FINNISH RED CROSS BLOOD SERVICE AND FACULTY OF BIOSCIENCES, DEPARTMENT OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES, DIVISION OF GENETICS, UNIVERSITY OF HELSINKI, FINLAND

Genetics of T cell co-stimulatory receptors -CD28, CTLA4, ICOS and PDCD1 in immunity and transplantation

Katri Haimila

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biosciences of the University of Helsinki in the Nevanlinna Auditorium of the Finnish Red Cross Blood Service, Kivihaantie 7, 00310 Helsinki, on April 24th 2009, at 12 o'clock noon.

Helsinki 2009

ACADEMIC DISSERTATIONS FROM THE FINNISH RED CROSS BLOOD SERVICE NUMBER 54

Supervisors

Docent Jukka Partanen, PhD
Research and Development
Finnish Red Cross Blood Service
Helsinki, Finland

Docent Päivi Saavalainen, (née Holopainen) PhD
Department of Medical Molecular Biology
University of Helsinki
Helsinki, Finland

Reviewers

Professor Minna Nyström, PhD
University of Helsinki
Helsinki, Finland

Docent Robert Hermann, MD, PhD
University of Turku
Turku, Finland

Opponent

Docent Pentti Tienari, MD, PhD
University of Helsinki
Helsinki, Finland

Custos

Professor Katarina Pelin, PhD
University of Helsinki
Helsinki, Finland

ISBN 978-952-5457-20-9 (print)
ISBN 978-952-5457-21-6 (pdf)
ISSN 1236-0341
http://ethesis.helsinki.fi
Helsinki 2009
Yliopistopaino

CONTENTS

1. List of original publications 7
2. Abbreviations 8
3. Abstract 9
4. Review of the literature 12
4.1 Basics of T cell immunology 12
4.1.1 T cell maturation 12
4.1.2 T cell recognition of antigens 14
4.1.3 Role of co-stimulatory receptors in T cell activation 14
4.1.3.1 CD28 is an essential co-stimulator 15
4.1.3.2 CTLA4 has an inhibitory function 15
4.1.3.3 ICOS induces cytokine expression 16
4.1.3.4 PDCD1 down-regulates T cell proliferation 17
4.1.4 Therapeutic potential of co-stimulatory receptors 17
4.1.5 The genes encoding co-stimulatory receptors CD28, CTLA4, ICOS and PDCD1 18
4.1.5.1 The co-stimulatory receptor genes belong to the same gene family 19
4.1.5.2 Variation within the genes 20
4.2 Immune related diseases and transplantation 21
4.2.1 Autoimmune disorders often co-occur 21
4.2.1.1 Autoimmunity develops as a result of self-reactive T cell activation 22
4.2.1.2 HLA association with autoimmune diseases 23
4.2.2 Coeliac disease 23
4.2.2.1 CD is common and has strong hereditary component 24
4.2.2.2 Challenges of CD diagnosis 25
4.2.2.3 Pathogenesis –dietary gluten leads to a small intestinal lesion 26
4.2.2.4 Complex genetics of coeliac disease 26
4.2.3 Co-stimulatory receptor genes in autoimmunity diseases 28
4.2.4 Immunodeficiencies 30
4.2.4.1 IgA deficiency 30
4.2.4.2 Common variable immunodeficiency 31
4.2.5 Co-stimulatory receptor genes in immunodeficiencies 32
4.2.6 Kidney transplantation 32
4.2.6.1 T cells react to non-self organ 33
4.2.7 Co-stimulatory receptor genes in organ transplantation 34
4.3 Genetic studies of multifactorial diseases 35
4.3.1 Genetic polymorphisms 35
4.3.2 A variety of genetic markers 35
4.3.3 The design of genetic linkage and association studies 37
4.3.3.1 Recruitment of a representative study population 37
4.3.3.2 Gathering informative genetic data 39
4.3.3.3 Ethical aspects 39
4.3.4 Complex diseases 40

4.3.5 Organ transplantation as a multifactorial trait 42
4.3.6 Linkage and linkage disequilibrium 42
4.3.7 Linkage disequilibrium and haplotypes 42
4.3.7.1 Linkage disequilibrium block as a tool in genetic studies 43
4.3.7.2 Measuring linkage disequilibrium 44
4.3.8 Linkage analysis 46
4.3.9 Association analysis 46
4.3.9.1 Adequate matching of controls is essential 47
4.3.9.2 Association study in family material 48
4.3.10 The purpose of genetic disorder studies 48
5. Aims of the study 50
6. Materials and methods 51
6.1 Study subjects 51
6.1.1 Coeliac disease patients 51
6.1.2 Kidney transplantation patients 52
6.1.3 IgA deficiency patients 52
6.1.4 Common variable immunodefieciency patients 53
6.1.5 Healthy controls 53
6.2 Methods 53
6.2.1 DNA extraction (I, II, III, IV) 53
6.2.2 HLA typing (II, III, IV) 53
6.2.3 Cell preparation and cDNA amplification (I) 54
6.2.4 Sequencing (I) 54
6.2.5 Genotyping 55
6.2.5.1 Microsatellite genotyping (II) 55
6.2.5.2 SNP genotyping 56
6.3 Statistical analyses and computer softwares 57
6.3.1 Data quality check 57
6.3.2 Linkage (III) 57
6.3.3 Association (II, III, IV) 57
6.3.4 Survival (IV) 58
6.3.5 Linkage disquilibrium (I, II, III) 58
6.3.6 Power (II, III) 58
6.3.7 Meta analysis (III) 58
6.4 Ethics 58
7. Results and discussion 59
7.1 Polymorphism screening of the gene region 2q33 (I and II) ... 59
7.2 Linkage disequilibrium and haplotype analyses of the T cell co-stimulatory receptor genes (I, II and III) 61
7.2.1 The linkage disequilibrium structure of the 2q33 region (I, II, III) 61
7.2.2 Strong linkage disequilibrium covers the PDCD1 gene (II) 63
7.2.3 The ICOS haplotypes observed in the Finnish population (I, II) 63
7.2.4 Identification of haplotype-tagging SNPs (II, III) 64
7.2.5 Extended haplotypes in European populations (II, III) 65

7.3 Association studies 66
7.3.1 Coeliac disease (II and III) 66
7.3.2 IgA deficiency (III) 72
7.3.3 Common variable immunodeficiency (III).....73
7.3.4 Meta-analysis (III).....74
7.3.5 Discrepant association results for the 2q33 region in CD ..75
7.3.6 Kidney transplantation patients (IV).....78
7.4 Analysis of linkage in coeliac disease (III) 79
7.5 Statistical significance of the results.....80
7.6 Functional effects of the variants in the 2q33 region.....81
7.6.1 Genetic variants 82
7.6.2 Soluble CTLA4.....83
8. Conclusions 86
9. Acknowledgements 88
10. References 90

1. LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, which are referred to in the text by Roman numerals (I-IV).

I **Haimila** KE, Partanen JA, Holopainen PM. Genetic polymorphism of the human *ICOS* gene. Immunogenetics. 2002 Mar;53(12):1028-32.

II **Haimila** K, Smedberg T, Mustalahti K, Mäki M, Partanen J, Holopainen P. Genetic association of coeliac disease susceptibility to polymorphisms in the *ICOS* gene on chromosome 2q33. Genes and Immunity. 2004 Mar;5(2):85-92.

III **Haimila** K*, Einarsdóttir E*, de Kauwe A, Koskinen L, Pan-Hammarström Q, Kurppa K, Mustalahti K, Ziberna F, Not T, Vatta S, Ventura A, Korponay-Szabo I, Ádány R, Pocsai Z, Széles G, Dukes E, Partanen J, Kaukinen K, Mäki M, Koskinen S, Hammarström L, Saavalainen P. The shared *CTLA4-ICOS* risk locus in celiac disease, IgA deficiency and common variable immunodeficiency. Genes and Immunity. 2009 Mar;10(2):151-61.

IV **Haimila** K, Turpeinen H, Kyllönen LE, Salmela KT, Partanen J Genetic variation in inducible costimulator gene predisposes to delayed graft function in kidney transplantation. Transplantation. 2009 Feb 15;87(3):393-6.

* These authors have contributed equally.

2. ABBREVIATIONS

Ag	antigen
APC	antigen presenting cell
B7RP-1	B7 related protein 1
CD	coeliac disease
CNV	copy number variation
CTL	cytotoxic T lymphocyte
CTLA4	cytotoxic T lymphocyte associated antigen 4
CVID	common variable immunodeficiency
DGF	delayed graft function
DH	dermatitis herpetiformis
DNA	deoxyribobonucleic acid
EmA	endomysial antibodies
ESPGHAN	the European Society for Pediatric Gastroenterology, Hepatology and Nutrition
HLA	human leukocyte antigen
HtSNP	haplotype-tagging SNP
ICOS	inducible co-stimulator
Ig	immunoglobulin
IgAD	Immunoglobulin A deficiency
IL	interleukine
INDEL	insertion or deletion polymorphism
kb	kilo bases
LD	linkage disequilibrium
MHC	major histocompatibility complex
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
PDCD1	programmed cell death 1
sCTLA4	soluble cytotoxic T lymphocyte associated antigen 4
SNP	single nucleotide polymorphisms
TCR	T cell receptor
Treg	regulatory T cell
tTG	tissue transglutaminase

3. ABSTRACT

Co-stimulatory signals are essential for the activation of naïve T cells and productive immune response. Naïve T cells receive first, antigen-specific signal through T cell receptor. Co-stimulatory receptors provide the second signal which can be either activating or inhibitory. The balance between these positive and negative signals determines the outcome of an immune response. The co-stimulatory receptor CD28 is crucial for T cell activation; whereas cytotoxic T lymphocyte associated antigen 4 (CTLA4) mediates critical inhibitory signal. Inducible co-stimulator (ICOS) augments cytokine expression and plays role in immunoglobulin class switching. Programmed cell death 1 (PDCD1) acts as negative regulator of T cell proliferation and cytokine responses. The co-stimulatory receptor pathways are potentially involved in self-tolerance and thus, they provide a promising therapeutic strategy for autoimmune diseases and transplantation.

The genes encoding co-stimulatory receptors CD28, CTLA4 and ICOS are located adjacently in the chromosome region 2q33. The *PDCD1* gene maps further, to the region 2q37. Polymorphism of the genes has been under screening during the last decade: all four genes contain several variants and they have been explored in a number of genetic association studies. *CTLA4* and *PDCD1* are associated with the risk of a few autoimmune diseases although results are rather contradictory.

Identifying the true risk variant is complicated because there is strong linkage disequilibrium (LD) on the 2q33 region. The whole gene of *CD28* exists in its own LD block but *CTLA4* and the 5' part of *ICOS* are within a same LD block. A recombination hot spot in the first intron of *ICOS* splits 3' part of *ICOS* into its own LD block; also *PDCD1* exists in a completely separate LD block due to its further location. Extended haplotypes covering the 2q33 region can be identified and especially conserved *CTLA4* haplotypes are frequent.

This study focuses on immune related conditions like coeliac disease (CD) which is a chronic inflammatory disease with autoimmune features. Immunoglobulin A deficiency (IgAD) and common variable immunodeficiency (CVID) belong to the group of primary antibody deficiencies characterised by reduced levels of immunoglobulins. Immunodeficiencies co-occur often with coeliac disease. Renal transplantation is needed in the end stage kidney diseases. Transplantation causes strong immune response which is tried to suppress with drugs. All these conditions are multifactorial with complex genetic background and multiple environmental factors affecting the outcome.

In this study, we have screened *ICOS* for polymorphisms by sequencing the exon regions. We detected 11 new variants and determined their frequencies in Finnish population. We have measured linkage disequilibrium on the 2q33 region in Finnish as well as other European populations and observed conserved haplotypes. We analysed genetic association and linkage of the co-stimulatory receptor gene region aiming to study if it is a common risk locus for immune diseases. The 2q33 region was replicated to be linked to coeliac disease in Finnish population and *CTLA4-ICOS* haplotypes were found to be associated with CD, IgAD and CVID – being the first non-HLA risk locus common for CD and immunodeficiencies. We also showed association between *ICOS* and the outcome of kidney transplantation.

Our results suggest new evidence for *CTLA4-ICOS* gene region to be involved in susceptibility of coeliac disease. The earlier published contradictory association results can be explained by involvement of both *CTLA4* and *ICOS* in disease susceptibility. The pattern of variants acting together or located in several risk haplotypes rather than a single polymorphism may confer the disease risk. The genes of co-stimulatory receptors may predispose also to immunodeficiencies as well as decreased graft survival and delayed graft function. Consequently, the present study indicates that like the well established HLA locus,

the co-stimulatory receptor genes predispose to variety of immune disorders. However, functional studies are required to identify primary risk variants and ultimately to understand the biological mechanisms of them.

4. REVIEW OF THE LITERATURE

4.1 Basics of T cell immunology

T cells play a fundamental role in the initiation and regulation of the highly specific and long-lasting immune response to antigens. In general, T cells can be divided into two broad subsets of CD4+ helper T cells and CD8+ cytotoxic T cells based on their function and expression of surface markers. Naïve T cells, which have never encountered antigens, proliferate and differentiate into effector cells as a result of activation. Effector T cells are able to migrate to the site of infection or inflammation where they usually eliminate the source of the antigen. CD4+ effector cells secrete a large variety of cytokines, activate macrophages to kill phagocytosed microbes, and in addition, they help B cells to differentiate into antibody secreting cells. CD8+ effector cells, also known as cytotoxic T lymphocytes (CTLs), can directly kill infected and tumor cells. CTLs are also able to secrete cytokines to activate phagocytes and induce inflammation. Both CD4+ and CD8+ T cells can exist as memory T cells, an expanded population of antigen-specific cells that can respond rapidly to antigen upon a secondary encounter.

4.1.1 T cell maturation

T cell precursors emerging from the bone marrow migrate to the thymus. During the development of a T cell, it produces a randomly generated T cell receptor (TCR). This receptor must recognise both an immunogenic peptide and the molecule that presents it: the major histocompatibility complex (MHC) known in humans as the human leukocyte antigen (HLA). The immature, double-positive (CD4+CD8+) T cell undergoes a process of positive selection (Figure 1), whereby those clones showing no TCR recognition of self-HLA-molecules are neglected to die. The selected cells, which show sufficient affinity for HLA-peptide complexes, are allowed to differentiate into single-positive, CD4+ or CD8+, T cells. This process serves to generate a self MHC-restricted T cell repertoire that can recognise when

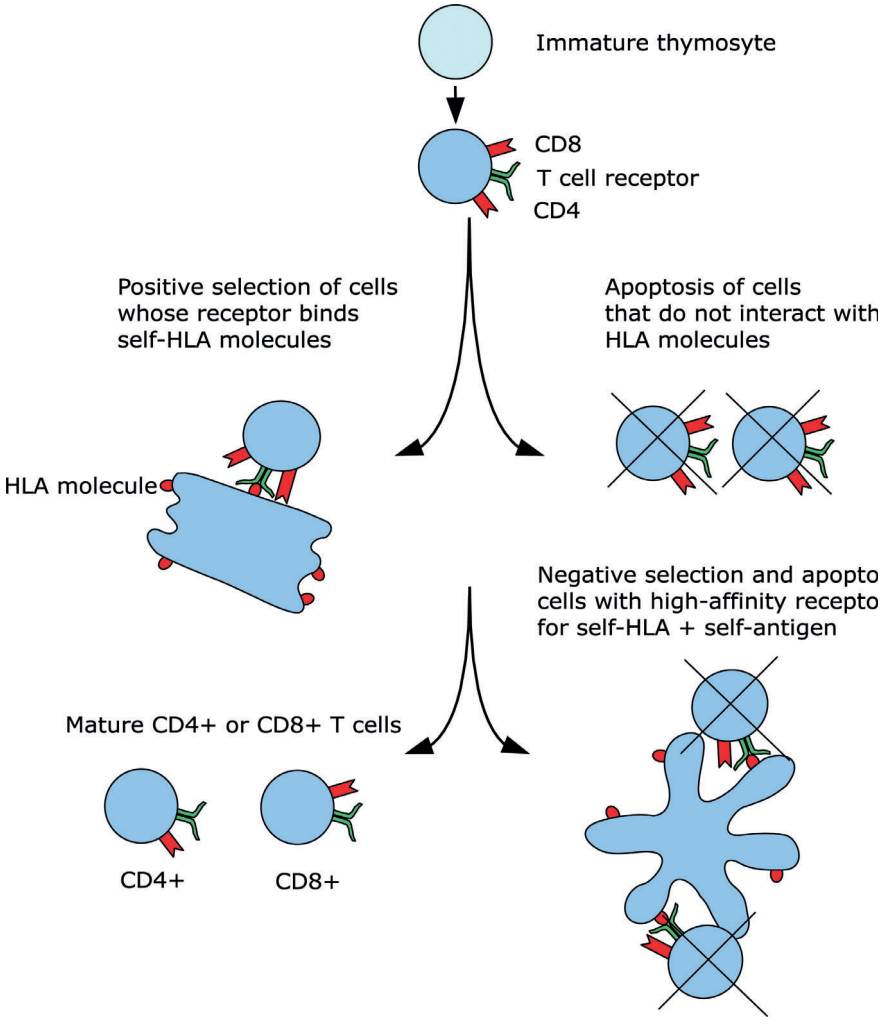


Figure 1. Positive and negative selection of T cells. The processes of thymic selection result in mature T cells are both self-HLA restricted and tolerant of the body's own peptide antigens.

the individual's own HLA molecules (and hence, cells) are presenting peptide antigen. Immature T cells also undergo a negative selection process (Figure 1), whereby clones whose receptors bind too strongly to self antigens complexed with HLA molecules are eliminated from the repertoire by apoptotic cell death. Such cells are highly prone to producing destructive autoimmune reactions. Only 3% of T cell precursors that enter the thymus survive

both positive and negative selection (Goldrath & Bevan 1999). After acquiring functional maturation (ability to produce cytokines or cytolytic molecules), the successful clones are allowed to exit the thymus into the circulation and peripheral lymphoid tissues.

4.1.2 T cell recognition of antigens

Discrimination between self and non-self is an essential ability of T cells. The function of HLA molecules is to present foreign antigens to T cells via their receptors. T cells can only recognise antigenic peptides bound to and displayed by self-HLA molecules; this is called self HLA restriction. Microbe-associated proteins are first processed into peptide antigens before binding to the cleft of HLA molecules. These molecules are expressed at the cell surface; class I molecules are expressed by most cell types, whereas class II molecules are only expressed by specialised antigen presenting cells (APCs).

4.1.3 Role of co-stimulatory receptors in T cell activation

Naïve T cells, which have not yet encountered antigen, need two signals for their full activation (Lafferty & Cunningham 1975). The first one is an antigen-specific signal provided by interaction of the TCR with peptide-HLA complexes on the APC. T cells receive the second signal for activation via co-stimulatory receptors (Figure 2 and Table 1). The act of engaging such receptors is called co-stimulation because it functions together with antigen to stimulate T cells. Following these two activating signals, a number of pathways are triggered leading to expansion of the antigen-specific T cell pool and differentiation of T cells into effector cells.

In the absence of co-stimulation, naïve T cells that encounter antigens will not respond, they either dye by apoptosis or enter a state of unresponsiveness called anergy. Thus, co-stimulatory signals are essential for the activation of naïve T cells and a productive immune response. Co-stimulatory receptor pathways can be either

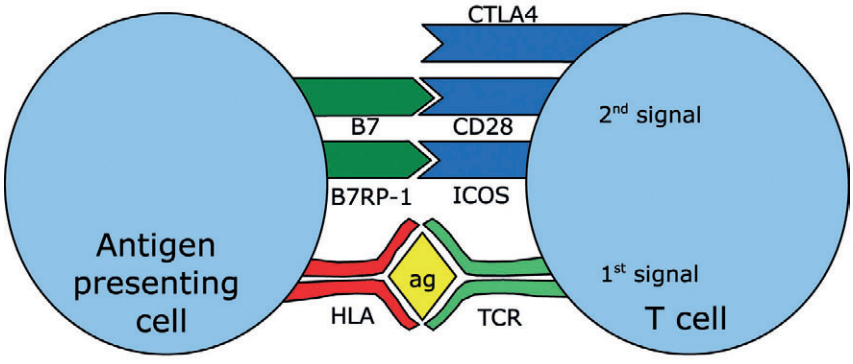


Figure 2. Naïve T cells need two signals for their activation. The first signal is an antigen-specific provided by TCR and the second one is mediated by co-stimulatory receptors.

activating or inhibitory, and the balance between these positive and negative signals can determine the outcome of an immune response. The best characterised co-stimulation pathway is the interaction between CD28 and B7 which provides a positive signal for T cell activation (Linsley & Ledbetter 1993).

4.1.3.1 CD28 is an essential co-stimulator

The CD28 pathway is crucial for T cell activation; signalling through CD28 increases cytokine production in T cells, by enhancing transcriptional activity and stabilising messenger RNA (Thompson et al 1989). CD28 ligation also reduces the number of engaged TCRs that are needed for proliferation or effective cytokine production, thereby lowering the threshold for T cell activation (Viola & Lanzavecchia 1996). CD28 is expressed constitutively on T cells and it binds to ligands B7-1 (CD80) and B7-2 (CD86) found primarily on APCs. The ligands have distinct but overlapping functions; B7-2 may mediate initial T cell activation, while B7-1 may be more important for maintaining the immune response (Vincenti & Luggen 2007).

4.1.3.2 CTLA4 has an inhibitory function

Cytotoxic T lymphocyte associated antigen 4 (CTLA4) mediates a critical inhibitory signal for T cell activation.

Table 1. T cell co-stimulatory receptors and their ligands.

Receptor	Alias	Expression	Major function	Stimulation	Ligand	Alias	Expression
CD28		T cells	Co-stimulation of naive cells	Activating	B7-1	CD80	Dendritic cells, B cells, macrophages
CTLA4		T cells	Negative regulation of immune responses, self-tolerance	Inhibitory	B7-2	CD86	Dendritic cells, B cells, macrophages
ICOS		T cells	Co-stimulation of effector T cells	Activating	B7RP-1	ICOS-L, B7h, B7H2, GL50	Dendritic cells, B cells, macrophages, other cells
PDCD1	PD-1	T cells, B cells, myeloid cells	Negative regulation of T cells	Inhibitory	PD-L1	B7-H1	Dendritic cells, B cells, macrophages, other cells
					PD-L2	B7-DC	

CTLA4 binds with higher affinity to the same B7 ligands as CD28. It is induced on T cells after their activation and functions in the termination of immune responses; CTLA4 ligation raises the activation threshold for T cell. CTLA4 decreases interleukine 2 (IL2) and IL2 receptor expression and it arrests T cells at the G1 phase of the cell cycle (Vincenti & Luggen 2007). The CTLA4 pathway may have an important role in peripheral T cell tolerance (Yamada et al 2002). Principal evidence for an inhibitory function of CTLA4 was obtained from CTLA4 knockout mice. These CTLA4 deficient ($CTLA4^{-/-}$) mice develop a fatal lymphoproliferative disorder with multiorgan autoimmune disease (Tivol et al 1995, Waterhouse et al 1995).

4.1.3.3 ICOS induces cytokine expression

Inducible co-stimulator (ICOS) plays a critical independent role in T cell activation that is synergistic with CD28 signalling. ICOS augments effector T cell cytokine responses; in particular, it appears to superinduce production of the anti-inflammatory cytokine IL10 (Hutloff et al 1999). ICOS expression is enhanced on activated T cells by CD28 co-stimulation (Beier et al 2000). ICOS binds B7 related protein 1, (B7RP-1) which is expressed constitutively by B cells and macrophages (Yoshinaga et

al 1999) but can also be induced on non-lymphoid cells by inflammatory stimuli (Swallow et al 1999). ICOS knockout mice have reduced CD4+ T cell responses and increased risk of experimental autoimmune encephalomyelitis (Dong et al 2001) as well as defects in immunoglobulin (Ig) class switching and germinal centre formation (McAdam et al 2001).

4.1.3.4 PDCD1 down-regulates T cell proliferation

Programmed cell death 1 (PDCD1 or PD-1) acts as a negative regulator of T cell proliferation and cytokine responses (Freeman et al 2000). This inhibitory receptor is constitutively expressed in the thymus (Nishimura et al 1996) and is induced on activated T and B cells. PDCD1-deficient mice have an autoimmune-like phenotype suggesting that the PDCD1 pathway may be involved in self-tolerance (Keier et al 2008).

4.1.4 Therapeutic potential of co-stimulatory receptors

Blockade of the CD28 co-stimulatory pathway provides a promising therapeutic strategy for autoimmune diseases and transplantation. CTLA4-Ig is a fusion protein which consists of the extracellular binding domain of CTLA4 linked to a modified Fc domain of human antibody IgG. The Fc domain mediates complement activation and interacts with Fc cell surface receptors. The CTLA4 fragment defines the specific target of the fusion antibody, that is the B7 ligands. It was developed to selectively interrupt full T cell activation by blocking the interaction of CD28 and B7 ligands (Vincenti & Luggen 2007).The use of CTLA4-Ig is effective in inducing long-term allograft survival in solid organ transplantation in mouse, rat and primate models (Snanoudj et al 2006). The first clinical trial with CTLA4-Ig in human renal transplantation showed promise although widely used immunosuppressive drug called cyclosporine was still more effective in preventing acute rejection (Vincenti 2005).

