 10). The relative luciferase levels were significantly higher in pCTLA-T^{-318} transfections than pCTLA-C^{-318}; 21.8 ± 2.27 versus 15.8 ± 2.05 ($P=0.005$, paired t-test, n=5). All transfections were performed in triplicates and the results shown are the means of 5 independent experiments. These results confirm a previous report where higher luciferase expression was found in constructs containing the T allele in the promoter (Wang et al. 2002). We confirm similar allelic differences in a larger genomic context; 662 bp of the promoter versus 329 bp.

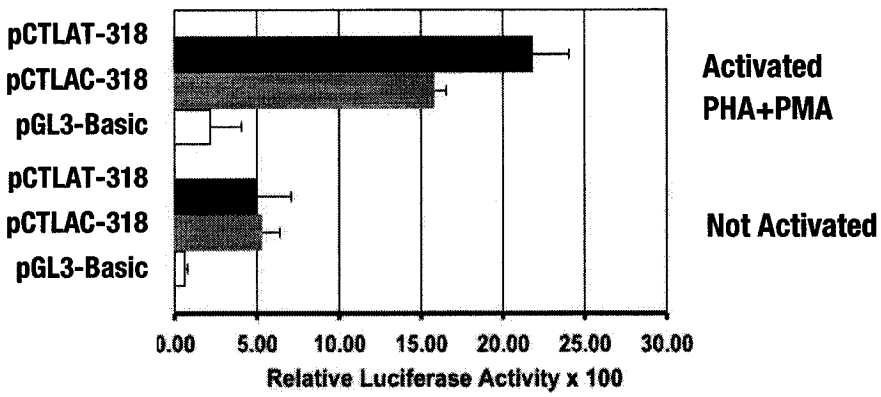


Figure 10: Higher *CTLA4* expression from the promoter –318 T allele in vitro.
Luciferase reporter plasmids driven by the *CTLA4* 5' flanking sequence, each corresponding to one allele of the –318 SNP; pCTLA-T^{-319} and pCTLA-C^{-318}, were expressed in the Jurkat T cell line. pCTLA-T^{-318} induces a ~4.4 fold increase in luciferase expression, reported as 21.8 ± 2.27, versus a 3 fold induction under the control of pCTLA-C^{-318}, 15.8 ± 1.08 (*$P=0.006$, paired t-test).

The next section confirms that the differential transcriptional regulation of *CTLA4* promoter alleles *in vitro* was reproducible in human lymphocytes *in vivo*.

The -318T allele of the *CTLA4* promoter is associated with higher expression in human lymphocytes.

In order to determine whether this *in vitro* effect reflects *in vivo* transcription, individuals heterozygous at the promoter -318(C/T) and at the signal peptide (+49) (A/G) were selected (n=8). Due to tight linkage disequilibrium, the haplotypes in the vast majority of such individuals are -318T;+49**A**;+6230G/-319C;+49**G**;+6230G, allowing us to use the transcribed +49A>G polymorphism to distinguish transcripts derived from the two promoter alleles. This was confirmed in each individual by examining genotypes of family members. We quantified relative mRNA expression levels by comparing the ratio of +49 A/G in RNA with that of the corresponding +49 A/G DNA ratio. This was accomplished by sequencing of an RT-PCR product (as described in the Materials and Methods and Figure 9). When steady-state RNA levels of +49A and +49G were compared within the same individuals, statistically significant higher expression was detected from the -318T;+49**A**;+6230G haplotype in activated PBMCs: 0.39± 0.006 vs. 0.29 ± 0.005 (± SEM) for +49A and +49G, respectively (P=3.34 x 10^{-9}, paired t-test, Figure 11A, lower graph). A similar difference was also detected in activated PBMCs and is depicted in Figure 11 (0.39 ± 0.004 *vs.* 0.29 ± 0.012 (± SEM) for +49A and +49G respectively, P=1.3x10^{-4}, paired t-test), almost precisely mirroring the *in vitro* effect shown in the previous section. Whereas the +49 A/G ratio in DNA was 1.01± 0.021

SEM, in the RNA of activated PBMCs it was 1.32 ± 0.05 SEM and 1.36 ± 0.02 SEM in inactivated cells, clearly indicating increased transcription from the -319T; +49**A**; +6230G haplotype. Because of the LDs described above, all -318C>T and 49A>G heterozygous individuals examined here were G/G homozygous at the 6230 3' polymorphism and, therefore, these results cannot be attributed to possible effects of +6230 on the soluble isoform.

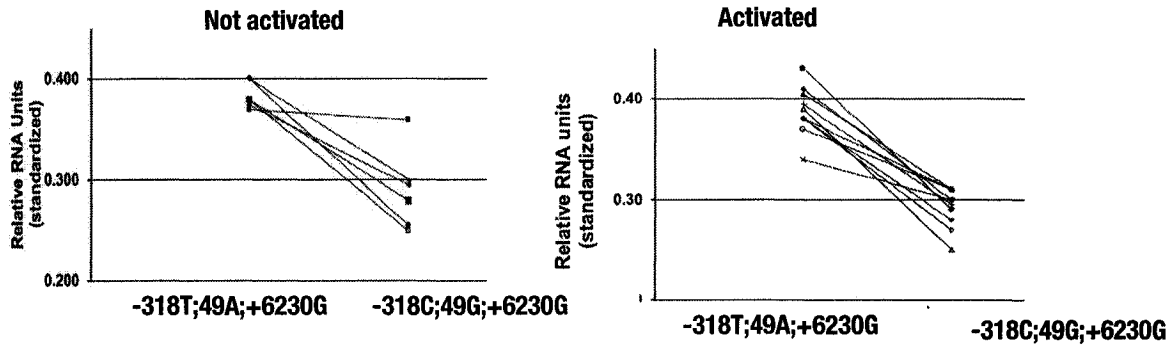


Figure 11: Allelic imbalance at the *CTLA4* promoter locus in human lymphocytes ex-vivo correlates with in vitro data; the -318T promoter allele induces higher *CTLA4* mRNA levels

PBMCs from individuals heterozygous at the –318 locus and the +49 locus were used to assess allelic variation at the promoter (Figure 2 and methods) There is consistently higher RNA levels due to the T alleles than the C alleles (P=1.3 x 10^{-4}, n=8) in inactivated PBMCs. Similar data is obtained in activated PBMCs. Each data point represents an average of at least two independent experiments per individual and the lines connect data for each individual.

4.6 Discussion

Due to the haplotype structure of the human genome, determination of the polymorphism(s) responsible for genetic association of a haplotype with a phenotypic trait must often rely on complementing genetic studies with functional ones. Ultimately, however, the only way to distinguish between independent functional effects of the -318 and +6230 SNPs vs. the functional effect of some other polymorphism for which they are merely markers can be done only by a systematic association studies of all polymorphisms in the LD block. This was actually done by Ueda et al. in their Graves' disease dataset (Ueda et al. 2003) and failed to identify a polymorphism with an effect stronger than that of +6230, after a resequencing effort that is unlikely to have missed a SNP responsible for the effect (Johnson et al. 2001; Ueda et al. 2003).

The alternative explanation is that -318C>T polymorphism has an independent functional effect. Both the genetic and the functional effects described in the Results section are independent of any functional effect the 3'flanking SNP may have because both were observed in individuals homozygous at +6230. In this study we confirm the ~30% increase in expression at the -318T allele described in previous *in vitro* studies and directly demonstrate that this expression difference is reproducible in human lymphocytes, the cells involved in diabetes autoimmunity. It is conceivable that a single base substitution in a critical region for transcription factor binding, such as that of the *LEF1* factor, may affect transcriptional efficiency, the regulation of which is an important determinant in the pathogenesis of disease. Such is the case of the *TNF* promoter, where

allelic specific binding of the transcription factor *OCT1* determines survival outcome in malaria infections (Knight et al. 1999).

A study comparing lymphocyte expression levels in 44 individuals heterozygous at the promoter vs. 186 homozygotes (C/C) found an effect in a small number of myasthenia gravis patients but not in normal controls or in patients with multiple sclerosis (Ligers et al. 2001). This inconsistent result is not surprising, as between-individuals comparisons may easily miss subtle differences in the noise introduced by inter-subject variation due to genetic loci in *trans* and the individual's immune and other environmental experience. Our approach solves the problem by eliminating biological (as opposed to technical) noise by comparing transcriptional effects within the same cell sample from the same heterozygous individual and is becoming the standard for evaluating the subtle transcriptional effects responsible for complex phenotypes (Cowles et al. 2002; Yan et al. 2002).

The assumption that a disease-marker association must be reduced to LD with a single functional polymorphism is often implicit in discussions of complex traits but it is quite plausible that more than one SNP may be contributing functional effects, which, each in isolation might not have been substantial enough to detect. Ueda et al. (Ueda et al. 2003) addressed this possibility and, in the case of T1D, could not rule out an effect from the 5' end of *CTLA4*. They did not examine the -318C>T polymorphism.

The possibility of two independent effects at the *CTLA4* locus suggests yet another complexity that must be considered when dissecting association haplotypes: In the presence of two functional polymorphisms on the same LD block it is possible that the predisposing allele of one is in LD with the protective allele of the other, in which

case the genetic effect of the two SNPs would be blunted by an increase in the frequency of haplotypes that carry one predisposing and one protective allele. In the case of *CTLA4*, under the two-effect hypothesis, the CG haplotype exerts its genetic effect against a background of partially predisposing TG and CA haplotypes, because the nonpredisposing TA is virtually absent as a result of the LD. This diminishes the genetic effect of each SNP examined separately.

Although -318C>T appears to confer higher risk than +6230G>A (OR=1.67 vs. 1.29, respectively), it is overshadowed in the haplotype association because of the low frequency of the minor T allele. Nevertheless, for individuals carrying that allele, it must be taken into account in any future algorithms used to determine T1D risk from a panel of genotypes.

The recent association of the +6230G>A 3' flanking SNP with Graves' disease and Type 1 Diabetes reported in the largest and most comprehensive SNP mapping and association study to date raises some important issues (Ueda et al. 2003). The authors find increased susceptibility to autoimmune disease associated with the +6230 G allele. They propose that the predisposing +6230 G allele mediates the lower expression of a soluble *CTLA4* isoform, *sCTLA4*. A molecular mechanism to explain such effects remains unclear. Although Ueda et al. (Ueda et al. 2003) show evidence of the presence of the 3'SNP in Northern blot transcripts, no clear molecular model was proposed. If an alternative polyadenylation site indeed exists that includes +6230G>A in the mRNA, this SNP might affect RNA stability and/or the efficiency of the alternative splicing that creates *sCTLA4*. The possibility of an effect through a 3' enhancer cannot be eliminated either, as we are not aware of any conclusive evidence that *sCTLA4* is transcribed from

the same promoter as the full-length form.

Ueda et. al., reported highest association in the 3'UTR of *CTLA4* with Graves' disease and Autoimmune Thyroiditis after testing 108 SNPs. They subsequently tested only 9 SNPs in the 3,600 T1D families, *not including* the –318 promoter SNP (Ueda et al. 2003). Its effect partially explains the weaker association they found in the 3'end (odds ratio = 1.15). Our data clearly point to independent effects of the promoter *and* 3'end. After accounting for the promoter SNP, the odds ratio at 3' becomes substantially stronger. It is worth noting that Ueda et al. found no association with the –318C>T in Graves' disease, a disorder of humoral immunity whose pathogenesis is likely quite different from the T-cell mediated destruction of the b-cells involved in T1D (Ueda et al. 2003). If the role of the soluble form is more important in humoral immune dysregulation, it may overshadow the promoter effect. This scenario also provides an attractive explanation of the much weaker effect of +6230G>A in T1D than in Graves' (Ueda et al. 2003).

As in all disease association studies with candidate polymorphisms, our findings will need to be replicated in independent DNA sets. A poor reproducibility record in association reports is the result of multiple hypothesis testing without statistical correction. Unlike those reports, this study focused on only four SNPs following genetic evidence of an effect from the 5' end of the gene in addition to the 3' end.

To summarize, taken together with previous work, the results presented here raise the possibility of a model in which disease susceptibility is modulated through non-mutually exclusive allele-specific transcriptional and, possibly, post-translational mechanisms. These new, *Homo-Sapiens* specific alleles may have arisen in evolution

because of the need for better protection of humans against infection, an advantage that would have easily counterbalanced a small increase in the risk of autoimmunity.

4.7 Footnotes

Figure 1 was reproduced in part and reprinted from "Mechanisms of Genetic Susceptibility to Type 1 Diabetes: beyond the HLA", Vol 81, No 3, pp 187-195, Copyright (2004), with written permission from Elsevier.

5.0 CHAPTER FIVE

Allelic effects on gene regulation at the autoimmunity-predisposing *CTLA4* locus: a re-evaluation of the 3' +6230G>A polymorphism.

Suzana Anjos, Wei Shao, Luc Marchand, Constantin Polychronakos

Accepted for publication in Genes and Immunity, Dec 2004

5.1 Contributions of Authors

Suzana M. Anjos : Collection and extraction of RNA samples, experimental design and strategy, hypothesis development, supervision and training of summer student, manuscript preparation.
Wei Shao : Experimental data acquisition in the allelic imbalance assay.
Luc Marchand : Allelic imbalance assay design and experimental data acquisition with respect to the *ICOS* gene data.
Constantin Polychronakos : Research Director, hypothesis development, manuscript editing and preparation.

In light of our findings in the previous section, and the reported correlation between alleles at the 3' +6230G>A and expression levels of the soluble isoform of *CTLA4* (Ueda et al. 2003) we undertook this study in order to address the allelic variation at the 3'SNP in full-length and soluble *CTLA4* isoforms, using (1) more reliable methodologies, and (2) accounting for haplotypes given our compelling evidence of a functional polymorphism at the –318C>T promoter SNP.
We also asked if this 3'flanking SNP altered the expression of the downstream *ICOS* gene and if it was part of the most abundant *CTLA4* transcript.

5.2 Abstract

Genetic variation at a linkage disequilibrium block encompassing the *CTLA4* gene influences susceptibility to autoimmunity but identifying the polymorphism(s) responsible for this effect has been challenging. Recently, a single-nucleotide polymorphism (SNP) located 3' to the known polyadenylation site of *CTLA4* (+6230G>A) and strongly associated with autoimmune disease, was reported to regulate levels of soluble *CTLA4* isoform (*sCTLA4*) but not the full-length isoform. The purpose of the present study is to define the mechanistic effect of the 3'SNP on the isoforms of *CTLA4* (alternative splicing vs. polyadenylation vs. effects on RNA stability). However, using allele-specific single nucleotide primer extension we found no difference between mRNA transcripts derived from either +6230G>A allele in 11 heterozygous individuals, in either of the two known *CTLA4* isoforms. We also found no effect of this polymorphism on *ICOS*, a putative downstream target. In addition, repeated attempts at 3' RACE were unsuccessful in amplifying any contiguous sequence past the known *CTLA4* polyadenylation site and no such sequence was found in the EST databases. We conclude that the mechanism of the observed association of the +6230 SNP with autoimmune disease remains to be determined but does not involve modulation of steady-state mRNA of any known *CTLA4* isoform.

5.3 Introduction

Genetic susceptibility to Type 1 Diabetes (T1D) is a complex trait. In addition to the Human Leukocyte Antigen (HLA) and the *INS-VNTR*, a locus encompassing a 300 kb region on chromosome 2q23 has been consistently reported to be associated with T1D. It includes several genes encoding immunoregulatory proteins, such as *CD28*, the cytotoxic T-lymphocyte antigen-4 (*CTLA4*) and the inducible costimulator (*ICOS*). Although any one of these genes could harbor the causative variant, most of the evidence provided by association studies in multiple ethnic groups (Nistico et al. 1996; Marron et al. 1997; Van der Auwera et al. 1997) as well as fine-mapping linkage disequilibrium (LD) studies (Marron et al. 2000; Ueda et al. 2003) strongly suggest that one or more polymorphisms within or around the *CTLA-4* gene region are responsible for the disease association. Associated single nucleotide polymorphisms (SNPs) include potentially functional variants such as a –318 C>T promoter SNP, a +49A>G base substitution resulting in a Threonine (+49A) to Alanine (+49G) amino acid change in the signal peptide, and an $(AT)^n$ repeat in the 3'UTR, all of which are in tight linkage disequilibrium (LD). *CTLA4* SNPs have also been associated with several autoimmune endocrinopathies, notably Graves' disease (Vaidya and Pearce 2004) (where the association is the strongest) and autoimmune thyroiditis (Vaidya and Pearce 2004). Since all the polymorphisms listed above are in tight LD, their effects are difficult to disentangle from one another and it has been unclear which is the disease-predisposing polymorphism.

Recently, fine mapping in and around the *CTLA4* gene region based on the testing of 108 SNPs in a Graves' disease case-control study, excluded *CD28* and narrowed down the association to the 3'region of *CTLA-4* and 5'end of *ICOS* (a 100kb region of LD)

(Ueda et al. 2003). Within this 3'region, one polymorphism, +6230G>A (located 3' to the known polyadenylation site of *CTLA4*) conferred the highest risk to Graves' disease and autoimmune thyroiditis (Ueda et al. 2003). However, in T1D, the +6230G>A SNP conferred considerably lower relative risk (OR=1.15) and additional effects at the 5'end of *CTLA-4* could not be ruled out. Of potentially functional single-nucleotide polymorphisms (SNPs) at the 5' end of *CTLA4*, regression analysis excluded the non-synonymous +49A>G but the most proximal promoter SNP, –318C>T, was not studied.

As functional explanation for the genetic effect, Ueda et. al. found lower *CTLA4* expression levels from the chromosome carrying the G allele at +6230G>A (Ueda et al. 2003). This effect was confined to an alternatively spliced form of *CTLA4*, lacking the transmembrane domain and encoding a secreted soluble protein (s*CTLA4*). The two alleles were equal in the full-length (*flCTLA4*) form. The mechanism remains unclear but may involve differential splicing efficiencies and/or alternative polyadenylaton (Ueda et al. 2003). In relation to the latter, Ueda et. al. showed some evidence for the existence of transcripts containing the +6230 nucleotide, situated nearly 0.5 kb downstream of the known polyadenylation site (Ueda et al. 2003). The molecular structure of those transcripts was not defined.

In a recent study of nearly 500 T1D Canadian family trios, we have confirmed the association of +6230A>G with T1D and also found a strong association with the most proximal promoter SNP, the –318C>T. This SNP is the only one of the *CTLA4* 5' flanking SNPs which is located in a phylogenetically conserved region and its under-transmitted T allele (frequency=0.09) is predicted to disrupt a response element for *TCF1/LEF1*, a pair of transcription factors important in lymphocyte development (Staal

and Clevers 2003). We found that this genetic effect is independent of the +6230 SNP and we confirmed a previous report (Wang et al. 2002) of increased *in vitro* transcription of the –318T allele (Anjos et al. 2004). This transcriptional effect is relevant *in vivo* as we showed that, in heterozygous subjects, there was increased *CTLA4* mRNA transcribed from the T-allele (Anjos et al. 2004) an effect we hypothesized may be mediated by the disruption of the predicted *TCF1/LEF1* site.

The purpose of this study was to 1) confirm the allelic effect *in cis* of the +6230G>A SNP on the abundance of the soluble isoform by single-nucleotide primer extension on RT-PCR products of heterozygous lymphocyte RNA samples; 2) separate this effect from that of the promoter and 3) explore possible mechanisms (effects on alternative splicing vs. effects on alternative polyadenylation, vs. differential transcription start site for soluble vs. full-length) underlying the effects (if any).

We report no observable effect of the +6230A>G SNP on the expression of either s*CTLA4* or fl*CTLA4*. There was also no appreciable effect on the expression of *ICOS*.