The impact of the ICOS co-stimulation pathway to emerging rejection episodes has been demonstrated by anti-ICOS therapy (Özkaynak et al 2001). Anti-ICOS antibody treatment has also been studied together with anti-CD40L and CTLA4-Ig in animal models of transplantation; the animals displayed less signs of chronic rejection (Snanoudj et al 2006).

An alternative therapeutic strategy employing the co-stimulation mechanism was developed with an agonistic anti-CD28 antibody. TGN1412A is a genetically engineered anti-CD28 monoclonal antibody that is capable of effectively stimulating the proliferation and activation of regulatory T cells, as well as inducing the release of anti-inflammatory cytokines. Upon its development, TGN1412 was considered to be promising new treatment for leukaemia and serious autoimmune diseases, and early testing demonstrated that it was well-tolerated in non-human primates. However, the first trial in humans (in London 2006) resulted in disaster: all six volunteers became seriously ill with life-threatening adverse events triggered by an unexpected cytokine storm within hours of receiving the drug. Fortunately none died but all volunteers had residual defects (Dowsing & Kendall 2007).

4.1.5 The genes encoding co-stimulatory receptors CD28, CTLA4, ICOS and PDCD1

CD28 and *CTLA4* are neighbouring genes on the chromosome 2q33 and were identified in the late 1980s (Aruffo & Seed 1987, Dariavach 1988). A decade later, *ICOS* was identified in very close proximity (Hutloff et al 1999). However, it was only in 2003 that the 300 kb block of chromosome 2q33 that spans these genes was fully sequenced and the three functional genes were confirmed with exact distances between genes (Ueda et al 2003). The fourth member of this gene family, *PDCD1* that maps to 2q37 was identified in 1994 (Shinohara et al 1994) (Figure 3).

Figure 3. Structure of the genes on the chromosome region 2q33-37. Black boxes denote coding regions and open boxes untranslated regions of the exons. Intron regions are indicated by horizontal lines between exons.

4.1.5.1 The co-stimulatory receptor genes belong to the same gene family

The lengths and compositions of the exons of *CD28*, *CTLA4* and *ICOS* are well conserved. *CD28* and *CTLA4* include four exons: exon 1 encodes the leader sequence, exon 2 the ligand binding domain, exon 3 the transmembrane segment and exon 4 determines the cytoplasmic tail of the protein (Lee et al 1990, Harper et al 1991). *ICOS* and *PDCD1* contain five exons; exons 1-4 are parallel with those of *CD28* and *CTLA4*, while the short exon 5 encodes the cytoplasmic domain together with exon 4 (Ling et al 2001, Finger et al 1997).

Remarkable homology (over 20 %) between the co-stimulatory receptor genes (Harper et al 1991, Ling et al 1999) strongly suggests that the genes belong to the same gene family which is the result of gene duplications. The duplications are proposed have taken place before speciation, but relatively recently to account for the homologies between the molecules (Harper et al 1991). The homology between the mouse and human *CTLA4* genomic sequence is 70% for both non-coding and coding regions with complete amino acid conservation of the intracellular domain (Ling et al 1999). Interestingly, linkage analysis and congenic mapping in mice have identified a susceptibility locus for type 1 diabetes in the chromosome 1. This locus, Idd5.1, includes the Cd28, Ctla4 and Icos genes (Lamhamedi-Cherradi et al 2001).

4.1.5.2 Variation within the genes

Besides the gene transcript containing complete exons, which codes for the full length form of the co-stimulatory receptors, several alternatively spliced variants of messenger ribonucleic acid (mRNA) exist. For *CD28*, concurrent presence of eight different mRNA isoforms has been demonstrated in human T cells (Deshpande et al 2002). Alternative splicing of gene transcripts enhances gene diversity and may play a role in regulation of the molecule, but the function and importance of splice variants of *CD28* remain unknown. A soluble form of CTLA4 (sCTLA4) lacks exon 3, which encodes the transmembrane domain. Soluble CTLA4 is present in serum and is immunologically active; it is able to bind B7 ligands (Magistrelli et al 1999). The soluble molecule functions to suppress T cell activation, similar to membrane bound CTLA4 (Oaks et al 2000). It has therefore been hypothesised to play a role in the development of autoimmune diseases (Gough et al 2005), but this is yet to be confirmed. This far, no alternatively spliced isoforms of *ICOS* have been found, but there are five variable transcripts of *PDCD1*. It has been proposed that one of *PDCD1* transcripts will encode a soluble form of the receptor where such a soluble receptor could be involved in the maintenance of peripheral tolerance and prevention of autoimmunity (Nielsen et al 2005).

All four co-stimulatory receptor genes are polymorphic: *CD28* contains one microsatellite and 50 single nucleotide polymorphisms (SNPs), *CTLA4* includes one microsatellite and four dimorphisms, *ICOS* contains two microsatellites in intron 4 and 31 SNPs and *PDCD1* includes 15 SNPs (http://www.hapmap.org/, Ueda et al 2003). However, only two of all these polymorphisms are in a coding region and lead to a change of amino acid: +49A/G (Ala-Thr) in the leader sequence of *CTLA4*, and the PD1.8 (Ala-Val) in exon 5 of *PDCD1*.

A few of the polymorphisms observed in the 2q33-37 region have been demonstrated to be functional variants. For example, a correlation has been suggested between

the leader sequence polymorphism in *CTLA4* and T cell proliferation (Kouki et al 2000). Another *CTLA4* variant, called CT60, 6 kilo bases (kb) downstream from the transcription starting site has been proposed to affect mRNA levels of soluble CTLA4 (Ueda 2003). Expression of ICOS is also affected by polymorphisms in the gene (Kaartinen et al 2007); such polymorphisms have also been found to correlate with IL10 secretion by the T cell (Castelli et al 2007). The *PDCD1* polymorphism, PD1.3, alters a binding site for a transcription factor located in an intronic enhancer, thereby disrupting binding of the transcription factor and possibly causing aberrant regulation of PDCD1 (Prokunina et al 2002).

There is strong linkage disequilibrium (LD) throughout the 2q33 region (Holopainen & Partanen 2001); not only within genes but also between them. *CD28* and the 5'end of *ICOS* exist in their own LD blocks, and between them, *CTLA4* and the 3' part of *ICOS* are within a same LD block (Ueda 2003). Due to its location much further downstream on 2q37, *PDCD1* exists in a completely separate LD block. Linkage disequilibrium is fairly strong within the gene (Nielsen et al 2003).

4.2 Immune related diseases and transplantation

4.2.1 Autoimmune disorders often co-occur

Autoimmune diseases are initiated by loss of immunologic tolerance to self-antigens. Evidence for genetic contribution to the development of autoimmune diseases is strong (Heward & Gough 1997) but generally susceptibility genes are not yet identified except for *HLA*.

The presence of one autoimmune disease condition in an individual or family member predisposes that individual to a higher risk of a second autoimmune condition. For example, type I diabetes patients and their family members have autoimmune thyroid disease and rheumatoid arthritis more often than the general population (Somers et al 2006). Coeliac disease patients show a similar pattern of

increased incidence of type I diabetes, thyroid disorders and Sjögren's syndrome (Collin et al 1994). The appearance of multiple autoimmune disorders in an individual or families indicates that the common autoimmune diseases share at least some aetiological mechanisms.

4.2.1.1 Autoimmunity develops as a result of self-reactive T cell activation

There are three cardinal features of a T cell: specificity of antigen recognition, memory of previous antigen encounter and lack of reactivity against self. The third feature is highly important because self-reactive T cells can be dangerous for the body by generating destructive autoimmune responses. Thus, T cells showing too high affinity for self-peptides are eliminated through negative selection in the thymus. However, some self-reactive T cells do escape the stringent thymic processes and require controlling in the periphery by other regulatory mechanisms, of which the primary mechanism involves specialised regulatory T cells (Treg) (Romagnani 2006).

T regs slow down pathogenic autoimmune responses and maintain immune homeostasis. Acting through dendritic cells, they obstruct the differentiation and effector functions of autoreactive T cells. Tregs undergo expansion in the periphery which appears to be induced by exposure to self-antigen. The development of a Treg in the periphery can be influenced by the presence of co-stimulatory receptors, their ligands and cytokines expressed by other cells in the surrounding environment. The affinity and/or the expression level of the TCR are also thought to impact on Treg development (Pakravan et al 2007). CD28 has been shown to be involved in the development and homeostasis of Treg cells, whereas CTLA4 is not necessary for the development of these cells, but it is involved in their regulatory function. Treg cells constitutively express a high level of intracellular CTLA4 and low level of surface CTLA4; surface expression of CTLA4 is raised after the cells are activated (Yi et al 2006).

4.2.1.2 HLA association with autoimmune diseases

HLA molecules have a fundamental role in T cell immunology and are thus linked with T cell-mediated autoimmune diseases.

The HLA complex occupies a 4 mega base region on chromosome 6p21 containing about 200 genes, many of which have immunological function. Classical *HLA* genes are highly polymorphic; genetic polymorphisms located in exons 2 and 3 of these genes can alter amino acids in the peptide binding groove and thereby increase the repertoire of peptides that can be presented. HLA molecules can present autoimmune peptides, thus triggering pathogenetic pro-inflammatory cascades.

Strong linkage disequilibrium occurs within the *HLA* region. As a result, extended *HLA* haplotypes are frequent in population groups. The highly conserved *A*1-*B*8-*DR*3-*DQ*2 haplotype is a commonly carried haplotype that is also known to be associated with autoimmunity, e.g. coeliac disease, IgA deficiency, type 1 diabetes and myasthenia gravis (Price et al 1999).

Coeliac disease can be used as a model for exploring the complex genetics of *HLA*-associated autoimmune diseases because its *HLA* association is well established and relatively straightforward. In addition, the principal environmental trigger, gluten of wheat, as well as auto-antigen are known.

4.2.2 Coeliac disease

Also called gluten sensitive enteropathy, coeliac disease (CD), is a chronic inflammatory disease with autoimmune features that develops in genetically susceptible individuals. It is characterised by an immune response to ingested gluten that leads to inflammation, villous atrophy and crypt hyperplasia in the proximal region of the small intestine. Inheritance of coeliac disease does not follow a Mendelian pattern; rather CD is a complex genetic disorder with multiple susceptibility genes and multiple environmental

triggers. The primary environmental factor is dietary gluten from wheat, barley or rye. The only available therapy is a lifelong gluten free diet which settles the intestinal damage and leads to disappearance of symptoms.

4.2.2.1 CD is common and has strong hereditary component

Initially, CD was thought to occur only in Western countries where gluten-containing grains are an important part of the diet. But coeliac disease is actually a worldwide disorder, being as common in Latin America, Africa and Asia as it is in Europe (Cataldo & Montalto 2007). The prevalence of CD is approximately 1% in general population (Mäki et al 2003, Dube et al 2005), but among first degree family members of coeliac patients, the prevalence is as high as 16% (Dube et al 2005). A concordance rate in monozygotic twins is 70-85%, in dizygotic twins 11-20% and among *HLA*-identical siblings about 30% (Greco et al 2002, Nistico et al 2006). These differences in concordance rates indicate that there is a strong genetic component to CD, and that carriage of *HLA* susceptibility allele is necessary but not sufficient for disease development. In addition, the concordance rate in monozygotic twins reveals that, environmental factors are also crucial for development of CD.

The classical (or symptomatic) coeliac phenotype refers to presentations with fully developed gluten-induced villous atrophy of the small intestine and classic features of intestinal malabsorption. Coeliac disease can also be defined as silent (or asymptomatic), when patients have no gastrointestinal symptoms but are found to have gluten-induced villous atrophy. Atypical CD is often found when another presentation such as iron deficiency, osteoporosis, short stature or infertility is examined (Dewar & Ciclitira 2005). For most patients, gastrointestinal symptoms are often generalised or subtle, including abdominal pain, diarrhoea, steatorrhoea, malabsorption syndrome and bloating. Common non-gastrointestinal symptoms are weight loss, fatigue, aphthous ulcers, arthralgia and

myalgia; patients may even suffer from depression or neurological symptoms (Hopper et al 2007). Dermatitis herpetiformis (DH) is a cutaneous manifestation of coeliac disease, where patients present with a blistering skin rash on the elbows, knees, buttocks and scalp. DH patients show similar enteropathy than CD patients and their condition resolves upon commencement of a strict gluten free diet (Reunala 2001).

4.2.2.2 Challenges of CD diagnosis

Serological screening is often used as the first step in the diagnosis of CD, where individuals are tested for endomysial antibodies (EmA) and antibodies against a defined autoantigen called tissue transglutaminase (tTG). HLA typing can assist the diagnosis, especially in unclear cases. Histological demonstration of small bowel villous atrophy by biopsy remains the critical method for diagnosing coeliac disease (Hopper et al 2007).

ESPGHAN (the European Society for Pediatric Gastro-enterology, Hepatology and Nutrition) has established standard criteria for CD diagnosis (Walker-Smith et al 1990). The main criteria are an intestinal biopsy showing histological evidence of a flat mucosa while on a gluten-containing diet and full clinical remission together with disappearance of positive antibodies following commence-ment of a gluten-free diet.

Diagnosis of coeliac disease is often missed because many patients do not demonstrate the classical gastrointestinal symptoms. The proportion of undiagnosed affected individuals is notable, and because significant risks and complications are known to correlate with untreated disease, family members of known coeliac patients and other risk groups are encouraged to be screened for CD (Torres et al 2007).

4.2.2.3 Pathogenesis –dietary gluten leads to a small intestinal lesion

Three factors are necessary in pathogenesis of CD: genetic susceptibility; the immune system and dietary gluten. As such, presentation of gluten peptide by certain HLA molecules to gluten-specific T cells and their subsequent activation are seen as critical events in the chain of pathogenesis.

The healthy small bowel mucosa is covered with fine finger-like structures called villi. The villi are enclosed with a single layer of epithelial cells and each epithelial cell is covered with microvilli that have a critical role in digestion and absorption. In coeliac disease, lymphocytes migrate to the surface epithelium leading to complete loss of villous structure and hypertrophy of the mucosa, which results in the characteristic flat mucosa (Kagnoff 2007).

Gluten is a water-soluble protein that consists of an alcohol-soluble fraction termed gliadin in wheat, hordein in barley and secalin in rye. Gluten peptides, which are rich in proline and glutamine, can resist enzymatic digestion in the gastrointestinal tract and are thereby available to cross the intestinal epithelial cell barrier to access the immune system (Shan et al 2002). Gluten peptides can react with tissue transglutaminase which deamidates certain glutamine residues in the peptides to glutamic acid. This process of deamidation introduces a negative charge into the gluten peptides and allows them to bind more efficiently to HLA-DQ2 or DQ8 molecules. After binding gluten epitope, the HLA molecules trigger T cell activation and subsequent production of inflammatory cytokines leading to tissue damage (Kagnoff 2007).

4.2.2.4 Complex genetics of coeliac disease

Association of the *HLA* locus with coeliac disease has been known for over 30 years (Falchuk et al 1972, Stokes et al 1972). Thus far, *HLA-DQA1* and *DQB1* genes (Sollid et al 1989) are the only known genetic variants predisposing to CD. Approximately 90% of coeliac patients carry the

HLA-DQ2 molecule, 6% carry the HLA-DQ8 molecule and the remaining patients possess one half of DQ2 (either *DQA1**05 or *DQB1**02 but not both alleles) (Karell et al 2003).

The association with the *HLA* locus was found at a time when only serological methods were in use. HLA-specific antibodies in lymphocytotoxicity tests revealed the molecules DQ2 and DQ8, but the genes encoding them remained unrecognized. Due to rudimentary beginning, the imprecise serological codes of DQ2 and DQ8 are still commonly used even though it is known that at the allelic level the susceptibility variants are *DQA1**05-*DQB1**02 (DQ2) and *DQB1**0302 (DQ8) (Sollid & Thorsby 1993). The HLA-DQ2 molecules can be encoded by genes on the same haplotype (in *cis*), or more rarely by genes which are located on separate haplotypes (in *trans)*.

The *HLA* genes are thought to contribute to approximately 1/3 to 1/2 of the genetic risk for CD development (Petronzelli et al 1997, Bevan et al 1999). However, the *HLA-DQB1**02 and *DQB1**0302 alleles are common in the healthy individuals, carried by approximately 30% of Finnish population. Hence, the genetic risk of coeliac disease cannot be explained by *HLA* alone, and other, non-*HLA* genetic variants predisposing to disease have been searched for. Several low-risk genes are supposed to exist, on their own, they have a much smaller influence than the *HLA*.

Genome wide linkage studies have revealed a couple of potential susceptibility loci for coeliac disease. The most interesting loci are 2q33 including *CTLA4*, 19p13 containing *MYO9B* and 5q31-33 (van Heel et al 2005, Monsuur & Wijmenga 2006). Many other candidate genes have been tested for association on the basis of known pathogenetic and immunological mechanisms of coeliac disease but no convincing association has been found.

4.2.3 Co-stimulatory receptor genes in autoimmunity diseases

Due to the inhibitory role of CTLA4 in T cell activation, it is a good candidate susceptibility gene in autoimmunity and is thus the most studied gene of the co-stimulatory receptor family. There are growing number of publications reporting association of *CTLA4* with several autoimmune diseases including coeliac disease, type 1 diabetes, systemic lupus erythematosus and rheumatoid arthritis (Table 2). These positive findings in different autoimmune diseases and several populations suggest that *CTLA4* might be a general autoimmunity locus.

Despite the number of reported positive associations, the results for the 2q33 region are confusing because no disease causing variants have been specifically identified and different studies report differing variants as demonstrating the strongest association. Currently, it is still impossible to say where the primary polymorphism lies in the 2q33 region. The distance between the genetic markers used and the actual risk factor may be long due to the strong LD on the region. Thorough examination of polymorphisms within the 2q33 region is necessary to be able to identify the primary variant, which may be located in a gene neighbouring *CTLA4*; thus far, most studies have restricted to their search to *CTLA4* only. It is also possible that there is an additional susceptibility locus in the region, or perhaps distinct variants for different autoimmune diseases.

CD28 and *ICOS* have not been found to associate with any autoimmune disease to date. But *PDCD1* is associated with systemic lupus erythematosus (Prokunina et al 2002, Ferreiros-Vidal et al 2004, Thorburn et al 2007), lupus nephritis (Prokunina et al 2004), type 1 diabetes (Nielsen et al 2003, Ni et al 2007), rheumatoid arthritis (Lin et al 2004, Kong et al 2005) and Graves' disease (Newby et al 2007).

There are also several studies showing no association (Perez de Nanclares et al 2004, Lorentzen et al 2005,

Table 2. Examples of autoimmune diseases showing association with *CTLA4*.

Disease	Abbreviation	Population	Reference
Coeliac disease	CD	French	Djilali-Saiah et al, 1998
		Finnish	Holopainen et al, 1999
		British	King et al, 2002
		British	Hunt et al, 2005
		Irish	Brophy et al, 2006
Type 1 diabetes	T1D	multiethnic	Nisticò et al, 1996
		multiethnic	Marron et al, 1997
		Russian	Chistiakov et al, 2001
		Swedish	Einarsdottir et al, 2003
		Dutch	Zhernakova et al, 2006
Systemic lupus erythematosus	SLE	Japanese	Ahmed et al, 2001
		Korean	Hudson et al, 2002
		Portugues	Barreto et al, 2004
		British	Graham et al, 2006
Rheumatoid arthritis	RA	British	Vaidya et al, 2002
		Spanish	Rodríguez et al, 2002
		Chinese	Lee et al, 2003
		multiethnic	Plenge et al, 2005
Grave's disease	GD	British	Ueda et al, 2003
		British	Vaidya et al, 2003
		Japanese	Furugaki et al, 2004
		Han	Chen et al, 2008
Addison's disease	AD	British	Kemp et al, 1998
		British	Kotsa et al, 1997
		Norwegian/British	Blomhoff et al, 2004
Autoimmune hypothyroidism	AIH	British	Kotsa et al, 1997
		British	Nithiyananthan 2002
		British	Ueda et al, 2003
		Japanese	Furugaki et al, 2004
Primary biliary cirrhosis	PBC	Caucasian	Agarwal et al, 2000
		Chinese	Fan et al, 2004
		multiethnic	Juran et al, 2008
Multiple sclerosis	MS	Swedish	Ligers et al, 1999
		multiethnic	Kantarci et al, 2003

Greve et al 2007); that is, a reported association can often not be replicated in a different study. The failure to replicate findings may be due to variable allele frequencies in different populations, which result in a lack of statistical power for its part. For example, frequencies of *CTLA4* alleles fluctuate between South and North Europe (Holopainen et al 2004). Also, gene-environment interactions vary between populations, such as diet or viruses in childhood and may contribute to the failure to replicate genetic findings in different populations. Usually an individual susceptibility gene has a relative risk under 2; therefore, studies with only a few hundred families typically lack statistical power to detect a locus.

4.2.4 Immunodeficiencies

The family of antibody deficiencies is a heterogeneous group of primary (or congenital) immunodeficiencies. In these disorders, the typical abnormality is in antibody synthesis, which leads to reduced serum levels of one particular immunoglobulin (Ig) isotype or can affect multiple isotypes.

4.2.4.1 IgA deficiency

Immunoglobulin A deficiency (IgAD) is the most common form of inherited immunodeficiency in humans, affecting approximately 1 in 600 individuals in Western countries. Most IgAD cases are sporadic, but also large families with deficiencies have been identified.

IgAD is characterized by the absence or very low level (<7 mg/dL) of serum IgA. Most individuals with IgAD are totally asymptomatic. Individuals with symptomatic IgAD are predisposed to recurrent infections, especially sinopulmonary and gastrointestinal infections and have an increased incidence of lymphoid and nonlymphoid malignancies. IgAD often co-occurs with atopy, asthma and autoimmune disorders particularly those associated with the gastrointestinal tract (Latiff & Kerr 2007).

Immunoglobulin A is the second most prevalent antibody in the serum (normally 2-3 mg/ml) after IgG. IgA found in bodily secretions is called secretory IgA and is an important defence mechanism against inhaled and ingested pathogens on mucosal surfaces such as the respiratory, gastrointestinal and genitourinary tracts (Woof & Kerr 2006). IgA is known to down-regulate cell-mediated responses, therefore, during an absence of IgA, autoimmunity may occur (Koskinen 1996).

IgA deficient individuals do not demonstrate disease-causing variations in the *IgA* genes themselves; thus the deficiency appears to be the result of mutations in the genes controlling IgA production. The *HLA* haplotypes *A*1-*B*8-*DR*3, *A*28-*B*14-*DR*1 and *B*44-*DR*7 are more frequent in IgA-deficient families than in the general population, but a

HLA association is not very clear (De la Concha et al 2002). A number of non-*HLA* genes have also been studied; one of the most interesting candidates is the gene encoding the tumour necrosis factor receptor family member, TACI, which mediates isotype switching in B cells (Castigli et al 2005).

Interestingly, some individuals initially present with IgAD and then develop common variable immunodeficiency (CVID) at a later stage. IgAD and CVID are known to co-occur in the same families (Hammarström et al 2000). This would suggest that IgAD and CVID have a common genetic trigger.

4.2.4.2 Common variable immunodeficiency

CVID is characterized by a deficiency in all immunoglobulin isotypes. It is a clinically and immunologically heterogeneous immunodeficiency and has a complex genetic background. CVID has an estimated prevalence of 1 in 25 000 in Caucasians with relatively late onset. Most cases of CVID are sporadic, but approximately 10–20% of CVID cases are familial; with most multiplex CVID families showing an autosomal dominant mode of inheritance (Salzer & Grimbacher 2006).

Patients with CVID suffer from recurrent bacterial infections, particularly of the upper respiratory tract, as well as autoimmunity and malignancies. Chronic respiratory tract infections cause permanent lung damage resulting in bronchiectasis, emphysema and fibrosis. CVID is diagnosed on the basis of an impaired ability to produce specific antibodies: patients usually have reduced serum levels of IgG, IgA and often IgM. The accepted treatment for CVID is immunoglobulin replacement therapy (Blanco-Quirós et al 2006).

The molecular basis of CVID remains unknown. The immunological heterogeneity in CVID may reflect the diversity of mechanisms that lead to the deficiency. The production of immunoglobulins is based on the interaction of antigen-presenting cells with antigen-specific T and

B cells within the peripheral lymphoid tissues. Defects in B cell survival, B cell activation after antigen receptor cross-linking, T cell signalling and cytokine expression have all been observed in CVID patients. Some individuals with CVID have a global isotype switching defect. Class-switch or heavy-chain isotype switching is a process of deoxyribonucleic acid (DNA) rearrangement in B cells which changes the Ig isotype from IgM to IgG, IgE or IgA, still maintaining the antigen specificity of the receptor. Thus, disease causing mutations may be hidden in the genes encoding molecules involved in class switching pathway (Blanco-Quirós et al 2006).