5.4 Materials and Methods

Sample collection

Lymphocyte samples were obtained from peripheral blood of individuals, most of whom were non-diabetic parents of children with T1D, followed at the diabetic clinic of the Montreal Children's Hospital. Informed consent was obtained for all individuals involved in this study, approved by the Research Ethics Boards of the hospitals where recruitment took place. Thymus samples were obtained from human fetal tissue recovered from pregnancy terminations for reasons other than maternal or fetal disease. Fetal age was determined by foot length and was 10-20 weeks. Human tissue collection was approved by local ethics committees and informed consent obtained in each case. Tissues were flash frozen and stored at -70°C until nucleic acid extraction. Extraction of nucleic acids has been described elsewhere (McCann et al. 2001).

Genomic DNA was extracted from whole blood by standard phenol-chloroform methods. RNA from whole blood was extracted using QIAamp RNA Blood Mini kit (Qiagen, Germany). PBMCs were obtained from whole blood by standard Ficoll gradient and 5-8x10^6 cells were cultured in RPMI 1640 supplemented with 5% heat-inactivated FBS, 1 mM sodium pyruvate, 2μM β-mercaptoethanol. Total RNA was isolated from PBMCs using the Rneazy Mini Kit (Qiagen, Germany) with on column Dnase treatment or using Trizol reagent.

PCR, RT-PCR, and 3'Rapid Amplification of cDNA ends (3' RACE)

Reverse transcription was performed under standard conditions using random primers and 250-1,000 ng of total RNA template. Primers for exon 1 of *CTLA4* (includes the +49A>G polymorphism) in DNA were: forward, 5'ATGGCTTGCCTTG GATTTCA-3' (C1 in Figure 12) and reverse, 5'-CTTTGCAGAAGACAGGGATG-3' (C-1, Figure 12) and PCR conditions have been described elsewhere (Anjos et al. 2004). Primers used to distinguish between the soluble and full-length *CTLA4* isoforms were located in, or overlapped, the alternatively spliced exon as indicated in Figure 12. Primer sequences have been described by Ueda et. al., (Ueda et al. 2003). PCR conditions for amplification of soluble *sCTLA4* are: 94°C, 5min, followed by 30 cycles of 94°C 20 sec; 56°C 20 sec and 72°C 20 sec and a final extension of 6 minutes at 72°C using Taq Polymerase (Invitrogen). The full-length *CTLA4* isoform was amplified using the same conditions as described for the *sCTLA4*, with an annealing temperature of 60°C.

Primers for amplification of *ICOS* 3'end (NCBI accession: NM_012092) are 5'-TC CCAGAGGCTGAAGTCACC-3' and 5'-GGGCATGCAGACAGGAAGTA-3' for amplification of DNA and 5'-AGAGCAGTGAACACAGCCAAA-3' and 5'-GGG CATGCAGACAGGAAGTA-3' for cDNA amplification under standard PCR conditions. In all cases, cDNA samples did not amplify detectable products by PCR in the absence of reverse transcriptase. The RT-PCR products were used to assess allelic imbalance in samples heterozygous for the rs1559931 SNP at the 3' UTR of *ICOS*.

We performed the reverse transcriptase step of 3'-RACE using an equal mixture of the GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT(C, A, and G) primers at 0.5μg/μl and 2μg's of RNA under standard RT conditions. For the 3'RACE-RT-PCR, the

reverse adapter primer 5'-GACTCGAGTCGACATCGA-3'(3'RACE-AP in Figure 12) was used in conjunction with a sense primer straddling part of the 5'untranslated region and exon 1 of *CTLA4,* 5'-ATGGCTTGCCTTGGATTTCA-3' (3'RACE-1, Figure 12) for the first round of nested PCR. 2μls of cDNA were added to a reaction containing 200ng each of primers, 25μmoles dNTPs in a final volume of 50μls. 0.5 units of Taq Polymerase were added following a hot start at 94°C 3 min; 72°C 3min. The cycling conditions were: 94°C 30 sec; 60°C 30 sec; 72°C 2 min 30 sec; 30 cycles, with a final extension at 72°C for 10min. In the second round of nested PCR a forward primer in the known 3'UTR of *CTLA4* (NCBI accession NM_005214), 5'-GTATGACCTTCTAG GAAGCTCCAGTTC-3'(3'RACE-2, Figure 12) was used along with the same reverse primer used in the first round of PCR. 1μl of PCR from the first round was used as template for the second round of the nested PCR under the following conditions: 95°C 3min after which 0.5 units of Taq polymerase are added and 30 cycles at 95°C 30sec, 56.8°C 30sec and 72°C 2min 30sec. A final extension of 10min at 72°C ends the PCR. Under these conditions a single fragment of the expected ~250bp is amplified.

Single nucleotide primer extension (SNuPe)

5μls of amplified samples were incubated with 1U shrimp alkaline phosphatase, SAP (Applied Biosystems, Foster City CA) and 2U of Exonuclease I (Applied Biosystems, Foster City CA) for 1 hour at 37°C followed by 15 min at 75°C prior to primer extension reactions. Primer extension was carried out with the SNaPshot Multiplex Kit (PE Applied Biosystems, Foster City CA) with extension primers,

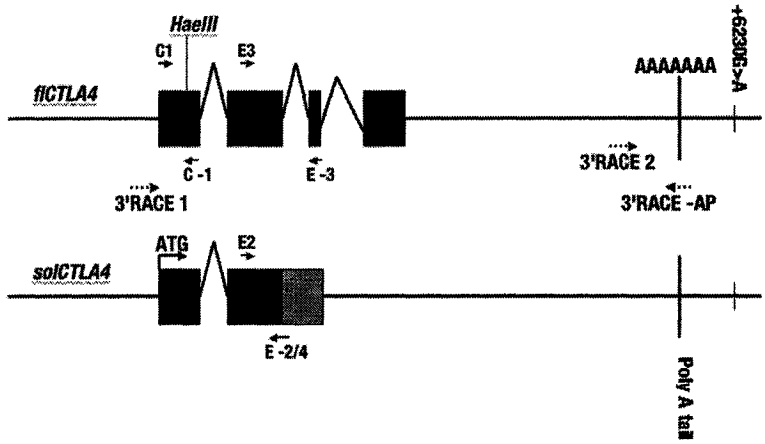


Figure 12: Schematic representation of the primers used in this study and their location in the *CTLA4* gene region.

C1 and C -1 are the sense and antisense primers, respectively used in the amplification of PCR derived from DNA in *CTLA4*. E2 and E -2/4 are the primers used for the RT-PCR of the soluble isoform, while the E3 and E -3 are those used in the RT-PCR of the full-length isoform. The 3'RACE-1 and 3'RACE-2 are the forward primers used in the first, and second rounds of nested RT-PCR, respectively and 3'RACE-AP is the antisense "adaptor" primer used in both rounds of the nested RT-PCR. Primer sequences are given in the text. Black boxes denote exons but not drawn to scale

5'-AGGCTCAGCTGAACCTGACT-3' flanking *CTLA4* 49A>G (rs# 231775), 5'-TTCCAGCTAAGAATAATCATATTCCA-3' flanking *ICOS* 3'SNP +2042G>A (rs# 1559931) and 5'-GATTTCTTCACCACTATTTGGGATATAAC-3' flanking *CTLA4* +6230G>A (rs# 3087243).

Reactions were performed in a total volume of 10μl, containing 2μl of treated PCR product, 4.5μl of SnaPshot kit, 2.5μl of water and 2.5μM extension primer. Primer extension thermocycling conditions consisted of an initial step of 95°C 2min, followed by 25 cycles of 95°C for 5s, 43°C for 5s, and 60°C for 5s. Following primer extension, reaction products were treated with 1.0 U SAP for 1 hour at 37°C and then for 15 min at 75°C. Aliquots of 1μl SnaPshot reaction product were combined with 9μl of Hi-Di Formamide and loaded onto a 3100 DNA sequencer (Applied Biosystems, CA). We electrophoresed products on a 36-cm capillary array at 60°C and processed the data by using Genescan Analysis version 3.7 software (Applied Biosystems, CA). Peak heights representing allele-specific extended primers were determined using Genescan (Applied Biosystems, CA) and were used to give a ratio of allelic representation.

Allele Expression Assay

All individuals were initially genotyped in order to identify heterozygotes for the marker polymorphisms (Anjos et al. 2004). The transcribed signal peptide polymorphism, 49A>G, was used to distinguish RNA alleles. We confirmed phase with the –318C>T (rs#5742909) and +6230A>G loci by genotyping family members. We compared expression of alleles in three genotypes, all heterozygous at +49A>G: the common –318C;49G;+6230G vs. –318C;49A;+6230A, the less common –318T;49A;+6230G vs. –318C;49G;+6230G, and the rare –318C;+49A;+6230G vs. –318C;+49G;+6230G. In

order to achieve isoform-specific amplification we used different primers for DNA and RNA, which may introduce artifacts related to secondary structure during the SNuPe reaction (Bray et al. 2003). We corrected this by digesting our PCR and RT-PCR products with a restriction enzyme (*Hae III*, Invitrogen) in order to generate fragments of identical size prior to SNuPe. Results were identical whether this digestion step was included or omitted. Each cDNA sample was assayed alongside its corresponding heterozygote genomic DNA to ensure uniformity in reaction and assay conditions and samples were tested four times in this manner by using two separate RT reactions. The ratio of one allele over the other (A/G for *CTLA4* and T/C for *ICOS*) was the readout for the allelic imbalance assay. The ratios observed in genomic DNA represent 1:1 stoichiometry of the two alleles but where *cis-acting* variation exists mRNA originating from one chromosome will be expressed at a higher level than that from its sister chromosome and this will be detected by changes in the ratio of abundance of each allele. To account for differences in probe and fluorochrome efficiencies, the cDNA sample ratio was divided by the average genomic ratio for that assay batch (Bray et al. 2003).

We validated the assay by mixing, in predetermined proportions, PCR products derived from DNA homozygous for each allele and performing primer extension analysis on 17.5ng/μl of mixed DNA. In doing so, we show that there is excellent correlation between observed and expected allelic ratios (Figures 13A and B).

Statistical Analysis

DNA measurements in a given assay were used to estimate the technical variance of the assay and the lower limit of detectable allelic imbalance for that assay. Mean ± 2

SD for the DNA ratios ranged from 0.97 to 1.09 for the *CTLA4* allelic imbalance SNP and from 1.04 to 1.16 for *ICOS*. Therefore, RNA results outside these ranges were considered statistically significant. Differences in RNA ratios between different genotypes were evaluated pairwise by individual t-test, given the small number of comparisons made.

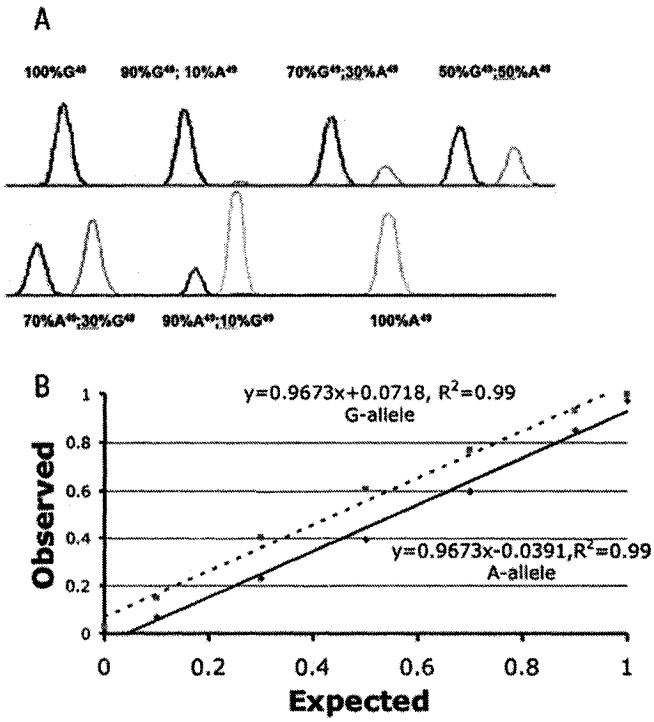


Figure 13: Validation of the SNuPe methodology.

(A) We validated the method by mixing, in defined proportions (10:90, 20:80 etc), PCR products derived from homozygotes for each allele at +49 and performing SNuPe. The black chromatogram represents the +49G allele and the grey the +49A allele. Peak heights are proportional to amount present in the sample and known relative amounts of samples mixed are plotted against the observed in (b) with strikingly good correlation between the two. In **(b)** the *x-axis* denotes the known mixed proportions of each allele plotted against the observed on the *y-axis* upon subjecting the mix to SNuPe, plotted for each allele.

5.5. Results

No allelic variation attributable to +6230G>A in expression of either *sCTLA4* or *flCTLA4*.

We first addressed results of a previous report that the level of *CTLA4* mRNA transcribed from the susceptible haplotype (+49G;+6230G) in unstimulated $CD4^+$ T cells from 3 heterozygous individuals was 2.0-2.5 fold lower than in those from the protective haplotype (+49A;+6230A) but only in *sCTLA4* and not in *flCTLA4* transcripts(Ueda et al. 2003).

To eliminate the effect of the -318C>T promoter SNP (not taken into account by Ueda et. al., (Ueda et al. 2003)), we selected individuals heterozygous at the marker SNP, +49A>G and at 6230G>A, but homozygous at –318C>T. Due to strong Linkage Disequilibrium (LD), in the vast majority of individuals with this combination of genotypes the A or G allele at the indicator +49A>G SNP marks, respectively, an A or G allele at +6230. We confirmed this to be the case in all the 11 samples we used, by genotyping family members. Thus, when referring to an allelic ratio of A/G, we are comparing the –318C;+49A;+6230A haplotype against the –318C;+49G;+6230G in steady-state mRNA isolated from resting PBMCs. The mean allelic ratio of A/G in fl*CTLA4* was 1.14 ± 0.02 (95% CI: 1.10-1.18), not statistically different from that of the s*CTLA4* transcript, 1.12 ± 0.04 (95% CI: 1.05-1.19). Although a pairwise t-test between *flCTLA4* and genomic DNA ratios is statistically significant ($p=0.017$) this difference is unlikely to be biologically significant (Table 11 and Figure 14). Individuals heterozygous at +49A>G but homozygous at +6230G>A (–318C;+49A;+6230G haplotype against –318C;+49G;+6230G) gave results statistically indistinguishable from the above (Table

11 and Figure 14). Thus our data suggest that there is no allelic variation in either s*CTLA4* or *flCTLA4* mRNA expression level attributable to the +6230G>A SNP.

Table 11: Summary of difference in allelic variation of *CTLA4* haplotypes and statistical significance as measured by the 2-tailed ttest in resting T cells.

Haplotype*	Number of heterozygotes	Average allelic ratio (A/G) ± SEM with 95%CI in fl*CTLA4*	Avg. allelic ratio (A/G) in s*CTLA4*	Avg allelic ratio (A/G) in genomic DNA (1:1)
-318C;**+49A**;+6230A *vs.* -318C;**+49G**;+6230G	11	1.14 ± 0.02 (1.10-1.18) **p= 0.0017* ***p=n.s*	1.12 ± 0.04 (1.05-1.19) *§p = n.s.*	1.03 ± 0.03 (1.01-1.05)
-318C;**+49A**;+6230G *vs.* -318C;**+49G**;+6230G	3	1.13 ± 0.02 (1.09-1.17) **p= n.s* ***p=0.02*	1.01 ± 0.003 (1.006-1.014) *§p = n.s.*	1.02 ± 0.004 (1.016-1.024)
-318T;**+49A**;+6230G *vs.* -318C;**+49G**;+6230G	4	1.35 ± 0.06 (1.23-1.47) **p= 0.01* ***p=n.s.*	1.21 ± 0.23 (1.17-1.25) *§p = 0.02*	1.03 ± 0.03 (1.00-1.06)

Marker SNP is bolded. *p-value calculated for *flCTLA4* vs. DNA; §: p value calculated for *sCTLA4* vs DNA and, ** p-value for *flCTLA4* vs *sCTLA4*
CI: Confidence interval

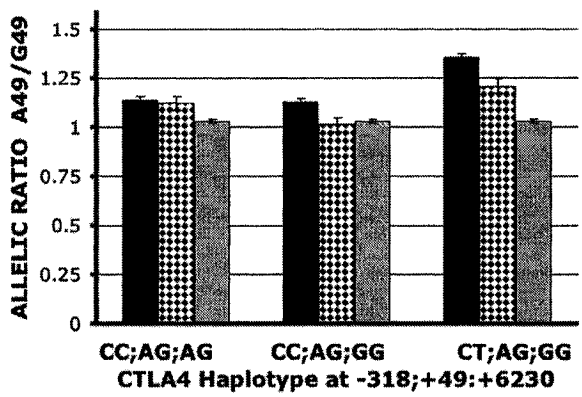


Figure 14: Relative allelic abundance by *CTLA4* haplotype shows significant mRNA expression differences due to the –318C>T promoter SNP and none at the +6230G>A SNP.

Solid bar represents the mean ± SEM A/G allelic ratio for *flCTLA4* transcripts, the dotted bar denotes the same in *sCTLA4* transcripts and the grey bar represents the mean±SEM A/G ratio in *CTLA4* DNA. Shown are the results from three independent trials of SNuPe on n=11 individuals for CC;AG:AG haplotypes, n=3 for CC;AG;GG haplotypes and n=4 for CT;AG;GG haplotypes. The individual means ± SEM are summarized in Table 11 along with the statistical analysis.

Allelic variation in individuals heterozygous at the –318C>T SNP but homozygous at the +6230G>A SNP

To confirm our previous results, and as an additional validation of the method, we also investigated allelic variation of mRNA levels in individuals heterozygous at the promoter -318C>T SNP and at the indicator +49A>G but homozygous at +6230G>A.

We had previously reported ~40% higher expression of the –318T allele vs. the –318C allele in mRNA *without* distinguishing between full-length and soluble *CTLA4* transcripts (Anjos et al. 2004). In this study, we confirm a significantly higher expression from the –318T;+49A;+6230G haplotype than –318C;+49G;+6230G in *flCTLA4* transcripts, where the mean allelic ratio of A/G is 1.36 ± 0.06 (±SEM) (95% CI, 1.24-1.47) *vs.* that in genomic DNA, 1.03 ± 0.01 (±SEM) (95% CI, 1.01-1.02, *p=0.01*) (Table 11, Figure 14) in resting T cells. Similar differences were observed in activated PBMCs, (data not shown). The *sCTLA4* transcript also showed statistically higher expression of the –318T;+49A;+6230G haplotype than the -318C;+49G;+6230G haplotypes in genomic DNA (*p=0.02*) (Table 11 and Figure 14). This suggests that the –318C>T SNP, which is located in a highly conserved sequence and is predicted *in silico,* to disrupt a transcription factor binding site, affects transcription of the *flCTLA4* and *sCTLA4* isoforms and is indirect evidence that the two transcripts originate from the same promoter. This effect is consistent with the independent T1D association with -318C>T but does not explain the genetic effect of +6230G>A (Anjos et al. 2004).