Genetic linkage studies in CVID families have suggested several candidate loci for the condition, but the major disease-causing variant has not been found. Mutations in the *ICOS* and *TACI* genes explain approximately 15% of CVID cases (Castigli & Geha 2006).

4.2.5 Co-stimulatory receptor genes in immunodeficiencies

ICOS is the only co-stimulatory receptor gene which has been studied in immunodeficiencies. A very rare homozygous deletion of *ICOS* has been reported in individuals with CVID (Grimbacher et al 2003, Salzer et al 2004). All nine cases appear to descend from a common founder and carry the same large deletion (Salzer et al 2004). Apart from the absence of ICOS expression, the phenotype of the CD4+ T cells in ICOS-deficient patients is normal; however, functionally the cells produce only low amounts of IL10 and IL17 upon stimulation compared to ICOS-intact individuals. In addition, germinal centre formation in ICOS-deficient patients is disturbed resulting in a reduction of switched memory B cells and lack of plasma cells in peripheral lymphoid organs (Warnatz et al 2006).

4.2.6 Kidney transplantation

Renal transplantation is the treatment of choice for patients with end stage kidney disease. Transplantation not only

improves the quality of life but also prolongs life when compared to the alternative renal replacement treatment, dialysis. Approximately 50 individuals in every million are considered potential candidates for kidney transplantation in Europe in year. Deceased donors are the major source of organs in Finland, but sometimes suitable family members of patients volunteer to donate their kidney. Over the past decades, the acute rejection of kidney transplants rate has fallen dramatically and the 1-year graft survival rate has increased to 90% using deceased donors and 95% using living related donors. This increase in graft survival is largely due to advances in immunosuppressant medication (Yates & Nicholson 2006).

4.2.6.1 T cells react to non-self organ

T cells are the primary cellular contributors to allograft rejection, due mainly to their ability to discriminate between self and non-self. A donor's HLA molecules act both as antigen presenting molecules and as direct foreign antigens in a transplant recipient. Activation of recipient T cells may occur through two pathways: directly or indirectly. In the direct pathway, recipient T cells recognise the allogenic HLA molecules of the donor as foreign and generate a response against the cells expressing the foreign molecule. Alternatively, the indirect pathway of allorecognition typically involves graft-derived antigens being presented to the recipient's T cells by the recipient's own APCs. Generally, the direct pathway appears to be more important in acute graft rejection. Regardless of the pathway, T cell activation leads to humoral and cell-mediated immune responses, causing acute rejection and allograft damage (Trivedi 2007).

As a result of a transplant recipient's robust immune response, lifelong immunosuppression is mandatory after transplantation. Current immunosuppressive therapy has been designed to disrupt cytokine and interleukine (IL) networks or limit cell proliferation. Monoclonal antibodies against the IL2 receptor interrupt the IL2 pathway, cal-cineurin inhibitors prevent transcription of the IL2 gene,

and anti-proliferative medications inhibit T cell and B cell proliferation. However, it is crucial to find a balance between prevention of acute rejection and toxicity induced by the immunosuppressants. The side effects of immunosuppressive therapy are serious, including increased risk of diabetes, infectious complications, cardiovascular disease and malignancy (Zand 2005). Hence, the focus of therapeutic development for transplantation has turned towards novel agents with greater tissue selectivity such as blockade of T cell co-stimulation.

Besides acute rejection, delayed graft function (DGF) is a risk factor for allograft loss as well as poor patient outcome in the long term. In renal transplantation, DGF has been defined as the failure of the transplanted kidney to function immediately, resulting in weak urine secretion and therefore a need of dialysis within the first week after surgery. Long cold ischemia time, receipt of an organ from a donor of older age, hypertension, high plasma renin or angiotensin II levels are all known factors contributing to DGF (Blumenfeld et al 2001). The success of transplantation is also dependent on donor source, duration on dialysis while awaiting the transplant, HLA matching and surgical achievements.

4.2.7 Co-stimulatory receptor genes in organ transplantation

CTLA4 gene polymorphisms have been examined related to kidney transplantation in a few studies (Slavchena et al 2001, Dmitrienko et al 2005, Gendzekhadze et al 2006, Wisniewski et al 2006). All the associations published are with different *CTLA4* markers: CTLA4(AT)n (Slavchena et al 2001), *CTLA4*+49 (Gendzekhadze et al 2006) and *CTLA4*-318 (Wisniewski et al 2006). They all associated with an increased incidence of acute rejection. *CTLA4* polymorphisms have also been found to be associated with success of liver transplantation (Slavchena et al 2001, Marder et al 2003, de Reuver et al 2003, Tapirdamaz et al 2006, Muro et al 2008).

4.3 Genetic studies of multifactorial diseases

Multifactorial diseases are common and thus, they are under great interest. Susceptibility genes for diseases are often searhed by linkage and association studies. Identifying susceptibility genes is challenging due to the complex aetiology of multifactorial diseases so, genetic studies must be carefully designed. In genetic studies, samples from either affected individuals, cases, and matched controls for them or families including one or several affected members are examined for genetic polymorphisms.

4.3.1 Genetic polymorphisms

Most genetic variation is neutral with no obvious biological effects, but some of the polymorphisms do affect the phenotype of an individual, such as susceptibility to particular diseases. However, non-functional polymorphisms are useful markers in genetic studies. A genetic marker is any DNA sequence that is polymorphic and can be used for mapping a disease locus; this is based on the hypothesis that markers located near a disease predisposing locus are segregating with the disease.

Polymorphisms are scattered throughout the genome but their density can vary in different regions. A variation hotspot is a locus that contains even dozens of times higher levels of polymorphisms than an average region (Mills et al 2006). Genetic stability is maintained by several repair mechanisms (Jeggo & Löbrich 2006). However, some replication errors can remain in the genome as mutations. If they reach frequency of 1% in the general population, they are called a polymorphism (Vogel & Motulsky 1996). A high level of variation may indicate a lack of functionally important sequence in that region or positive selective pressure for diverse alleles for a biological purpose (The International SNP Map Working Group 2001).

4.3.2 A variety of genetic markers

Genetic markers are essential for DNA profiling of individuals or for genetic linkage and association studies.

Ideal markers are abundant, co-dominant, highly polymorphic, and distributed throughout the genome (Bennett 2000).

A highly useful genetic marker type is microsatellite, which is a short (one to five base pairs) tandem repeat. There are approximately 30 000 microsatellites scattered throughout the human genome (Bennett 2000) and they can be very informative for genome wide linkage studies because they have multiple alleles; the number of repeats varies from one individual to another. The most common type of genetic polymorphism is a change in a single nucleotide position: a single nucleotide polymorphism (SNP). They are diallelic and as genetic markers, they are therefore less informative than microsatellites. However, to their advantage SNPs are technically easier to work with and less expensive to genotype. Interestingly, some SNPs can directly affect phenotypes (Shastry 2002). Currently, more than 11 million SNPs have been identified in the human genome (www.ncbi.nlm.nih.gov/SNP). These polymorphisms are found more frequently in the exons of genes than microsatellites are.

Third type of genetic marker is an insertion or deletion (INDEL) polymorphism. These are found throughout the genome, with at least 1.6 million INDELs estimated to exist in humans. Some of these are expected to have an impact on human gene function. Various studies screening for INDELS have been performed and interest in these polymorphisms is growing (Mills et al 2006).

Copy number variation (CNV) results from gains and losses of genomic regions that are between 1000 base pairs in length to several megabases long. At least 12% of the human genome is subject to CNV, so this type of variation also significantly contributes to genetic diversity and most of the difference between two randomly selected genomes is attributed to CNV. Copy number variation has the potential to alter gene dosage and disturb regulation of gene expression, thus it is capable of altering phenotype of an individual (Redon et al 2006).

It is impossible to genotype every polymorphisms that exist within the human genome in a genetic study; therefore it is necessary to select an informative subset of markers for such study. Choosing both an appropriate type of marker and a sufficient number of markers are important aspects of the design of a genetic study.

4.3.3 The design of genetic linkage and association studies

A beneficial genetic linkage or association study of human multifactorial diseases needs to carefully define the phenotype of interest in order to minimize sources of error and ambiquity. Importantly, it is inappropriate to perform a genetic study before determining that the feature under examination actually has a genetic component. In addition, prior to commencing, aims of the study should be clear and a hypothesis determined in order to select the most suitable strategy for the investigation.

4.3.3.1 Recruitment of a representative study population

Subject enrolment is perhaps the most critical component of the study design as size and suitability of the study population strongly affect the power of the study. It is also judicious to collect detailed information on family and disease history, as well as environmental exposures or impacting factors.

Prospective genetic studies are perhaps most optimal as they allow for the collection of subject information relevant to the new study and base-line samples prior to surfacing of the disease phenotype; also, study controls can be selected from the same population. However, prospective studies require a lot of time and resources. In addition, these types of studies are only possible for relatively common diseases (Kraft & Hunter 2005). Therefore, retrospective subject recruitment is the more frequent practice. But retrospective studies can have several disadvantages, including survival bias (patients having a lethal form of the disease may die before they can be enrolled in the study),

selection bias (controls do not adequately represent the population from which cases are collected), or recall bias (where a patient's report about their disease history is influenced by their diagnosis) (Kraft & Hunter 2005).

If the generation of a disease appears to be sporadic, the only option for an investigator is to collect the cases and try to find sufficiently matched controls for them. A major weakness of this kind of case-control study design is the potential for confounding factors influencing the outcome of the study, such as age, ethnicity, environment or sex. Hence, to avoid false positive results, controls must adequately match the recruited cases

Family-based study design is immune to population stratification bias because family members are used as controls. In addition to traditionally used twins nuclear families and sibling pairs are also effective study subjects. Extended pedigrees contain lots of genetic information but they are not easy to find. A major disadvantage of a family-based study is the difficulty to collect DNA samples from all parents, especially for late-onset diseases.

The frequency of the variant genotype as well as the risk ratio of the gene determines the required size of the study population to achieve sufficient statistical power. If the risk of disease is only marginally greater for those who carry the genetic variant and frequency of this allele in the general population is low, then the number of patients required for an informative study will be enormous. Genotyping more controls than cases increases the statistical power of the study, or in other words, decreases the number of affected subject needed (Lee 2004).

The selection of particular ethnic groups or isolated populations to the study can make association studies easier. Such populations often have a limited number of founders and can thus be assumed to be genetically more homogeneous. Also, etiological and phenotypic heterogeneity is reduced (Ellsworth & Manolio 1999).

4.3.3.2 Gathering informative genetic data

There are two general ways to approach disease gene discovery: genome-wide screening and candidate gene study. Genome-wide screening attempts to search the entire genome using markers that are evenly spaced throughout the genome and find a region linked or associated to the disease in question. It does not require any prior knowledge about the function of genes or the biology of the disease. However, if an interesting candidate gene is suggested to be functionally related to the disease, the faster and less expensive approach is to study that particular gene directly.

Markers to be typed must be informative enough and cover the whole region of interest. Polymorphisms in coding or promoter regions of gene are recommended because they can be directly associated with the disease (Daly & Day 2001) but random non-coding variations are more abundant and can be useful markers. Rare variants with a minor allele frequency less than 5%, are likely to be population specific (Risch 2000) and uninformative. But polymorphisms common to all major ethnic groups may be also useless because they are assumed to be neutral and in weak linkage disequilibrium with causal variants (Collins et al 1999). For replication studies, standard sets of markers have been proposed because then it is easier to compare and combine the data from different studies.

Genotype errors will reduce the power of the study so it is important to choose a trustful genotyping method and follow good laboratory practise. Sample size and the amount of DNA needed must be considered when choosing a genotyping method. Ultimately however, the decision will depend on the money available to the investigator.

4.3.3.3 Ethical aspects

From the planning stages until the publication of results, a study should be aware that it is meeting certain ethical standards. Full information about the nature of the study and the purpose for which the data will be used should be

given to all study participants. Samples should be collected in a humane and ethical way, respecting an individual's rights and cultural or religious beliefs. All samples should be coded so that the study subjects are not identifiable and private data should remain confidential. Usually, study subjects will not benefit financially by participating in the study, but the results of the investigation should be reported publicly and the patients should benefit about the study at some point in the future (Endacott 2007).

4.3.4 Complex diseases

An appropriate study design is particularly important for the study of complex diseases because overall susceptibility is the sum of multiple genetic and environmental risk factors, where each of these risk factors usually has only a modest effect on susceptibility (Figure 4). Typical examples of complex diseases include autoimmune diseases such as diabetes, coeliac disease and asthma, as well as neurological conditions like schizophrenia, bipolar disorder and autism.

Studies with twins, or comparisons between first-degree relatives and the general population, do suggest that complex diseases are heritable although they do not exhibit classic Mendelian inheritance. However, it is difficult to separate shared genetics from shared environmental influences in these studies. Complicating the matter further is the fact that the responsible environmental triggers are still largely unknown for most complex diseases. Environmental factors proposed to contribute to disease risk include exposure to chemicals or pathogens, lifestyle and diet. Also, late or variable age of onset and variable disease progression are hallmarks of complex diseases.

One characteristics of complex disease is locus heterogeneity, where different variations in different genes can create the same phenotype. Locus heterogeneity is a challenge to genetic mapping because each such variant is derived from a distinct ancestor and is thus located on a different ancestral haplotypes. The search for disease susceptibility genes can also be difficult due to a

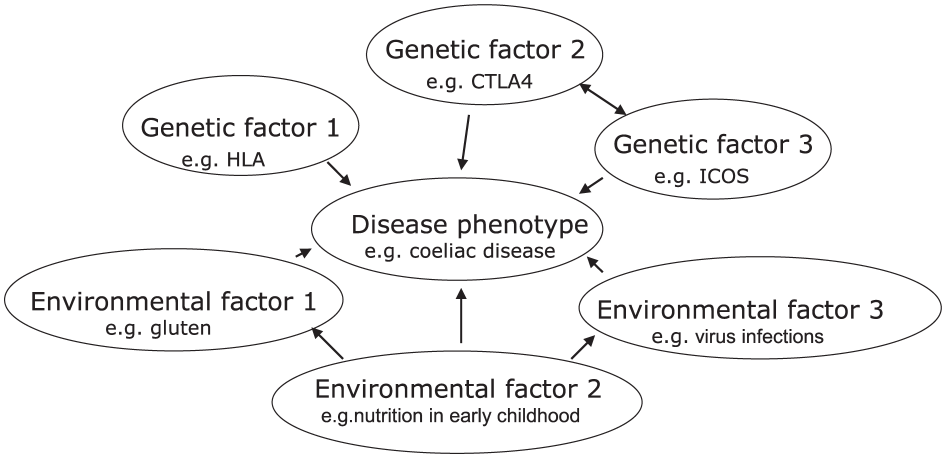


Figure 4. Development of complex disease, e.g. coeliac disease, is an interaction of many factors.

phenomenon known as incomplete penetrance, where the predicted phenotype is not always present in individuals carrying the disease predisposing genetic variation.

Defining a phenotype for complex diseases can be problematic due to multiple manifestations, subclinical forms of the condition or variable age of onset. In addition, environmental factors can modify incidence rates (Ellsworth & Manolio 1999). Nevertheless, clear phenotype definition and uniform diagnostic criteria will lead to more informative genetic results when investigating complex diseases. Although highly specific diagnostic criteria for defining disease can yield a more homogeneous group of study subjects, in other hand it may also limit both patient recruitment and universality of the results for other disease subgroups (Dawn Teare & Barrett 2005).

The "common disease, common variant" hypothesis means that genetic susceptibility to common complex traits is influenced by relatively common polymorphisms (Chakravarti 1999). So, susceptibility alleles are also common among healthy population and individuals carrying susceptibility allele will not necessarily develop the disease, carriers simply have an increased risk of disease.

4.3.5 Organ transplantation as a multifactorial trait

Although transplantation is not a disease, it is a multifactorial condition. A unique feature of transplantation is that genetic and environmental factors of both the donor and recipient will influence the outcome. Several genes appear to affect the outcome of transplantation, and they also indirectly predispose to the need for transplantation (for example, diabetes is the major reason for kidney transplantation). Environmental factors are numerous: the original disease causing the requirement of kidney replacement, donor matching, surgical aspects, condition of the donor organ prior to transplantation etc. Today, immunosuppressive drugs are highly effective and can repress variance in immune responses which genes may cause; this phenomenon, from a geneticist's point of view, can be compared to incomplete penetrance. Definition of outcome phenotypes like rejection or function of organ is complicated, too.

4.3.6 Linkage and linkage disequilibrium

Linkage analysis is traditionally the first phase of the genetic mapping, since it can be used to identify a genomic region that might contain a disease related gene. The linked region is often large, thus fine mapping of the disease causing variant is required. Association tests based on linkage disequilibrium (LD) can be used to narrow the region.

4.3.7 Linkage disequilibrium and haplotypes

Linkage disequilibrium means the non-independence of alleles at two different loci. If particular alleles are detected together more frequently than is expected by chance, these alleles are said to be in linkage disequilibrium (Pritchard & Przeworski 2001). This phenomenon is essential for genetic studies as it facilitates fine mapping of susceptibility genes. In association studies, disease variants can be discovered through the LD with markers locating nearby (Devlin & Risch 1995). When a new mutation, a possible disease

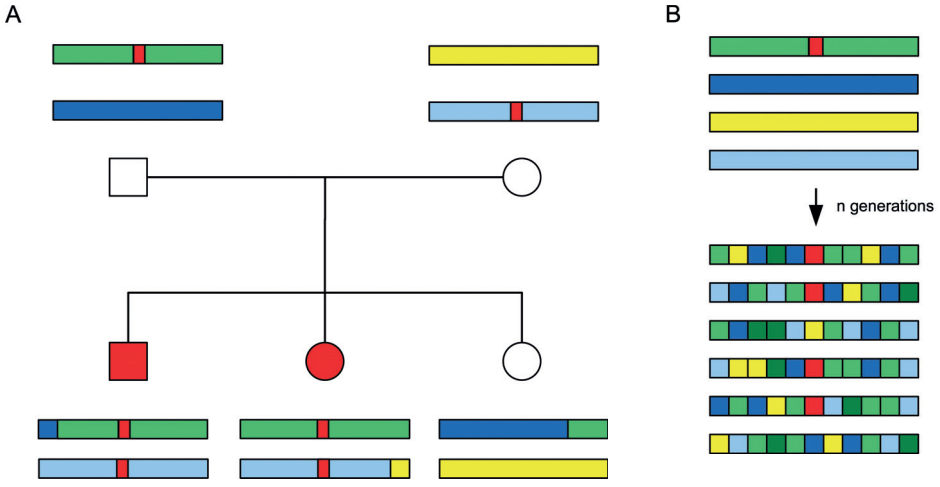


Figure 5. Disease causing variant (red box in the figure) is always located in a haplotype (green, yellow, dark blue and light blue oblongs) with other, pre-existing genetic variants (not marked into the figure). Recombination divides haplotypes and thus eliminates linkage and association over generations. Linkage analysis aims to identify haplotypes that are inherited intact within the family revealing often large regions (A). Association studies explore adjacent variants over many generations so historic recombinations have split the region very small (B).

causing allele, arises, it co-occurs always with certain flanking markers; thus, the disease allele and alleles of the flanking markers are in complete LD. Through time, recombinations delimit the complete LD and little by little alleles of markers further away from the disease locus are in weaker LD with the disease causing allele (Figure 5). Besides time and physical distance, also population history, population isolation, a distance from the centromere, and a presence of recombination hot spots affect the level of LD. Therefore, extent of LD varies from one genomic region to another as well as between different populations (Cardon & Bell 2001).

4.3.7.1 Linkage disequilibrium block as a tool in genetic studies

Haplotype is a combination of alleles at different loci along the same chromosome that are inherited as a unit. The offspring inherits one haplotype from both parents

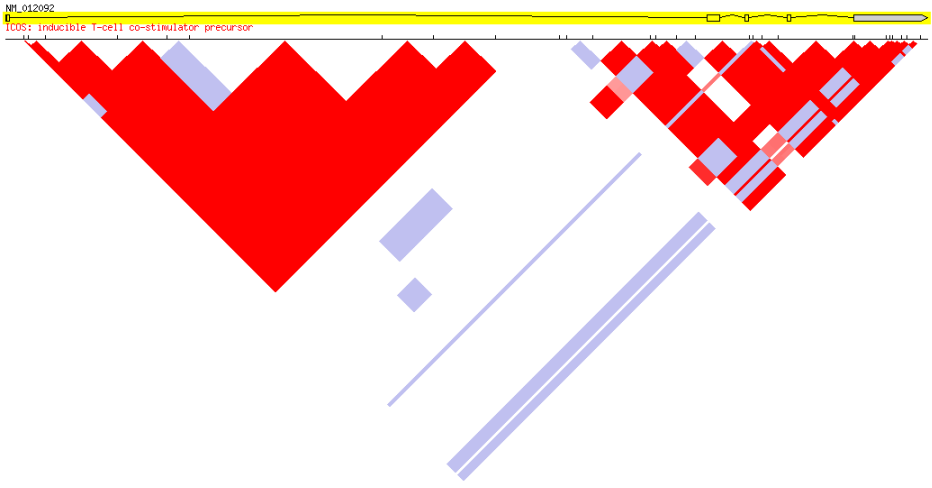


Figure 6. An example of linkage disequilibrium blocks. The *ICOS* gene is divided into two blocks of strong LD. Red colour refers linkage disequilibrium of D'=1 and LOD≥2, blue D'=1 and LOD<2, and white D'<1 and LOD<2.

and so the genetic variation is segregated in haplotypes through generations. Recombinations delimit ancestral haplotypes and create at the same time new haplotypes. Genome sites with lots of recombinations and no LD are called recombination hot spots and they flank the regions of strong LD and of low recombination rates. Hence, the genome is divided to LD blocks (Daly et al 2001, Goldstein 2001, Johnson et al 2001) (Figure 6). Each block encloses only a few conserved haplotypes. When LD blocks are defined and haplotypes identified, it is possible to select particular haplotype-tagging SNPs (htSNPs) which serve as substitutes to the variation of a whole block (Johnson et al 2001). Thus, the region can be scanned with a limited set of tagging markers. The use of htSNPs increases efficiency of the mapping and power of the test. The information of the block is reached by a smaller set of representative markers as well as the number of tests is reduced by testing haplotypes instead of every single SNP (Clark 2004).

4.3.7.2 Measuring linkage disequilibrium

Several measures of the level of LD have been formulated (Devlin & Risch 1995); two commonly used are D'

(Lewontin 1964) and r^2 (Hill & Roberson 1968). Both D' and r^2 measures range from 0 to 1. D' describes the difference between observed and expected probabilities of the alleles in two loci and it is directly related to the recombination rate. D' equals 1 if at least one allele at each locus is completely associated with an allele at the other locus (known as complete disequilibrium, Figure 7A); whereas r^2 equals 1 if there exists two haplotypes for two biallelic loci. The measure r^2 entails that each allele at each locus is completely associated with one particular allele at the other locus and allele frequencies at both loci are identical (known as the perfect LD, Figure 7B). Thus, r^2 is allele frequency dependent and D' is better when analysing multiple allele data but on the other hand, D' depends on sample size (Ardlie et 2002).

The International HapMap project aims to create a genome-wide map of LD. It is a multi-ethnic map based on four different populations (The International HapMap Consortium 2003). The HapMap aids human geneticists by defining LD blocks and pinpointing haplotype tagging SNPs. Although it is easy to use the HapMap for designing a mapping study, a researcher should be familiar with the detailed LD structure of his/her population of interest. For example, a LD map of the Finnish population may be quite different to that of mixed populations. Due to isolated location of Finland and a bottle neck in the population history, LD blocks are extensive in some chromosome regions (Arcos-Burgos & Muenke 2002).

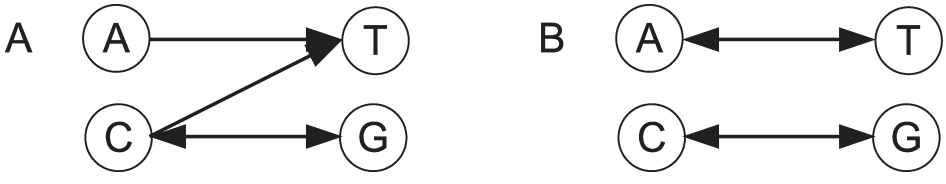


Figure 7. If an allele (A in the figure) is completely associated with an allele at the other locus (T in the figure) it is referred as complete disequilibrium (A). If each allele (A and C in the figure) is completely associated with one particular allele at the other locus (T and G in the figure) and allele frequencies at both loci are identical, linkage disequilibrium is perfect (B).

4.3.8 Linkage analysis

The aim of linkage study is to pinpoint the chromosomal location that shows evidence of containing a disease gene. Markers close to the disease gene co-segregate with the disease within the family. Linkage would be suggested if affected members of a family share significantly more alleles identical by descent than expected by chance (Dawn Teare & Barrett 2005). The most commonly used version of the linkage method in complex diseases is the affected sib pair test due to easiness of collection the patients. But the higher the number of affected individuals in a pedigree, the more informative the family is, and therefore extended families are very valuable.

In linkage analysis, genetic markers cover the whole genome or a large candidate region. Markers do not need to be functionally relevant; regularly spaced microsatellites can be used in an initial screen of the whole genome.

Linkage analysis has been the main method of localising the chromosomal region of disease genes. Association tests with a denser set of markers have then been used to further localise the disease-predisposing gene because linkage disequilibrium extends over much smaller distances than linkage (Dawn Teare & Barrett 2005). Linkage analyses have been successful in the studies of single gene disorders. But, for complex diseases it is typical that each genetic variant has only a modest contribution; there is a lack of clear genetic segregation of variants in large families and thus, association tests may be more efficient in complex diseases than linkage studies (Hirschhorn & Daly 2005).

4.3.9 Association analysis

Association studies compare allele frequencies of genetic markers in groups of affected and unaffected individuals in a population. The aim is to identify genetic variations that increase the risk of the disease and are therefore at a higher frequency among cases than among unaffected controls.

Association mapping localises genes or genetic variants involved in disease susceptibility through association with nearby markers. LD between a marker and a disease causing variant is assumed; it can be obtained with a dense enough set of markers (Kruglyak 1999). It is important to understand the structure and extent of LD across the region of interest to be able to choose an optimal marker set and to interpret the results of an association study correctly. Use of functional SNPs, which may have a direct biological influence to the trait under study, is naturally highly recommended (Risch & Merikangas 1996, Daly & Day 2001) but it is possible to choose also random markers which are more widespread and can reveal the disease locus through LD (Collins et al 1997). The most practical strategy is to combine different kind and number of polymorphisms considering statistical, technical and economical aspects. The allele frequencies of the disease locus and of markers used in the study should be alike, so the LD between them may be as complete as possible.

4.3.9.1 Adequate matching of controls is essential

Association studies are used especially for multifactorial disorders. The power of association analysis in detecting genetic factors with a low relative risk can be greater than that of linkage studies (Risch & Merikangas 1996). It is, however, crucial to match the control group very carefully for ethnicity, demography, age, sex etc., because any systematic allele frequency difference between cases and controls can result in false positive disease association. A significant reason for numerous false-positive association results may be a poorly matched control group; other causes for a lack of replication may be weak study design, small sample size, over-interpretation of results and publication bias towards positive results (Cardon & Bell 2001).

If the association study gives a positive result but the linkage study does not, the association found may be spurious, because no familial segregation can be detected. Thus, supportive evidence for a positive association result

is needed from linkage analysis or functional assays. It is also important to attempt to replicate the results in other populations.

In association studies, multiple testing is a common practice which creates a risk of associations by chance. Multiple testing must be taken account when interpreting the results, or preferably the results should be corrected appropriately for multiple testing. In addition to multiple markers, analysing small subgroups of the study population is another form of multiple testing (Cardon & Bell 2001).

4.3.9.2 Association study in family material

Family-based association studies have been developed to avoid spurious associations due to imperfectly matched controls. In family-based studies, the parental alleles not transmitted to the affected child are used as control alleles. There are several positive features in family-based studies like a possibility to check consistency in genotyping of family members and to detect maternal versus paternal effects. Use of families also facilitates haplotype association studies and determination of LD (Thomson 1995).

The most commonly used test in family-based association studies is the transmission disequilibrium test (TDT) which takes advantage of heterozygous parental genotypes providing a joint test of linkage and association (Spielman et al 1993). While the test surveys only heterozygous parents, it wastes some genotype information leading to a loss of power. A major limitation of TDT testing is the requirement of collecting samples from parents of affected individuals which can be difficult especially for late-onset disorders. A solution to non-available parental samples is a variation of the test using genotypes of siblings (Curtis 1997, Martin et al 1997).

4.3.10 The purpose of genetic disorder studies

Knowledge on genetic basis of a disease results better understanding of its pathogenesis. Genetic studies help clarify those immune mechanisms which are common for autoimmune diseases and those which are unique

to a particular disease. The genetic base would provide potential targets for effective treatment and promote design of new therapies. The focus of new treatments is to maximise target specificity while minimising unwanted side effects.

Knowledge of disease causing variants would improve the diagnostics and enable cost-effective screening of people at risk. It would also increase the understanding of why some patients do not respond to currently available treatments while others do. More tailored therapeutics on the basis of genotype of each patient could be developed if the exact functional variants would be identified.

In addition, awareness of disease mechanisms might provide a set of useful biomarkers to determine the risk of developing and progression of a disease. Possibility to predict the risk of disorder might enable avoidance of the severe damage or even prevention of the disease. Prediction would be especially important when early diagnosis and intervention positively impact the patients. For transplantation patients, early treatment affects significantly the outcome and fine tuning of the treatment might lead to a better graft and patient survival. But, until an effective preventive treatment will be available, the identification of individuals at increased risk of a disease includes ethical problems.

At the moment, the significance of *HLA* in the risk of autoimmune disease or rejection of organ transplant is known. The main *HLA* risk haplotypes are identified for many autoimmune diseases and thus, it is possible to distinguish on the basis of *HLA* alleles those individuals who, for example, have a very low risk to develop coeliac disease. Exclusion is easier to do than a positive prediction. There are a number of unknown genetic and environmental factors that also influence the outcome, so it is impossible to differentiate which carriers of the *HLA* risk haplotype will develop the disease.

5. AIMS OF THE STUDY

The purpose of this thesis was to investigate the polymorphism in the genes of co-stimulatory receptors and to determine the role of the polymorphism in susceptibility to immunological disorders with the following specific aims:

I To study the variation in the *ICOS* gene by sequencing all exons of the gene and to determine the frequencies of polymorphisms found in Finnish population.

II To determine whether an association exists between the 2q33 gene region and the genetic susceptibility to coeliac disease by screening markers on the *CD28*, *CTLA4*, *ICOS* and *PDCD-1* genes in Finnish families with coeliac disease

III To study whether IgAD, CVID and coeliac disease would share a common genetic risk factor within *CTLA4* or *ICOS*.

IV To examine the effect of genetic variation in *CD28*, *CTLA4*, *ICOS* and *PDCD-1* on the outcome of kidney transplantation.

6. MATERIALS AND METHODS

6.1 Study subjects

Study I. Altogether 63 Finnish individuals were studied.

Study II. A total of 106 Finnish coeliac disease trio families were included in the study. All families consisted of one randomly selected affected child and both parents.

Study III. The study comprised 278 Finnish coeliac families including 619 patients as well as 385 Hungarian coeliac families including 657 patients. A cohort of independent CD cases, including 181 patients from Finland, 259 from Hungary and 135 from Italy as well as 231 Swedish and 44 Finnish IgAD individuals and 87 Swedish CVID patients were studied. The study also included matched controls: 267 from Sweden, 178+89 from Finland, and 184 from Italy.

Study IV. A cohort of 678 renal transplant recipients was selected for the study.

6.1.1 Coeliac disease patients

The Finnish CD patients and their families were recruited on a voluntary basis with the help of the Finnish Coeliac Society by advertising in the patients' newsletter. The samples were collected at the University of Tampere. The prior diagnoses were re-evaluated by scrutinizing the medical records. Antibody screenings were performed to reveal the asymptomatic cases. Diagnoses were based on a small-bowel biopsy (or skin biopsy from dermatitis herpetiformis patients) and/or coeliac autoantibody positivity.

The Hungarian coeliac families were collected at the Heim Pal Children's Hospital, Budapest, and at the University of Debrecen, Department of Paediatrics. The diagnoses were based on an initial small bowel biopsy and/or autoantibody positivity.

The Italian CD material consisted of independent single cases. Untreated coeliac patients were diagnosed at the

Burlo Garofolo Children's Hospital, Trieste. Diagnoses were based on the ESPGHAN criteria including intestinal biopsies (Walker-Smith et al 1990). The patients' serum samples tested positive for both anti-transglutaminase and anti-endomysium antibodies.

All the coeliac disease patients from Finland, Hungary and Italy carried known *HLA* disease susceptibility alleles DQB1*02 and/or DQB1*0302.

6.1.2 Kidney transplantation patients

Kidney recipients underwent transplantation between 1999 and 2003 at the Renal Transplant Unit of Helsinki University Hospital, Helsinki, Finland. The mean age of the adult patients was 50 years and 63% of them were male. Three most common diagnoses before transplantation were glomerulonephritis, type I diabetes and polycystic kidney. Eighty-two per cent of the patients received immunosuppressive therapy containing cyclosporine/ tacrolimus, azathioprine/ mycophenolate mofetil and methylprednisone and 18% received additional antibody induction therapy (anti-thymocyte globulin, basiliximab or daclizumab). Acute rejection was determined by clinical and biopsy findings (Banff criteria) in 17% of the patients. Delayed graft function was defined as serum creatinine >500 μmol/l throughout the first week, more than one dialysis session or oliguria (<1 l per day) for more than two days after transplantation. The mean of patient survival was 1491 days (from 3 to 2607 days) and of graft survival 1471 days (from 1 to 2607 days).

6.1.3 IgA deficiency patients

Diagnoses of Swedish individuals with IgAD were established in accordance with accepted recommendations (ESID criteria, www.esid.org), on the basis of the measurements of Ig levels in multiple independent blood samples. Patients had a serum IgA level of less than 0.07 mg/L but normal serum IgG and IgM. Selective IgA deficiency in Finnish blood donors has been defined by IgA levels lower than 0.05 mg/L.

6.1.4 Common variable immunodefieciency patients

Swedish patients with CVID were diagnosed on the account of measurements of Ig levels in multiple independent blood samples. Patients had a marked decrease of IgG as well as an apparent decrease in at least one of the isotypes IgM or IgA.

6.1.5 Healthy controls

The study comprised healthy unrelated volunteers representing the populations of Finland, Sweden, Italy and Hungary. Controls were demographically matched with patients.

6.2 Methods

6.2.1 DNA extraction (I, II, III, IV)

DNA was extracted from citrate or EDTA anticoagulated whole blood samples using commercial methods (QIAamp DNA Blood Mini Kit, Qiagen Inc, CA, USA; DNA Isolation Kit, PelFreez Clinical Systems, Brown Deer, WI, USA; FlexiGene DNA kit, Qiagen Inc, CA, USA).

6.2.2 HLA typing (II, III, IV)

HLA-DQ typing of coeliac disease patients was performed using commercial kits (OLERUP SSP AB, Saltsjöbaden, Sweden or DELFIA, PerkinElmer Life Sciences Wallac Finland Oy, Turku, Finland).

HLA- A, B and *DRB1* alleles of kidney patients were determined before transplantation either by serological assay based on complement-dependent cytotoxicity (Biotest, Dreieich, Germany) or by PCR-SSO method (sequence-specific oligonucleotides) (Inno-LiPA kit, Innogenetics, Gent, Belgium).

HLA specificities of the Swedish IgAD individuals were determined using commercial PCR-SSP (sequence-specific primers) kits (OLERUP SSP AB, Saltsjöbaden, Sweden).

The Swedish controls and the controls for IgA levels were genotyped for the SNP rs2187668 previously reported to tag the *HLA-DRB1**03-*DQA1**05-*DQB1**02 haplotype (van Heel et al 2007, Monsuur et al 2008). Genotyping was performed using the TaqMan On Demand assay by Applied Biosystems (Applied Biosystems, Foster City, CA, USA).

For a long time determining of HLA specifities rested on serological typing before more accurate PCR-based methods were developed. Routine high throughput HLA typing kits are mainly based on either sequence specific primers or sequence specific oligos used as probes. HLA typing is so complex that the use of validated commercial kits is reasonable. But HLA typing is also so expensive that it is tempting to try to develop techniques capable of defining only a certain haplotype or HLA-allele of interest.

6.2.3 Cell preparation and cDNA amplification (I)

Peripheral blood mononuclear cells were isolated by by Ficoll-Paque

(Amersham Pharmacia Biotech AB, Sweden) gradient centrifugation.

The cells were stimulated with anti-hCD28 mAb and with plate-bound anti-hCD3 mAb (R&D Systems, USA). Total RNA was isolated from activated cells using RNAzol B-solution (Wak-Chemie Medical, Germany) and complementary DNA was prepared using OmniScript RT-PCR kit (Qiagen, Germany). *ICOS* cDNA was PCR-amplified with *ICOS*-specific primers 5'GAAGTCAGGCCTCTGGTATTTCT3' and 5'TGCAAACAACTCAGGGAACAC3'.

6.2.4 Sequencing (I)

At the time of the study, the genomic reference sequences were available in public databases for exon 1 and the 3' end of the last exon. These regions, together with 250 bp of intron 1, were sequenced from genomic DNA. The remaining regions of the exons were sequenced from a 2.3-kb fragment of cDNA in six partially overlapping fragments, using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and ABI

PRISM 310 sequencer (Applied Biosystems). Some samples were sequenced by the SEQLAB Sequence Laboratories in Göttingen, Germany.

Sequencing is the best method to identify new gene mutations because it reveals the location and type of the mutation at the same time. Dideoxy DNA sequencing is the most commonly used sequencing technique; it is simple and sensitive.

6.2.5 Genotyping

In addition to the new polymorphisms we had identified, we also sought useful markers (Figure 8) for genotyping from publications and genetic maps of public databases. The markers were selected according to their minor allele frequency (as high as possible, at least 0.1), tagging properties (linkage disequilibrium) and location (potentially functional polymorphism, coverage of the gene region).

The coding of polymorphisms has been confusing: some SNP identifiers have given a name based on the location of the polymorphism (for example ICOSivs4+1070G/T) but some have used personal codes (for example CT61). The National Center for Biotechnology Information single nucleotide polymorphism dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/) uses so called rs numbers as identifiers for SNPs (for example rs231770) and these rs numbers have become established, which has enabled the comparison of the markers studied in different publications.

6.2.5.1 Microsatellite genotyping (II)

For microsatellite genotyping, standard PCR methods with fluorescent labelled primers were used, and the size of the alleles were determined by electrophoresis using ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Fluorescence-based microsatellite genotyping with semi-automate is an efficient high-throughput method. PCR-amplification can be multiplexed and different dyes can

be used to label microsatellite alleles. Multiplex PCR optimization and designing the primers are the most challenging part of the otherwise simple method.

6.2.5.2 SNP genotyping

6.2.5.2.1 SNaPshot (I, II, III)

SNaPshot is a minisequencing method (ABI PRISM SNaPshot Multiplex Kit, Applied Biosystems) that is based on the extension of a T-tailed primer immediately adjacent to the SNP under study using fluorescently labelled ddNTPs. The labelled extension products were visualized by electrophoresis on ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). Primers have to be designed to bind the region immediately adjacent to the SNP but the binding site can be on either the sense or antisense strand. Due to the length of tailed primers and ddNTPs labelled with different dyes, it is possible to pool PCR products and genotype up to ten SNPs simultaneously. SNaPshot is a medium throughput genotyping method. The analysis of the results was time consuming because appropriate analysis software was not available.

6.2.5.2.2 Restriction fragment length polymorphism, RFLP (I, II, IV)

RFLP is a common and traditional method for analyzing digested PCR products on a gel. The amplified PCR product containing a polymorphism is digested with a restriction enzyme at specific nucleotide sequences resulting in a certain fragment pattern visualized by gel electrophoresis.

The RFLP method is a robust and easy method, which does not need expensive equipment but it is also laborious and slow.

6.2.5.2.3 Sequence specific primers, SSP (II, III)

Allele-specific primers designed specific to the polymorphism under study are employed in PCR amplification producing a PCR fragment if the sample contains the allele matching the primer. Each PCR reaction includes also a

primer pair amplifying a certain fragment from a general gene locus common for all the samples which functions as an internal positive amplification control (Olerup & Zetterquist 1992).

6.2.5.2.4 Sequenom MassARRAY system (III, IV)

The Sequenom MassARRAY system (Sequenom Inc., San Diego, CA, USA) is based on the MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) technology using mass spectrometry and primer extension chemistry. Primers are designed to extend beyond the SNP site. Following PCR amplification, primer extension products are analysed by chip-based MALDI-TOF mass spectrometry. Genotyping with this method was performed either at the Finnish Genome Center, Helsinki, or at the MAF core facility in Karolinska Institute, Stockholm.

The technique enables theoretically even 36-plex amplification, so it is quick, effective and cheap, but on the other hand, the assay requires expensive specialised equipment and the designing of primers is challenging.

6.3 Statistical analyses and computer softwares

6.3.1 Data quality check

Deviations from Mendelian inheritance in the family materials were determined with PEDCHECK 1.1. (O'Connell & Weeks 1998). All the genotype frequencies were examined for Hardy-Weinberg equilibrium.

6.3.2 Linkage (III)

MERLIN 1.0.1 (Abecasis et al 2002) was used to perform non-parametric linkage analysis in the coeliac pedigrees.

6.3.3 Association (II, III, IV)

The transmission/disequilibrium test (TDT) was calculated by GENEHUNTER 2.1 (Kruglyak 1998) to investigate genetic linkage in the presence of allelic association in the family materials. The TDT compares transmitted versus non-transmitted alleles from heterozygous parents to

the affected offspring. Haploview 4.0 (Barrett et al 2005) was used to estimate haplotypes. Statistical analyses for association tests were performed using SPSS 15.0 (SPSS Inc., Chicago, IL, USA), StatsDirect (StatsDirect Ltd. Cheshire, UK) or Haploview 4.0. Comparisons of the allele and genotype frequencies were made using Fisher's exact test.

6.3.4 Survival (IV)

Graft survival rates were calculated using Kaplan-Meier method and statistical significance was estimated using the log rank test.

6.3.5 Linkage disquilibrium (I, II, III)

Test of linkage disequilibrium was performed by ARLEQUIN program package (Schneider et al 2000) or by Haploview 4.0 (Barrett et al 2005).

6.3.6 Power (II, III)

Power in the genetic studies was estimated using the online Genetic Power Calculator (http://statgen.iop.kcl.ac.uk/cgi-bin/powercalc/dtdt.cgi) (Purcell et al 2003).

6.3.7 Meta analysis (III)

The Cochran-Mantel-Haenszel (CMH) function in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) was used to perform meta-analysis of combined case-control datasets. Forest plots were produced with the "forestplot" function in rmeta 2.14 on the R platform (http://cran.rproject.org/web/packages/rmeta/index.html).

6.4 Ethics

The study has been approved by the ethical committees of the Helsinki University Central Hospital (IV), Tampere University Hospital (II, III), Huddinge University Hospital (III), Heim Pal Children's Hospital (III), University of Debrecen (III), Burlo Garofolo Children's Hospital (III) and Finnish Red Cross Blood Service (I, III).

7. RESULTS AND DISCUSSION

7.1 Polymorphism screening of the gene region 2q33 (I and II)

At the time of the study, the entire sequence of the *ICOS* gene was not known yet. The genomic reference sequences for the exon 1 with a fragment of the first intron and for the 3' end of the last exon were available in public databases. The reference sequences of *CD28* and *CTLA4* were accessible but only a few polymorphisms in the genes had been identified (Polymeropoulos et al 1991, Harper et al 1991, Deichmann et al 1996, Marron et al 2000). We began the study by sequencing all exons of *ICOS* as well as the 5'-flanking sequence upstream from the transcription initiation site of both *ICOS* and *CD28* in order to screen for genetic variation. We found 11 new single nucleotide polymorphisms in the *ICOS* gene: one in the 5'-flanking region, one in intron 1 and nine in the 3-untranslated region of the last exon. We identified also a novel SNP in the 5'-flanking region of the *CD28* gene as well as a microsatellite marker (CTTT)n with 14 alleles (Figure 8).

To estimate the allele frequencies in the general Finnish population, unrelated, apparently healthy individuals were genotyped for all the polymorphisms we had identified. The observed minor allele frequencies varied from 0.02 to 0.31. All polymorphisms were found to be in Hardy–Weinberg equilibrium.

Since 2001 several other screenings for polymorphism have been published. Ihara et al sequenced the exons and exon/intron boundaries of *ICOS* in ten individuals and found two microsatellite markers in intron 4, which were the first published polymorphisms in the gene (Ihara et al 2001). Haaning Andersen et al sequenced 10 Danish individuals for entire the gene except intron 1 and identified 11 novel polymorphisms in addition to the microsatellites found by Ihara et al (2001) and the SNPs published by us (2002) (Haaning Andersen et al 2003). Similar allele

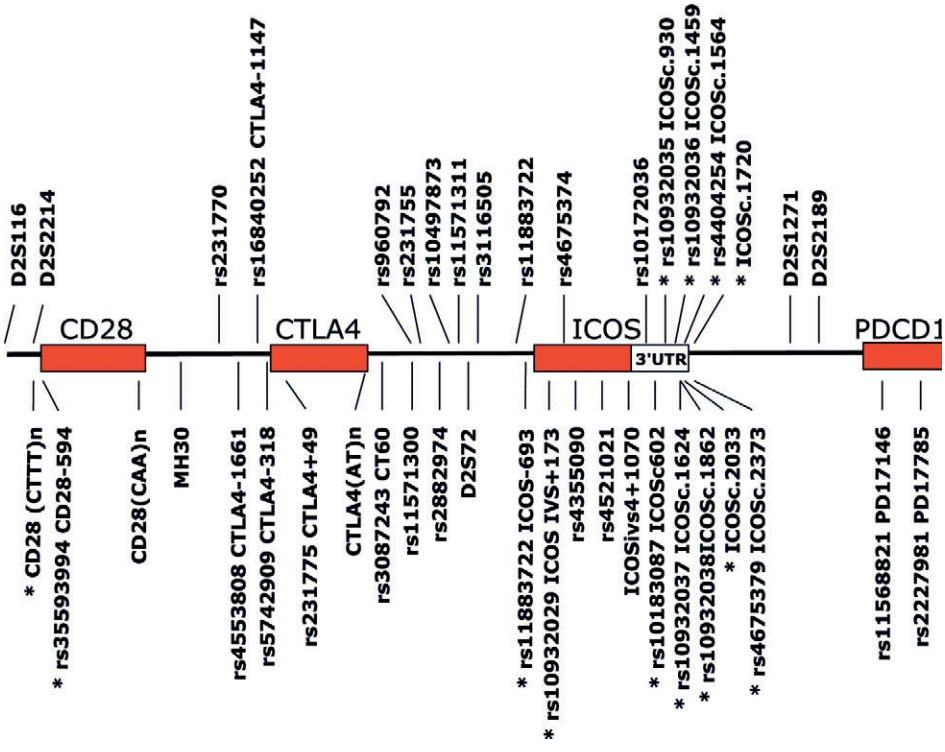


Figure 8. Location of markers identified in the studies I and II (marked with *) and genotyped in the studies II, III and IV. Markers for genotyping were selected based on their minor allele frequency, tagging properties and location. Distances are not to scale. 3'UTR=3' untranslated region

and genotype frequencies were detected in the Finnish and Danish populations. No polymorphisms that altered amino acids were found in the *ICOS* gene in any of these screening studies.

The gene encoding PDCD1 was completely sequenced in ten unrelated individuals in 2002 and seven SNPs were identified (Prokunina et al 2002). Nielsen et al found additional eight SNPs by sequencing but seven of them were extremely rare (Nielsen et al 2003).

Ueda at el conducted an impressive study by sequencing the region of 330 kb containing *CD28*, *CTLA4* and *ICOS* in 2003. They reported 108 SNPs covering the 2q33 region, mostly located between the genes (Ueda et al 2003).

Nowadays, there are over 100 confirmed polymorphisms in the *CD28*, *CTLA4*, *ICOS* and *PDCD1* genes and much more in intergenic regions available in the database of the The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/projects/SNP/).

7.2 Linkage disequilibrium and haplotype analyses of the T cell co-stimulatory receptor genes (I, II and III)

7.2.1 The linkage disequilibrium structure of the 2q33 region (I, II, III)

An exact test of linkage disequilibrium between all pairs of polymorphic sites was performed using the Arlequin program (Schneider et al 2000) or Haploview (Barrett et al 2005). Three blocks of strong LD were observed: one including *CD28*, another covering *CTLA4* and the 5' region of *ICOS* up to the first intron and the third block extending from the exon 2 to the end of *ICOS* starting (Figure 9). Only much weaker LD was detected between intron 1 markers and the other polymorphisms located in the last exon of *ICOS* so the gene was split into two blocks. Our findings of the LD structure of *ICOS* supported those seen in British population by Ueda et al (Ueda et al 2003). In the study III, we measured the amount of, and explored the structure of LD in the genotyping data of the *CTLA4-ICOS* region in Hungarian, Swedish and Italian control populations in addition to the Finnish individuals and found that the differences between populations were small (data not shown). This finding was consistent with the results by others: populations worldwide share similar LD patterns in the *CTLA4* region (Ramirez-Soriano et al 2005, Brophy et al 2006, Butty et al 2007). As expected from the migration history, LD is less extensive in Afirican populations and blocks of strong LD are longest in European and Asian populations (Butty et al 2007).