Characterization of the +6230G>A-containing transcript

In the study by Ueda et. al., there are two transcripts detected by Northern blot analysis in resting PBMCs: a 4.1kb and 2kb transcript (Ueda et al. 2003). Though it is possible that the larger transcript contains the 3'SNP this is not clear from this experiment. In the same study, 3'UTR-mapping with non-contiguous RT-PCR amplicons (Ueda et al. 2003) reveals that there are transcribed sequences 3' to the known polyadenylation site of *CTLA4*, one of which contains +6230, but not that these are part of an alternative polyadenylated *CTLA4* transcript that also contains the coding sequences. We searched the EST database (http://www.ncbi.nlm.nih.gov/BLAST) by probing the Blast algorithm with 310 bp of genomic sequence immediately 5' to the only known polyadenylation site (1621-1931bp in NM_005214) and obtained 20 hits of which 12 were matches to *CTLA4* sequences. All of these ESTs stopped at the known polyadenylation site and *none* included the +6230G>A SNP, nor any other downstream sequence. Thus, if a transcript exists that contains the +6230 SNP, its abundance must be very low, at least in the tissues and activation states covered by the EST libraries in the public databases. Our several attempts to find alternative transcripts generated by alternative polyadenylation by 3' RACE were unsuccessful in obtaining sequences of alternatively polyadenylated *CTLA4*.

No effect of +6230A>G on *ICOS* expression.

Having failed to find an effect of allelic variation at +6230 on *CTLA4*, we considered the possibility that transcriptional modulation of *ICOS* could explain its

association with Graves' disease and T1D. In an attempt to rule out allelic variation at *ICOS* due to remote regulatory effects from +6230, we examined expression of *ICOS* in thymus samples heterozygous for a marker SNP (rs1559931) in the 3'UTR. The A allele (frequency: 0.31) showed a 20-40% higher expression than the G allele (frequency: 0.69) in all samples, indicating that the polymorphism responsible for this allelic imbalance in *ICOS* is in perfect LD with rs1559931. Since the 3' UTR of *ICOS* is outside the LD block containing +6230, this effect could not be attributed to +6230. This is more directly demonstrated by the presence of this effect in all +6230 homozygotes (Figure 15). Thus, the attribution of the genetic association to this SNP reported by Ueda et. al., (Ueda et al. 2003) cannot be explained by allelic expression effects on either *ICOS* or *CTLA4*.

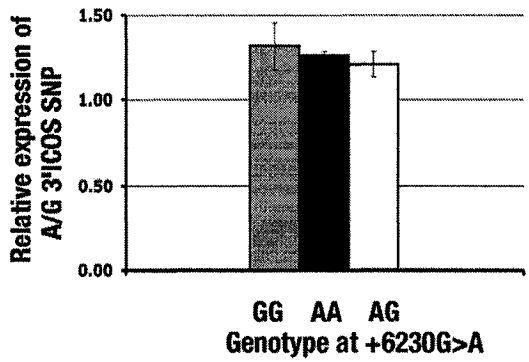


Figure 15: No correlation between genotype at *CTLA4* +6230 and allele specific expression of *ICOS* in the thymus.

Each data point represents the mean relative expression of alleles T/C of an *ICOS* 3'UTR SNP (rs1559931) quantified by SNuPe and is the result of 2 independent trials on each individual plotted according to +6230 genotype and includes at least 3 individuals per genotype at +6230 as shown.

5.6 Discussion

Genetic evidence supporting the association of *CTLA4* with Type 1 Diabetes is substantial (Anjos and Polychronakos 2004). Identification of the disease-causing polymorphism or haplotype has proven to be far more challenging, due to the extent of linkage disequilibrium in the region. To this end, detailed functional analysis is required. So far, several studies have addressed biological function of common polymorphisms of the *CTLA4* gene, particularly at the signal peptide (Thr17Ala). This disease-associated coding polymorphism has been linked with increased proliferation of T cells (Kouki et al. 2000) through a mechanism which may involve inefficient *N*-linked glycosylation leading to less mature CTLA-4 at the cell-surface (Anjos et al. 2002). Although coding polymorphisms are likely to have functional implications, there is an obvious bias towards their identification for practical reasons, as regulatory regions are often ill-defined and can extend well far beyond the open reading frame. Their effect cannot be underestimated as cis-acting variation is thought to account for 25% to 35% of the inter-individual differences which determine susceptibility to or protection from most complex diseases and have already been identified in several diseases (Buckland 2004; Pastinen and Hudson 2004). In the case of the *CTLA4* gene and autoimmune disease, compelling genetic evidence supports +6230G>A as the causative variant (Ueda et al. 2003). Less certain is the biological mechanism involved. Our data do not confirm the previously reported effect on *sCTLA4* mRNA, a discrepancy almost certainly due to technical considerations. Ueda et. al., (Ueda et al. 2003) used allele-specific restriction digestion with densitometric quantification of bands, a technique less sophisticated than fluorescent primer extension, which was not validated by assaying known allele mixtures as we did

in Figure 13b. It is worth noting that the results of the study given as the methodological reference for this method used by Ueda et. al., evaluating a *TNF*-alpha promoter polymorphism for which no allelic differences were found (Kaijzel et al. 2001), have been contradicted in another report (Wilson et al. 1997). In addition, Ueda et al. examined only three +6230G>A heterozygotes and their result may be a statistical artifact of small numbers. Our findings are based on a much larger number with many tightly concordant replicate determinations of each sample.

Thus, the functional basis for the *CTLA4* association with autoimmunity proposed by Ueda et al., needs re-evaluation. In addition to our failure to confirm the functional effect on which it was based, it lacks a clear molecular model to explain it. The SNP is not in a conserved region (more likely to harbor a functional domain (Hudson 2003; Pastinen and Hudson 2004)) and at present it is not clear whether it is indeed part of the transcript encoding either *CTLA4* splicing isoform. If, as the genetic data suggest, it is the functional variant mostly responsible for the *CTLA4* association with autoimmunity, it must influence biology via an alternative mechanism not addressed in our study, such as remote effects on transcription of some gene other than *CTLA4*, or it may affect *CTLA4* expression under a physiologic activation state of T-cells not elicited in the artificial *in vitro* activation we employed in this study. In our search of the EST database we did not find one single EST containing this SNP or any sequence upstream of the known polyadenylation end of *CTLA4*. Our data suggests that allelic variation exists at *CTLA4* as a result of the –318 promoter polymorphism and we previously showed that this genetic effect was independent of the 3'SNP (Anjos et al. 2004). However, because of the LD structure of the region, this effect cannot explain the +6230 association thus, the

possibility that +6230 may be merely a marker for a yet unexamined variant should be re-examined. An attractive alternative explanation would be that +6230 modifies some enhancer element in *ICOS*. We did, indeed, find evidence of allele-specific differential expression of this gene in heterozygotes for a 3'UTR SNP. However, it was present in +6230G>A homozygotes and thus cannot be due to this SNP and neither can it explain its genetic effect.

Allelic-specific expression has typically been studied in relatively uncommon situations like genomic imprinting (Giannoukakis et al. 1993; Reik and Walter 2001) but it is now accepted as a source of phenotypic variation amongst individuals which cannot be accounted for by protein-altering polymorphisms (Buckland 2004; Morley et al. 2004). This phenotypic variation is at the core of complex human diseases since it is thought that subtle alterations in gene expression modulate risk to disease progression. However, if such an effect underlies the association of the +6230 SNP with T1D, the nature of that effect remains elusive.

5.7 Acknowledgements

We would like to thank Dr. Cindy Goodyear for collection of the thymus samples and Marylène Rousseau for their technical expertise and help in the SNuPe assay.

6.0 CHAPTER SIX: Discussion

6. Discussion

In the time it took to complete the contents of this thesis, many advances took place in the field of Human Genetics, including the completion of the draft sequence of the Human Genome in 2000. In the four years since, technical advances in high-throughput genotyping and bioinformatics tools for data analysis, the emergence of Genomics as a discipline encompassing Regulatory Genetics, Comparative Genomics among others, have evolved in response. The efforts in narrowing down the causative functional variants leading to predisposition to complex disease have continued with a renewed optimism. All of these developments have enabled the work presented in this thesis and evolved in parallel.

The challenges of identifying the loci associated with Type 1 diabetes, and the genes mediating the genetic effects, have been facilitated with this new knowledge though the progress has been less spectacular than originally hoped for. For the *CTLA4* gene, the third candidate gene to be identified in T1D susceptibility, advances have been made in the last years, and the next sections discuss our contribution to this knowledge.

6.1 The *CTLA4* signal peptide

Our work on the autoimmunity-associated *CTLA4* signal peptide Ala17 allele was the first to address the molecular mechanism underlying the genetic linkage and association studies of *CTLA4* with autoimmune diseases and one of the first to address the post-targeting effects of the signal peptide.

6.1.1 Proposed Molecular Model

Our model arose from computer-simulated predictions of altered signal peptide endpoints, such as hydrophobicity, alpha-helix propensity and cross-species comparative sequence analysis, to a demonstrable effect on function at the level of the ER post-translational machinery. In this model, the signal peptide bearing the disease-associated Ala17 is translocation competent, is cleaved following insertion into the ER but the efficiency and/or extent of *N*-linked Glycosylation is altered. We showed that this *N*-linked Glycosylation intermediate is targeted for proteasome-mediated degradation suggesting that it would not make it to the cell-surface as a partially glycosylated intermediate (Chapter 3, (Anjos et al. 2002)). This study contributed to the accumulating evidence that signal peptides have functions beyond ER-targeting and membrane anchoring, and the amino acid sequence encoded by them influences events such as protein folding (Li et al. 1996), ER-maturation (Rutkowski et al. 2003) and may alter interactions of the folding protein with modifier enzymes (Chen et al. 2001).

Our cell-surface quantification experiments in *COS1* cells show significantly reduced levels at the cell-surface of the CTLA-4Ala17 allele. Our allele-specific co-transfection experiments in a T-cell model (Appendix I), though not of a quantitative nature, show differential localization across cell layers (Appendix I). As our work did not evaluate the function of the mature CTLA4Ala17 reaching the cell surface, we do not know whether any incompletely glycosylated molecules that may reach the surface membrane are fully functional CTLA-4 glycoproteins. However, it is clear that a significant amount of the aberrantly *N*-linked glycosylated intermediate is targeted for degradation by the ER quality control machinery as we have indirectly shown

6.2 Cell-surface targeting of CTLA-4

The studies preceding ours dealt with biological endpoints of CTLA-4 allele-dependent function such as proliferation, response to cytokines and cell surface quantification but did not address the molecular mechanism underlying these differences (Kouki et al. 2000; Maurer et al. 2002; Xu et al. 2002). In one study, differences between alleles at the cell surface were not detected in transfected Jurkats (Xu et al. 2002). This finding is not surprising given the very low cell surface expression of CTLA-4, coupled with its intricate cycling pattern. At any given time, in activated T cells there is merely a fraction of surface CTLA-4 barely detectable above background. In our attempts at quantifying cell surface differences of CTLA4 in stimulated PBMCs from individuals homozygous for signal peptide allele by flow cytometry (data not shown), we encountered difficulties associated with inter-individual comparisons. The only robust way to detect such subtle effects in vivo is comparison of the two alleles in the same heterozygous sample. However, at the protein level it is not possible to distinguish between the two allelic derived proteins because the signal peptide has been cleaved.

We devised a co-transfection assay in Jurkat T-cells, to simultaneously quantify total CTLA-4 from one signal peptide allele (detectable through YFP fluorescence), *vs.* cell surface CTLA-4 from the same allele (by immunofluorescence). In this experiment we detected a trend for less Ala17 at the cell-surface, but it did not reach statistical significance (Appendix I).

6.3 Future Studies

Several questions remain unanswered with regard to the CTLA-4 signal peptide. First, further investigation into the mechanism of altered *N*-glycosylation via the signal peptide amino acid change would be of great interest. Though we did not find signal peptide cleavage differences between CTLA-4 allelic signal peptide forms, the timing of signal peptide cleavage has been recently linked to *efficiency* of glycosylation (Rutkowski et al. 2003). Association with the core translocation channel, Sec61α and Sec61β particularly, predict the ultimate timing of cleavage and thus the efficiency of glycosylation, and cross-linking studies can address this in a qualitative manner (Rutkowski et al. 2003). Second, if a proportion of aberrantly glycosylated molecules make it past the ER, it is not known whether *N*-linked glycosylation has effects on (i) CTLA-4 compartmentalization (ii) cycling to the cell-surface and, (iii) signaling. Since CTLA-4 is heavily glycosylated but the role of glycosylation in its case it not known this may be an important regulatory mechanism. Defects in glycosylation are common human and animal diseases, including Alzheimers' disease (Katayama et al. 2004), Prion-mediated diseases (Otvos and Cudic 2002) and Diabetes complications (Stitt et al. 2004) etc.

6.4 Regulatory Variation at *CTLA4*

Our work dealing with *cis*-acting variation at *CTLA4* focused on the promoter and 3'UTR regions using targeted genetic and functional tools. Since previous work had relied primarily on the non-synonymous signal peptide polymorphism, few aspects of

regulatory regions and T1D were known, and what was known was somewhat fragmented.

6.4.1 The *CTLA4* promoter

Based on our data we describe a diabetes-associated haplotype on which both protective and susceptibility alleles reside on the same chromosome. This haplotype contains the diabetes-associated +6230G allele and the protective –318T allele at the promoter and is significantly under-transmitted despite a low frequency of 8.6% (P=0.029) (Chapter 4, (Anjos et al. 2004)). It is the first time, to our knowledge that a disease-associated haplotype is described to contain both protective and predisposing alleles.

We also showed that these alleles contribute independently to the genetic effect, an important distinction in light of the report of an autoimmunity-associated +6230G>A 3' SNP accounting for the genetics in Graves' disease, and to a lesser extent in T1D (Ueda et al. 2003). From our own data it appears that the –318T allele at the promoter offers a dominant protective effect, as haplotype TDT analysis covering 97% of the variability at the *CTLA4* locus and containing the susceptibility +6230G allele resulted in the –1661G;-658C;318T;+49A;+6230G haplotype to be significantly *under*-transmitted (Chapter 4, (Anjos et al. 2004)). It is conceivable that the individuals bearing this haplotype (8.6% of the population) the –318T allele offers protection over the effect of the predisposing +6230G-allele. This remains to be investigated further.

Given our genetic effect at the promoter, we followed up with functional studies *in vitro* where the –318T allele induces significantly higher transcription in a luciferase reporter system validating previous work where a smaller promoter context was used

(Wang et al. 2002) and *in vivo* in mRNA from individual heterozygous at the promoter (Anjos et al. 2004).

We hypothesize that the transcriptional effect at the *CTLA4* promoter is mediated directly through disruption of the transcription factor consensus binding site of the Lymphoid Enhancer Factor-1 (*LEF1*) by the –318T allele. If modulation of this interaction indeed has significant effects on T1D risk, the question arises of whether additional T1D risk may be imparted by polymorphisms of *LEF1*, or its homologue *TCF1*. As a first step towards addressing this possibility, we studied possible association of two non coding intronic SNPs at *LEF1* (rs867565 and rs1693115) but found no association with T1D in our dataset (data not shown).

6.4.2 Future Studies

To date, our own genetic association study (Chapter 4,(Anjos et al. 2004)) and that of Lee et. al., (Lee et al. 2001) report a positive association at the promoter, while two others have not (Donner et al. 1998; Ihara et al. 2001) though likely due to lack of power since they used a cohort of 109 T1D families, and 160 patients vs. 200 controls, respectively. Clearly, more studies are required in different cohorts to confirm this association. Notably, in the largest (most recent) association study to date in T1D where 9 SNPs were evaluated by the TDT test, the promoter –318C>T was not tested (Ueda et al. 2003).

Having demonstrated an increase in transcriptional activity associated with the T-allele at the proximal *CTLA4* promoter begs the question: what is the mechanism of increased transcription? Based on computer-based algorithms, the prediction is that a

T-allele at this location disrupts a response element for the products of *LEF1*, and the closely related T-cell Factor-1, (*TCF1*). Both proteins are members of the High Mobility Group (HMG) transcription factors involved in various stages of Development and in T-cell differentiation. Whether or not the *CTLA4* promoter interacts with *LEF1* remains to be verified experimentally by electrophoretic mobility shift assays (EMSA) (which we have attempted rather exhaustively with mixed results), and more ideally via the Chromatin Immunoprecipitation assay (CHIP) allowing for *in vivo* assaying of DNA/protein interactions in their chromosomal/chromatin context. In addition, it is not known what effect *LEF1* has on *CTLA4* transcription (if it does, indeed bind to the *CTLA4* promoter) though our data can best be explained if we hypothesize that it represses transcription. *LEF1* is known to act as a repressor in the absence of Wnt signals (van Noort and Clevers 2002) and an activator in the presence of Wnt/β-Catenin as in the case of the APC, cyclin D1 and c-myc genes (He et al. 1998; Shtutman et al. 1999; Hovanes et al. 2001) and it remains a very interesting possibility to determine if *CTLA4* is a *LEF1* target.

6.4.3 The *CTLA4* 3'UTR dinucleotide repeat

The 3'UTR of *CTLA4* contains an $(AT)^n$ microsatellite repeat sequence ranging from 84-131 bp, though two thirds of Caucasian chromosomes can be accounted for by the 88bp and 106bp alleles, in LD with the +49A and +49G alleles, respectively (Holopainen and Partanen 2001) and our own data (Appendix II). The more common alleles were reported to be associated with T1D in various studies (Nistico et al. 1996;

Marron et al. 1997) and its location in the 3'UTR, a region typically associated with regulating mRNA stability compelled us and others to assess the effect of an expanded repeat on mRNA stability.

In a study by Wang et. al., the 86 bp allele was associated with a 50% increase in steady-state *CTLA4* mRNA when compared to the longer alleles at the 3'UTR and a longer $(AT)^n$ allele was associated with decreased mRNA stability in a mRNA disappearance assay designed to compare across homozygous samples (*n=8*) (Wang et al. 2002). These data compelled us to investigate the effect of the dinucleotide repeat on mRNA stability using a more suitable study design. PBMCs from heterozygous individuals for the common alleles (88 bp vs. 106bp) were treated with actinomycin D for specified intervals and allelic-specific transcripts were quantified. We found no difference in the mRNA disappearance rate between the two alleles (Appendix II).

6.4.4 The alternatively spliced soluble *CTLA4* isoform

The SNP in the 3'flanking region of *CTLA4*, +6230G>A is the focus of our latest effort in dissecting the contributions of individual SNPs within the linkage disequilibrium block that encompasses the *CTLA4* region. Though Ueda et. al., provide strong evidence in favor of this SNP bearing the highest association with Graves' disease, its contribution to T1D is less striking and the 5'end of *CTLA4* cannot be ruled out (Ueda et al. 2003).

It was proposed that, functionally, the effect of this 3'SNP involved the soluble *CTLA4* isoform (*sCTLA4*). The disease-associated allele, G at +6230 was reported to be associated with lower expression of the *sCTLA4* isoform, while having no effect on the

full-length isoform in unstimulated T cells from (n=3) individuals heterozygous at +6230G>A (Ueda et al. 2003). Our study addresses the functional contribution of the 3'SNP in a larger dataset, accounting for haplotype structure, given the prior evidence of effects at the –318C>T promoter SNP. We found no difference in the allele-specific expression of *CTLA4* soluble or full-length mRNA attributable to the 3'SNP in inactivated (or activated) PBMCs' from heterozygous individuals (Chapter 5, (Anjos S.M. et al. 2005)). It has been suggested that the effect of the +6230G>A is to regulate *in cis*, the alternative splicing and/or polyadenylation of the soluble *CTLA4* isoform (Ueda et al. 2003), though the nature of the transcript harboring this 3'SNP remains a bit of a mystery and our own efforts in characterizing it have come up empty-handed (3'rapid-amplification of cDNA ends, RACE, data not shown). In addition, in our exhaustive search of the expressed-sequence tag (EST) databases, no transcript was found which included the known 3'UTR of *CTLA4 and* the sequence near or encompassing the 3'SNP (Chapter 5, (Anjos S.M. et al. 2005)). Of course, the possibility that this transcript may be a rare transcript expressed at a particular time in T-cell development or under certain T-cell activation conditions remains plausible. Because of these uncertainties, the possibility of the 3'SNP being a modifier *in cis* for the downstream *ICOS* gene was examined in our study and promptly excluded (Chapter 5, (Anjos S.M. et al. 2005)). Clearly, this requires more investigation.