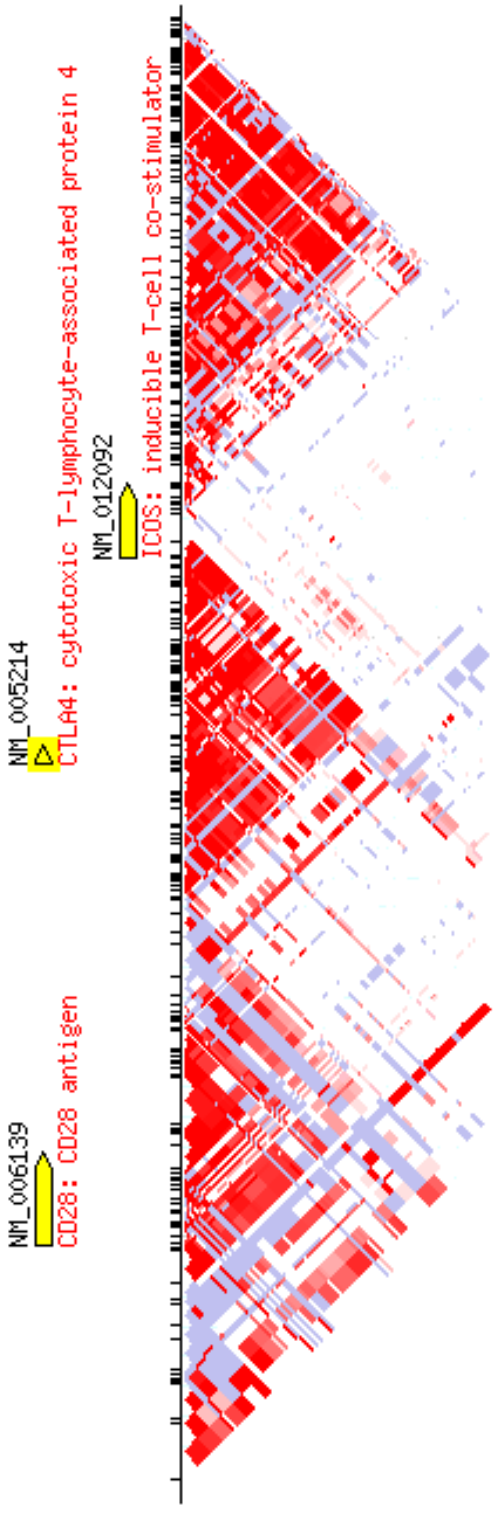


Figure 9. Two recombination hot spots divide the region 2q33 into three blocks of strong linkage disequilibrium. The NM accession numbers refer to mRNAs in the Reference Sequence database of NCBI (http://www.ncbi.nlm.nih.gov/RefSeq/).

7.2.2 Strong linkage disequilibrium covers the PDCD1 gene (II)

The *PDCD1* gene is located further, in 2q37, thus LD does not extend from 2q33 to *PDCD1*. Within the gene, there is strong LD and some of the SNPs are in complete linkage disequilibrium with each other. Only a few tagging SNPs are needed for genotyping in association studies (Prokunina et al 2002, Nielsen et al 2003).

7.2.3 The ICOS haplotypes observed in the Finnish population (I, II)

Due to the strong linkage disequilibrium, conserved haplotypes could be observed. In the study I, we constructed seven putative haplotypes containing eight *ICOS* SNPs using the genotype information from the Finnish unaffected individuals homozygous for most of the markers genotyped. The most common haplotype represented 63% of 126 chromosomes, and obvious fragments derived from this haplotype were detected in up to 26% of the remaining haplotypes. We examined haplotypes covering the same markers in Finnish families in the study II. Four out of seven putative haplotypes could be confirmed in family data (Table 3), besides them 18 other, rare haplotypes were observed. The putative haplotypes were built by limiting the number of different haplotypes as small as possible, but further studies revealed the LD pattern and haplotype diversity to be more complex.

Table 3. The haplotypes constructed in study I and confirmed in a family data in study II. The alleles associated with coeliac disease marked in bold (study II) and underlined (study III).

IVS1+173T/C rs10932029	c.602A/C rs10183087	c.930G/A	c.1459T/C	c.1564C/T rs4404254	c.1624C/T rs10932037	c.1862A/G	c.2373G/C rs4675379	Number	Frequency in study I	Frequency in study II	Transmitted	Non transmitted
T	A	G	A	T	C	A	G	223	0.63	0.70	134	89
C	A	G	A	T	C	A	G	20	0.14	0.06	7	13
T	**C**	A	T	C	**T**	G	**G**	19	0.10	0.06	**17**	**2**
T	C	G	A	C	C	A	C	9		0.03	6	3

7.2.4 Identification of haplotype-tagging SNPs (II, III)

The presence of strong linkage disequilibrium allows the definition of haplotype-tagging SNPs (htSNPs). They represent given haplotypes in a LD block, so with a small subset of the SNPs it is possible to capture the full haplotype information. However, the LD structure and haplotype information of the gene region of interest must be known before representative htSNPs can be selected. International HapMap project collects information on the LD structure of different populations enabled researchers to select an optimal subset of SNPs for genotyping in association studies (International HapMap consortium 2003). HapMap data of the 2q33 region were not available when we started this study so we have systematically genotyped markers, measured LD and sieved informative tagging SNPs for further studies. The best subset of htSNPs can only be defined using empirical data from particular study population, because haplotype and allele frequencies differ, although the LD structure is similar between populations (Ramirez-Soriano et al 2005, Butty et al 2007).

Johnson et al examined the LD pattern in *CTLA4* by genotyping nine SNPs in a set of UK controls (Johnson et al 2001). They defined five haplotype-tagging SNPs to detect 87% of haplotypes among Europeans (population values between 70 and 96%). Ramirez-Soriano et al identified three common htSNPs for ten different populations of European origin by genotyping 17 SNPs in *CTLA4* (Ramirez-Soriano et al 2005). This set of three htSNPs includes two tagging SNPs (rs4553808 and rs11571317) originally identified by Johnson et al and this set detects 71% of the European haplotypes. Both Ramirez-Soriano et al and Johnson et al have focused only on *CTLA4* but Butty et al also examined eight SNPs in *ICOS* in addition to seven SNPs in *CTLA4* (Butty et al 2007). They genotyped control sets of eight different European populations and constructed extended haplotypes. They did not choose

htSNPs but frequent SNPs located in the middle of the gene. We have selected htSNPs mainly in the Finnish population and aimed to detect haplotypes covering the whole *CTLA4-ICOS* region. We have chosen quite a large number of htSNPs in order to observe nearly 100% of the European haplotypes. HtSNPs of Ramirez-Soriano et al are all located at the 5' end of *CTLA4* but we have been interested in detecting haplotypes extending up to the beginning of *ICOS*.

Because of the high-throughput technologies available nowadays, careful designing of markers to be typed aiming to limit the number of SNPs may seem worthless. But, if a reduction in SNP number was achieved for each of the hundreds or thousands of genes to be typed, savings would be remarkable and would allow increasing the number of genes to be typed. In addition, use of random markers instead of carefully characterised htSNPs may lead to a false negative result; if for example, all genotyped SNPs represent only one LD block and the functional SNP is located in another block. Typing common markers frequently used in previous association studies may lead to the same situation because they may not be able to give information of a longer haplotype harbouring associated variant.

7.2.5 Extended haplotypes in European populations (II, III)

The discovery of strong LD across the 2q33 region has led to exploring of haplotypes. Ramirez-Soriano et al identified 6 to 25 different *CTLA4* haplotypes in ten European populations; 80% of haplotypes are common (frequency >5%) (Ramirez-Soriano et al 2005). Butty et al found in an overlapping set of European populations similar haplotype diversity and frequency when examining *CD28* and *ICOS* in addition to *CTLA4* (Butty et al 2007). Although the SNP frequency differences among populations are not large, haplotype diversity in *CTLA4* and especially in the *ICOS* promoter region is considerably different between continental population groups (the highest in

Africa, as expected). In almost all populations, one or two dominant haplotypes account for approximately half of the chromosomes (Butty et al 2007).

Brophy et al identified an extended haplotype covering *CD28*, *CTLA4*, and the *ICOS* promoter to be protective when studying susceptibility to coeliac disease (Brophy et al 2006). We have found similar extended haplotypes reaching from *CD28* to *ICOS* in the study II. Also, Butty et al reported long *CD28-CTLA4-ICOS* haplotypes to be frequent in European and Asian populations. Butty et al speculate that extended haplotypes may suggest selective events. The extended haplotypes dominate in the postmigration populations, which would have confronted very varied environmental challenges compared with the original environment in Africa and adaptation to these challenges would have affected the variation at the costimulatory receptor region (Butty et al 2007).

7.3 Association studies

7.3.1 Coeliac disease (II and III)

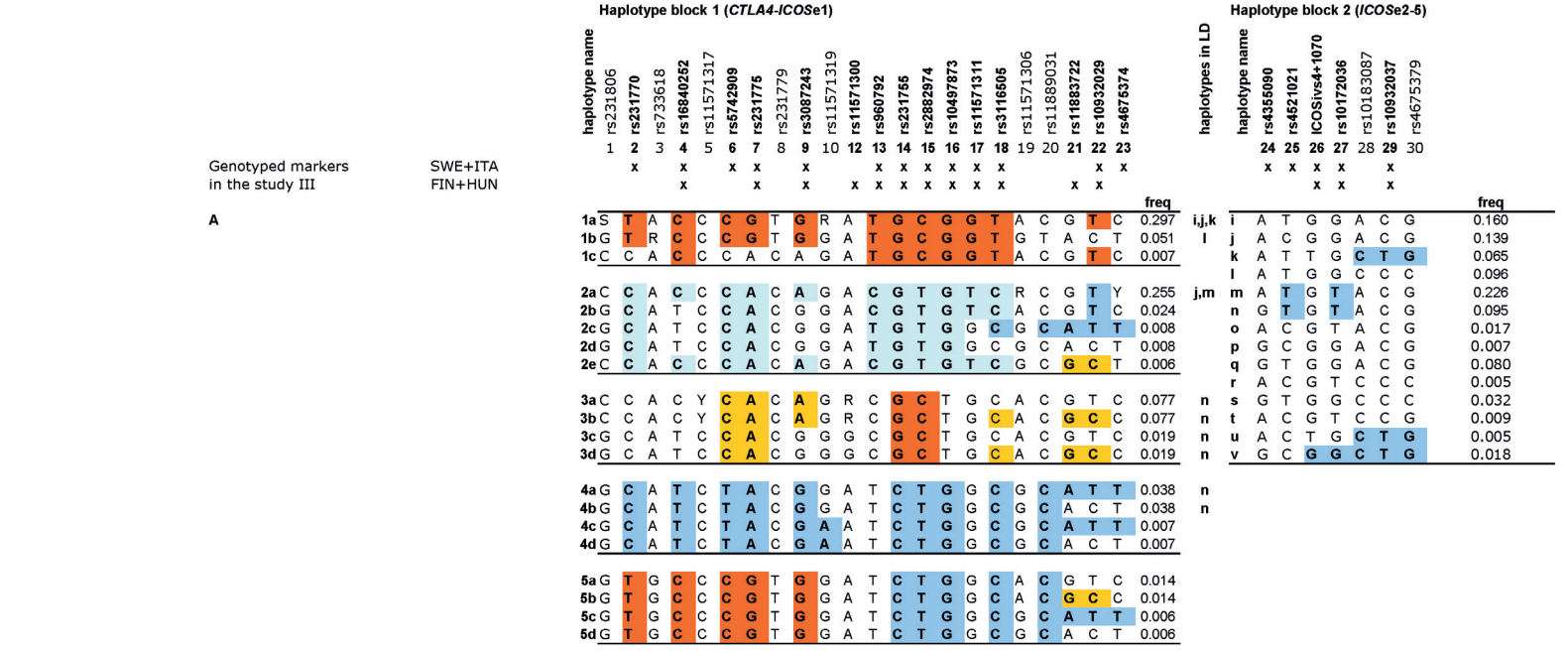
To analyse the possible association between coeliac disease and genetic variation in the 2q33 region, we genotyped 106 Finnish CD trio families (including parents and an affected offspring) for 25 markers covering all four co-stimulatory receptor coding genes (Figure 8). Transmission disequilibrium test revealed statistically significant results for four SNPs (rs10932029, rs10183087, rs10932037 and rs4675379) (Table 4) in the *ICOS* gene (P= 0.005, RR 1.28, 95% CI -1.32-4.37; P= 0.036, RR 1.29, 95% CI 1.25-1.79: P= 0.034, RR 1.88, 95% CI 1.84-3.1; P= 0.028, RR 1.25, 95% CI -1.61-4.54, respectively). The most significant finding was for rs10932029: the allele T was transmitted 35 times to an affected child versus 15 non-transmissions (P= 0.005). The haplotype including all the four *ICOS* alleles associated with CD, showed a transmitted:not transmitted ratio of 17:2 (P= 0.0006 by TDT, RR 1.14, 95% CI 0.54–2.0). This implied that the individual associated alleles formed a haplotype, which showed transmission preference to affected children.

To further examine the association found in Finnish CD families, we genotyped a denser SNP panel covering the *CTLA4–ICOS* region in Italian case-control, Hungarian family and case-control, as well as new Finnish family and case-control study materials. The Finnish family material was partially overlapping in these two studies. We found association between six markers in the *CTLA4-ICOS* region with coeliac disease in the Italian population (P= 0.0009-0.0209; Table 4). No association was observed in the Hungarian CD case-control material. TDT analysis of 278 Finnish and 385 Hungarian families combined showed association with the *ICOS* marker rs10932029*T in CD (P= 0.048, OR 1.41, 95% CI 0.93-2.14, Table 4).

Even stronger association was found with haplotypes compared with the single marker associations in all study populations except the Hungarian case-control material (Table 4). The G-T-G haplotype of the markers rs231755, rs2882974 and rs10497873 was observed (P= 0.0006, OR 1.81, 95% CI 1.28-2.54) in the Italian case-control set. The haplotype G-G-T of the adjacent markers ICOSivs4+1070, rs10172036 and rs10932037 in the *ICOS* gene was associated in the Finnish case-control material (P= 0.0009, OR 4.65, 95% CI 1.74-12.43).

Interestingly, we found association with *ICOS* rather than with *CTLA4* in the Finnish CD material in the study II and despite genotyping additional intergenic tagging SNPs, the strongest association was still found with SNPs in the 3’ end of *ICOS* in the Finnish case-control material in the study III. The marker rs10932037 in the end of *ICOS* was included in the original as well as the further study showing significant association with CD in both studies (P= 0.034, RR 1.88, 95%CI 1.84-3.1 and P= 0.0048, OR 1.98, 95%CI 1.22-3.19, respectively). In the study II, the most significant finding (P= 0.005) was for the marker rs10932029 located in the first intron of *ICOS*. The same marker showed association (P= 0.048) in Finnish and Hungarian CD families in the study III. Interestingly, the *ICOS* promoter haplotype containing the marker

Table 4. Systematic haplotype comparison of the genetic association and gene expression results of the current study with earlier publications. (A) Haplotypes over 0.5% frequency in the two haploblocks *CTLA4-ICOS*e1 and *ICOS*e2-5 were constructed from the genotypes of HapMap CEU families and the data for the markers used in the Study III (marked with x) and those of others. The markers showing association to IgAD, CVID or CD in our materials are marked in bold. The haplotypes of the two blocks with the strongest LD between each other are also indicated. The allele ambiguity code S indicates C/G, R G/A and Y indicates C/T. (B) Significant associations found in the current study for single markers and haplotypes are indicated with the most significant P values and classified into the most likely risk and protective haplotype groups. (C) Association reported in previous CD studies. Marker alleles associated with risk in our study and those of others are marked in two shades of blue, protective alleles are marked in two shades of orange.



B	Risk	Prot																						
SWE IgAD	haplotypes	haplotypes																					lowest P value (*)	OR(95%CI)
single markers (all)	4	1	C	T		A				C	T*			C				T					0.002	1.50 (1.17-1.93)
single markers (HLA+)	2		C			A	A		C		T		T	C*				T		T			0.002	2.09 (1.31-3-33)
single markers (HLA-)	4	1,3			T					C	T*												0.034	1.41 (1.03-1.93)
haplotype	2,4,5										T	G											0.0015	1.50 (1.17-1.93)
haplotype										G	C												0.0015	0.67 (0.52-0.86)
haplotype	4,5									C	T	G											0.042	1.40 (1.01-1.95)
haplotype			T	C	C	G	G		T	G	C	G	G	T		T							0.005	0.69 (0.53-0.89)
SWE CVID																								
single markers				T																			0.0315	1.63 (1.04-2.55)
haplotype	2	1		T	C	A	G		T	G													0.0064	2.66 (1.28-5.53)
ITA CD																								
single markers	2	1,3	C			A			C		T		T*	C									0.0009	1.85 (1.29-2.67)
haplotype	2									G	T	G											0.0006	1.81 (1.28-2.54)
haplotype										G	C	G											0.005	0.59 (0.42-0.86)
haplotype	2		C	C	C	A	A		C	G	T	G	T	C		T							0.020	1.57 (1.07-2.30)
haplotype			T	C	C	G	G		T	G	C	G	G	T		T							0.008	0.61 (0.43-0.88)
FIN CD cases																								
single markers	k,u,v																				T		0.005	1.98 (1.22-3.19)
haplotype	v	3,5																	G	G		T		0.0009	4.65 (1.74-12.43)
FIN+HUN CD families																								
single markers	1-5	3,5														T							0.048	1.41 (0.93-2.14)
haplotype	2,4,5													C	A	T	T						0.0007	2.30 (1.47-3.61)
haplotype		3,5													G	C							0.022	0.58 0.38-0.88)
haplotype		2,3,5												C	G	C	C						0.035	0.60 (0.39-0.93)
haplotype	2,4,5	3,5												C	A	T							0.033	1.61 (1.07-2.40)
haplotype														C	G	C							0.026	0.60 (0.39-0.91)
haplotype	2,4,5												G	C	A	T							0.040	1.61 (1.06-2.46)
haplotype															G	C	C						0.029	0.58 (0.37-0.89)
All cases combined																								
single markers	2,k,u,v										T*		T	C									0.0005	1.27 (1.11-1.46)

C																							
single markers	2,3,4					A																	Djilali-Saiah et al 1998
single markers	2,3,4					A																	Naluai et al 2000
single markers	2,3,4					A																	Mora et al 2003
single markers+haplotypes	k,u,v															T					C	T	G	Haimila et al 2004
haplotype	4			T		A	G	A																Amundsen et al 2004
single markers	1-5	1,2,3					G																	van Belzen et al 2004
CD risk (HLA high risk)	1-5						G																	Zhernakova et al 2005
CD risk (early onset)	4					A	G																	Zhernakova et al 2005
haplotype	1,5				C	G	G																	Hunt et al 2005
haplotype					C	A	A																	Brophy et al 2006
single markers	2,3,4					A																		Capilla et al 2007
single markers	1,2,4,5																T							van Heel et al 2007

Table 5. The genome wide linkage studies and one genome wide association study (van Heel et al) in coeliac disease. Abbreviations: fam=families, asp=affected sib pair, ctrl=control, Stat=statistics, Map=physical map position of the best marker, MMLS=maximum multipoint lod score, Mb=mega bases, LOD=logarithm of odds, NPL=non-parametric linkage, HLOD= heterogeneity LOD score, Zlr= linkage score of the Z_{lr} statistics. The table is modified from Saavalainen 2007.

Author	Population	Samples	Stat/Map	1p	1q	2p	2q	3p	3q	4p	4q	5p	5q	6p	6p(HLA)	6p	6q	7p	7q	8p	8q	9p	9q	10p	10q
Zhong et al	Irish	15 fam:	MMLS						1.0				0.9	4.7	4.4	1.4			3.0						
1996		45 asp	Mb						188				173	15	33	51			121						
*Greco et al	Italian	39+71 asp	MLS										1.8	1.0	3.5										
1998			Mb										180	14	36										
*King et al	UK	16 fam:	LOD	1.7	1.1		1.4			1.3	1.6						1.7	0.7			1.4		0.8		1.6
2000		47 aff	Mb	7	197		183			7	144						80	18			132		92		111
*Naluai et al	Swedish&	70+36	NPL				2.0	1.8					1.8		4.4							1.8			
2001	Norwegian	fam(asp)	Mb				107	13					135									18			
*Liu et al	Finnish	60 fam:	LOD	1.9						2.1			1.7		6.0				1.0			1.1			
2002		86 asp	Mb	18						18			135						96			25			
Woolley et al	Finnish	9 fam:	LOD			1.2				1.2			1.2							1.9					
2002	isolate	23 aff	Mb			3				13			180							7					
Popat et al	Northern	24 fam	NPL							1.9					2.3										
2002	European		Mb							40															
Neuhausen et al	US	62 fam	HLOD					3.2	1.6			1.9			2.2						1.5			1.7	
2002			Mb					3	119			22			37						74			19	
Van Belzen et al	Dutch	82 fam:	MMLS		1.7								1.1		8.1		2.7				2.3				
2003		101 asp	Mb		83								174				108				118				
Van Belzen et al	Dutch	1 large	LOD/*NPL*						1.1						2.3		*4.6*					2.6			*4.2*/1.3
2004		pedigree	Mb						104						31		159					35/37			57/130
Babron et al	meta-analysis		Zlr				2.0				2.2		4.4		9.6			1.7			2.0				2.2
2003	of (*)		Mb				195				124		150		25			46			107				128
Rioux et al	Finnish	54 fam	NPL				2.8								3.0									1.9	
2004			Mb				169																	17	
Garner et al	US	160 fam	NPL		2.0				2.0						5.5				2.6				1.3		2.1
2006			Mb		192				145						25				141				135		74
van Heel et al	UK	778/1422	$\log_{10}$(p)	-5	-5	-5	-5	-5	-7	-5	-7		-5		-19	-5	-5				-6	-6		-6	
2007		case/ctrl	Mb	57/	165	23/	102/136/	2	180/	32	123		151			53	93/				103/	1/		6	
				62		30	193/240		190								129				134	28			

Author	Population	Samples	Stat/Map	11p	11q	12p	12q	13q	14q	15q	16p	16q	17q	18p	18q	19p	19q	20p	20q	21q	22q	Xp	Xq
Zhong et al 1996	Irish	15 fam: 45 asp	MMLS Mb	3.9 39						2.1 94						0.8 3	1.8 60	2.7 16					
*Greco et al 1998	Italian	39+71 asp	MLS Mb		1.8 126																		
*King et al 2000	UK	16 fam: 47 aff	LOD Mb	1.6 36	1.2 120		1.2 23		1.1 26	0.9 64		2.6 82	1.2 28	1.0 6	1.1 76		1.1 62				1.5 44		
*Naluai et al 2001	Swedish& Norwegian	70+36 fam(asp)	NPL Mb	1.8 11	2.4 123								1.8 66									1.9 43	
*Liu et al 2002	Finnish	60 fam: 86 asp	LOD Mb									1.4 53											
Woolley et al 2002	Finnish isolate	9 fam: 23 aff	LOD Mb	1.2 36						3.1 26													1.2 103
Popat et al 2002	Northern European	24 fam	NPL Mb													2.1 3							
Neuhausen et al 2002	US	62 fam	HLOD Mb			1.4 13	1.6 66	1.4 30							1.5 76								
Van Belzen et al 2003	Dutch	82 fam: 101 asp	MMLS Mb													4.8 21		1.5 7					
Van Belzen et al 2004	Dutch	1 large pedigree	LOD/*NPL* Mb	*5.5* 5						1.36 64		1.2 80				1.2 17							
Babron et al 2003	meta-analysis of (*)		Zlr Mb		2.5 120		1.9 127		2.5 92	1.9 26			2.1 75									2.0 43	
Rioux et al 2004	Finnish	54 fam	NPL Mb																				
Garner et al 2006	US	160 fam	NPL Mb																				
van Heel et al 2007	UK	778/1422 case/ctrl	$\log_{10}$(p) Mb	-5 4/ 45	-5 111			-5 74	-6 81	-5 33/ 80	-5 13				-6 54/ 74		-5 35	-5 6/ 16					

rs10932029 was more significantly associated than the marker solely in the study II (P= 0.0006) as well as in the study III (P= 0.0007) as seen in Table 4.

Several genome wide linkage studies have been performed in coeliac disease, but this far only one genome wide association study has been published (Table 5). The *HLA* region (CELIAC 1) is the most significant finding in almost all of them. The 2q (CELIAC 3) region has been reported to be linked to CD in three independent studies although the results are not statistically very significant. Other genomic regions observed to be linked in a few populations are 5q (CELIAC 2), 11p and 19p. All these linked regions contain functionally potential candidate genes, but no strong evidence has been presented for any of them. A recent large meta-analysis of genome wide linkage studies covered 11 autoimmune diseases (including CD) and over 18 000 patients (Forabosco et al 2009). The study aimed to identify common susceptibility genes for different autoimmune diseases. Suggestive evidence of linkage (P=0.005) was detected only on chromosome 16p, in addition to highly significant linkage (P< 0.0001) on the *HLA* region. No evidence for linkage was observed at *CTLA4*, but the authors calculated that power to detect common low-risk genes in their analysis was only 1% for genome wide levels of significance (Forabosco et al 2009). Two follow-up studies have been performed after the first genome wide association study (Hunt et al 2008, Romanos et al 2008) and association of six loci (in chromosomes 1,3,4,6 and 12) replicated. Non-*HLA* susceptibility genes in coeliac disease are still waiting for confirmation.

7.3.2 IgA deficiency (III)

We genotyped 231 Swedish IgAD individuals as well as 267 matched controls for 18 markers in the *CTLA4* and *ICOS* region and for the known autoimmunity risk haplotype *HLA-DRB1**03–*DQB1**02 aiming to identify a possible disease-associated polymorphism. We found association of IgAD with 7 SNPs, of which the *CTLA4* marker rs2882974 demonstrated the highest significance (P= 0.0015, OR

1.50, 95% CI 1.17-1.93). The reanalysis after stratifying all of the Swedish IgAD patients and controls on the basis of *HLA-DR*3-*DQ*2 positivity, revealed the *CTLA4* association to be even stronger (P= 0.0014, OR 2.10, 95% CI 1.33-3.32) in patients carrying the *HLA* risk haplotype (Table 4). The G-C-G haplotype of rs231755-rs2882974-rs10497873 was significantly associated with protection from IgAD (P= 0.005, OR 0.69, 95% CI 0.53-0.90), and the T-G haplotype of rs2882974-rs10497873 was significantly associated with IgAD risk (P= 0.0015, OR 1.50 95% CI 1.17-1.93). No association was detected in a small set of Finnish IgAD individuals (n=44), which may be attributed to the limited power in the genetic analyses.

HLA-stratified two-locus analysis provided evidence of epistasis, that is, genetic interaction between the *CTLA4-ICOS* and *HLA* loci. Epistasis indicates that proteins encoded by associated genes probably act in the same or related pathways, which is natural in the case of *CTLA4-ICOS* and *HLA* because of their co-operative roles in regulation of T cell activation. Further studies, especially functional, are still needed to reveal the detailed mechanisms for this interaction related to the susceptibility of IgAD.

This was the first study of association between co-stimulator receptor genes and IgAD although *ICOS* is very interesting candidate gene for IgAD because of its role in class switching. Furthermore, the ICOS molecule is a superinducer of IL10 secretion (Hutloff et al 1999), which is interesting because plasma level of IL10 is high in patients with both CD and IgAD (decreasing after a gluten free diet) (Cataldo et al 2003) and in addition, IL10 is involved in the regulation of IgA secretion (Defrance et al 1992, Kitani et al 1994). Our study was also the first to report shared non-*HLA* risk variants between IgAD and CD.

7.3.3 Common variable immunodeficiency (III)

A rare *ICOS* deletion leading to absence of ICOS expression has been reported in CVID patients in a few families (Grimbacher et al 2003, Salzer et al 2004).

However, common variations in the *ICOS* gene have not been analysed in CVID, and therefore we genotyped 87 Swedish CVID patients for 18 markers in *CTLA4-ICOS*. We observed weak association (P= 0.0315, OR 1.63, 95% CI 1.04-2.55) between the T allele of the CTLA4 marker rs16840252 and CVID. The T-C-A-G-T-G haplotype of rs16840252-rs5742909-rs231775-rs3087243-rs960792-rs231755 was significantly associated with the risk of CVID (P= 0.0064, OR 2.66, 95% CI 1.28-5.53) as seen in Table 4.

7.3.4 Meta-analysis (III)

To further explore the association, we combined Finnish, Italian and Hungarian CD as well as Finnish IgAD, Swedish IgAD and CVID materials with a large European case-control material including 942 patients and 1126 controls. Association of the intergenic markers rs2882974, rs11571311 and rs3116505 (P= 0.0005, OR 1.27, 95% CI 1.11-1.46; P= 0.0141, OR 1.22 95% CI 1.04-1.43; and P= 0.0069, OR 0.82, 95% CI 0.71-0.95, respectively) was found with IgAD, CVID and CD (Figure 10).

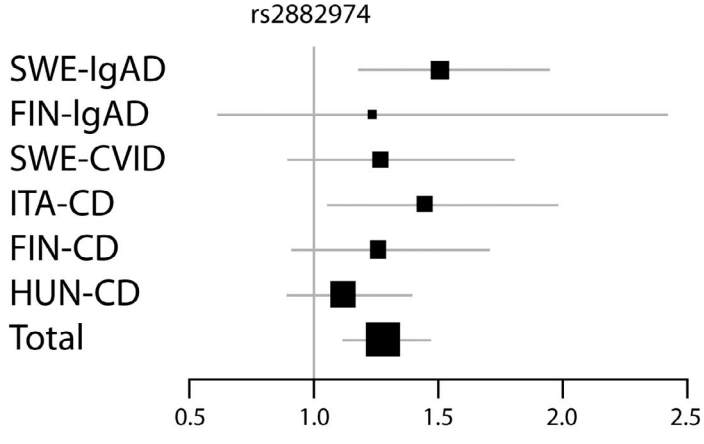


Figure 10. Meta-analysis of association in the combined case-control material. The size of the squares indicates the size of each study population, grey lines refer to 95% CI of odds-ratios, X axis refers to OR. The most significant association (P= 0.0005) was found with the intergenic marker rs2882974.

7.3.5 Discrepant association results for the 2q33 region in CD

Results of genetic association studies of the *CTLA4* region in coeliac disease have been somewhat contradictory (Table 6). Pioneer studies suffered from limited sample sizes and low number of polymorphisms leading to insufficient statistical power and ignorance of linkage disequilibrium, but during recent years such shortcomings should have been avoided.

If a poor study design is not taken into consideration, still several reasons for contradictory results can be suggested. For example, gene regulatory or environmental factors can modify the expression of a gene and lead to different levels of association in different populations. The degree of linkage disequilibrium varies between marker and disease alleles; also, significant differences in allele frequencies and haplotype composition exist between geographical groups, which points out the need of a very well matched control population. The modest effect of susceptibility genes in complex diseases complicates the finding of true risk variants and explains contradictory results.

Because the published association results of *CTLA4-ICOS* region in coeliac disease seem to be contradictory we compared systematically haplotypes from our studies with those reported earlier. We constructed haplotypes from the genotypes of HapMap CEU families (build 36, http://www.hapmap.org/) and from the data of our studies as well as of earlier publications. We picked common haplotypes (frequency >0.5%) of the two haploblocks: first one extending from the *CTLA4* gene until the first exon of *ICOS* (*CTLA4-ICOS*e1) and the second one containing the exons from 2 to 5 of *ICOS* (*ICOS*e2-5) (Table 4A). Haplotypes are divided into groups of risk and protective haplotypes on the basis of published association results (Table 4B and C).

The *CTLA4-ICOS*e1 risk haplotype groups 2 and 4 both carry the allele most commonly associated (also the most studied marker in the gene region) with coeliac disease,

rs231775*A (CTLA4+49*A). However, this allele is also carried in the protective *CTLA4-ICOS*e1 haplotype group 3, which indicates that it is not the primary functional disease variant and perhaps explains the failures of several studies to replicate this association. A risk haplotype C-G-G of *CTLA4* markers (rs5742909, rs231775 and rs3087243) was reported in the UK population (Hunt et al 2005). Discordantly, this haplotype is frequent in our protective *CTLA4-ICOS*e1 haplotype group 1 but it is also found in *CTLA4-ICOS*e1 haplotype group 5 which further shares risk alleles with *CTLA4-ICOS*e1 haplotype group 4. The most common *CTLA4-ICOS*e1 haplotype (frequency 0.30) in Europe (referred as 1a in Table 4A), carries this shorter C-G-G haplotype, but it is also in LD with *ICOS*e2-5 haplotype (k in Table 4A) carrying the alleles associated with CD risk in Finland. In a large genome wide association study in the UK population, the only *CTLA4-ICOS* variant showing association with CD was rs4675374*T, located in the first intron of *ICOS* (Van Heel et al 2007) and carried in our risk *CTLA4-ICOS*e1 haplotype groups 2 and 4. In addition, a recent study from Ireland reported the protective effect of a C-A-A haplotype of *CTLA4* in coeliac disease (Brophy et al 2006); this shorter haplotype occurs frequently in our risk *CTLA4-ICOS*e1 haplotype group 2 but also exists in the protective groups 1 and 3. Hence, our results perhaps offer an explanation to the apparently conflicting results of previous publications (Table 4C).

Often different disease-causing alleles are suggested to predominate in different populations. Instead of the earlier explanation of allelic heterogeneity, our findings suggest that the risk variants can be found on several haplotype backgrounds. The true risk variants for coeliac disease in the 2q33 region are possibly in LD with the associated alleles previously published. On account of our results, denser genotyping of whole 2q33 region can be recommended. Also, an analysis based on haplotypes could be advantageous. Analysis of haplotypes instead of single SNPs is relevant because genes produce proteins

Table 6. The published genetic association studies on the 2q33 region in coeliac disease.

Number of markers										
centromeric	CD28	intragenic region	CTLA4	intragenic region	ICOS	telomeric	associated marker	lowest P value	size of material	country	author
			1				CTLA4+49 A	<0.0001	101 cases, 130 controls	France	Djilali-Saiah et al 1998
3			2			2	D2S116	0,02	100 families	Finland	Holopainen et al 1999
			1					ns	132 families	Italy/Tunisian	Clot et al 1999
4			1			1	CTLA4+49 A	0.007	107 families	Norway/Sweden	Naluai et al 2000
4			3	1			D2S2214	0.007	166 families	UK	King et al 2002
			1				CTLA4+49 A	0.02	62 trio families	Sweden	Popat et al 2002
			2					ns	41 families	Spain/Basque	Martin-Pagola et al 2003
			1				CTLA4+49 A	0.03	86 cases, 144 controls	Italy	Mora et al 2003
			1				CTLA4+49 A	0.03	113 trio families	Italy	Mora et al 2003
				2				ns	149 trio families	UK	King et al 2003
				2				ns	149 cases, 100 controls	UK	King et al 2003
7		1	4	6	2	2	CTLA4-1147 T	0.02	100 multiplex families	Norway/Sweden	Amundsen et al 2004
		1	2	5				ns	225 simplex families	Norway/Sweden	Amundsen et al 2004
3	1		3	1		1	D2S2214	<0.01	796 families	Europe	Holopainen et al 2004
			1					ns	54 families	Finland	Rioux et al 2004
			1	1			CT60 G	0.048	215 cases, 215 controls	Netherlands	van Belzen et al 2004
			5	1	1		CTLA4 haplotype	0.0007	340 cases, 973 controls	UK	Hunt et al 2005
		1	1	2				ns	310 cases, 900 controls	Netherlands	Zhernakova et al 2005
				1				ns	90 families	Spain	Rueda et al 2005
	3		5	1	7		CTLA4-658 C	0.007	394 cases, 421 controls	Ireland	Brophy et al 2006
			2	2	1		CTLA4+49 A	0.01	168 cases, 186 controls	Spain	Capilla et al 2007
			1	1				ns	120 cases, 231 controls	Slovakia	Dallos et al 2008

corresponding to haplotypes and variation in a population is transmitted as haplotypes (Clark 2004).

7.3.6 Kidney transplantation patients (IV)

A total of 678 renal patients were genotyped for 13 markers covering the whole *CD28-CTLA4-ICOS* region as well as the *PDCD1* gene to study the possible association between these genetic markers and the outcome of kidney transplantation. We found statistically significant association between the *ICOS* marker rs10932037 and graft survival (P= 0.026) in addition to association with delayed or non functioning of the graft with the *ICOS* markers rs10183087 and rs4404254 (P= 0.020, OR= 5.8 and P= 0.019, OR= 5.8).

Previously, three reports of genetic association between *CTLA4* and acute rejection on kidney transplantation patients have been published (Slavcheva 2001; Gendzekhadze 2006; Wisniewski 2006). We did not detect association with acute rejection, perhaps due to a low number of acute rejection episodes (16.8%), although the total number of patients was relatively high. A Canadian study on acute rejection, graft and patient survival related to kidney transplantation did not succeed in replicating the association with *CTLA4* (Dmitrienko 2005). This study as well as those three with positive association findings (Slavcheva 2001; Gendzekhadze 2006; Wisniewski 2006) suffered from a lack of power because of small sample sizes (from 63 to 167 patients). Furthermore, for these previous studies the rejection patients were selected for case-control type of comparisons whereas we examined all the samples available from the patients transplanted between 1999 and 2003.

These four previous studies of transplantation suffered from the same faults that autoimmune disease studies commonly do: the association analyses were limited solely to markers in the *CTLA4* gene, leaving out *ICOS* and other neighbouring genes. Besides, all the associations reported are with different *CTLA4* markers (rs231775, (AT)n, rs5742909). We examined markers across the

entire gene cluster and our results suggest that *ICOS*, rather than *CTLA4*, is the genetic factor affecting the renal transplantation outcome.

Most genetic association studies in kidney or more generally organ transplantation are focussed on cytokine genes. They are potential candidate genes on the basis of their function, and variations in these genes may cause differences in their production, binding or signalling, which in turn may affect the immune responses of transplantation patients. Although at least a dozen of cytokine association studies have been published, the results are controversial and the effect of genetic variation on the outcome of transplantation still bargain for confirmation (Kruger et al 2008). The *TNF*-α gene is one the most frequently studied cytokine genes and variation in the gene has been shown to have functional effects (Wilson et al 1997). However, association studies suffer from similar shortcomings as those related to the *CTLA4* region in autoimmune diseases. Over ten polymorphisms have been identified in the *TNF*-α gene, but only one or two of them are included in the association studies. Furthermore, the gene is located in the region of strong linkage disequilibrium in the middle of several polymorphic genes which encode molecules functioning in the immune response, e.g. *HLA* genes. Besides cytokine genes, also toll-like receptor and complement genes are under study as well as genes related to drug metabolism. The results are not indisputable, so further studies are required (Kruger et al 2008).

7.4 Analysis of linkage in coeliac disease (III)

In order to detect linkage between the 2q33 region and coeliac disease we analysed 13 markers in 133 Finnish and 141 Hungarian CD families. Two earlier studies have revealed significant linkage of 2q33 to coeliac disease in Finnish families (Holopainen et al 1999, Rioux et al 2003). We detected linkage between the *CTLA4-ICOS* gene region and coeliac disease in Finnish families (LOD 2.21, P= 0.0007) but not in Hungarian families (LOD 0.52, P= 0.06). When Finnish and Hungarian family materials were

combined, stronger linkage (LOD 2.38, P= 0.0005) was observed. The set of Finnish families partially overlap with that used in the study by Holopainen et al 1999 but the overlapping part is only small so it does not wholly explain the positive result. Furthermore, our aim was not to replicate the original linkage result but to fine-map the association.

The positive linkage and association results of Finnish material support the role of the *CTLA4-ICOS* region in the susceptibility in CD, although linkage is not very strong. Association studies are better than linkage analysis for genes with a low relative risk because association analysis has a greater power to detect gene effects. Linkage analysis requires high relative risk and at least intermediate allele frequencies, which are not present in 2q33 (Risch 2000).

7.5 Statistical significance of the results

Association studies very often suffer from limited statistical power leading to inability to draw conclusions to any direction. If association has not been detected it does not mean that association does not exist; the study population may have been too small or optimal markers have not been explored. We found statistically significant associations for *CTLA4-ICOS* with coeliac disease, IgA deficiency and common variable immunodeficiency in four different European populations but the P values observed were not indisputable.

The statistical power is greatest when alleles are common, because such alleles offer the widest genotype variability. We chose markers with as great minor allele frequencies as possible for genotyping. The size of the study population affects mostly the power. The number of patients needed for an analysis capable of detecting minor susceptibility loci with low relative risk is very large. For example, we analysed 106 coeliac trio families in the study II and we assumed relative risk to be 2 and the risk allele frequency 0.3, so TDT had 40% power to detect a risk allele. After that we have collected hundreds of new CD cases and families. In the study III, when we analysed the 663 Finnish and

Hungarian CD families, the TDT had 48% power to detect a risk allele, if the risk allele frequency is estimated to be 0.14 and relative risk is 1.28 (values based on our analyses with rs10932029 in the study II). In the study IV of kidney transplantation, we had many more controls than cases (approximately 5:1 regarding rejection and 3:1 regarding graft function) which increased the power. On the other hand, if the association would not have been detected with our reasonably sized study population, it may not have had any clinical significance.

When a large number of different polymorphisms are studied concomitantly, a problem of multiple testing appears. The correcting for multiple testing is needed because with each test, the possibility of a false-positive result increases. The correcting can be done for example by computer simulation; a permutation of 1000 shuffles of the data is recommended to estimate a threshold of 0.05 (Churchill & Doerge 1994). Other commonly used option is the Bonferroni correction, which is based on the amount of markers and correlation between them. Our results did not remain statistically significant when their significance was evaluated by correction methods for multiple testing. For example, in the study IV, we tested 13 markers against three end points in eight stratification groups; thus we performed approximately 312 tests and we should divide the threshold of 0.05 with 312 to get a suitable threshold for significant result under multiple testing. On the other hand, our study is not a random genome wide analysis but instead we have chosen probable candidate genes; thus correction is not absolutely required and besides, the Bonferroni correction is too conservative in the presence of linkage disequilibrium (Rice et al 2008).

7.6 Functional effects of the variants in the 2q33 region

The strong linkage disequilibrium covering the 2q33 region complicates the identification of the primary genetic variant to affect the disease susceptibility. Numerous genetic association studies have been performed in complex

diseases with fairly discordant results, so proceeding with association studies does not seem to help reaching the ultimate goal. Thus, functional studies are required to identify disease causing genetic factors and to understand their biological mechanism.

7.6.1 Genetic variants

Genomic DNA variants that alter physiological, biochemical, and immunological mechanisms involved in a disease are by their nature primary risk factors. Functional relevance for certain *CTLA4* and *ICOS* variations has been suggested.

First functional studies of this locus focused on *CTLA4* and mainly on its exon 1 SNP rs231775 (CTLA4+49A/G) and promoter SNP rs5742909 (CTLA4−318C/T) which appeared to have an influence on the expression level of CTLA4 protein (Kouki et al 2000, Ligers et al 2001, Maurer et al 2002, Wang et al 2002, Anjos et al 2002). Functional effects have also been found further downstream. The markers rs231775 as well as the 3' UTR microsatellite (AT) n have been reported to be involved in the regulation of IgE levels. Individuals carrying AA genotype of rs231775 as well as individuals homozygous for the allele of the microsatellite with the fewest repeats had the highest total serum IgE levels (Howard et al 2002).

All the polymorphisms identified in the *ICOS* gene are silent variants; they do not change the amino acid sequence of the ICOS molecule and thereby, it must be variation in gene transcription, splicing or transcript stability that is associated with autoimmune susceptibility. Silent polymorphisms may change the motif of functional DNA binding sites and thus their affinity for the relevant transcription factors. Functional studies have revealed that stimulated peripheral blood mononuclear cells from individuals homozygous for the ICOS-1413 A produce higher levels of IL4, IL5, IL13 and TNF-α compared with cells from ICOS-1413GG homozygotes. Also, the ICOS-693AA and ICOS-1413AA genotypes were associated with increased total serum IgE levels (Shilling et al 2005).

Differences in secretory levels may be explained by the potential differential transcription factor binding. But, these SNPs are not necessarily primarily responsible for the associated phenotypes because they are in LD with other polymorphisms nearby.

A recent Italian study showed that the homozygous carriers of the *ICOS* haplotype +1459A, +1564T and +2373G display decreased ICOS expression and increased IL10 secretion (Castelli et al 2007). Similarly, a study on human CD4+ T cells demonstrated that rs10932037 CC (ICOS c.1624) homozygotes have increased mRNA levels compared with CT and TT individuals and besides, rs10932029 (ICOS IVS1+173) associated with *CTLA4* mRNA expression: homozygotes for the T allele have higher expression of both *CTLA4* splicing isoforms (Kaartinen et al 2007).

In conclusion, the variation in the *ICOS* gene can affect the expression of the molecule and the level of ICOS expression can influence differentiation and level of cytokine production.

7.6.2 Soluble CTLA4

The soluble form of CTLA4 (sCTLA4) is capable of binding B7-1 and B7-2 ligands on antigen presenting cells, potentially blocking down-regulation of T cell activation through the B7 interaction with CTLA4 expressed on the disease-specific CD4+ T cells. Thus, sCTLA4 itself as well as ratio of soluble and full length (flCTLA4) forms of CTLA4 have been increasingly under interest.

The mRNA ratio of sCTLA4 to flCTLA4 is reported to be higher in the homozygous rs3087243*A individuals compared with heterozygous and GG homozygous individuals. It is suggested that the alleles of rs3087243 (CT60) affect the efficiency of the splicing and production of sCTLA4 (Ueda et al 2003). Anjos et al could not replicate this finding (Anjos et al 2005) but Atabani et al showed the effect of rs3087243 on the sCTLA4:flCTLA4 ratio and also found the same marker to be associated with a 30–40%

increase in Treg frequency in the peripheral blood (Atabani et al 2005). We analysed polymorphisms within *CTLA4-ICOS* against mRNA expression of CTLA4 and ICOS. The intragenic marker rs11571300*A was associated with increased levels of both the flCTLA4 (P= 0.009) and sCTLA4 (P= 0.029) isoforms upon stimulation. In addition, rs231755*G (also intragenic) was associated with higher sCTLA4 levels (P= 0.036) as well as rs10932029*T allele. The alleles of rs11571311*T and rs10932037*T (ICOS c.1624) correlated with lower ICOS expression levels in stimulated cells (P= 0.046, Figure 11).

The markers rs10183087 and rs4404254 in the *ICOS* gene were found to predispose to the delayed graft function and rs10932037 to the decreased graft survival in kidney transplantation patients in the study IV. Our results together with the expression studies suggest that rs10932037 CC homozygotes likely have a higher ICOS expression level and better graft survival; and rs4404254 TT homozygotes express ICOS at a lower level and they may suffer from delayed graft function more easily.

The SNPs that associate with CD (in study III) differ between the Italian (rs11571311*T showing the strongest association) and the Finnish (rs10932037*T) case-control materials, as well as being located in different haplotype blocks. Despite this, both variants associated with lower ICOS expression levels in stimulated cells. The rs10932029*T allele, conferring CD risk in Finnish and Hungarian families, appeared to produce higher sCTLA4 levels.

The decreased expression levels of ICOS observed for the variants associated with CD and IgAD risk are highly interesting. A relationship between genotype and immune function is suggested but a role in disease development is not indicated. Nevertheless, further dissection of the biological relevance of the *CTLA4-ICOS* alleles associated with risk of IgAD and coeliac disease is required in order to determine if any of these variants are true disease susceptibility alleles. Our findings support the hypothesis

that there may be several CD risk haplotypes within the *CTLA4-ICOS* region with similar biological effects. Increased expression of the soluble molecule compared to membrane-bound CTLA4 or to ICOS could play a role in pathogenesis of autoimmune diseases. Indeed, elevated levels of sCTLA4 are detected in autoimmune disorders (Pawlak et al 2005).

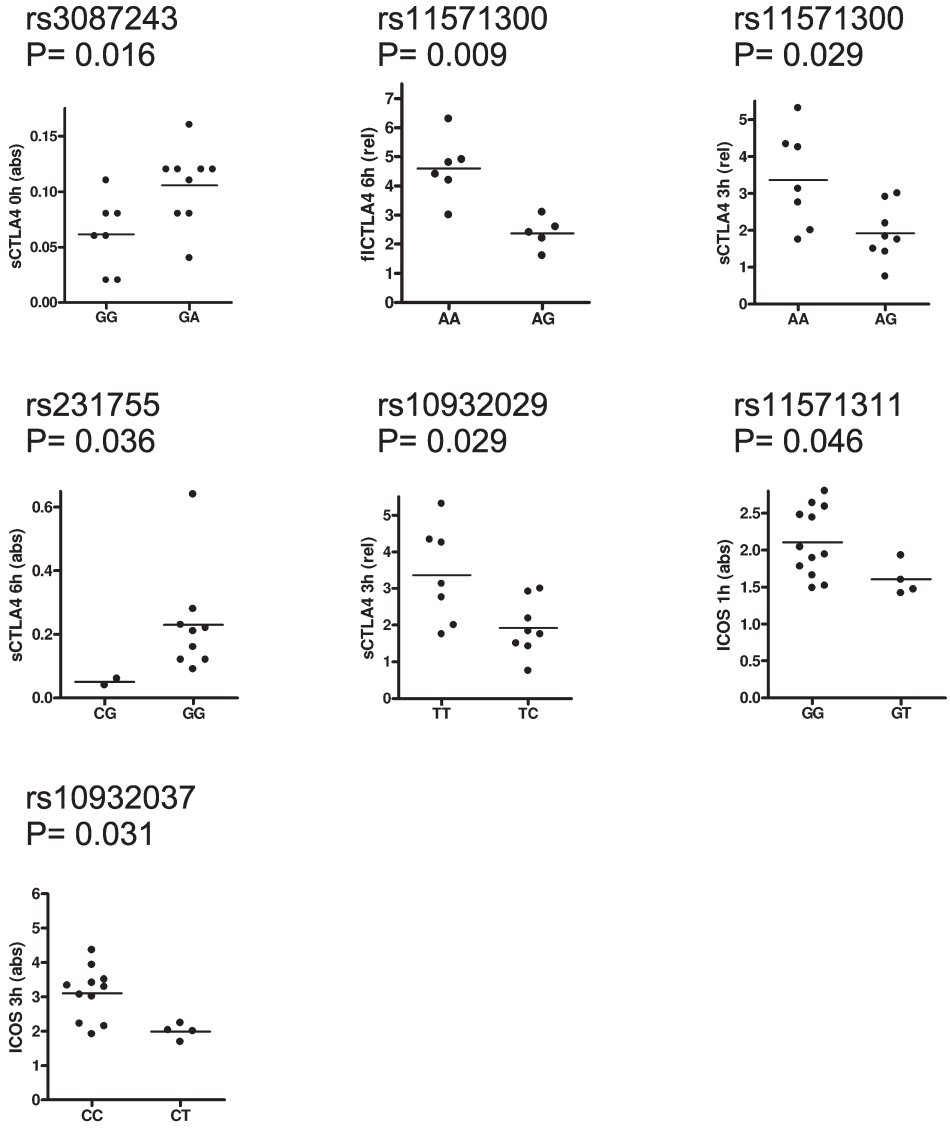


Figure 11. Effect of genotypes to the expression levels of sCTLA4, flCTLA4 and ICOS.