6.4.5 Future Directions

With regards to the contribution of the +6230G>A SNP, work remains to be done to characterize the transcript harboring it. If indeed it affects polyadenylation and/or

splicing also remains to be determined experimentally. More importantly, more studies are need in other populations, with particular attention paid to haplotypes structures to confirm the association of this SNP with T1D.

Though large sample sizes are needed for the detection of effects with low penetrance expected from loci involved in complex disease susceptibility (Risch 2000), smaller populations stratified according to age of disease onset, presence of auto-antibodies such as GAD65, IA-2 and/or insulin, and *HLA* alleles may indeed prove to be highly informative. This has already been done in certain studies, where smaller, carefully controlled and more homogeneous populations are more informative than larger ones (Awata et al. 1998; Abe et al. 1999; Hayashi et al. 1999; Abe et al. 2001; Mochizuki et al. 2003). In our own efforts to address confounding factors, we are currently genotyping all families at the *HLA* class II region.

6.5 Evaluating the hypothesis using transgenic mouse models

The ultimate experiment, given the accumulated knowledge on the CTLA4 gene in autoimmunity, would be a transgenic mouse model where high risk and low risk CTLA4 haplotypes would be introduced into the mouse. If I were designing such an experiment, based on the work presented here, I would consider comparing haplotypes containing the –1661A;-658C;-319C;+49A;+6230G alleles, vs. –1661G;-658C;-319T;+49A;+6230G introduced into the mouse, using human BAC constructs containing the haplotypes and then crossed against the *CTLA4* knockout These two haplotypes are the most significantly over-transmitted and under-transmitted, respectively and collectively account for 40% of the variability at the CTLA4 locus. Crosses against the

diabetes-prone NOD mouse, with engineered class I and class III VNTR alleles (transgenic mouse strains developed in our lab, results not yet published) would present a model where all three confirmed candidate genes can interact *in vivo*.

6.6 Implications for Type 1 Diabetes (and autoimmune disease)

We have provided direct and indirect evidence for a functional role of the *CTLA4* gene in susceptibility to Type 1 Diabetes. Based on our data, the emphasis is on disease predisposing or protecting *haplotypes* rather than single SNPs or markers as modulators of disease susceptibility. This remains to be proven in future studies, but if it is correct it could have important implications for the treatment of autoimmune diseases.

6.6.1 *CTLA4* in autoimmune endocrine diseases: a common susceptibility gene?

Based on our data, and that of others, it is highly probable that the susceptibility to AITD (including GD and AIH) and T1D conferred by the *CD28/CTLA4/ICOS* locus is a result of a polymorphism or polymorphisms within the *CTLA4* gene itself. Could this be a common susceptibility gene in autoimmune (endocrine) diseases?

Let us examine the facts. Both T1D and AITD are organ-specific T-cell mediated diseases sharing a similar pathogenesis involving T-cell infiltrates resulting in the destruction of the target organ (pancreatic islets in T1D and the thyroid in AITD). There is a well recognized association between T1D and AITD (Jenkins and Weetman 2002). 20% of Type 1 Diabetics have antibodies against the thyroid, with 50% of these progressing to clinical AITD (Kordonouri et al. 2002). While in normal controls there are 0% of antibodies against the beta cells of the pancreas, in children with AITD, 2.3% also

have beta cell autoantibodies (Bright et al. 1982) and the presence of thyroid antibodies in diabetics was found to increase with age (3.7% in <5 years of age *versus*, 25.3% in 15-20 years of age) in a German population (Holl et al. 1999).

There is evidence of a shared genetic susceptibility to AITD and T1D in families (Einarsdottir et al. 2003). In fact two immune-regulatory genes, *HLA* class II and *CTLA4*, contribute to the susceptibility to both diseases.

6.6.2 Treatment outcome influenced by CTLA4 polymorphisms

CTLA4 genotypic effects on expression may also be a factor in determining treatment outcome, given that costimulatory pathways are being exploited in the treatment of autoimmune diseases, and graft versus host disease (reviewed in (Salomon and Bluestone 2001)) as well as in immune-mediated tumor therapy (Chambers et al. 2001; Demaria et al. 2003; Phan et al. 2003) with promising results. Recent findings that the G/G genotype at *CTLA4* +49A>G SNP is associated with increased rejection of organ liver transplantations (P=0.0055, n=211 liver transplant patients) (Marder et al. 2003) suggests that allelic or haplotypic differences are important factors in therapeutic interventions.

Understanding the immuno-modulatory effects of such polymorphisms on expression of *CTLA4* is important if therapies are to be designed to fine-tune CTLA-4 function. For instance, the competition between CTLA-4 and CD28 for each of its ligands, B7-1/2, may be influenced by the altered glycosylation due to the signal peptide polymorphism. In addition, at least some of the function of regulatory T cells (Tregs express CTLA-4 constitutively) is mediated through CTLA-4 binding with B7-1/2

receptors (Lin et al. 1998; Grohmann et al. 2002), such that T effector activity could be determined by *CTLA4* genotype.

Clearly, understanding the regulatory mechanisms controlling *CTLA4* expression, modified *in cis*, along with the allelic specific regulation occurring at the post-translational level, will enable the development of drugs and protocols aimed at "tricking" the immune system to recognize nonself as self, or to restrain autoimmune reactions, which will provide true targeted therapies (perhaps custom haplotype-designed) to cure disease.

6.7 Summary

The original research presented in this dissertation focused mainly on ascertaining the functional differences between common alleles at the *CTLA4* gene and their relation with susceptibility to autoimmune diabetes. This work has resulted in evidence favoring a novel hypothesis to describe the contribution of *CTLA4* to type 1 diabetes. Namely, that common disease-associated polymorphisms at *CTLA4* demonstrably alter its expression. Post-translationally, they do so through effects of a signal peptide amino acid substitution on *N*-linked glycosylation resulting in reduced mature CTLA-4 at the cell-surface. Transcriptionally, we provide evidence that allelic variation at the *CTLA4* promoter modulates *CTLA4* expression in a dominant, allele-dependent manner and through TDT analysis we identified a haplotype containing both protective (-318 T) and disease-predisposing (+6230G) alleles whose contributions are independent from each other. Furthermore we have presented evidence against the involvement of the +6230G>A 3'SNP in susceptibility to Type 1 Diabetes via allele-dependent effects on the expression of either soluble or full-length *CTLA4* isoforms, or via modulatory effects *in cis* on the expression of downstream *ICOS* gene. Specifically, the main contributions are outlined below:

6.7.1 Main Contributions

1. One of the first demonstrations of the post-targeting effects of signal peptides (Chapter 3, (Anjos et al. 2002)).
2. CTLA-4 G-allele (Ala17) is inefficiently *N*-glycosylated in the ER and is, in part, targeted for degradation (Chapter 3, (Anjos et al. 2002)).

3. Decreased cell-surface expression of mature CTLA-4 derived from the Ala17 signal peptide shown in *COS1* cells (Chapter 3, (Anjos et al. 2002).
4. Differential co-localization of allelic specific CTLA-4 fluorescently tagged constructs in a co-transfection system in *COS1* cells (Chapter 3, (Anjos et al. 2002))
5. Differential spatial orientation of CTLA-4 signal peptide allelic specific constructs in a co-transfection T-cells (Appendix I, *unpublished data*).
6. The *CTLA4* –318C>T proximal promoter SNP is highly associated with T1D protection by the TDT in our family dataset (Chapter 4, (Anjos et al. 2004)).
7. We confirm the association of +6230G>A with T1D by the TDT on our dataset, albeit modest (Chapter 4, (Anjos et al. 2004)).
8. The highly protective promoter –318T allele is found almost exclusively on a chromosome containing the predisposing +6230G allele at the 3'end and this haplotype is significantly under-transmitted, suggesting that the promoter polymorphism overrides the 3'SNP (Chapter 4, (Anjos et al. 2004)).
9. TDT analysis demonstrates independent genetic effects of the promoter and the 3'SNP on T1D (Chapter 4, (Anjos et al. 2004)).
10. The –318T allele at the *CTLA4* promoter induces higher transcriptional levels *in vitro* compared with the –318C allele in transfected T-cells (Chapter 4, (Anjos et al. 2004)).
11. Higher –318T-allele specific *CTLA4* mRNA expression obtained *in vivo* in individuals heterozygous at the –318C>T SNP but homozygous at the +6230G>A SNP (Chapter 4, (Anjos et al. 2004)).

12. Evidence against any effect of the most common alleles at the $(AT)^n$ microsatellite on *CTLA4* mRNA stability (Appendix II, *manuscript in preparation for submission*).
13. No detectable effect of the +6230G>A on the soluble or full-length isoforms of *CTLA4* (Chapter 5, (Anjos S.M. et al. 2005)).
14. Confirmation of –318C>T allele-specific promoter effects in soluble and full-length *CTLA4* isoforms using a different methodology (Chapter 5, (Anjos S.M. et al. 2005)).
15. No modulatory effects *in cis,* of the 3'SNP +6230G>A on the expression of downstream gene *ICOS* were found, that could xplain the T1D association (Chapter 5, (Anjos S.M. et al. 2005)).

A1. APPENDIX ONE

EVALUATING THE EFFECT OF THE DIABETES-ASSOCIATED *CTLA4* SIGNAL PEPTIDE POLYMORPHISM ON ITS FOCAL LOCALIZATION IN T CELLS

Suzana M. Anjos, Constantin Polychronakos

A1.1 Contributions of authors:

Suzana M. Anjos: Hypothesis development, experimental data, confocal microscopy, image analysis and manuscript preparation

Constantin Polychronakos: Research Director, hypothesis development, manuscript editing and preparation.

A1.2 Introduction

Though CTLA-4 and CD28 share extensive homology their patterns of expression and co-localization are markedly different. CTLA-4 is upregulated on the cell-surface of activated T cells (Lindsten et al. 1993) while CD28 is constitutively expressed on naïve as well as activated cells (Gross et al. 1992). CTLA-4 is a cell-surface receptor *per se*, but has an intricate cycling pattern keeping it stored in the endosomal compartments via association of its intracellular tail and the medium chain of the clathrin pit adaptor protein AP-2 (Bradshaw et al. 1997; Chuang et al. 1997; Shiratori et al. 1997; Zhang and Allison 1997). Thus, surface expression of CTLA-4 is tightly restricted suggesting a potential regulatory point for controlling its inhibitory function. It has also been established that upon TCR engagement to the APC there is a re-orienting of the constitutively expressed CD28 co-stimulatory receptor towards the immunological synapse with a fraction of CTLA-4 eventually doing the same (Linsley et al. 1996). In addition, this focal orientation towards the TCR/APC synapse is dependent on the strength of the TCR signal strength, with cells receiving stronger signals being more susceptible to CTLA-4 mediated negative signaling (Egen and Allison 2002). It has been recently shown that this focal localization is important for the binding of the co-stimulatory ligands B7-1/2 and may be a factor in determining the outcome of the immune response (Pentcheva-Hoang et al. 2004).

We have previously shown that inefficient *N*-linked glycosylation of the autoimmunity-predisposing Alanine allele in the signal peptide of CTLA-4 is the mechanism of altered trafficking and decreased CTLA-4 at the cell-surface reported by us and others (Kouki et al. 2000; Anjos et al. 2002; Maurer et al. 2002). In our original

study, we qualitatively and quantitatively assessed levels of intracellular and cell-surface CTLA-4 alleles in a co-transfection system in *COS1* cells, a monkey kidney cell line (Anjos et al. 2002). When each fluorescently-tagged allele was co-transfected into the same cell, the allelic-specific CTLA-4 proteins colocalized to different compartments, and there was quantitatively less CTLA-4 Ala at the cell-surface in the *COS1* cells (Anjos et al. 2002). In the present study we examined the distribution of CTLA-4 alleles co-transfected in a more relevant T-cell system (Jurkat T cells) by Z-sectioning on a confocal microscope, and quantified the relative amounts of allelic-specific CTLA-4 surface expression vs. total in transfected Jurkats by flow cytometry.

We report that the CTLA-4 allelic-specific proteins display different patterns of localization in T cells and that there is a small decrease in cell-surface levels of CTLA-4 (Ala17) at the cell-surface of T-cells (though P did not reach significance). Because CTLA-4 intracellular cycling and focal localization is an important determinant of the outcome of the inhibitory response, the differences we report may explain the increased proliferation of T cells in individuals with the CTLA-4 (Ala17) allele (Kouki et al. 2000; Maurer et al. 2002), as well as the allele-dependent effects on *IL-2* (Maurer et al. 2002) and ultimately the association with autoimmunity.

A1.3 Materials and Methods

Plasmids and plasmid DNA preparation

Full-length YFP/CFP CTLA-4 (Ala17 and Thr17) fusion proteins have been described in detail by (Anjos et al. 2002). Similarly, carboxyl-tail truncated fluorescent fusion proteins have also been described by (Anjos et al. 2002). Plasmid DNA was purified on Qiagen (Germany) maxi-prep columns following the instructions provided.

Cell culture and transfections

Jurkat T cells were obtained from American Type Culture Collection (ATCC, Manassas, VA) and maintained in RPMI 1640 supplemented with 10% FBS, 1 mM Sodium Pyruvate, 1 mM MEM and 100 units/ml penicillin/streptomycin at 37°C, 5% CO_2 (cell-culture reagents from Invitrogen, Carlsbad, CA). 5 X 10^5 cells were transfected in triplicates with 2 μg of CTLA-4 plasmid DNA using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) as per manufacturer's instructions. All transfections were performed with the same plasmid DNA maxiprep preparation. The cells were activated with 1 μg/ml PHA and 50 ng/ml PMA (Sigma) 4 hours following transfection for 48 hours. Although PHA and PMA are physiologically irrelevant stimuli they were found to be specifically required for constitutive expression under the *CMV* promoter in the Jurkat cell line (Schifferli 1996). Sterile coverslips in 6-well plates were incubated with α-CD3 mAb (1μg/ml) (Pharmingen, SanDiego) for 4 hours at 37°C, 5% CO_2 after which they were rinsed with sterile PBS. 5X10^5 T cells were immobilized on the α-CD3-coated coverslips for the duration of the transfection.

Flow Cytometry

48 hours following transfection, Jurkats were washed twice with sterile ice-cold PBS/0.05% sodium azide and stained with saturating concentrations of Phycoethyrin (PE)-conjugated anti-CTLA-4 mAb (Pharmingen, SanDiego). Appropriate isotype-matched immunoglobulin IgG1-PE was used.CTLA-4 at the cell-surface was stained by immunofluorescence, while total CTLA-4 was detected directly through YFP fluorescence. Ten thousand viable cells were measured within 1 hour of staining without fixation with a FACScan flow cytometer (Becton Dickinson) and analyzed with WinMDI software package available freely online at http://facs.scripps.edu/software .html. Mean fluorescence intensity was obtained and the data is reported as a ratio of mean fluorescence intensity cell surface (anti-CTLA-4-PE/ mean fluorescence intensity total (YFP).

Confocal Microscopy

Cells were washed gently with 4% paraformaldehyde for 30 minutes at ambient temperature. The fixed cells were mounted on standard microscope slides using conventional mounting media. Cells were visualized confocal microscopy (LSM 510, Zeiss Axiophot, Germany). Simultaneous double acquisitions were performed using the 458nm and 514nm laser lines to excite Cyan Fluorescent Protein (CFP) and Yellow Fluorescent Protein (YFP) respectively using 63X oil immersion Neofluar objectives. The fluorescence was selected with double fluorescence dichroic mirror and a band pass filter of 480-520nm for CFP and a longpass filter of 560nm for YFP. Z-sectioning was performed at 1μm slices.

A1.4 Results and Discussion

CTLA-4 alleles do not colocalize across T cell layers

We investigated the colocalization of CTLA-4 allelic-specific proteins tagged with different colored yellow and cyan fluorescent protein in Jurkat T cells. By coating the coverslips with anti-CD3 antibody, T cells are immobilized onto the coverslip and the anti-CD3/TCR interaction mimicks the APC/TCR interface. It was our objective to address the orientation and spacial distribution of CTLA-4 signal peptide allelic proteins in T cells. In our dual-transfection assay YFP-CTLA-4 (Ala17) and CFP-CTLA-4 (Thr17) co-localize to different compartments and orient themselves differentially with respect to the TCR/anti-CD3 interface (Figure 1A, YFP, red and CFP, green, overlay in yellow). Though a significant amount of CTLA-4 from each allele co-localizes at the immunological synapse (Figure 1A, panels G-K), there are distinct differences across the different cell layers with regards to localization (Fig 1A, panels A-F). To demonstrate that the differences reported are not a result of the inherent differences in the properties of the fluorophores we repeated the transfections using reciprocal CTLA-4 fusion proteins. In the reciprocal transfection, where YFP-CTLA-4 (Thr17) and CFP-CTLA-4 (Ala17) were introduced into the same cell, the two proteins co-localize at the synapse to a great extent (Figure 1B, panels P-T) but distinct localization patterns become clear as the Z-slicing cuts away from the coverslip (Figure 1B, panels V-AB).

It was known that polarization of CTLA-4 towards the TCR/APC interface occurs when a T cell receives a stimulatory signal (Linsley et al. 1996), and it was recently shown that this polarization is also accompanied by its expression at the cell surface and is dependent of the strength of the signal at the TCR (Egen and Allison 2002). Thus, the

balance between co-stimulatory and inhibitory signal is regulated in part, by localization of the proteins involved (Egen and Allison 2002), while the actual binding capacity of the B7-1/2 ligands to CTLA-4 is dictated at least in part by its accumulation at the immunological synapse. Our data suggest that there is alternate spatial orientation of the individual CTLA-4 allelic derived proteins and according to the studies above may in turn affect their inhibitory abilities. This may in fact explain the increased T cell proliferative capability of the autoimmunity-predisposing Ala17 variant reported by (Kouki et al. 2000; Maurer et al. 2002) and the effects on IL-2 down-regulation (Maurer et al. 2002).

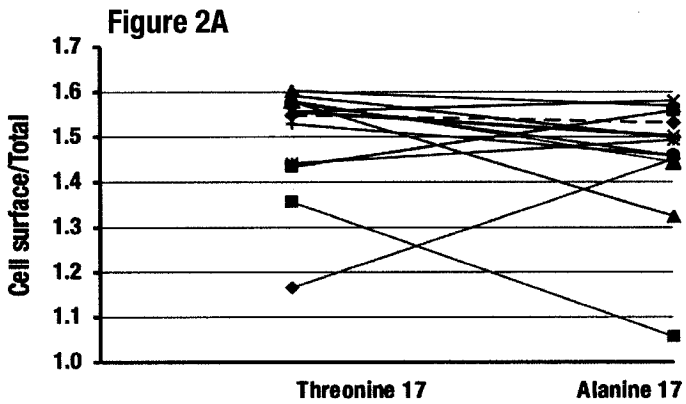
In addition, should the inefficient *N*-linked glycosylation of the CTLA-4 Ala 17 affect the kinetics of protein co-translation (via the signal peptide described by (Rutkowski et al. 2003)) this may influence the co-localization pattern of CTLA-4 (Ala17) as the kinetics of CTLA-4 polarization determine its transit from the back of the cell to site of the TCR/APC interface following activation where it likely undergoes a rapid turn-over (Egen and Allison 2002).