8. CONCLUSIONS

During this decade, the knowledge of the 2q33 region has increased markedly. The research started by basic screening of polymorphisms and exploring the gene structure. The *CD28* and *CTLA4* genes were the first to be identified and the genetic studies have focused on them. *ICOS* and *PDCD1* have been taken into account only during the recent years.

This study has explored the linkage disequilibrium structure on the chromosome region of 2q33 and revealed its haplotype composition. We found strong linkage disequilibrium divided into three blocks. Interestingly, the promoter region and first exon of *ICOS* belong to the same LD block with *CTLA4*; linkage disequilibrium complicates discrimination between effects of genetic variants in these genes.

Most of the earlier association studies in autoimmune diseases have focused their analysis on only a few polymorphisms in the *CTLA4* gene, even though strong linkage disequilibrium is known to cover the region. This study was the first one in coeliac disease to analyse markers systematically across the entire *CD28-CTLA4-ICOS* gene cluster and extending also to the *PDCD1* gene. Whereas there was no evidence for involvement of specific genetic variants in *CD28* or *PDCD1*, we found linkage and/or association of the *CTLA4* and *ICOS* genes with coeliac disease and IgA deficiency. Interestingly, CD and IgAD share the same *HLA* risk haplotype *B*8-*DR*3-*DQ*2 as well as the same association with the *CTLA4-ICOS* region. Several haplotypes in the *CTLA4-ICOS* locus affect the disease risk, thus it is still difficult to conclude whether the risk variant in coeliac disease is in *CTLA4* or *ICOS*, or if rather a combination of variants within both genes leads to association of the *CTLA4-ICOS* region.

Genetic defects at the 2q33 locus appear to be involved in the development of several chronic inflammatory conditions, especially autoimmunity, but the precise risk

genetic factor is yet to be identified. Pathogenesis may be influenced through low expression levels of ICOS and/or comparatively higher levels of membrane-bound CTLA4 or soluble CTLA4.

Contradictory association results may be frustrating, but it still is important to identify non-*HLA* susceptibility genes, because although they have only a minor effect on the disease risk on their own, put together they all together have even stronger effect than the *HLA* genes. The association findings of this study must be followed-up in genetic studies in other populations but the LD structure must be taken into account more carefully in future studies instead of just genotyping some commonly used markers. Even more importantly, true disease risk variants must be confirmed by functional assays. The genes of T cell co-stimulatory receptors are highly interesting but the final evidence for their role in the pathogenesis of autoimmune diseases still remains to be found. Also, more complex analyses of many variants simultaneously are required to test joint contributions to the disease risk.

Despite improved immunosuppressive medicaments, new organ preservation techniques, and decreased rejection rates, the improvement in long-term kidney allograft survival has been modest. There is growing interest in immunogenetics; if genetic factors determining transplantation patients for example as low- and high-responders could be found. In this study, we found variants of the *ICOS* gene to predispose to the delayed graft function and the decreased graft survival in kidney transplantation patients.

Functional variation in the *CD28-CTLA4-ICOS* region may not be due to a polymorphism in a single gene but rather a few haplotypes carrying a pattern of variations that act together. The involvement of both *CTLA4* and *ICOS* in disease susceptibility would explain the conflicting association and linkage reports from different populations. This study suggests that like the well established *HLA* locus, the co-stimulatory receptor genes predispose to a variety of immune disorders.

9. ACKNOWLEDGEMENTS

This work was carried out at the Finnish Red Cross Blood Service in Helsinki. The financial support of the Finnish Red Cross Blood Service research grant, the Commission of the European Communities, the Sigrid Juselius Foundation and the University of Helsinki is acknowledged.

I wish to express my deep gratitude to all those who made this work possible, especially:

I would like to thank the Director of the institute, Docent Jukka Rautonen, MD, PhD, the former and present Department Heads Docent Jukka Partanen, PhD, Elina Vahtera, PhLic, and Elina Honkavaara, MHA, for providing excellent research facilities and for a positive attitude towards research.

I address my warmest thanks to my supervisors Docent Jukka Partanen, PhD, and Docent Päivi Saavalainen, PhD, for sharing their extensive knowledge as well as for their optimism, great patience and wealth of ideas. They have introduced me to the fascinating world of science and have asked significant questions as well as given expert answers.

I am grateful to Professor Minna Nyström, PhD, and Docent Robert Hermann, MD, PhD, the excellent reviewers of this thesis, for their time and valuable comments.

I express my gratitude to all the co-authors for their collaboration and offering their competence while preparing manuscripts. Special mention goes to the Coeliac Disease Study group led by Professor Markku Mäki in Tampere for their energetic drive. Also, the clinicians in the Helsinki University Central Hospital and abroad are thanked for their valuable help with the sample collection. I am grateful to skilful post-docs Hannu Turpeinen and Elísabet Einarsdóttir as well as to "co-PhD-students" Niina, Kati, Noora, Tanja and Jyrki for their help and friendship.

My special thanks are directed to Niina Woolley for being my personal language coach for years and to Andrea de Kauwe for her expert language consultancy.

I warmly thank every present and former member of Coeliac disease and Transplantation genetics study groups as well as workmates, especially "young researchers" at the Finnish Red Cross Blood Service for creating the most enjoyable working atmosphere. It has been great to share ups and downs with such obliging and cheerful people during all those laborious hours in the lab but also during fantastic moments after work.

I address my very sincere thanks to my parents, grandparents and parents in-law for their care and for being a part of my life. I am grateful to my dear friends and godchildren for always encouraging me and providing balance in my life.

Finally, my sweetest thanks are saved for my beloved husband Miikka for his versatile help, endless support and love.

Helsinki, February 2009

Katri Hiltunen

10. REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. 2002 Jan;30(1):97-101.

Agarwal K, Jones DE, Daly AK, James OF, Vaidya B, Pearce S, et al. CTLA-4 gene polymorphism confers susceptibility to primary biliary cirrhosis. J.Hepatol. 2000 Apr;32(4):538-541.

Ahmed S, Ihara K, Kanemitsu S, Nakashima H, Otsuka T, Tsuzaka K, et al. Association of CTLA-4 but not CD28 gene polymorphisms with systemic lupus erythematosus in the Japanese population. Rheumatology (Oxford) 2001 Jun;40(6):662-667.

Amundsen SS, Naluai AT, Ascher H, Ek J, Gudjonsdottir AH, Wahlstrom J, et al. Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease. Tissue Antigens 2004 Nov;64(5):593-599.

Anjos S, Nguyen A, Ounissi-Benkalha H, Tessier MC, Polychronakos C. A common autoimmunity predisposing signal peptide variant of the cytotoxic T-lymphocyte antigen 4 results in inefficient glycosylation of the susceptibility allele. J.Biol.Chem. 2002 Nov 29;277(48):46478-46486.

Anjos SM, Shao W, Marchand L, Polychronakos C. Allelic effects on gene regulation at the autoimmunity-predisposing CTLA4 locus: a re-evaluation of the 3' +6230G>A polymorphism. Genes Immun. 2005 Jun;6(4):305-311.

Arcos-Burgos M, Muenke M. Genetics of population isolates. Clin. Genet. 2002 Apr;61(4):233-247.

Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat.Rev.Genet. 2002 Apr;3(4):299-309.

Aruffo A, Seed B. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. Proc.Natl.Acad.Sci.U.S.A. 1987 Dec;84(23):8573-8577.

Atabani SF, Thio CL, Divanovic S, Trompette A, Belkaid Y, Thomas DL, et al. Association of CTLA4 polymorphism with regulatory T cell frequency. Eur.J.Immunol. 2005 Jul;35(7):2157-2162.

Babron MC, Nilsson S, Adamovic S, Naluai AT, Wahlstrom J, Ascher H, et al. Meta and pooled analysis of European coeliac disease data. Eur.J.Hum.Genet. 2003 Nov;11(11):828-834.

Barreto M, Santos E, Ferreira R, Fesel C, Fontes MF, Pereira C, et al. Evidence for CTLA4 as a susceptibility gene for systemic lupus erythematosus. Eur.J.Hum.Genet. 2004 Aug;12(8):620-626.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005 Jan 15;21(2):263-265.

Beier KC, Hutloff A, Dittrich AM, Heuck C, Rauch A, Buchner K, et al. Induction, binding specificity and function of human ICOS. Eur.J.Immunol. 2000 Dec;30(12):3707-3717.

Bennett P. Demystified ... microsatellites. Mol.Pathol. 2000 Aug;53(4):177-183.

Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, et al. Contribution of the MHC region to the familial risk of coeliac disease. J.Med.Genet. 1999 Sep;36(9):687-690.

Blanco-Quiros A, Solis-Sanchez P, Garrote-Adrados JA, Arranz-Sanz E. Common variable immunodeficiency. Old questions are getting clearer. Allergol.Immunopathol.(Madr) 2006 Nov-Dec;34(6):263-275.

Blomhoff A, Lie BA, Myhre AG, Kemp EH, Weetman AP, Akselsen HE, et al. Polymorphisms in the cytotoxic T lymphocyte antigen-4 gene region confer susceptibility to Addison's disease. J.Clin.Endocrinol. Metab. 2004 Jul;89(7):3474-3476.

Blumenfeld JD, Catanzaro DF, Kinkhabwala M, Cheigh J, Hartono C, Serur D, et al. Renin system activation and delayed function of the renal transplant. Am.J.Hypertens. 2001 Dec;14(12):1270-1272.

Brophy K, Ryan AW, Thornton JM, Abuzakouk M, Fitzgerald AP, McLoughlin RM, et al. Haplotypes in the CTLA4 region are associated with coeliac disease in the Irish population. Genes Immun. 2006 Jan;7(1):19-26.

Butty V, Roy M, Sabeti P, Besse W, Benoist C, Mathis D. Signatures of strong population differentiation shape extended haplotypes across the human CD28, CTLA4, and ICOS costimulatory genes. Proc.Natl. Acad.Sci.U.S.A. 2007 Jan 9;104(2):570-575.

Capilla A, Donat E, Planelles D, Espinos C, Ribes-Koninckx C, Palau F. Genetic analyses of celiac disease in a Spanish population confirm association with CELIAC3 but not with CELIAC4. Tissue Antigens 2007 Oct;70(4):324-329.

Cardon LR, Bell JI. Association study designs for complex diseases. Nat.Rev.Genet. 2001 Feb;2(2):91-99.

Castelli L, Comi C, Chiocchetti A, Nicola S, Mesturini R, Giordano M, et al. ICOS gene haplotypes correlate with IL10 secretion and multiple sclerosis evolution. J.Neuroimmunol. 2007 May;186(1-2):193-198.

Castigli E, Geha RS. Molecular basis of common variable immunodeficiency. J.Allergy Clin.Immunol. 2006 Apr;117(4):740-6.

Castigli E, Wilson SA, Garibyan L, Rachid R, Bonilla F, Schneider L, et al. TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat.Genet. 2005 Aug;37(8):829-834.

Cataldo F, Lio D, Marino V, Scola L, Crivello A, Corazza GR, et al. Plasma cytokine profiles in patients with celiac disease and selective IgA deficiency. Pediatr.Allergy Immunol. 2003 Aug;14(4):320-324.

Cataldo F, Montalto G. Celiac disease in the developing countries: a new and challenging public health problem. World J.Gastroenterol. 2007 Apr 21;13(15):2153-2159.

Chakravarti A. Population genetics--making sense out of sequence. Nat.Genet. 1999 Jan;21(1 Suppl):56-60.

Chen PL, Fann CS, Chang CC, Wu IL, Chiu WY, Lin CY, et al. Family-based association study of cytotoxic T-lymphocyte antigen-4 with susceptibility to Graves' disease in Han population of Taiwan. Genes Immun. 2008 Mar;9(2):87-92.

Chistiakov DA, Savost'anov KV, Nosikov VV. CTLA4 gene polymorphisms are associated with, and linked to, insulin-dependent diabetes mellitus in a Russian population. BMC Genet. 2001;2:6.

Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994 Nov;138(3):963-971.

Clark AG. The role of haplotypes in candidate gene studies. Genet. Epidemiol. 2004 Dec;27(4):321-333.

Clot F, Fulchignoni-Lataud MC, Renoux C, Percopo S, Bouguerra F, Babron MC, et al. Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations. Tissue Antigens 1999 Nov;54(5):527-530.

Collin P, Reunala T, Pukkala E, Laippala P, Keyrilainen O, Pasternack A. Coeliac disease--associated disorders and survival. Gut 1994 Sep;35(9):1215-1218.

Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc.Natl.Acad.Sci.U.S.A. 1999 Dec 21;96(26):15173-15177.

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 1997 Nov 28;278(5343):1580-1581.

Curtis D. Use of siblings as controls in case-control association studies. Ann.Hum.Genet. 1997 Jul;61(Pt 4):319-333.

Dallos T, Avbelj M, Barak L, Zapletalova J, Pribilincova Z, Krajcirova M, et al. CTLA-4 gene polymorphisms predispose to autoimmune endocrinopathies but not to celiac disease. Neuro Endocrinol.Lett. 2008 Jun;29(3):334-340.

Daly AK, Day CP. Candidate gene case-control association studies: advantages and potential pitfalls. Br.J.Clin.Pharmacol. 2001 Nov;52(5):489-499.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat.Genet. 2001 Oct;29(2):229-232.

Dariavach P, Mattei MG, Golstein P, Lefranc MP. Human Ig superfamily CTLA-4 gene: chromosomal localization and identity of protein sequence between murine and human CTLA-4 cytoplasmic domains. Eur.J.Immunol. 1988 Dec;18(12):1901-1905.

Dawn Teare M, Barrett JH. Genetic linkage studies. Lancet 2005 Sep 17-23;366(9490):1036-1044.

De la Concha EG, Fernandez-Arquero M, Gual L, Vigil P, Martinez A, Urcelay E, et al. MHC susceptibility genes to IgA deficiency are located in different regions on different HLA haplotypes. J.Immunol. 2002 Oct 15;169(8):4637-4643.

de Reuver P, Pravica V, Hop W, Boor P, Metselaar HJ, Hutchinson IV, et al. Recipient ctla-4 +49 G/G genotype is associated with reduced incidence of acute rejection after liver transplantation. Am.J.Transplant. 2003 Dec;3(12):1587-1594.

Defrance T, Vanbervliet B, Briere F, Durand I, Rousset F, Banchereau J. Interleukin 10 and transforming growth factor beta cooperate to induce anti-CD40-activated naive human B cells to secrete immunoglobulin A. J.Exp.Med. 1992 Mar 1;175(3):671-682.

Deichmann K, Heinzmann A, Bruggenolte E, Forster J, Kuehr J. An Mse I RFLP in the human CTLA4 promotor. Biochem.Biophys.Res.Commun. 1996 Aug 23;225(3):817-818.

Deshpande M, Venuprasad K, Parab PB, Saha B, Mitra D. A novel CD28 mRNA variant and simultaneous presence of various CD28 mRNA isoforms in human T lymphocytes. Hum.Immunol. 2002 Jan;63(1):20-23.

Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 1995 Sep 20;29(2):311-322.

Dewar DH, Ciclitira PJ. Clinical features and diagnosis of celiac disease. Gastroenterology 2005 Apr;128(4 Suppl 1):S19-24.

Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, Mougenot JF, Bach JF, Caillat-Zucman S. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. Gut 1998 Aug;43(2):187-189.

Dmitrienko S, Hoar DI, Balshaw R, Keown PA. Immune response gene polymorphisms in renal transplant recipients. Transplantation 2005 Dec 27;80(12):1773-1782.

Dong C, Juedes AE, Temann UA, Shresta S, Allison JP, Ruddle NH, et al. ICOS co-stimulatory receptor is essential for T-cell activation and function. Nature 2001 Jan 4;409(6816):97-101.

Dowsing T, Kendall MJ. The Northwick Park tragedy--protecting healthy volunteers in future first-in-man trials. J.Clin.Pharm.Ther. 2007 Jun;32(3):203-207.

Dube C, Rostom A, Sy R, Cranney A, Saloojee N, Garritty C, et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. Gastroenterology 2005 Apr;128(4 Suppl 1):S57-67.

Einarsdottir E, Soderstrom I, Lofgren-Burstrom A, Haraldsson S, Nilsson-Ardnor S, Penha-Goncalves C, et al. The CTLA4 region as a general autoimmunity factor: an extended pedigree provides evidence for synergy with the HLA locus in the etiology of type 1 diabetes mellitus, Hashimoto's thyroiditis and Graves' disease. Eur.J.Hum. Genet. 2003 Jan;11(1):81-84.

Ellsworth DL, Manolio TA. The emerging importance of genetics in epidemiologic research II. Issues in study design and gene mapping. Ann.Epidemiol. 1999 Feb;9(2):75-90.

Endacott R. Clinical research 2: Legal and ethical issues in research. Accid.Emerg.Nurs. 2007 Jul;15(3):166-169.

Falchuk ZM, Strober W. HL-A antigens and adult coeliac disease. Lancet 1972 Dec 16;2(7790):1310.

Fan LY, Tu XQ, Cheng QB, Zhu Y, Feltens R, Pfeiffer T, et al. Cytotoxic T lymphocyte associated antigen-4 gene polymorphisms confer susceptibility to primary biliary cirrhosis and autoimmune hepatitis in Chinese population. World J.Gastroenterol. 2004 Oct 15;10(20):3056-3059.

Ferreiros-Vidal I, Gomez-Reino JJ, Barros F, Carracedo A, Carreira P, Gonzalez-Escribano F, et al. Association of PDCD1 with susceptibility to systemic lupus erythematosus: evidence of population-specific effects. Arthritis Rheum. 2004 Aug;50(8):2590-2597.

Finger LR, Pu J, Wasserman R, Vibhakar R, Louie E, Hardy RR, et al. The human PD-1 gene: complete cDNA, genomic organization, and developmentally regulated expression in B cell progenitors. Gene 1997 Sep 15;197(1-2):177-187.

Freeman GJ, Long AJ, Iwai Y, Bourque K, Chernova T, Nishimura H, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J.Exp.Med. 2000 Oct 2;192(7):1027-1034.

Furugaki K, Shirasawa S, Ishikawa N, Ito K, Ito K, Kubota S, et al. Association of the T-cell regulatory gene CTLA4 with Graves' disease

and autoimmune thyroid disease in the Japanese. J.Hum.Genet. 2004;49(3):166-168.

Garner CP, Ding YC, Steele L, Book L, Leiferman K, Zone JJ, et al. Genome-wide linkage analysis of 160 North American families with celiac disease. Genes Immun. 2007 Mar;8(2):108-114.

Gendzekhadze K, Rivas-Vetencourt P, Montano RF. Risk of adverse post-transplant events after kidney allograft transplantation as predicted by CTLA-4 +49 and TNF-alpha -308 single nucleotide polymorphisms: a preliminary study. Transpl.Immunol. 2006 Nov;16(3-4):194-199.

Goldrath AW, Bevan MJ. Selecting and maintaining a diverse T-cell repertoire. Nature 1999 Nov 18;402(6759):255-262.

Goldstein DB. Islands of linkage disequilibrium. Nat.Genet. 2001 Oct;29(2):109-111.

Gough SC, Walker LS, Sansom DM. CTLA4 gene polymorphism and autoimmunity. Immunol.Rev. 2005 Apr;204:102-115.

Graham DS, Wong AK, McHugh NJ, Whittaker JC, Vyse TJ. Evidence for unique association signals in SLE at the CD28-CTLA4-ICOS locus in a family-based study. Hum.Mol.Genet. 2006 Nov 1;15(21):3195-3205.

Greco L, Corazza G, Babron MC, Clot F, Fulchignoni-Lataud MC, Percopo S, et al. Genome search in celiac disease. Am.J.Hum.Genet. 1998 Mar;62(3):669-675.

Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, et al. The first large population based twin study of coeliac disease. Gut 2002 May;50(5):624-628.

Greve B, Simonenko R, Illes Z, Peterfalvi A, Hamdi N, Mycko MP, et al. Multiple sclerosis and the CTLA4 autoimmunity polymorphism CT60: no association in patients from Germany, Hungary and Poland. Mult. Scler. 2008 Mar;14(2):153-158.

Grimbacher B, Hutloff A, Schlesier M, Glocker E, Warnatz K, Drager R, et al. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. Nat.Immunol. 2003 Mar;4(3):261-268.

Haaning Andersen AD, Lange M, Lillevang ST. Allelic variation of the inducible costimulator (ICOS) gene: detection of polymorphisms, analysis of the promoter region, and extended haplotype estimation. Tissue Antigens 2003 Apr;61(4):276-285.

Hammarström L, Vorechovsky I, Webster D. Selective IgA deficiency (SIgAD) and common variable immunodeficiency (CVID). Clin.Exp. Immunol. 2000 May;120(2):225-231.

Harper K, Balzano C, Rouvier E, Mattei MG, Luciani MF, Golstein P. CTLA-4 and CD28 activated lymphocyte molecules are closely related in both mouse and human as to sequence, message expression,

gene structure, and chromosomal location. J.Immunol. 1991 Aug 1;147(3):1037-1044.

Heward J, Gough SC. Genetic susceptibility to the development of autoimmune disease. Clin.Sci.(Lond) 1997 Dec;93(6):479-491.

Hill, WG, Roberson, A. Normal 0 21 Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38:226-131.

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat.Rev.Genet. 2005 Feb;6(2):95-108.

Holopainen P, Arvas M, Sistonen P, Mustalahti K, Collin P, Maki M, et al. CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. Tissue Antigens 1999 May;53(5):470-475.

Holopainen P, Naluai AT, Moodie S, Percopo S, Coto I, Clot F, et al. Candidate gene region 2q33 in European families with coeliac disease. Tissue Antigens 2004 Mar;63(3):212-222.

Holopainen PM, Partanen JA. Technical note: linkage disequilibrium and disease-associated CTLA4 gene polymorphisms. J.Immunol. 2001 Sep 1;167(5):2457-2458.

Hopper AD, Hadjivassiliou M, Butt S, Sanders DS. Adult coeliac disease. BMJ 2007 Sep 15;335(7619):558-562.

Howard TD, Postma DS, Hawkins GA, Koppelman GH, Zheng SL, Wysong AK, et al. Fine mapping of an IgE-controlling gene on chromosome 2q: Analysis of CTLA4 and CD28. J.Allergy Clin.Immunol. 2002 Nov;110(5):743-751.

Hudson LL, Rocca K, Song YW, Pandey JP. CTLA-4 gene polymorphisms in systemic lupus erythematosus: a highly significant association with a determinant in the promoter region. Hum.Genet. 2002 Oct;111(4-5):452-455.

Hunt KA, McGovern DP, Kumar PJ, Ghosh S, Travis SP, Walters JR, et al. A common CTLA4 haplotype associated with coeliac disease. Eur.J.Hum.Genet. 2005 Apr;13(4):440-444.

Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat.Genet. 2008 Apr;40(4):395-402.

Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, Anagnostopoulos I, et al. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. Nature 1999 Jan 21;397(6716):263-266.

Ihara K, Ahmed S, Nakao F, Kinukawa N, Kuromaru R, Matsuura N, et al. Association studies of CTLA-4, CD28, and ICOS gene polymorphisms

with type 1 diabetes in the Japanese population. Immunogenetics 2001 Aug;53(6):447-454.

Jeggo PA, Löbrich M. Contribution of DNA repair and cell cycle checkpoint arrest to the maintenance of genomic stability. DNA Repair (Amst) 2006 Sep 8;5(9-10):1192-1198.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nat.Genet. 2001 Oct;29(2):233-237.