CTLA-4 allelic differences in cell-surface expression in transfected T-cells

We had previously shown decreased (C-terminus lacking) CTLA-4 (Ala17) on the cell surface of COS1 cells (Anjos et al. 2002). In this study we investigated the cell-surface expression of transfected truncated C-terminus CTLA-4 constructs fused to YFP in a more relevant T cell system, simultaneously detecting total CTLA-4 expression via YFP fluorescence and cell-surface expression with antibody staining against surface CTLA-4. Untransfected Jurkats do not express surface or intracellular CTLA-4, and PHA

and PMA addition, required for expression under the CMV promoter in Jurkats, results in an upregulation of CTLA-4 protein expression (data not shown), both well documented facts (Perkins et al. 1996; Xu et al. 2002). By quantifying the cell-surface expression relative to the total, we account and correct for differences which may arise as a result of transfection efficiency. All experiments were performed in replicates of five except for the first one (in duplicates), and mean fluorescence intensity was measured. Though we detect a trend towards a lower expression of the Ala17-derived CTLA-4 it did not reach statistical significance (Figure 2A and B) MFI ratio for the Ala17 vs. Thr 17, 1.45± 0.079 vs. 1.50 ± 0.072 (± SEM), $P>0.05$.

In light of the results obtained in the co-localization experiments above by confocal microscopy, such a result is not entirely surprising. Based on this data (Maurer et al. 2002), there are distinct differences in trafficking between the two allelic-proteins of CTLA-4 in T-cells and we have shown that the mechanism underlying these is inefficicent *N*-linked glycosylation (Anjos et al. 2002). Because we have evidence that the two proteins, once past the endoplasmic reticulum (ER), exhibit different patterns of localization, the assumption that some of the differentially glycosylated Ala17 makes it beyond the ER, is a valid one. It has become increasingly clear that this intricate intracellular localization and cycling pattern, coupled with a rapid turnover at the cell-surface is an added layer of post-translational regulation (Linsley et al. 1996; Egen and Allison 2002) recently shown to influence ligand binding (Pentcheva-Hoang et al. 2004). Given these findings, though cell-surface differences may not be readily detectable due to rapid turnover, differences in intracellular trafficking are just as important in determining the outcome of the inhibitory response.



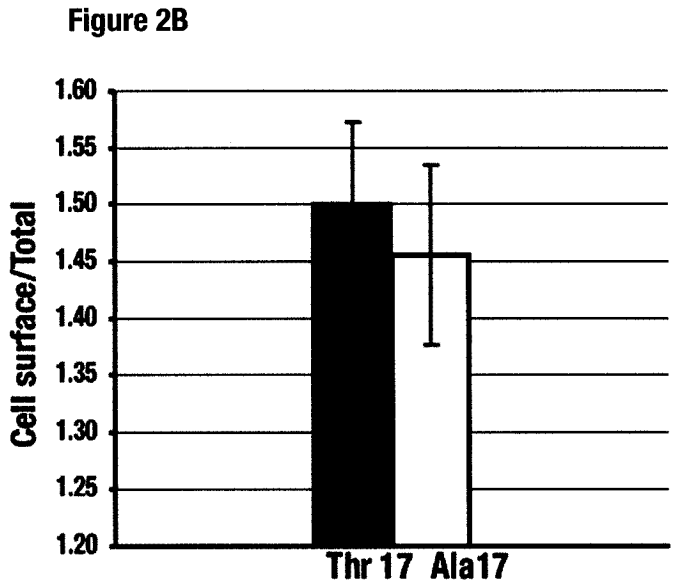


Figure 1: CTLA-4 signal peptide allelic proteins do not co-localize in T cells

(A) α-CD3 immobilized Jurkat T cells were co-transfected with YFP-CTLA-4Gly (Ala17) shown in red and CFP CTLA-4Gly (Thr17), in green and expression induced by PHA and PMA. The cells were analyzed by laser-scanning confocal microscopy and subjected to 1 μm Z-sectioning. Panels A-L show the Z-slices from the top-of-the-cell to the coverslip plane with the TCR/CD3 interface appearing in panels I-J. Focal localization of CTLA-4 receptors is apparent in panels H-J, with most of the fluorescence concentrated at one end of the cell (the sight of TCR/CD3 engagement). Though both

Figure 1A

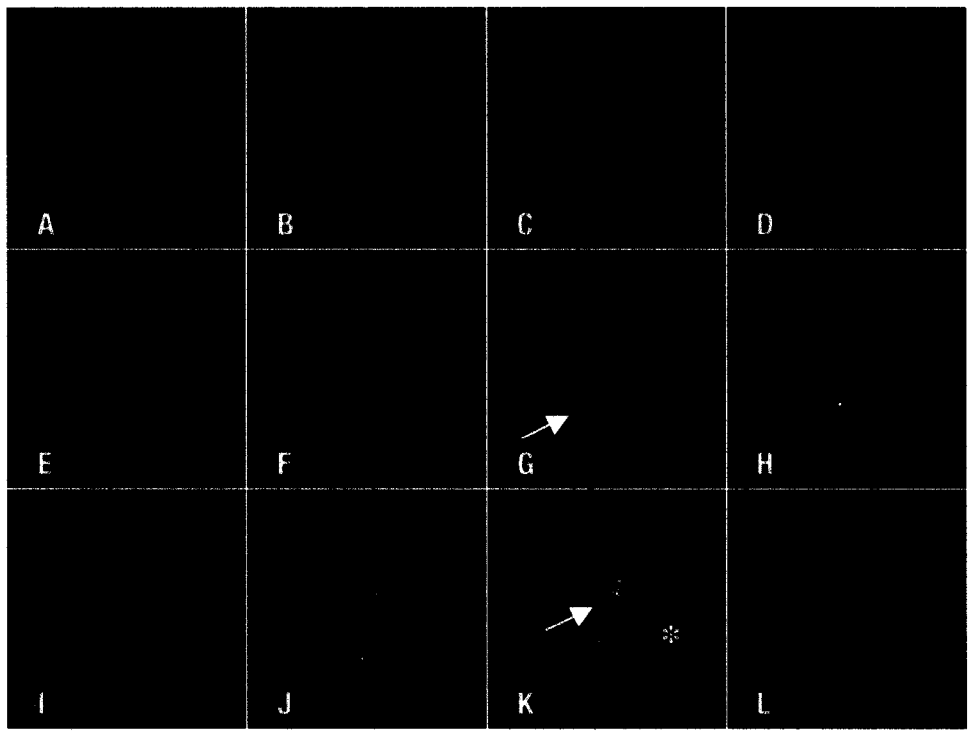


Figure 1B

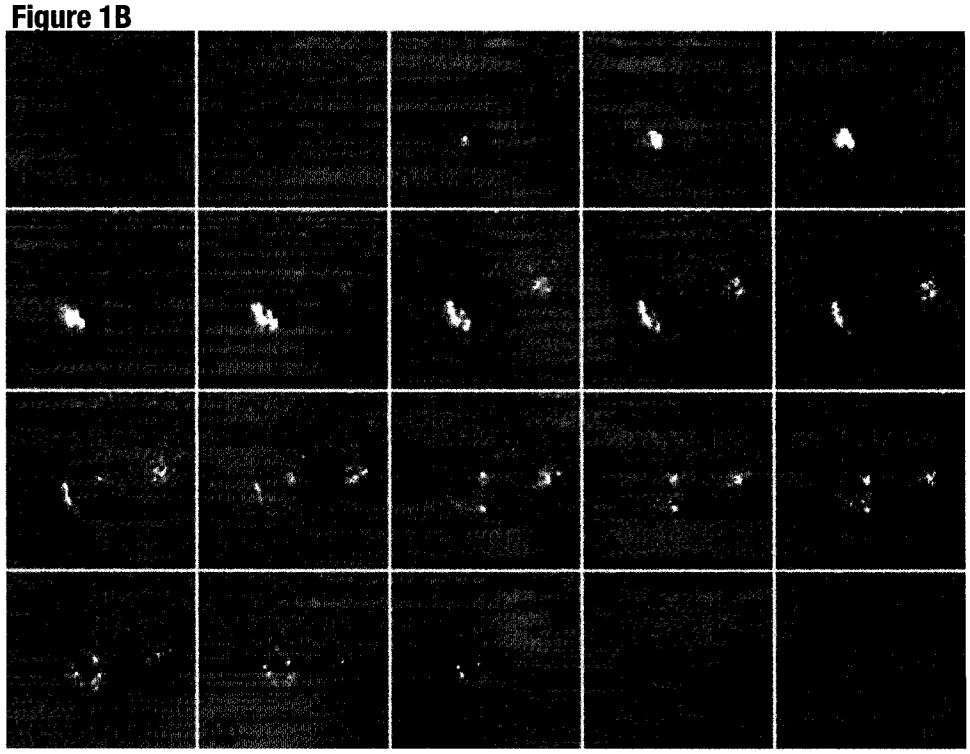


Figure 2: Quantification of cell-surface expression of CTLA-4 signal peptide allelic variants by flow cytometry

Jurkat T cells were transfected with either YFP-CTLA-4Gly (Ala17) or YFP-CTLA-4Gly (Thr17) and prepared for flow cytometry by immunofluorescence staining against cell-surface CTLA-4 upon activation with PHA/PMA. Total CTLA-4 expression (intracellular + cell-surface) was detected with YFP fluorescence. The mean fluorescence intensity (MFI) is reported as the ratio of cell-surface (CTLA-4-PE) MFI /Total (YFP) MFI. In ***(A)*** individual transfections performed in one experiment are connected by the lines, with each data point representing the relative amount of each allele. The colors indicate the replicate transfections within one single experiment. ***(B)*** Summarizes the mean ratio across the multiple experiments for cell-surface/total CTLA-4, with the ratio of Ala17 at 1.45± 0.079 and the Thr17 ratio, 1.50 ± 0.072 (± SEM, n=3 independent experiments), though $P>0.05$ there is a trend towards lower CTLA-4 expression of Ala17.

A2. APPENDIX TWO

The dinucleotide $(AT)^n$ repeat in the 3'UTR of the diabetes associated gene, *CTLA4* does not alter the stability of the messenger RNA.

.

Suzana M. Anjos, Constantin Polychronakos

Manuscript in preparation for submission

A2.1 Contributions of Authors

Suzana Anjos: Hypothesis development, experimental data, and manuscript preparation

Constantin Polychronakos: Research Director, hypothesis development and manuscript editing.

A2.2 Introduction

The polymorphic $(AT)^n$ repeat in the *CTLA4* 3' UTR could have allelic specific effects on mRNA stability. Attempts to determine the role of these putative regulatory polymorphisms in the existing literature have been fragmentary, show inconsistencies, and often use suboptimal methodology. We undertook this study to provide a thorough and integrated answer using the most reliable methodology, which compares the two alleles within heterozygous individuals, thus avoiding all the interference from extrinsic factors (genetic influences in trans, and the individual's immune experience) that obscure allelic differences in expression when comparisons are made amongst, rather than within individuals.

A2.3 MATERIALS AND METHODS

Subjects, DNA and RNA isolation and Genotyping

Blood was obtained from individuals with informed consent. Genomic DNA was extracted from whole blood using standard phenol-chloroform protocols while RNA was isolated from PBMCs using the Rneazy Mini Kit (Qiagen, Germany). On-column Dnaseing (Qiagen) was done for all RNA samples.

The 3'UTR $(AT)^n$ repeat was amplified using primers designed by (Kemp et al. 1998). PCR conditions: 95 °C 3 min; 35 cycles at 95 °C 30 sec; 55.8 °C 30 sec, 72 °C 30 sec and a 10 min extension at 72 °C using Taq Polymerase (Invitrogen) in a T Gradient thermocycler (Biometra, Goettingen, Germany). The products were resolved by PAGE (10%; 19:1 bis:acrylamide) and were denoted as 88bp, 106bp (the two most common alleles) or other, where "other" refers largely to alleles >106 bp, but also to a few between 88-106 bp.

RT-PCR and quantification of allelic differences by automated sequencing

Reverse-transcription was performed under standard conditions using random primers. In each case, 250 ng-1 μg of total RNA template was used. PCR conditions for amplification of the first exon of *CTLA4* which includes the +49A>Gpolymorphism were described above except an antisense primer 5'-GCTGGGCCACGTGCATTGCT-3' which spanned exons 1 and 2 was used to avoid contamination by genomic DNA. PCR amplicons were purified by the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and were sequenced.

Allele-specific quantification by sequencing was done by comparing ratios of peak heights (equivalent to peak area) of both alleles in one single reaction in matched heterozygous RNA and DNA-derived PCR products as describe in detail in Chapter 2. Differences deviating from one in RNA suggest allele-specific differences in expression. All RNA and DNA peak heights are reported as a ratio of peak height at SNP/ average peak height of adjacent bases and this is referred to herein as standardized RNA units.

Statistics

Functional data were compared with the 2-tailed, paired Student t-test. The paired test was used when comparing DNA and RNA from the same individual, or in vitro expression of each allelic constructs in each of the 5 experimental replications, to account for variation among experiments that affected both alleles equally.

Actinomycin D treatments

In the mRNA stability studies, Ficoll isolated PBMCs were activated for 20-24 hours with 5 μg/ml ConA and 10 nM PMA and treated with 20 μg/ml actinomycin D (Sigma) every 2 hours up to 10 hours. Cells were harvested at 10 hours and RNA was isolated as described above.

A2.4 RESULTS

The $(AT)^n$ 3'UTR repeat in *CTLA4* has no effect on steady-state mRNA expression

Considerable evidence exists to support the role of 3'UTR AU-rich sequences in the disappearance rate of the particular messenger RNA (Chen and Shyu 1995); length of the AU region is among one of the factors which determine mRNA stability.

In order to determine if common *CTLA4* AT repeat alleles, (denoted here as 88 bp, 106 bp or other), had an effect on mRNA levels and stability we took advantage of the linkage disequilibrium between the signal peptide and the 3'UTR. We also chose individuals heterozygous (A/G) at the signal peptide (+49) but homozygous for the –318C>T promoter allele, to exclude interference from transcriptional effects. Steady-state +49A;88 bp RNA was compared to the +49G;106 bp or "other") in the same individual. We did not find significant differences between 88 bp alleles and all other alleles (Figure 1). The ratio of +49A;88 bp RNA over +49G;106 bp (or other) was (0.92 ± 0.03), nearly identical to that in the DNA ratio (0.90 ± 0.03) with minimal variability around the mean (Table 1). Since individuals with –318T were excluded, the shortest (88 bp) allele, was in cis to +49A in almost all cases, because of the LD (Holopainen and Partanen 2001). Phased haplotypes were confirmed by genotyping family members in all cases (Table 2). Allelic ratios were identical whether the 3 individuals with the "other" allele on the +49G chromosome were included or excluded. The *88 and *106 bp alleles are in tight LD with, respectively, +49A and +49G and, combined, account for 67% of all alleles (Holopainen and Partanen 2001) confirmed in our own data). Given the clear lack of any functional effect of the two major alleles, it is unlikely that the remaining 33% of

alleles at the 3'UTR with lengths above 106 bp or between 88 and 106, which are found indiscriminately on either +49A or +49G chromosomes, can explain the association found at the coding SNP through LD (Table 2).

In the next section we sought to confirm the absence of allelic effects by directly measuring differential mRNA stability.

Table 1: Regulatory variation at the 3'UTR

Stratified by length of $(AT)^n$ in individuals heterozygous at the +49 locus and homozygous at the –318 locus.

Genotype at 3'UTR	RNA (+49A) ± SEM	RNA (49G) ± SEM	*P*
88 + 106	0.38 ± 0.04 (n=5)	0.37 ± 0.07 (n=5)	*0.73**
88 + 106 and other	0.40 ± 0.03 (n=8)	0.40 ± 0.04 (n=8)	*0.97*

Table 2: Frequency of *CTLA4* haplotypes (n=154, chromosomes).

Haplotypes			# of Chromosomes (%)
-318	+49	(AT)	
C	A	88	n=71 (46%)
C	A	106	n=2 (1%)
C	A	Other*	n=8 (5%)
T	A	Other	n=7 (5%)
C	G	88	n=3 (2%)
C	G	106	n=32 (21%)
C	G	Other	n=31 (20%)

*Most "other" are >106bp

The length of the $(AT)^n$ repeat does not alter *CTLA4* mRNA stability

We investigated allele-specific differences in the stability of the messenger RNA by treating in vitro activated PBMCs with the transcription attenuator, actinomycin D for 0-10 hours. PBMCs were obtained from individuals heterozygous at the signal peptide locus but homozygous at the –318 promoter locus (C/C). All individuals carried an 88 bp allele in their 3'UTR *in cis* to the +49 A allele (confirmed by family genotypes). Relative RNA levels in +49 A/G were quantified and compared to their corresponding DNA +49 A/G ratios (by sequencing). The combined data from 4 independent experiments in 4 different individuals can be found in Figure. The individual data points represent the average of 4 experiments. There is clearly no consistent difference in mRNA stability arising from differences in alleles at the 3'UTR. This is in agreement with our steady-state results.

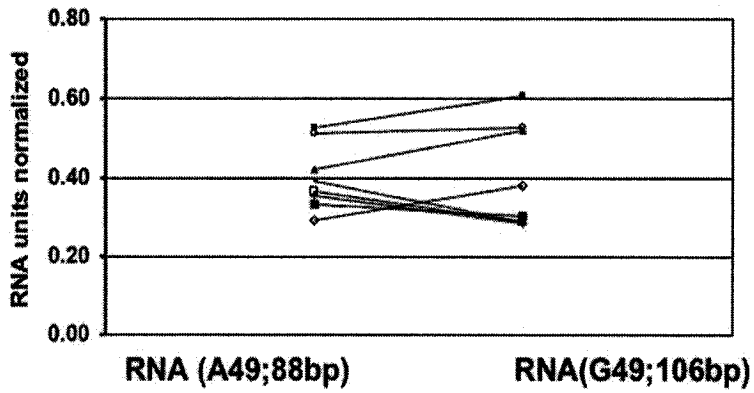


Figure 3: Allelic steady-state mRNA does not depend on length of the 3'UTR $(AT)^n$ repeat.

PBMCs from individuals heterozygous at the 3'UTR (88 bp+106 bp or other), heterozygous at the +49 locus and homozygous for the –318C allele in the promoter were chosen. The +49A>G SNP was used to distinguish the two alleles. Relative RNA attributable to the 49A;88bp is compared to 49G;106bp (or other) in 8 individuals. There is no significant difference between the relative levels of RNA associated with the short allele *versus* any of the longer alleles. Lines connect the values for each individual.

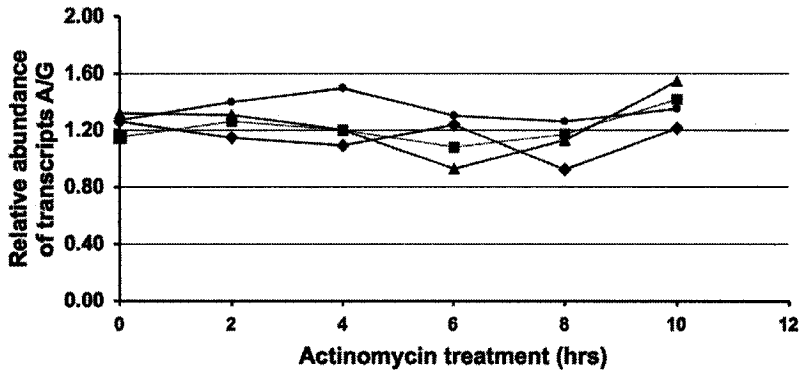


Figure 4: The length of the *CTLA4* 3'UTR repeat does not affect mRNA stability

PBMCs from 4 individuals heterozygous at the +49 locus, heterozygous at the 3'UTR, (88 vs. 106 bp or other alleles) and homozygous at the promoter –318 locus (C/C) were activated and treated with the transcription attenuator actinomycin D for 0-10 hours. Relative RNA was quantified in an allele-specific manner as described in Fig 3, and the ratios of A49;88 bp/G49;106 bp (or other) are plotted against time of treatment. There is no change over time of actinomycin D treatment.

A2.5 Discussion

Regulation of mRNA decay is an important mechanism by which the level of gene expression is controlled. AU-rich elements (ARE's) commonly associated with mRNA degradation process, particularly the AUUUA pentamer and the nonamer UUAUUUA(U/A)(U/A), are not present in the *CTLA4* 3' UTR. Since these are not always required for a functional ARE, as is the case, for example, of the zif278, c-jun, c-myc genes (Chen and Shyu 1995), it is conceivable that the dinucleotide repeat might play this role for *CTLA4*.

Wang et al. found a ~50% increase in steady-state *CTLA4* mRNA level for the 86 bp allele, compared to longer alleles in heterozygous individuals (Wang et al. 2002). They also show a faster RNA disappearance curve after actinomycin D treatment in an individual *homozygous* for a long AT repeat compared to one homozygous for a shorter one. Our data in this report did not reproduce either finding. The more likely explanation for the discrepancy is technical: Wang et al., (Wang et al. 2002). present data that appear to be from a single individual. Moreover, they compared RNA disappearance curves between two individuals homozygous for different alleles, an approach invalid for reasons detailed in Chapter 2, section. For the steady-state mRNA study in heterozygotes, they used the microsatellite itself to distinguish the two RNA alleles. Relative quantification of microsatellite alleles is subject to artifacts such as polymerase slippage and large differences in amplification efficiency between alleles differing by tens of base pairs that may not be completely corrected by normalizing to genomic DNA. Specifically for this reason we used the +49A>G SNP as a proxy, an assay we validated with a serial mixture experiment.

Alternatively, the longer alleles in Wang et. al. may have been different from those in our study (the shorter allele was the same in both studies). It is possible that among the longer cluster there are certain alleles that do shorten *CTLA4* half-life. Even so, a glance at the haplotype frequencies (Holopainen and Partanen 2001) makes it obvious that this is a very unlikely explanation for the genetic association. The two alleles we found identical in terms of RNA stability, 88 and 106 bp, account for 2/3 of all chromosomes and are in tight linkage disequilibrium with A and G respectively at the coding polymorphism (Holopainen and Partanen 2001). In the largest published study of genetic association (transmissions from a total of 850 heterozygous parents) there was highly significant distortion in the transmission from parents heterozygous for either the 88 or the 106 bp but not from those heterozygous for other alleles (Marron et al. 1997).

A3. APPENDIX III

COPYRIGHT APPROVALS

ELSEVIER

11 January 2005 Our ref: HG/ND/Jan05/j031

Suzana Anjos
McGill University Health Center
The Montreal Childrens Hospital research center
4060 Ste-Catherine St west, room 414
Montreal, Quebec
H3Z 2Z3
CANADA

Dear Ms Anjos

MOLECULAR GENETICS & METABOLISM, Vol 81, Issue 3, 2004, Anjos et al, "Mechanisms of genetic....."

As per your letter dated 10 January 2005, we hereby grant you permission to reprint the aforementioned material at no charge **in your thesis** subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

   "Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier".

3. Reproduction of this material is confined to the purpose for which permission is hereby given.

4. This permission is granted for non-exclusive world **English** rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form. Should you have a specific electronic project in mind please reapply for permission.

5. This includes permission for the National Library of Canada to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely

Helen Gainford
Rights Manager

From: jcuthbert@asbmb.org
Subject: **Re: Fw: Copyright permission (JBC Feedback Form)**
Date: January 11, 2005 7:51:32 AM EST
To: suzana.anjos@mail.mcgill.ca

---

gratulations on completing your dissertation, quite an accomplishment! The ASBMB offers free membership to new Ph.D's. Please contact us once your ree has been granted so that we can send you the necessary forms. If you are an author on the paper you do not need permission from the journal.

cy Rodnan/ASBMB

0/2005 03:49 PM

To Joan Cuthbert/ASBMB@ASBMB
cc
Subject Fw: Copyright permission (JBC Feedback Form)

ı,
se respond.
ık you.
cy
cy J. Rodnan
ctor of Publications
rican Society for Biochemistry and Molecular Biology
) 634-7145
(301)634-7126
nan@asbmb.org

r.asbmb.org
Forwarded by Nancy Rodnan/ASBMB on 01/10/2005 03:46 PM -----
zana Anjos" <suzana.anjos@mail.mcgill.ca>

0/2005 03:43 PM

To <jbc-feedback@highwire.stanford.edu>
cc <suzana.anjos@mail.mcgill.ca>
Subject Copyright permission (JBC Feedback Form)

---
ments sent via JBC Feedback Page

ME: Suzana Anjos USER NAME: (not signed in)
AIL: suzana.anjos@mail.mcgill.ca
ADDRESS: 70.48.227.49
STNAME: Ottawa-HSE-ppp4069036.sympatico.ca
EVIOUS PAGE: http://www.jbc.org/help/
OWSER: Mozilla/4.0 (compatible; MSIE 5.23; Mac_PowerPC)
OMOTIONAL USE: Not granted

MENTS:
eed to obtain copyright permission from you to reproduce
manuscript published in the Journal of Biological
mistry in 2002. it is for the purpose of a doctoral
sertation.
you let me know how to do this.
nks
njos

From: permissions@nature.com
Subject: **Permission granted**
Date: January 10, 2005 2:22:30 PM EST
To: suzana.anjos@mail.mcgill.ca
Cc: permissions@nature.com

---

e of Request: 10-Jan-2005
e of Intended Publication: -- (Unscheduled)

ana Anjos
Gill University Health Center Research Institute
0 Ste-Catherine St-West room 414

ntreal Quebec
/ 1X1
nada
4) 412-4400 x22623
ana.anjos@mail.mcgill.ca

ar Suzana Anjos,

nk you for your request to use Item 1: figure, Number 1 from Nature, Vol. 423, pp. 2 to 2, "Association of the T-cell regulatory gene CTLA4 susceptibility to autoimmune disease" by Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D.B., ter, K.M., Smith, A.N., Di Genova, G., Herr, M.H., Dahlman, I., Payne, F., Smyth, D., Lowe, C., Twells, R.C., Howlett, S., Healy, B., Nutland, S., ce, H.E., Everett, V., Smink, L.J., Lam, A.C., Cordell, H.J., Walker, N.M., Bordin, C., Hulme, J., Motzo, C., Cucca, F., Hess, J.F., Metzker, M.L., gers, J., Gregory, S., Allahabadia, A., Nithiyananthan, R., Tuomilehto-Wolf, E., Tuomilehto, J., Bingley, P., Gillespie, K.M., Undlien, D.E., ningen, K.S., Guja, C., Ionescu-Tirgoviste, C., Savage, D.A., Maxwell, A.P., Carson, D.J., Patterson, C.C., Franklyn, J.A., Clayton, D.G., erson, L.B., Wicker, L.S., Todd, J.A. and Gough, S.C.:, et al. Print and electronic permission is granted for one-time publication in Functional section of the type 1 diabetes gene: CTLA-4/ (), by McGill (format: Print), with the following specifications:

Reproduction is requested for a total of not more than 10 Figures/Tables in single publication, and for no other items.

Reproduction is intended in a primary journal, secondary journal, CD-ROM or book.

Nature Publishing Group's permission for electronic reproduction is not unlimited: we may in future require you to link through to the figure on ure Online, rather than maintain an electronic version on your own web site.

Nature's copyright permission must be indicated next to the Figure/Table in print. In electronic form, this acknowledgement must be visible at same time as the Figure/Table, and must be hyperlinked to the journal's homepage or at least to the Nature home page (http:// w.nature.com/).

You must yourself obtain the consent of the author or, where such information is printed in a Nature Publishing Group journal, the tographer/illustrator.

Significant modifications will be faxed for approval by the authors and Nature Publishing Group.

\ny further usage will be requested in writing.

are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

From: aweber@endo-society.org
Subject: **Granted - Copyright Permission Request**
Date: January 10, 2005 2:56:09 PM EST
To: suzana.anjos@mail.mcgill.ca
Cc: aweber@endo-society.org

Endocrine Society
1 Connecticut Avenue, Suite 900
Chevy Chase, MD 20815-5817
Telephone: 301-941-0200
Fax: 301-941-0259
www.endo-society.org
Tax ID Number: 73-0531256

Form was submitted by: suzana.anjos@mail.mcgill.ca

Original author to publish in a Dissertation.

Date: January 10, 2005
Reference Number:

Name: Suzana Anjos
Organization: McGill University Health Center Research Institute
Department:
Address: 4060 Ste Catherine St West 414
City, State and Postal Code: Montreal, Quebec H3Z 2Z3
Country: Canada
Phone: (514) 412-4400 x22623
Fax:
Email: suzana.anjos@mail.mcgill.ca

Journal: The Journal of Clinical Endocrinology & Metabolism
Author Name: S. Anjos
Title: Association of the Cytotoxic T lymphocyte associated antigen-4 gene with Type 1 Diabetes: evidence for independent effects of two polymorphisms on the
Year: 2004
Volume: 89
Page Range: 6267-6265
Extract Reproduction:
Number of Copies: 1
Where will the figures appear:
Dissertation
Title: Functional Dissection of the Type 1 Diabetes associated gene: CTLA4
Publisher: McGill University

Additional Comments:

This is a request for the reproduction of the entire manuscript, not only the figures. It's for my doctoral dissertation.

The Endocrine Society grants permission to reproduce the abstract from the selected article stated above contingent upon the following conditions: 1) That you give proper credit to the author(s) and to include in your citation, the title of journal, title of article, volume, issue number, date, and page numbers. 2) That you include the statement *Copyright 2004, The Endocrine Society*. Please understand that permission granted for one-time use only. Permission must be requested separately for future editions, revisions, translations, derivative works, and promotional pieces.

Name: Adrienne Weber
Title: Journal Publications Coordinator
Date: January 10, 2005

REFERENCES

"Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group." 2000. Lancet **355**(9207): 873-6.

"The International HapMap Project." 2003. Nature **426**(6968): 789-96.

Abe, T., H. Takino, et al. (1999). "CTLA4 gene polymorphism correlates with the mode of onset and presence of ICA512 Ab in Japanese type 1 diabetes." Diabetes Res. Clin Pract **46**(2): 169-75.

Abe, T., Y. Yamaguchi, et al. (2001). "CTLA4 gene polymorphism contributes to the mode of onset of diabetes with antiglutamic acid decarboxylase antibody in Japanese patients: genetic analysis of diabetic patients with antiglutamic acid decarboxylase antibody." Diabet Med **18**(9): 726-31.

Ahmed, S., K. Ihara, et al. (2001). "Association of CTLA-4 but not CD28 gene polymorphisms with systemic lupus erythematosus in the Japanese population." Rheumatology (Oxford) **40**(6): 662-7.

Akamizu, T., M. M. Sale, et al. (2000). "Association of autoimmune thyroid disease with microsatellite markers for the thyrotropin receptor gene and CTLA-4 in Japanese patients." Thyroid **10**(10): 851-8.

Alegre, M. L., H. Shiels, et al. (1998). "Expression and function of CTLA-4 in Th1 and Th2 cells." J Immunol **161**(7): 3347-56.

Altshuler, D., J. N. Hirschhorn, et al. (2000). "The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes." Nat Genet. **26**(1): 76-80.

Anderson, M. S. and J. A. Bluestone (2005). "The NOD Mouse: A Model of Immune Dysregulation." Annu Rev Immunol. **23**: 447-85

Anjos S.M., Shao W., et al. (2005). "Allelic effects on gene regulation at the autoimmunity-predisposing CTLA4 locus: a re-evaluation of the 3' +6230G>A polymorphism." Genes and Immunity **6**: 305-11

Anjos, S. and C. Polychronakos (2004). "Mechanisms of genetic susceptibility to type I diabetes: beyond HLA." Mol Genet Metab **81**(3): 187-95.

Anjos, S. M., M. C. Tessier, et al. (2004). "Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block." J Clin Endocrinol Metab **89**(12): 6257-65.

Elgar, G. (1996). "Quality not quantity: the pufferfish genome." Hum Mol Genet **5** Spec No: 1437-42.

Ellgaard, L., M. Molinari, et al. (1999). "Setting the standards: quality control in the secretory pathway." Science **286**(5446): 1882-8.

Ellis, J. H., M. N. Burden, et al. (1996). "Interactions of CD80 and CD86 with CD28 and CTLA4." J Immunol **156**(8): 2700-9.
Fajans, S. S., G. I. Bell, et al. (2001). "Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young." N Engl J Med **345**(13): 971-80.

Fajardy, I., A. Vambergue, et al. (2002). "CTLA-4 49 A/G dimorphism and type 1 diabetes susceptibility: a French case-control study and segregation analysis. Evidence of a maternal effect." Eur J Immunogenet **29**(3): 251-7.

Fallarino, F., U. Grohmann, et al. (2003). "Modulation of tryptophan catabolism by regulatory T cells." Nat Immunol **4**(12): 1206-12.

Fehervari, Z. and S. Sakaguchi (2004). "CD4+ Tregs and immune control." J Clin Invest **114**(9): 1209-17.

Fingerhut, A., S. Reutrakul, et al. (2004). "Partial deficiency of thyroxinebinding globulin-Allentown is due to a mutation in the signal peptide." J Clin Endocrinol Metab **89**(5): 2477-83.

Fujisawa, T., H. Ikegami, et al. (1995). "Class I HLA is associated with age-at onset of IDDM, while class II HLA confers susceptibility to IDDM." Diabetologia **38**(12): 1493-5.

Fukazawa, T., T. Yanagawa, et al. (1999). "CTLA-4 gene polymorphism may modulate disease in Japanese multiple sclerosis patients." J Neurol Sci **171**(1): 49-55.

Gambineri, E., T. R. Torgerson, et al. (2003). "Immune dysregulation, polyendocrinopathy, enteropathy, and X-linked inheritance (IPEX), a syndrome of systemic autoimmunity caused by mutations of FOXP3, a critical regulator of T-cell homeostasis." Curr Opin Rheumatol **15**(4): 430-5.

Gavel, Y. and G. von Heijne (1990). "Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering." Protein Eng **3**(5): 433-42.

Genc, S., K. Genc, et al. (2004). "Analysis of cytotoxic T lymphocyte antigen-4 (CTLA-4) exon 1 polymorphism in patients with type 1 diabetes mellitus in a Turkish population." J Pediatr Endocrinol Metab **17**(5): 731-5.7

Giannoukakis, N., C. Deal, et al. (1993). "Parental genomic imprinting of the human IGF2 gene." Nat Genet **4**(1): 98-101.

Girvin, A. M., M. C. Dal Canto, et al. (2000). "A critical role for B7/CD28 costimulation in experimental autoimmune encephalomyelitis: a comparative study using costimulatory molecule-deficient mice and monoclonal antibody blockade." J Immunol **164**(1): 136-43.

Giscombe, R., X. Wang, et al. (2002). "Coding sequence 1 and promoter single nucleotide polymorphisms in the CTLA-4 gene in Wegener's granulomatosis." J Rheumatol **29**(5): 950-3.

Glazier, A. M., J. H. Nadeau, et al. (2002). "Finding genes that underlie complex

traits." Science **298**(5602): 2345-9.

Graham, J., W. A. Hagopian, et al. (2002). "Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes." Diabetes **51**(5): 1346-55.

Green, A., E. A. Gale, et al. (1992). "Incidence of childhood-onset insulin dependent diabetes mellitus: the EURODIAB ACE Study." Lancet **339**(8798): 905-9.

Grogg, D., S. Hahn, et al. (1992). "CD4+ T cell-mediated killing of major histocompatibility complex class II-positive antigen-presenting cells (APC). III. CD4+ cytotoxic T cells induce apoptosis of APC." Eur J Immunol **22**(1): 267-72.

Grohmann, U., C. Orabona, et al. (2002). "CTLA-4-Ig regulates tryptophan catabolism in vivo." Nat Immunol **3**(11): 1097-101.

Gross, J. A., E. Callas, et al. (1992). "Identification and distribution of the costimulatory receptor CD28 in the mouse." J Immunol **149**(2): 380-8.

Hadj Kacem, H., M. Bellassoued, et al. (2001). "CTLA-4 gene polymorphisms in Tunisian patients with Graves' disease." Clin Immunol **101**(3): 361-5.

Hanifi Moghaddam, P., P. de Knijf, et al. (1998). "Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry." Diabetes **47**(2): 263-9.

Hardison, R. C., J. Oeltjen, et al. (1997). "Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome."Genome Res **7**(10): 959-66.

Hayashi, H., I. Kusaka, et al. (1999). "Association of CTLA-4 polymorphism with positive anti-GAD antibody in Japanese subjects with type 1 diabetes mellitus." Clin Endocrinol (Oxf) **51**(6): 793-9.
8
He, T. C., A. B. Sparks, et al. (1998). "Identification of c-MYC as a target of the APC pathway." Science **281**(5382): 1509-12.

Heward, J. M., A. Allahabadia, et al. (1998). "No evidence for allelic association of a human CTLA-4 promoter polymorphism with autoimmune thyroid disease in either population-based case-control or family-based studies." Clin Endocrinol (Oxf) 49(3): 331-4.

Heward, J. M., A. Allahabadia, et al. (1999). "The development of Graves' disease and the CTLA-4 gene on chromosome 2q33." J Clin Endocrinol Metab 84(7): 2398-401.

Hirschhorn, J. N. (2003). "Genetic epidemiology of type 1 diabetes." Pediatr Diabetes 4(2): 87-100.

Holl, R. W., B. Bohm, et al. (1999). "Thyroid autoimmunity in children and adolescents with type 1 diabetes mellitus. Effect of age, gender and HLA type." Horm Res **52**(3): 113-8.

Holopainen, P. M. and J. A. Partanen (2001). "Technical note: linkage disequilibrium and disease-associated CTLA4 gene polymorphisms." J Immunol **167**(5):2457-8.

Holst, B., A. W. Bruun, et al. (1996). "Competition between folding and glycosylation in the endoplasmic reticulum." Embo J **15**(14): 3538-46.

Hoogendoorn, B., S. L. Coleman, et al. (2003). "Functional analysis of human promoter polymorphisms." Hum Mol Genet **12**(18): 2249-54.

Hoogendoorn, B., S. L. Coleman, et al. (2004). "Functional analysis of polymorphisms in the promoter regions of genes on 22q11." Hum Mutat **24**(1): 35-42.

Hoover, M. L. and R. T. Marta (1997). "Molecular modelling of HLA-DQ suggests a mechanism of resistance in type 1 diabetes." Scand J Immunol **45**(2): 193-202.

Horn, G. T., T. L. Bugawan, et al. (1988). "Allelic sequence variation of the HLA-DQ loci: relationship to serology and to insulin-dependent diabetes susceptibility."Proc Natl Acad Sci U S A **85**(16): 6012-6.

Hovanes, K., T. W. Li, et al. (2001). "Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer." Nat Genet **28**(1):53-7.

Hu, Y., Y. Nakagawa, et al. (1992). "Functional changes in salivary glands of autoimmune disease-prone NOD mice." Am J Physiol **263**(4 Pt 1): E607-14.
9

Huang, D., L. Liu, et al. (1998). "Genetic association of Ctla-4 to myasthenia gravis with thymoma." J Neuroimmunol **88**(1-2): 192-8.