Juran BD, Atkinson EJ, Schlicht EM, Fridley BL, Lazaridis KN. Primary biliary cirrhosis is associated with a genetic variant in the 3' flanking region of the CTLA4 gene. Gastroenterology 2008 Oct;135(4):1200-1206.

Kaartinen T, Lappalainen J, Haimila K, Autero M, Partanen J. Genetic variation in ICOS regulates mRNA levels of ICOS and splicing isoforms of CTLA4. Mol.Immunol. 2007 Mar;44(7):1644-1651.

Kagnoff MF. Celiac disease: pathogenesis of a model immunogenetic disease. J.Clin.Invest. 2007 Jan;117(1):41-49.

Kantarci OH, Hebrink DD, Achenbach SJ, Atkinson EJ, Waliszewska A, Buckle G, et al. CTLA4 is associated with susceptibility to multiple sclerosis. J.Neuroimmunol. 2003 Jan;134(1-2):133-141.

Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, et al. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. Hum.Immunol. 2003 Apr;64(4):469-477.

Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Annu.Rev.Immunol. 2008;26:677-704.

Kemp EH, Ajjan RA, Husebye ES, Peterson P, Uibo R, Imrie H, et al. A cytotoxic T lymphocyte antigen-4 (CTLA-4) gene polymorphism is associated with autoimmune Addison's disease in English patients. Clin.Endocrinol.(Oxf) 1998 Nov;49(5):609-613.

King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, et al. Coeliac disease: investigation of proposed causal variants in the CTLA4 gene region. Eur.J.Immunogenet. 2003 Dec;30(6):427-432.

King AL, Moodie SJ, Fraser JS, Curtis D, Reid E, Dearlove AM, et al. CTLA-4/CD28 gene region is associated with genetic susceptibility to coeliac disease in UK families. J.Med.Genet. 2002 Jan;39(1):51-54.

King AL, Yiannakou JY, Brett PM, Curtis D, Morris MA, Dearlove AM, et al. A genome-wide family-based linkage study of coeliac disease. Ann. Hum.Genet. 2000 Nov;64(Pt 6):479-490.

Kitani A, Strober W. Differential regulation of C alpha 1 and C alpha 2 germ-line and mature mRNA transcripts in human peripheral blood B cells. J.Immunol. 1994 Aug 15;153(4):1466-1477.

Kong EK, Prokunina-Olsson L, Wong WH, Lau CS, Chan TM, Alarcon-Riquelme M, et al. A new haplotype of PDCD1 is associated with rheumatoid arthritis in Hong Kong Chinese. Arthritis Rheum. 2005 Apr;52(4):1058-1062.

Koskinen S. Long-term follow-up of health in blood donors with primary selective IgA deficiency. J.Clin.Immunol. 1996 May;16(3):165-170.

Kotsa K, Watson PF, Weetman AP. A CTLA-4 gene polymorphism is associated with both Graves disease and autoimmune hypothyroidism. Clin.Endocrinol.(Oxf) 1997 May;46(5):551-554.

Kouki T, Sawai Y, Gardine CA, Fisfalen ME, Alegre ML, DeGroot LJ. 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. 2000 Dec 1;165(11):6606-6611.

Kraft P, Hunter D. Integrating epidemiology and genetic association: the challenge of gene-environment interaction. Philos.Trans.R.Soc. Lond.B.Biol.Sci. 2005 Aug 29;360(1460):1609-1616.

Kruger B, Schroppel B, Murphy BT. Genetic polymorphisms and the fate of the transplanted organ. Transplant.Rev.(Orlando) 2008 Apr;22(2):131-140.

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat.Genet. 1999 Jun;22(2):139-144.

Lafferty KJ, Cunningham AJ. A new analysis of allogeneic interactions. Aust.J.Exp.Biol.Med.Sci. 1975 Feb;53(1):27-42.

Lamhamedi-Cherradi SE, Boulard O, Gonzalez C, Kassis N, Damotte D, Eloy L, et al. Further mapping of the Idd5.1 locus for autoimmune diabetes in NOD mice. Diabetes 2001 Dec;50(12):2874-2878.

Latiff AH, Kerr MA. The clinical significance of immunoglobulin A deficiency. Ann.Clin.Biochem. 2007 Mar;44(Pt 2):131-139.

Lee CS, Lee YJ, Liu HF, Su CH, Chang SC, Wang BR, et al. Association of CTLA4 gene A-G polymorphism with rheumatoid arthritis in Chinese. Clin.Rheumatol. 2003 Sep;22(3):221-224.

Lee KP, Taylor C, Petryniak B, Turka LA, June CH, Thompson CB. The genomic organization of the CD28 gene. Implications for the regulation of CD28 mRNA expression and heterogeneity. J.Immunol. 1990 Jul 1;145(1):344-352.

Lee WC. Case-control association studies with matching and genomic controlling. Genet.Epidemiol. 2004 Jul;27(1):1-13.

Lewontin R.C. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 1964;49(January):49--67.

Ligers A, Teleshova N, Masterman T, Huang WX, Hillert J. CTLA-4 gene expression is influenced by promoter and exon 1 polymorphisms. Genes Immun. 2001 May;2(3):145-152.

Ligers A, Xu C, Saarinen S, Hillert J, Olerup O. The CTLA-4 gene is associated with multiple sclerosis. J.Neuroimmunol. 1999 Jun 1;97(1-2):182-190.

Lin SC, Yen JH, Tsai JJ, Tsai WC, Ou TT, Liu HW, et al. Association of a programmed death 1 gene polymorphism with the development of rheumatoid arthritis, but not systemic lupus erythematosus. Arthritis Rheum. 2004 Mar;50(3):770-775.

Ling V, Wu PW, Finnerty HF, Agostino MJ, Graham JR, Chen S, et al. Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches. Genomics 2001 Dec;78(3):155-168.

Ling V, Wu PW, Finnerty HF, Sharpe AH, Gray GS, Collins M. Complete sequence determination of the mouse and human CTLA4 gene loci: cross-species DNA sequence similarity beyond exon borders. Genomics 1999 Sep 15;60(3):341-355.

Linsley PS, Ledbetter JA. The role of the CD28 receptor during T cell responses to antigen. Annu.Rev.Immunol. 1993;11:191-212.

Liu J, Juo SH, Holopainen P, Terwilliger J, Tong X, Grunn A, et al. Genomewide linkage analysis of celiac disease in Finnish families. Am.J.Hum.Genet. 2002 Jan;70(1):51-59.

Lorentzen AR, Celius EG, Ekstrom PO, Wiencke K, Lie BA, Myhr KM, et al. Lack of association with the CD28/CTLA4/ICOS gene region among Norwegian multiple sclerosis patients. J.Neuroimmunol. 2005 Sep;166(1-2):197-201.

Magistrelli G, Jeannin P, Herbault N, Benoit De Coignac A, Gauchat JF, Bonnefoy JY, et al. A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. Eur.J.Immunol. 1999 Nov;29(11):3596-3602.

Marder BA, Schroppel B, Lin M, Schiano T, Parekh R, Tomer Y, et al. The impact of costimulatory molecule gene polymorphisms on clinical outcomes in liver transplantation. Am.J.Transplant. 2003 Apr;3(4):424-431.

Marron MP, Raffel LJ, Garchon HJ, Jacob CO, Serrano-Rios M, Martinez Larrad MT, et al. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. Hum. Mol.Genet. 1997 Aug;6(8):1275-1282.

Marron MP, Zeidler A, Raffel LJ, Eckenrode SE, Yang JJ, Hopkins DI, et al. 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 2000 Mar;49(3):492-499.

Martin ER, Kaplan NL, Weir BS. Tests for linkage and association in nuclear families. Am.J.Hum.Genet. 1997 Aug;61(2):439-448.

Martin-Pagola A, Perez de Nanclares G, Vitoria JC, Bilbao JR, Ortiz L, Zubillaga P, et al. No association of CTLA4 gene with celiac disease in the Basque population. J.Pediatr.Gastroenterol.Nutr. 2003 Aug;37(2):142-145.

Maurer M, Loserth S, Kolb-Maurer A, Ponath A, Wiese S, Kruse N, et al. A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation. Immunogenetics 2002 Apr;54(1):1-8.

McAdam AJ, Greenwald RJ, Levin MA, Chernova T, Malenkovich N, Ling V, et al. ICOS is critical for CD40-mediated antibody class switching. Nature 2001 Jan 4;409(6816):102-105.

Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 2006 Sep;16(9):1182-1190.

Monsuur AJ, Wijmenga C. Understanding the molecular basis of celiac disease: what genetic studies reveal. Ann.Med. 2006;38(8):578-591.

Mora B, Bonamico M, Indovina P, Megiorni F, Ferri M, Carbone MC, et al. CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease. Hum.Immunol. 2003 Feb;64(2):297-301.

Muro M, Rojas G, Botella C, Miras M, Campillo JA, Minguela A, et al. CT60 A/G marker of the 3'-UTR of the CTLA4 gene and liver transplant. Transpl.Immunol. 2008 Jan;18(3):246-249.

Mäki M, Mustalahti K, Kokkonen J, Kulmala P, Haapalahti M, Karttunen T, et al. Prevalence of Celiac disease among children in Finland. N.Engl.J.Med. 2003 Jun 19;348(25):2517-2524.

Naluai AT, Nilsson S, Gudjonsdottir AH, Louka AS, Ascher H, Ek J, et al. Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. Eur.J.Hum.Genet. 2001 Dec;9(12):938-944.

Naluai AT, Nilsson S, Samuelsson L, Gudjonsdottir AH, Ascher H, Ek J, et al. The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. Tissue Antigens 2000 Oct;56(4):350-355.

Neuhausen SL, Feolo M, Camp NJ, Farnham J, Book L, Zone JJ. Genome-wide linkage analysis for celiac disease in North American families. Am.J.Med.Genet. 2002 Jul 22;111(1):1-9.

Newby PR, Roberts-Davies EL, Brand OJ, Heward JM, Franklyn JA, Gough SC, et al. Tag SNP screening of the PDCD1 gene for association with Graves' disease. Clin.Endocrinol.(Oxf) 2007 Jul;67(1):125-128.

Ni R, Ihara K, Miyako K, Kuromaru R, Inuo M, Kohno H, et al. PD-1 gene haplotype is associated with the development of type 1 diabetes mellitus in Japanese children. Hum.Genet. 2007 Apr;121(2):223-232.

Nielsen C, Hansen D, Husby S, Jacobsen BB, Lillevang ST. Association of a putative regulatory polymorphism in the PD-1 gene with susceptibility to type 1 diabetes. Tissue Antigens 2003 Dec;62(6):492-497.

Nielsen C, Ohm-Laursen L, Barington T, Husby S, Lillevang ST. Alternative splice variants of the human PD-1 gene. Cell.Immunol. 2005 Jun;235(2):109-116.

Nishimura H, Agata Y, Kawasaki A, Sato M, Imamura S, Minato N, et al. Developmentally regulated expression of the PD-1 protein on the surface of double-negative (CD4-CD8-) thymocytes. Int.Immunol. 1996 May;8(5):773-780.

Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum. Mol.Genet. 1996 Jul;5(7):1075-1080.

Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut 2006 Jun;55(6):803-808.

Nithiyananthan R, Heward JM, Allahabadia A, Franklyn JA, Gough SC. Polymorphism of the CTLA-4 gene is associated with autoimmune hypothyroidism in the United Kingdom. Thyroid 2002 Jan;12(1):3-6.

Oaks MK, Hallett KM, Penwell RT, Stauber EC, Warren SJ, Tector AJ. A native soluble form of CTLA-4. Cell.Immunol. 2000 May 1;201(2):144-153.

O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am.J.Hum.Genet. 1998 Jul;63(1):259-266.

Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue Antigens 1992 May;39(5):225-235.

Pakravan N, Hassan AT, Hassan ZM. Naturally occurring self-reactive CD4+CD25+ regulatory T cells: universal immune code. Cell.Mol. Immunol. 2007 Jun;4(3):197-201.

Pawlak E, Kochanowska IE, Frydecka I, Kielbinski M, Potoczek S, Bilinska M. The soluble CTLA-4 receptor: a new marker in autoimmune diseases. Arch.Immunol.Ther.Exp.(Warsz) 2005 Jul-Aug;53(4):336-341.

Perez de Nanclares G, Martin-Pagola A, Ramon Bilbao J, Vazquez F, Castano L. No evidence of association of CTLA4 polymorphisms with Addison's disease. Autoimmunity 2004 Sep-Nov;37(6-7):453-456.

Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. Ann.Hum.Genet. 1997 Jul;61(Pt 4):307-317.

Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am.J.Hum.Genet. 2005 Dec;77(6):1044-1060.

Polymeropoulos MH, Xiao H, Rath DS, Merril CR. Dinucleotide repeat polymorphism at the human CTLA4 gene. Nucleic Acids Res. 1991 Jul 25;19(14):4018.

Popat S, Bevan S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, et al. Genome screening of coeliac disease. J.Med. Genet. 2002 May;39(5):328-331.

Popat S, Hearle N, Hogberg L, Braegger CP, O'Donoghue D, Falth-Magnusson K, et al. Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. Ann.Hum.Genet. 2002 Mar;66(Pt 2):125-137.

Price P, Witt C, Allcock R, Sayer D, Garlepp M, Kok CC, et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol.Rev. 1999 Feb;167:257-274.

Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am.J.Hum.Genet. 2001 Jul;69(1):1-14.

Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat. Genet. 2002 Dec;32(4):666-669.

Prokunina L, Gunnarsson I, Sturfelt G, Truedsson L, Seligman VA, Olson JL, et al. The systemic lupus erythematosus-associated PDCD1 polymorphism PD1.3A in lupus nephritis. Arthritis Rheum. 2004 Jan;50(1):327-328.

Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 2003 Jan;19(1):149-150.

Ramirez-Soriano A, Lao O, Soldevila M, Calafell F, Bertranpetit J, Comas D. Haplotype tagging efficiency in worldwide populations in CTLA4 gene. Genes Immun. 2005 Dec;6(8):646-657.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006 Nov 23;444(7118):444-454.

Reunala TL. Dermatitis herpetiformis. Clin.Dermatol. 2001 Nov-Dec;19(6):728-736.

Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. Adv.Genet. 2008;60:293-308.

Rioux JD, Karinen H, Kocher K, McMahon SG, Karkkainen P, Janatuinen E, et al. Genomewide search and association studies in a Finnish celiac disease population: Identification of a novel locus and replication of the HLA and CTLA4 loci. Am.J.Med.Genet.A. 2004 Nov 1;130A(4):345-350.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996 Sep 13;273(5281):1516-1517.

Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000 Jun 15;405(6788):847-856.

Rodriguez MR, Nunez-Roldan A, Aguilar F, Valenzuela A, Garcia A, Gonzalez-Escribano MF. Association of the CTLA4 3' untranslated region polymorphism with the susceptibility to rheumatoid arthritis. Hum.Immunol. 2002 Jan;63(1):76-81.

Romagnani S. Introduction: characterization and functions of human T regulatory cells. Microbes Infect. 2005 Jun;7(7-8):1015-1016.

Romanos J, Barisani D, Trynka G, Zhernakova A, Bardella MT, Wijmenga C. Six new celiac disease loci replicated in an Italian population confirm association to celiac disease. J.Med.Genet. 2008 Sep 19.

Rueda B, Zhernakova A, Lopez-Nevot MA, Gomez-Garcia M, Ortega E, Pinero A, et al. CTLA4/CT60 polymorphism is not relevant in susceptibility to autoimmune inflammatory intestinal disorders. Hum. Immunol. 2005 Mar;66(3):321-325.

Saavalainen P. Genetics of celiac disease. Abstract book. University Medical Center, Maribor, Slovenia; 2007.

Salzer U, Grimbacher B. Common variable immunodeficiency: The power of co-stimulation. Semin.Immunol. 2006 Dec;18(6):337-346.

Salzer U, Maul-Pavicic A, Cunningham-Rundles C, Urschel S, Belohradsky BH, Litzman J, et al. ICOS deficiency in patients with common variable immunodeficiency. Clin.Immunol. 2004 Dec;113(3):234-240.

Schneider, S, Roessli, D, Excoffier, L. arlequin: A Software for Population Genetics Data Analysis., version 2.0. 2000.

Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, et al. Structural basis for gluten intolerance in celiac sprue. Science 2002 Sep 27;297(5590):2275-2279.

Shastry BS. SNP alleles in human disease and evolution. J.Hum.Genet. 2002;47(11):561-566.

Shilling RA, Pinto JM, Decker DC, Schneider DH, Bandukwala HS, Schneider JR, et al. Cutting edge: polymorphisms in the ICOS promoter region are associated with allergic sensitization and Th2 cytokine production. J.Immunol. 2005 Aug 15;175(4):2061-2065.

Shinohara T, Taniwaki M, Ishida Y, Kawaichi M, Honjo T. Structure and chromosomal localization of the human PD-1 gene (PDCD1). Genomics 1994 Oct;23(3):704-706.

Slavcheva E, Albanis E, Jiao Q, Tran H, Bodian C, Knight R, et al. Cytotoxic T-lymphocyte antigen 4 gene polymorphisms and susceptibility to acute allograft rejection. Transplantation 2001 Sep 15;72(5):935-940.

Snanoudj R, de Preneuf H, Creput C, Arzouk N, Deroure B, Beaudreuil S, et al. Costimulation blockade and its possible future use in clinical transplantation. Transpl.Int. 2006 Sep;19(9):693-704.

Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J.Exp.Med. 1989 Jan 1;169(1):345-350.

Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. Gastroenterology 1993 Sep;105(3):910-922.

Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune diseases co-occurring within individuals and within families: a systematic review. Epidemiology 2006 Mar;17(2):202-217.

Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am.J.Hum.Genet. 1993 Mar;52(3):506-516.

Stokes PL, Asquith P, Holmes GK, Mackintosh P, Cooke WT. Histocompatibility antigens associated with adult coeliac disease. Lancet 1972 Jul 22;2(7769):162-164.

Swallow MM, Wallin JJ, Sha WC. B7h, a novel costimulatory homolog of B7.1 and B7.2, is induced by TNFalpha. Immunity 1999 Oct;11(4):423-432.

Tapirdamaz O, Pravica V, Metselaar HJ, Hansen B, Moons L, van Meurs JB, et al. Polymorphisms in the T cell regulatory gene cytotoxic T lymphocyte antigen 4 influence the rate of acute rejection after liver transplantation. Gut 2006 Jun;55(6):863-868.

The International HapMap Consortium. The International HapMap Project. Nature 2003 Dec 18;426(6968):789-796.

Thompson CB, Lindsten T, Ledbetter JA, Kunkel SL, Young HA, Emerson SG, et al. CD28 activation pathway regulates the production of multiple T-cell-derived lymphokines/cytokines. Proc.Natl.Acad.Sci.U.S.A. 1989 Feb;86(4):1333-1337.

Thomson G. Mapping disease genes: family-based association studies. Am.J.Hum.Genet. 1995 Aug;57(2):487-498.

Thorburn CM, Prokunina-Olsson L, Sterba KA, Lum RF, Seldin MF, Alarcon-Riquelme ME, et al. Association of PDCD1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. Genes Immun. 2007 Jun;8(4):279-287.

Tivol EA, Borriello F, Schweitzer AN, Lynch WP, Bluestone JA, Sharpe AH. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. Immunity 1995 Nov;3(5):541-547.

Torres MI, Lopez Casado MA, Rios A. New aspects in celiac disease. World J.Gastroenterol. 2007 Feb 28;13(8):1156-1161.

Trivedi HL. Immunobiology of rejection and adaptation. Transplant. Proc. 2007 Apr;39(3):647-652.

Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003 May 29;423(6939):506-511.

Vaidya B, Oakes EJ, Imrie H, Dickinson AJ, Perros P, Kendall-Taylor P, et al. CTLA4 gene and Graves' disease: association of Graves' disease with the CTLA4 exon 1 and intron 1 polymorphisms, but not with the promoter polymorphism. Clin.Endocrinol.(Oxf) 2003 Jun;58(6):732-735.

Vaidya B, Pearce SH, Charlton S, Marshall N, Rowan AD, Griffiths ID, et al. An association between the CTLA4 exon 1 polymorphism and early rheumatoid arthritis with autoimmune endocrinopathies. Rheumatology (Oxford) 2002 Feb;41(2):180-183.

van Belzen MJ, Meijer JW, Sandkuijl LA, Bardoel AF, Mulder CJ, Pearson PL, et al. A major non-HLA locus in celiac disease maps to chromosome 19. Gastroenterology 2003 Oct;125(4):1032-1041.

van Belzen MJ, Mulder CJ, Zhernakova A, Pearson PL, Houwen RH, Wijmenga C. CTLA4 +49 A/G and CT60 polymorphisms in Dutch coeliac disease patients. Eur.J.Hum.Genet. 2004 Sep;12(9):782-785.

van Belzen MJ, Vrolijk MM, Meijer JW, Crusius JB, Pearson PL, Sandkuijl LA, et al. A genomewide screen in a four-generation Dutch family with celiac disease: evidence for linkage to chromosomes 6 and 9. Am.J.Gastroenterol. 2004 Mar;99(3):466-471.

van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat.Genet. 2007 Jul;39(7):827-829.

van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. Best Pract.Res.Clin.Gastroenterol. 2005 Jun;19(3):323-339.

Vincenti F. Current use and future trends in induction therapy. Saudi J.Kidney Dis.Transpl. 2005 Oct-Dec;16(4):506-513.

Vincenti F, Luggen M. T cell costimulation: a rational target in the therapeutic armamentarium for autoimmune diseases and transplantation. Annu.Rev.Med. 2007;58:347-358.

Viola A, Lanzavecchia A. T cell activation determined by T cell receptor number and tunable thresholds. Science 1996 Jul 5;273(5271):104-106.

Vogel F MA. Human genetics - problems and approaches. 2nd edn ed. Berlin: Springer; 1986.

Walker-Smith JA, Guandalini S, Schmitz J, Shmerling D, Visakorpi J. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. 1990; 65: 909–11. Arch.Dis.Child. 1990;65:909-911.

Wang XB, Zhao X, Giscombe R, Lefvert AK. A CTLA-4 gene polymorphism at position -318 in the promoter region affects the expression of protein. Genes Immun. 2002 Jun;3(4):233-234.

Warnatz K, Bossaller L, Salzer U, Skrabl-Baumgartner A, Schwinger W, van der Burg M, et al. Human ICOS deficiency abrogates the germinal center reaction and provides a monogenic model for common variable immunodeficiency. Blood 2006 Apr 15;107(8):3045-3052.

Waterhouse P, Penninger JM, Timms E, Wakeham A, Shahinian A, Lee KP, et al. Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. Science 1995 Nov 10;270(5238):985-988.

Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc.Natl.Acad.Sci.U.S.A. 1997 Apr 1;94(7):3195-3199.

Wisniewski A, Kusztal M, Magott-Procelewska M, Klinger M, Jasek M, Luszczek W, et al. Possible association of cytotoxic T-lymphocyte antigen 4 gene promoter single nucleotide polymorphism with acute rejection of allogeneic kidney transplant. Transplant.Proc. 2006 Jan-Feb;38(1):56-58.

Woof JM, Kerr MA. The function of immunoglobulin A in immunity. J.Pathol. 2006 Jan;208(2):270-282.

Woolley N, Holopainen P, Ollikainen V, Mustalahti K, Mäki M, Kere J, et al. A new locus for coeliac disease mapped to chromosome 15 in a population isolate. Hum.Genet. 2002 Jul;111(1):40-45.

Yamada A, Salama AD, Sayegh MH. The role of novel T cell costimulatory pathways in autoimmunity and transplantation. J.Am.Soc.Nephrol. 2002 Feb;13(2):559-575.

Yates PJ, Nicholson ML. The aetiology and pathogenesis of chronic allograft nephropathy. Transpl.Immunol. 2006 Nov;16(3-4):148-157.

Yi H, Zhen Y, Jiang L, Zheng J, Zhao Y. The phenotypic characterization of naturally occurring regulatory CD4+CD25+ T cells. Cell.Mol.Immunol. 2006 Jun;3(3):189-195.

Yoshinaga SK, Whoriskey JS, Khare SD, Sarmiento U, Guo J, Horan T, et al. T-cell co-stimulation through B7RP-1 and ICOS. Nature 1999 Dec 16;402(6763):827-832.

Zand MS. Immunosuppression and immune monitoring after renal transplantation. Semin.Dial. 2005 Nov-Dec;18(6):511-519.

Zhernakova A, Eerligh P, Barrera P, Wesoly JZ, Huizinga TW, Roep BO, et al. CTLA4 is differentially associated with autoimmune diseases in the Dutch population. Hum.Genet. 2005 Oct;118(1):58-66.

Zhong F, McCombs CC, Olson JM, Elston RC, Stevens FM, McCarthy CF, et al. An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. Nat.Genet. 1996 Nov;14(3):329-333.

Özkaynak E, Gao W, Shemmeri N, Wang C, Gutierrez-Ramos JC, Amaral J, et al. Importance of ICOS-B7RP-1 costimulation in acute and chronic allograft rejection. Nat.Immunol. 2001 Jul;2(7):591-596.