Huang, D., R. Giscombe, et al. (2000). "Dinucleotide repeat expansion in the CTLA-4 gene leads to T cell hyper-reactivity via the CD28 pathway in myasthenia gravis." J Neuroimmunol **105**(1): 69-77.

Hudson, L. L., K. Rocca, et al. (2002). "CTLA-4 gene polymorphisms in systemic lupus erythematosus: a highly significant association with a determinant in the promoter region." Hum Genet **111**(4-5): 452-5.

Hudson, T. J. (2003). "Wanted: regulatory SNPs." Nat Genet **33**(4): 439-40.

Hurwitz, A. A., T. J. Sullivan, et al. (1997). "Specific blockade of CTLA-4/B7 interactions results in exacerbated clinical and histologic disease in an actively-induced model of experimental allergic encephalomyelitis." J Neuroimmunol **73**(1-2): 57-62.

Ihara, K., S. Ahmed, et al. (2001). "Association studies of CTLA-4, CD28, and ICOS gene polymorphisms with type 1 diabetes in the Japanese population." Immunogenetics **53**(6): 447-54.

Ito, M., Y. Oiso, et al. (1993). "Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes

insipidus." J Clin Invest **91**(6): 2565-71.

Jenkins, R. C. and A. P. Weetman (2002). "Disease associations with autoimmune thyroid disease." Thyroid **12**(11): 977-88.

Johansson, S., B. A. Lie, et al. (2003). "Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1." Genes Immun **4**(1): 46-53.

Johansson, S., B. A. Lie, et al. (2003). "HLA associations in type 1 diabetes: DPB1 alleles may act as markers of other HLA-complex susceptibility genes." Tissue Antigens 61(5): 344-51.

Johnson, G. C., L. Esposito, et al. (2001). "Haplotype tagging for the identification of common disease genes." Nat Genet 29(2): 233-7.

Julier, C., R. N. Hyer, et al. (1991). "Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility." Nature **354**(6349): 155-9.10

Kaijzel, E. L., J. P. Bayley, et al. (2001). "Allele-specific quantification of tumor necrosis factor alpha (TNF) transcription and the role of promoter polymorphisms in rheumatoid arthritis patients and healthy individuals." Genes Immun **2**(3): 135-44.

Karandikar, N. J., C. L. Vanderlugt, et al. (1996). "CTLA-4: a negative regulator of autoimmune disease." J Exp Med **184**(2): 783-8.

Karandikar, N. J., T. N. Eagar, et al. (2000). "CTLA-4 downregulates epitope spreading and mediates remission in relapsing experimental autoimmune encephalomyelitis." J Neuroimmunol **109**(2): 173-80.

Karaplis, A. C., S. K. Lim, et al. (1995). "Inefficient membrane targeting, translocation, and proteolytic processing by signal peptidase of a mutant preproparathyroid hormone protein." J Biol Chem **270**(4): 1629-35.

Karplus, K., C. Barrett, et al. (1998). "Hidden Markov models for detecting remote protein homologies." Bioinformatics **14**(10): 846-56.

Kasturi, L., J. R. Eshleman, et al. (1995). "The hydroxy amino acid in an Asn-XSer/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein." J Biol Chem **270**(24): 14756-61.

Katayama, T., K. Imaizumi, et al. (2004). "Induction of neuronal death by ER stress in Alzheimer's disease." J Chem Neuroanat **28**(1-2): 67-78.

Kato, M., E. Nanba, et al. (2000). "Sonic hedgehog signal peptide mutation in a patient with holoprosencephaly." Ann Neurol **47**(4): 514-6.

Kemp, E. H., R. A. Ajjan, et al. (1998). "A cytotoxic T lymphocyte antigen-4 (CTLA-4) gene polymorphism is associated with autoimmune Addison's disease in English patients." Clin Endocrinol (Oxf) **49**(5): 609-13.

Khalil, I., L. d'Auriol, et al. (1990). "A combination of HLA-DQ beta Asp57-negative and HLA DQ alpha Arg52 confers susceptibility to insulin-dependent diabetes mellitus." J Clin Invest **85**(4): 1315-9.

Khoury, S. J., E. Akalin, et al. (1995). "CD28-B7 costimulatory blockade by CTLA4Ig prevents actively induced experimental autoimmune encephalomyelitis and inhibits Th1 but spares Th2 cytokines in the central nervous system." J Immunol **155**(10):4521-4.

Kikuoka, N., S. Sugihara, et al. (2001). "Cytotoxic T lymphocyte antigen 4 gene polymorphism confers susceptibility to type 1 diabetes in Japanese children: analysis of association with HLA genotypes and autoantibodies." Clin Endocrinol (Oxf) **55**(5): 597-603.11

Kikutani, H. and S. Makino (1992). "The murine autoimmune diabetes model: NOD and related strains." Adv Immunol **51**: 285-322.

King, A. L., S. J. Moodie, et al. (2002). "CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families." J Med Genet **39**(1): 51-4.

Klitz, W., T. L. Bugawan, et al. (2002). "Association of CTLA-4 variation with type I diabetes in Filipinos." Immunogenetics **54**(5): 310-3.

Knight, J. C. (2005). "Regulatory polymorphisms underlying complex disease traits." J Mol Med. 83:97-109

Knight, J. C., I. Udalova, et al. (1999). "A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria." Nat Genet **22**(2):145-50.

Kordonouri, O., A. Klinghammer, et al. (2002). "Thyroid autoimmunity in children and adolescents with type 1 diabetes: a multicenter survey." Diabetes Car **25**(8): 1346-50.

Kotsa, K., P. F. Watson, et al. (1997). "A CTLA-4 gene polymorphism is associated with both Graves disease and autoimmune hypothyroidism." Clin Endocrinol (Oxf) **46**(5): 551-4.

Kouki, T., C. A. Gardine, et al. (2002). "Relation of three polymorphisms of the CTLA-4 gene in patients with Graves' disease." J Endocrinol Invest **25**(3): 208-13.

Kouki, T., Y. Sawai, et al. (2000). "CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease." J Immunol **165**(11): 6606-11.

Kristiansen, O. P., Z. M. Larsen, et al. (2000). "CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity?" Genes Immun **1**(3): 170-84.

Krokowski, M., J. Bodalski, et al. (1998). "CTLA-4 gene polymorphism is associated with predisposition to IDDM in a population from central Poland." Diabetes Metab **24**(3): 241-3.

Krummel, M. F. and J. P. Allison (1996). "CTLA-4 engagement inhibits IL-2 accumulation and cell cycle progression upon activation of resting T cells." J Exp Med. **183**(6): 2533-40.

Kunkel, T. A. (1985). "Rapid and efficient site-specific mutagenesis without phenotypic selection." Proc Natl Acad Sci U S A **82**(2): 488-92.
12

Kupfer, A. and S. J. Singer (1989). "Cell biology of cytotoxic and helper T cell functions: immunofluorescence microscopic studies of single cells and cell couples."Annu Rev Immunol **7**: 309-37.

Kyewski, B., J. Derbinski, et al. (2002). "Promiscuous gene expression and central T-cell tolerance: more than meets the eye." Trends Immunol **23**(7): 364-71.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-32.

Lagerstrom-Fermer, M., M. Nilsson, et al. (1995). "Amelogenin signal peptide mutation: correlation between mutations in the amelogenin gene (AMGX) and manifestations of X-linked amelogenesis imperfecta." Genomics **26**(1): 159-62.

Lan, M. S., C. Wasserfall, et al. (1996). "IA-2, a transmembrane protein of the protein tyrosine phosphatase family, is a major autoantigen in insulin-dependent diabetes mellitus." Proc Natl Acad Sci U S A **93**(13): 6367-70.

Lanza, F., C. De La Salle, et al. (2002). "A Leu7Pro mutation in the signal peptide of platelet glycoprotein (GP)IX in a case of Bernard-Soulier syndrome abolishes surface expression of the GPIb-V-IX complex." Br J Haematol **118**(1): 260-6.

Larsen, Z. M., O. P. Kristiansen, et al. (1999). "IDDM12 (CTLA4) on 2q33 and IDDM13 on 2q34 in genetic susceptibility to type 1 diabetes (insulin-dependent)." Autoimmunity **31**(1): 35-42.

Lee, K. M., E. Chuang, et al. (1998). "Molecular basis of T cell inactivation by CTLA-4." Science **282**(5397): 2263-6.

Lee, Y. H., S. J. Choi, et al. (2002). "No association of polymorphisms of the CTLA-4 exon 1(+49) and promoter(-318) genes with rheumatoid arthritis in the Korean population." Scand J Rheumatol **31**(5): 266-70.

Lee, Y. H., Y. R. Kim, et al. (2001). "Polymorphisms of the CTLA-4 exon 1 and promoter gene in systemic lupus erythematosus." Lupus **10**(9): 601-5.

Lee, Y. J., F. S. Lo, et al. (2001). "The promoter region of the CTLA4 gene is associated with type 1 diabetes mellitus." J Pediatr Endocrinol Metab **14**(4): 383-8.

Lee, Y. J., F. Y. Huang, et al. (2000). "Association of CTLA4 gene A-G polymorphism with type 1 diabetes in Chinese children." Clin Endocrinol (Oxf) **52**(2): 153-7.

Leung, H. T., J. Bradshaw, et al. (1995). "Cytotoxic T lymphocyte-associated molecule-4, a high-avidity receptor for CD80 and CD86, contains an intracellular localization motif in its cytoplasmic tail." J Biol Chem **270**(42): 25107-14.
13

Levin, L. and Y. Tomer (2003). "The etiology of autoimmune diabetes and thyroiditis: evidence for common genetic susceptibility." Autoimmun Rev **2**(6): 377-86.

Li, Y., J. J. Bergeron, et al. (1996). "Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport." Proc Natl Acad Sci U S A **93**(18): 9606-11.

Liblau, R., P. Gajdos, et al. (1991). "Intravenous gamma-globulin in myasthenia gravis: interaction with anti-acetylcholine receptor autoantibodies." J Clin Immunol **11**(3): 128-31.

Lie, B. A., H. E. Akselsen, et al. (2002). "Polymorphisms in the gene encoding thymus-specific serine protease in the extended HLA complex: a potential candidate gene for autoimmune and HLA-associated diseases." Genes Immun **3**(5): 306-12.

Lie, B. A., J. A. Todd, et al. (1999). "The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene." Am J Hum Genet **64**(3): 793-800.

Lie, B. A., L. M. Sollid, et al. (1999). "A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease." Tissue Antigens **54**(2): 162-8.

Lieberman, S. M. and T. P. DiLorenzo (2003). "A comprehensive guide to antibody and T-cell responses in type 1 diabetes." Tissue Antigens **62**(5): 359-77.

Ligers, A., C. Xu, et al. (1999). "The CTLA-4 gene is associated with multiple sclerosis." J Neuroimmunol **97**(1-2): 182-90.

Ligers, A., N. Teleshova, et al. (2001). "CTLA-4 gene expression is influenced by promoter and exon 1 polymorphisms." Genes Immun **2**(3): 145-52.

Like, A. A., D. L. Guberski, et al. (1991). "Influence of environmental viral agents on frequency and tempo of diabetes mellitus in BB/Wor rats." Diabetes **40**(2): 259-62.

Lin, H., J. C. Rathmell, et al. (1998). "Cytotoxic T lymphocyte antigen 4 (CTLA4) blockade accelerates the acute rejection of cardiac allografts in CD28-deficient mice: CTLA4 can function independently of CD28." J Exp Med **188**(1): 199-204

Lindsten, T., K. P. Lee, et al. (1993). "Characterization of CTLA-4 structure and expression on human T cells." J Immunol **151**(7): 3489-99.

Lindsten, T., K. P. Lee, et al. (1993). "Characterization of CTLA-4 structure and expression on human T cells." J Immunol **151**(7): 3489-99.
14

Ling, V., P. W. Wu, et al. (1999). "Complete sequence determination of the mouse and human CTLA4 gene loci: cross-species DNA sequence similarity beyond exon borders." Genomics **60**(3): 341-55.

Ling, V., P. W. Wu, et al. (2001). "Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches." Genomics **78**(3):155-68.

Link, H., J. B. Sun, et al. (1992). "Virus-reactive and autoreactive T cells are accumulated in cerebrospinal fluid in multiple sclerosis." J Neuroimmunol **38**(1-2): 63-73.

Linsley, P. S., J. Bradshaw, et al. (1996). "Intracellular trafficking of CTLA-4 and focal localization towards sites of TCR engagement." Immunity **4**(6): 535-43.

Liu, M. F., C. R. Wang, et al. (2003). "Increased expression of soluble cytotoxic T-lymphocyte-associated antigen-4 molecule in patients with systemic lupus erythematosus." Scand J Immunol **57**(6): 568-72.

Lohmueller, K. E., C. L. Pearce, et al. (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-82.

Lowe, R. M., J. Graham, et al. (2000). "The length of the CTLA-4 microsatellite (AT)N-repeat affects the risk for type 1 diabetes. Diabetes Incidence in Sweden Study Group." Autoimmunity **32**(3): 173-80.

Luhder, F., C. Chambers, et al. (2000). "Pinpointing when T cell costimulatory receptor CTLA-4 must be engaged to dampen diabetogenic T cells." Proc Natl Acad Sci U S A **97**(22): 12204-9.

Luhder, F., P. Hoglund, et al. (1998). "Cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) regulates the unfolding of autoimmune diabetes." J Exp Med **187**(3):427-32.

Ma, Y., X. Tang, et al. (2002). "CTLA-4 gene A/G polymorphism associated with diabetes mellitus in Han Chinese." Chin Med J (Engl) **115**(8): 1248-50.

Maclaren, N. K., M. S. Lan, et al. (2003). "Multiple autoantibodies as predictors of Type 1 diabetes in a general population." Diabetologia **46**(6): 873-4.

Magistrelli, G., P. Jeannin, et al. (1999). "A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells." Eur J Immunol **29**(11): 3596-602.15

Marder, B. A., B. Schroppel, et al. (2003). "The impact of costimulatory molecule gene polymorphisms on clinical outcomes in liver transplantation." Am JTransplant **3**(4): 424-31.

Marengere, L. E., P. Waterhouse, et al. (1996). "Regulation of T cell receptor signaling by tyrosine phosphatase SYP association with CTLA-4." Science **272**(5265):1170-3.

Marrack, P., P. Hugo, et al. (1993). "Death and T cells." Immunol Rev **133**: 119-29.

Marron, M. P., A. Zeidler, et al. (2000). "Genetic and physical mapping of a type 1 diabetes susceptibility gene (IDDM12) to a 100-kb phagemid artificial chromosome clone containing D2S72-CTLA4-D2S105 on chromosome 2q33." Diabetes **49**(3): 492-9.

Marron, M. P., L. J. Raffel, et al. (1997). "Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups." Hum Mol Genet **6**(8): 1275-82.

Martins, T. C. and A. P. Aguas (1999). "Mechanisms of Mycobacterium avium induced resistance against insulin-dependent diabetes mellitus (IDDM) in non-obese diabetic (NOD) mice: role of Fas and Th1 cells." Clin Exp Immunol 115(2): 248-54.

Martoglio, B. (2003). "Intramembrane proteolysis and post-targeting functions of signal peptides." Biochem Soc Trans **31**(Pt 6): 1243-7.

Martoglio, B. and B. Dobberstein (1998). "Signal sequences: more than just greasy peptides." Trends Cell Biol **8**(10): 410-5.

Mathis, D. and C. Benoist (2004). "Back to central tolerance." Immunity **20**(5): 509-16.

Matoba, S. and D. M. Ogrydziak (1998). "Another factor besides hydrophobicity can affect signal peptide interaction with signal recognition particle." J Biol Chem **273**(30): 18841-7.

Maurer, M., S. Loserth, et al. (2002). "A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation." Immunogenetics **54**(1): 1-8.

McCann, J. A., H. Zheng, et al. (2001). "Evidence against GRB10 as the gene responsible for Silver-Russell syndrome." Biochem Biophys Res Commun **286**(5): 943-8.

McCormack, R. M., A. P. Maxwell, et al. (2001). "Possible association between CTLA4 DNA polymorphisms and early onset type 1 diabetes in a UK population." Genes Immun **2**(4): 233-5.16

McDevitt, H., S. Singer, et al. (1996). "The role of MHC class II genes in susceptibility and resistance to type I diabetes mellitus in the NOD mouse." Horm Metab Res **28**(6): 287-8.

McLauchlan, J., M. K. Lemberg, et al. (2002). "Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets." Embo J **21**(15):3980-8.

McLeod, J. F., L. Kovacs, et al. (1993). "Familial neurohypophyseal diabetes insipidus associated with a signal peptide mutation." J Clin Endocrinol Metab **77**(3):599A-599G.

Melanitou, E., P. Fain, et al. (2003). "Genetics of type 1A (immune mediated) diabetes." J Autoimmun **21**(2): 93-8.

Metzler, W. J., J. Bajorath, et al. (1997). "Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28." Nat Struct Biol **4**(7):527-31.

Mijovic, C. H., A. H. Barnett, et al. (1991). "Genetics of diabetes. Trans-racial gene mapping studies." Baillieres Clin Endocrinol Metab **5**(2): 321-40.

Miyamoto, S., J. A. Chiorini, et al. (1996). "Regulation of gene expression for translation initiation factor eIF-2 alpha: importance of the 3' untranslated region." Biochem J **315** (Pt 3): 791-8.

Mochizuki, M., S. Amemiya, et al. (2003). "Association of the CTLA-4 gene 49 a/g polymorphism with type 1 diabetes and autoimmune thyroid disease in Japanese children." Diabetes Care **26**(3): 843-7.

Morahan, G., D. Huang, et al. (1996). "Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus." Science **272**(5269): 1811-3.

Morel, P. A., J. S. Dorman, et al. (1988). "Aspartic acid at position 57 of the HLA-DQ beta chain protects against type I diabetes: a family study." Proc Natl Acad Sci U S A **85**(21): 8111-5.

Morley, M., C. M. Molony, et al. (2004). "Genetic analysis of genome-wide variation in human gene expression." Nature **430**(7001): 743-7.

Munn, D. H., M. D. Sharma, et al. (2004). "Expression of indoleamine 2,3-dioxygenase by plasmacytoid dendritic cells in tumor-draining lymph nodes." J Clin Invest **114** (2): 280-90.17

Nakanishi, K., T. Kobayashi, et al. (1997). "Synchronous decline of serum soluble HLA class I antigen and beta-cell function in insulin-dependent diabetes mellitus." Clin Immunol Immunopathol **85**(3): 246-52.

Ng, D. T., J. D. Brown, et al. (1996). "Signal sequences specify the targeting route to the endoplasmic reticulum membrane." J Cell Biol **134**(2): 269-78.

Nielsen, H., J. Engelbrecht, et al. (1997). "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Eng **10**(1): 1-6.

Nistico, L., R. Buzzetti, et al. (1996). "The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry." Hum Mol Genet **5**(7): 1075-80.

Nithiyananthan, R., J. M. Heward, et al. (2002). "Polymorphism of the CTLA-4 gene is associated with autoimmune hypothyroidism in the United Kingdom." Thyroid **12**(1): 3-6.

Notkins, A. L. (2002). "Immunologic and genetic factors in type 1 diabetes." J Biol Chem **277**(46): 43545-8.

Oaks, M. K. and K. M. Hallett (2000). "Cutting edge: a soluble form of CTLA-4 in patients with autoimmune thyroid disease." J Immunol **164**(10): 5015-8.

Oaks, M. K., K. M. Hallett, et al. (2000). "A native soluble form of CTLA-4." Cell Immunol **201**(2): 144-53.

Ohinmaa, A., P. Jacobs, et al. (2004). "The Projection of Prevalence and Cost of Diabetes in Canada: 2000 to 2016." Canadian Journal of Diabetes **28**(2): 116-123.

Oldstone, M. B. (1990). "Viruses as therapeutic agents. I. Treatment of nonobese insulin-dependent diabetes mice with virus prevents insulin-dependent diabetes mellitus while maintaining general immune competence." J Exp Med **171**(6): 2077-89.

Onengut-Gumuscu, S. and P. Concannon (2002). "Mapping genes for autoimmunity in humans: type 1 diabetes as a model." Immunol Rev **190**: 182-94.

Ongagna, J. C., R. Sapin, et al. (2002). "Markers for risk of type 1 diabetes in relatives of Alsacian patients with type 1 diabetes." Int J Exp Diabetes Res **3**(1): 1-9.

Otvos, L., Jr. and M. Cudic (2002). "Post-translational modifications in prion proteins." Curr Protein Pept Sci **3**(6): 643-52.

Owerbach, D. and K. H. Gabbay (1993). "Localization of a type I diabetes susceptibility locus to the variable tandem repeat region flanking the insulin gene."Diabetes **42**(12): 1708-14.18

Owerbach, D., F. J. Naya, et al. (1997). "Analysis of candidate genes for susceptibility to type I diabetes: a case-control and family-association study of genes on chromosome 2q31-35." Diabetes **46**(6): 1069-74.

Panagiotopoulos, C., H. Qin, et al. (2003). "Identification of a beta-cell-specific HLA class I restricted epitope in type 1 diabetes." Diabetes **52**(11): 2647-51.

Paquette, J., N. Giannoukakis, et al. (1998). "The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans." J Biol Chem **273**(23):14158-64.

Parrott, C. L., N. Alsayed, et al. (1992). "ApoC-IIParis2: a premature termination mutation in the signal peptide of apoC-II resulting in the familial chylomicronemia syndrome." J Lipid Res **33**(3): 361-7

Pastinen, T. and T. J. Hudson (2004). "Cis-acting regulatory variation in the human genome." Science **306**(5696): 647-50.

Pastinen, T., R. Sladek, et al. (2004). "A survey of genetic and epigenetic variation affecting human gene expression." Physiol Genomics **16**(2): 184-93.

Pentcheva-Hoang, T., J. G. Egen, et al. (2004). "B7-1 and b7-2 selectively recruit ctla-4 and CD28 to the immunological synapse." Immunity **21**(3): 401-13.

Perez, V. L., L. Van Parijs, et al. (1997). "Induction of peripheral T cell tolerance

in vivo requires CTLA-4 engagement." Immunity **6**(4): 411-7.

Perkins, D., Z. Wang, et al. (1996). "Regulation of CTLA-4 expression during T cell activation." J Immunol **156**(11): 4154-9.

Perrin, P. J., J. H. Maldonado, et al. (1996). "CTLA-4 blockade enhances clinical disease and cytokine production during experimental allergic encephalomyelitis." J Immunol **157**(4): 1333-6.

Phan, G. Q., J. C. Yang, et al. (2003). "Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma." Proc Natl Acad Sci U S A **100**(14): 8372-8377.

Pitkaniemi, J., T. Hakulinen, et al. (2004). "Class I and II HLA genes are associated with susceptibility and age at onset in Finnish families with type 1 diabetes."Hum Hered **57**(2): 69-79.

Pociot, F. and M. F. McDermott (2002). "Genetics of type 1 diabetes mellitus." Genes Immun **3**(5): 235-49.19

Polychronakos, C. (2001). Programmed Cell Death in the Pathogenesis of Autoimmune Diabetes. Programmed Cell Death, Vol II. M. Mattson, S. Estus and V. Rangnekar, Elsevier. II: 55-79.

Poser, S., B. Stickel, et al. (1989). "Increasing incidence of multiple sclerosis in South Lower Saxony, Germany." Neuroepidemiology **8**(4): 207-13.

Pugliese, A. and G. S. Eisenbarth (2000). Chapter 7: Type I Diabetes Mellitus of Man: Genetic Susceptibility and Resistance. Type I Diabetes: Molecular, Cellular and Clinical Immunology. Oxford.

Pugliese, A., M. Zeller, et al. (1997). "The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR IDDM2 susceptibility locus for type 1 diabetes." Nat Genet **15**(3): 293-7.

Pugliese, A., R. Gianani, et al. (1995). "HLA-DQB1*0602 is associated with dominant protection from diabetes even among islet cell antibody-positive first-degree relatives of patients with IDDM." Diabetes **44**(6): 608-13.

Qiu, P., G. J. Soder, et al. (2003). "Quantification of single nucleotide polymorphisms by automated DNA sequencing." Biochem Biophys Res Commun **309**(2): 331-8.

Quandt, K., K. Frech, et al. (1995). "MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data." Nucleic Acids Res **23**(23): 4878-84.

Racchi, M., H. H. Watzke, et al. (1993). "Human coagulation factor X deficiency caused by a mutant signal peptide that blocks cleavage by signal peptidase but not targeting and translocation to the endoplasmic reticulum." J Biol Chem **268**(8): 5735-40.

Reik, W. and J. Walter (2001). "Genomic imprinting: parental influence on the genome." Nat Rev Genet **2**(1): 21-32.

Repaske, D. R., R. Medlej, et al. (1997). "Heterogeneity in clinical manifestation of autosomal dominant neurohypophyseal diabetes insipidus caused by a mutation encoding Ala-1-->Val in the signal peptide of the arginine vasopressin/neurophysin II/copeptin precursor." J Clin Endocrinol Metab **82**(1): 51-6.

Rewers, M., T. L. Bugawan, et al. (1996). "Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY)." Diabetologia **39**(7): 807-12.

Risch, N. (1987). "Assessing the role of HLA-linked and unlinked determinants of disease." Am J Hum Genet **40**(1): 1-14.
20

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.

Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-56.

Robinson, W. P., J. Barbosa, et al. (1993). "Homozygous parent affected sib pair method for detecting disease predisposing variants: application to insulin dependent diabetes mellitus." Genet Epidemiol **10**(5): 273-88.

Ronningen, K. S., A. Spurkland, et al. (1991). "Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus." Tissue Antigens **37**(3): 105-11.

Rosati, G., I. Aiello, et al. (1988). "Incidence of multiple sclerosis in the town of Sassari, Sardinia, 1965 to 1985: evidence for increasing occurrence of the disease." Neurology **38**(3): 384-8.

Rutkowski, D. T., C. M. Ott, et al. (2003). "Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen." J Biol Chem **278**(32): 30365-72.

Saeki, K., M. Zhu, et al. (2002). "Targeted disruption of the protein tyrosine phosphatase-like molecule IA-2 results in alterations in glucose tolerance tests and insulin secretion." Diabetes **51**(6): 1842-50.

Sakaguchi, S. (2004). "Naturally arising CD4+ regulatory t cells for immunologic self-tolerance and negative control of immune responses." Annu Rev Immunol **22**: 531-62.

Sale, M. M., T. Akamizu, et al. (1997). "Association of autoimmune thyroid disease with a microsatellite marker for the thyrotropin receptor gene and CTLA-4 in a Japanese population." Proc Assoc Am Physicians **109**(5): 453-61.

Salomon, B. and J. A. Bluestone (2001). "Complexities of CD28/B7: CTLA-4 costimulatory pathways in autoimmunity and transplantation." Annu Rev Immunol **19**: 225-52.

Salomon, B., D. J. Lenschow, et al. (2000). "B7/CD28 costimulation is essential

for the homeostasis of the CD4+CD25+ immunoregulatory T cells that control autoimmune diabetes." Immunity **12**(4): 431-40.
Salomon, B., L. Rhee, et al. (2001). "Development of spontaneous autoimmune peripheral polyneuropathy in B7-2-deficient NOD mice." J Exp Med **194**(5): 677-84.

Sanjeevi, C. B., C. DeWeese, et al. (1997). "Analysis of critical residues of HLADQ6 molecules in insulin-dependent diabetes mellitus." Tissue Antigens **50**(1): 61-5.21

Schifferli, K., Ciccarone, V (1996). "Optimization of cationic lipid reagent mediated transfection for suspension cell lines." FOCUS **18**(2): 45-47.

Schneider, H., S. da Rocha Dias, et al. (2001). "A regulatory role for cytoplasmic YVKM motif in CTLA-4 inhibition of TCR signaling." Eur J Immunol **31**(7): 2042-50.

Sekigawa, I., T. Naito, et al. (2004). "Possible mechanisms of gender bias in SLE: a new hypothesis involving a comparison of SLE with atopy." Lupus **13**(4): 217-22.

Shakin-Eshleman, S. H., S. L. Spitalnik, et al. (1996). "The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked coreglycosylation efficiency." J Biol Chem **271**(11): 6363-6.
Shiratori, T., S. Miyatake, et al. (1997). "Tyrosine phosphorylation controls internalization of CTLA-4 by regulating its interaction with clathrin-associated adaptor complex AP-2." Immunity **6**(5): 583-9.

Shizuru, J. A., C. Taylor-Edwards, et al. (1988). "Immunotherapy of the nonobese diabetic mouse: treatment with an antibody to T-helper lymphocytes." Science **240**(4852): 659-62.

Shtutman, M., J. Zhurinsky, et al. (1999). "The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway." Proc Natl Acad Sci U S A **96**(10): 5522-7.

Siggaard, C., S. Rittig, et al. (1999). "Clinical and molecular evidence of abnormal processing and trafficking of the vasopressin preprohormone in a large kindred with familial neurohypophyseal diabetes insipidus due to a signal peptide mutation." J Clin Endocrinol Metab **84**(8): 2933-41.

Simmonds, M. J. and S. C. Gough (2004). "Unravelling the genetic complexity of autoimmune thyroid disease: HLA, CTLA-4 and beyond." Clin Exp Immunol **136**(1):1-10.

Solimena, M., R. Dirkx, Jr., et al. (1996). "ICA 512, an autoantigen of type I diabetes, is an intrinsic membrane protein of neurosecretory granules." Embo J **15**(9): 2102-14.

Spielman, R. S., R. E. McGinnis, et al. (1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)."Am J Hum Genet **52**(3): 506-16.

Spiess, M. (1995). "Heads or tails--what determines the orientation of proteins in the membrane." FEBS Lett **369**(1): 76-9.22

Staal, F. J. and H. C. Clevers (2003). "Wnt signaling in the thymus." Curr Opin Immunol **15**(2): 204-8.

Stitt, A. W., N. Frizzell, et al. (2004). "Advanced glycation and advanced lipoxidation: possible role in initiation and progression of diabetic retinopathy." Curr Pharm Des **10**(27): 3349-60.

Sun, J. B., T. Olsson, et al. (1991). "Autoreactive T and B cells responding to myelin proteolipid protein in multiple sclerosis and controls." Eur J Immunol **21**(6):1461-8.

Sunthornthepvarakul, T., S. Churesigaew, et al. (1999). "A novel mutation of the signal peptide of the preproparathyroid hormone gene associated with autosomal recessive familial isolated hypoparathyroidism." J Clin Endocrinol Metab **84**(10): 3792-6.

Takara, M., I. Komiya, et al. (2000). "Association of CTLA-4 gene A/G polymorphism in Japanese type 1 diabetic patients with younger age of onset and autoimmune thyroid disease." Diabetes Care **23**(7): 975-8.

Thomas, P. D. and A. Kejariwal (2004). "Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects." Proc Natl Acad Sci U S A **101**(43): 15398-403.

Thomson, G., W. P. Robinson, et al. (1988). "Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus." Am J Hum Genet **43**(6): 799-816.

Tisch, R. and H. McDevitt (1996). "Insulin-dependent diabetes mellitus." Cell **85**(3): 291-7.

Tivol, E. A., F. Borriello, et al. (1995). "Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4." Immunity **3**(5): 541-7.

Tivol, E. A., S. D. Boyd, et al. (1997). "CTLA4Ig prevents lymphoproliferation and fatal multiorgan tissue destruction in CTLA-4-deficient mice." J Immunol **158**(11): 5091-4.

Todd, J. A. and M. Farrall (1996). "Panning for gold: genome-wide scanning for linkage in type 1 diabetes." Hum Mol Genet **5** Spec No: 1443-8.

Todd, J. A., C. Mijovic, et al. (1989). "Identification of susceptibility loci for insulin-dependent diabetes mellitus by trans-racial gene mapping." Nature **338**(6216):587-9.23

Todd, J. A., J. I. Bell, et al. (1987). "HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus." Nature **329**(6140):599-604.

Todd, J. A., J. I. Bell, et al. (1988). "HLA antigens and insulin-dependent diabetes." Nature **333**(6175): 710.

Tomer, Y. and T. F. Davies (2003). "Searching for the autoimmune thyroid disease susceptibility genes: from gene mapping to gene function." Endocr Rev **24**(5):694-717.

Tomer, Y., D. A. Greenberg, et al. (2001). "CTLA-4 and not CD28 is a susceptibility gene for thyroid autoantibody production." J Clin Endocrinol Metab **86**(4):1687-93.

Ueda, H., J. M. Howson, et al. (2003). "Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease." Nature **423**(6939): 506-1.

Undlien, D. E., S. T. Bennett, et al. (1995). "Insulin gene region-encoded susceptibility to IDDM maps upstream of the insulin gene." Diabetes **44**(6): 620-5.

Vafiadis, P., H. Ounissi-Benkalha, et al. (2001). "Class III alleles of the variable number of tandem repeat insulin polymorphism associated with silencing of thymic insulin predispose to type 1 diabetes." J Clin Endocrinol Metab **86**(8): 3705-10.

Vafiadis, P., S. T. Bennett, et al. (1996). "Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes." J Autoimmun **9**(3):397-403.

Vafiadis, P., S. T. Bennett, et al. (1997). "Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus." Nat Genet **15**(3): 289-92.

Vafiadis, P., S. T. Bennett, et al. (1998). "Divergence between genetic determinants of IGF2 transcription levels in leukocytes and of IDDM2-encoded susceptibility to type 1 diabetes." J Clin Endocrinol Metab **83**(8): 2933-9.

Vaidya, B. and S. Pearce (2004). "The emerging role of the CTLA-4 gene in autoimmune endocrinopathies." Eur J Endocrinol **150**(5): 619-26.

Vaidya, B., H. Imrie, et al. (1999). "Cytotoxic T lymphocyte antigen-4 (CTLA-4) gene polymorphism confers susceptibility to thyroid associated orbitopathy." Lancet **354**(9180): 743-4.

Vaidya, B., H. Imrie, et al. (1999). "The cytotoxic T lymphocyte antigen-4 is a major Graves' disease locus." Hum Mol Genet **8**(7): 1195-9.
24

Vaidya, B., S. H. Pearce, et al. (2002). "An association between the CTLA4 exon 1 polymorphism and early rheumatoid arthritis with autoimmune endocrinopathies." Rheumatology (Oxford) **41**(2): 180-3.

Van der Auwera, B. J., C. L. Vandewalle, et al. (1997). "CTLA-4 gene polymorphism confers susceptibility to insulin-dependent diabetes mellitus (IDDM) independently from age and from other genetic or immune disease markers. The Belgian Diabetes Registry." Clin Exp Immunol **110**(1): 98-103.

Van der Auwera, B., F. Schuit, et al. (1995). "Genetic susceptibility for insulindependent diabetes mellitus in Caucasians revisited: the importance of

diabetes registries in disclosing interactions between HLA-DQ- and insulin gene-linked risk. Belgian
Diabetes Registry." J Clin Endocrinol Metab **80**(9): 2567-73.

van Noort, M. and H. Clevers (2002). "TCF transcription factors, mediators of Wnt-signaling in development and cancer." Dev Biol **244**(1): 1-8.

von Heijne, G. (1985). "Signal sequences. The limits of variation." J Mol Biol **184**(1): 99-105.

Walunas, T. L. and J. A. Bluestone (1998). "CTLA-4 regulates tolerance induction and T cell differentiation in vivo." J Immunol **160**(8): 3855-60.

Walunas, T. L., C. Y. Bakker, et al. (1996). "CTLA-4 ligation blocks CD28-dependent T cell activation." J Exp Med **183**(6): 2541-50.

Walunas, T. L., D. J. Lenschow, et al. (1994). "CTLA-4 can function as a negative regulator of T cell activation." Immunity **1**(5): 405-13.

Wang, X. B., M. Kakoulidou, et al. (2002). "Abnormal expression of CTLA-4 by T cells from patients with myasthenia gravis: effect of an AT-rich gene sequence." J Neuroimmunol **130**(1-2): 224-32.

Wang, X. B., M. Kakoulidou, et al. (2002). "CDS1 and promoter single nucleotide polymorphisms of the CTLA-4 gene in human myasthenia gravis." Genes Immun **3**(1): 46-9.

Wang, X. B., X. Zhao, et al. (2002). "A CTLA-4 gene polymorphism at position -318 in the promoter region affects the expression of protein." Genes Immun **3**(4): 233-4.

Waterhouse, P., J. M. Penninger, et al. (1995). "Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4." Science **270**(5238): 985-8.

Weihofen, A. and B. Martoglio (2003). "Intramembrane-cleaving proteases: controlled liberation of proteins and bioactive peptides." Trends Cell Biol **13**(2): 71-8. 25

Weiner, J. H., P. T. Bilous, et al. (1998). "A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins." Cell **93**(1): 93-101.

Wicker, L. S., B. J. Miller, et al. (1986). "Transfer of autoimmune diabetes mellitus with splenocytes from nonobese diabetic (NOD) mice." Diabetes **35**(8): 855-60.

Wilberz, S., H. J. Partke, et al. (1991). "Persistent MHV (mouse hepatitis virus) infection reduces the incidence of diabetes mellitus in non-obese diabetic mice." Diabetologia **34**(1): 2-5.

Williams, G. and J. C. Pickup (2001). Handbook of Diabetes. Oxford, Blackwell Sciences Ltd.

Wilson, A. G., J. A. Symons, et al. (1997). "Effects of a polymorphism in the

human tumor necrosis factor alpha promoter on transcriptional activation." Proc Natl
Acad Sci U S A **94**(7): 3195-9.

Witt, H., W. Luck, et al. (1999). "A signal peptide cleavage site mutation in the cationic trypsinogen gene is strongly associated with chronic pancreatitis." Gastroenterology **117**(1): 7-10.

Xu, Y., P. N. Graves, et al. (2002). "CTLA-4 and autoimmune thyroid disease: lack of influence of the A49G signal peptide polymorphism on functional recombinant human CTLA-4." Cell Immunol **215**(2): 133-40.

Yan, H., W. Yuan, et al. (2002). "Allelic variation in human gene expression." Science **297**(5584): 1143.

Yanagawa, T., M. Taniyama, et al. (1997). "CTLA4 gene polymorphism confers susceptibility to Graves' disease in Japanese." Thyroid **7**(6): 843-6.

Yanagawa, T., T. Maruyama, et al. (1999). "Lack of association between CTLA-4 gene polymorphism and IDDM in Japanese subjects." Autoimmunity **29**(1): 53-6.

Yanagawa, T., Y. Hidaka, et al. (1995). "CTLA-4 gene polymorphism associated with Graves' disease in a Caucasian population." J Clin Endocrinol Metab **80**(1): 41-5.

Yung, E., P. S. Cheng, et al. (2002). "CTLA-4 gene A-G polymorphism and childhood Graves' disease." Clin Endocrinol (Oxf) **56**(5): 649-53.

Zhang, Y. and J. P. Allison (1997). "Interaction of CTLA-4 with AP50, a clathrin-coated pit adaptor protein." Proc Natl Acad Sci U S A **94**(17): 9273-8.

Zhao, J., L. Hyman, et al. (1999). "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis."Microbiol Mol Biol Rev **63**(2): 405-45.26