T-cell receptors and human leukocyte antigens in primary sclerosing cholangitis

Thesis for the degree of *Philosophiae Doctor* (PhD)

by

Eva Kristine Klemsdal Henriksen

Norwegian PSC Research Center and Research Institute of Internal Medicine,
Division of Surgery, Inflammatory Medicine and Transplantation,
Oslo University Hospital Rikshospitalet

Faculty of Medicine, University of Oslo

2017

(Colophon page)

TABLE OF CONTENTS

1 ACKNOWLEDGEMENTS

The work presented in this thesis was carried out at the Norwegian PSC Research Center (NoPSC) and Research Institute of Internal Medicine (RIIM), Oslo University Hospital (OUH), Rikshospitalet between 2013 and 2017. Financial support from Canica AS is gratefully acknowledged.

First, I would like to thank the patients enrolled in the various studies, whose contributions made this research possible.

I am sincerely grateful to my main supervisor, Tom H. Karlsen. Thank you for giving me the opportunity to work on this captivating disease. I am thankful for your kindhearted and patient guidance, and for trusting me with great responsibilities and independence. I admire your ability to facilitate research collaboration, which together with your humble and hard-working personality is truly inspirational.

I would also like to express my gratitude to my three excellent co-supervisors. Espen Melum, I have appreciated working as part of your research group, and I am thankful for your sincere enthusiasm and sense of detail. Thank you for generously sharing your extensive knowledge. Benedicte A. Lie, thank you for always finding the time to supervise and to encourage me through both frustrating and joyful times. I have enjoyed taking part in your weakly group meetings, in which you made me feel 'at home'. Evaggelia (Lia) Liaskou, our great scientific discussions throughout the work for this thesis have been invaluable. Thank you for always motivating me and for being my friend.

Kristian Holm, thank you for patiently teaching me how to run Perl scripts on my computer and for always being eager to perform the bioinformatics analyses we could think of. I am also very grateful for your enthusiasm during our many informal discussions.

Marte K. Viken, I have highly appreciated our scientific (and not-so-scientific) talks. Thank you for continuously helping and supporting me in any way possible.

I would like to thank the present and former members of NoPSC: Erik S., Kirsten M.B., Johannes R.H., Trine F., Xiaojun J., Mette V., Kristin K.J., Kristine W., Kristian B., Bjarte F., Sigrid N., Elisabeth S., Natalie L.B., Martin K., Laura V., Fei (Freeman) Z., Amandeep K.D., Christopher S.L., Gupta U., Brian K.C., Georg S., Kristian H., Anne P., Liv Wenche T., Merete T., Siv F., Hege D.S., Mona B., Julia F., Kristian A., Jarl Andreas A., Corey T. and Tonje B. Thank you for providing an excellent scientific environment where I have been constantly reminded on the

importance of PSC research. I would also like to thank the Head of the Institute, Bente Halvorsen, and the members of RIIM for providing a friendly and social working environment. A special thanks to my friends and fellow PhD students: Elisabeth S., Natalie L.B., Martin K., Laura V., Negar S., Ana Q.J., Huda A., and Aurelija A. I have truly appreciated our conversations throughout these last few years, sharing the ups and downs in the PhD life.

Next, I would like to thank the members of the Immunogenetics group at OUH Ullevål, including Siri T.F., Fatemeh K., Hanna H., Marte K.V., Ingvild G., Kari G., Line S., Asgeir L., Fatima H. and Maria D.V. I have highly appreciated being invited to your social get-togethers.

Each and one of our collaborators and co-authors are acknowledged for their contributions, providing biological material and scientific input to our studies. Gideon M. Hirschfield at the University of Birmingham (the United Kingdom), Andre Franke at Kiel University (Germany), Christopher L. Bowlus at the University of California Davis School of Medicine (the United States of America), Johanna Olweus at OUH Radiumhospitalet and Knut E.A. Lundin at OUH Rikshospitalet should be mentioned in particular.

To my lovely friends and dear family, thank you for your unconditional support throughout these years, and for your sincere interest in my research. More importantly, thank you for taking my mind off work when we are together. A special thanks to my mother for your wisdom, encouragement and for answering the phone calls day and night when I wish to talk with you.

To Jostein, my best friend and husband. I am most grateful for your patience and love.

Eva Kristine Klemsdal Henriksen
Oslo, 2017

2 ABBREVIATIONS

AH8.1	8.1 ancestral haplotype
AIH	autoimmune hepatitis
ALD	alcoholic liver disease
AMA	anti-mitochondrial antibodies
ANA	anti-nuclear antibodies
APC	antigen presenting cell
BCR	B-cell receptor
CCA	cholangiocarcinoma
CCL	CC chemokine ligand
CCR	CC chemokine receptor
CD	Crohn's disease
CD1	cluster-of-differentiation 1
CDR	complementarity-determining region
D	diversity
DNA	deoxyribonucleic acid
ERC	endoscopic retrograde cholangiography
FFPE	formalin-fixed, paraffin-embedded
GALT	gut-associated lymphoid tissues
gDNA	genomic DNA
HLA	human leukocyte antigen
IBD	inflammatory bowel disease
IL	interleukin
iNKT cell	invariant natural killer T cell
J	joining
KIR	killer immunoglobulin-like receptor
LD	linkage disequilibrium
LSEC	liver sinusoidal endothelial cell
MAdCAM-1	mucosal addressin cell adhesion molecule 1
MAIT cell	mucosal-associated invariant T cell
MG	myasthenia gravis
MHC	major histocompatibility complex
MR1	MHC-related protein 1
MRC	magnetic resonance cholangiography
mRNA	messenger RNA
MS	multiple sclerosis
NK cell	natural killer cell
pANCA	perinuclear anti-neutrophil cytoplasmic antibodies
PBC	primary biliary cholangitis
PCR	polymerase chain reaction
PSC	primary sclerosing cholangitis
RA	rheumatoid arthritis
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
T1D	type 1 diabetes
TCR	T-cell receptor
TG2	transglutaminase 2
Tregs	regulatory T cells
UC	ulcerative colitis
UDCA	ursodeoxocholic acid
V	variable
VAP-1	vascular adhesion protein 1

3 LIST OF PUBLICATIONS

Paper I

Liaskou E, Henriksen EK, Holm K, Kaveh F, Hamm D, Fear J, Viken MK, Hov JR, Melum E, Robins H, Olweus J*, Karlsen TH*, Hirschfield GM*. **High-throughput T-cell receptor sequencing across chronic liver diseases reveals distinct disease-associated repertoires**. Hepatology 2016;63:1608-1619.

*These authors contributed equally to this work.

Paper II

Henriksen EK, Jørgensen KK, Kaveh F, Holm K, Hamm D, Olweus J, Melum E, Chung BK, Eide TJ, Lundin KEA, Boberg KM, Karlsen TH, Hirschfield GM, Liaskou E. **Gut and liver T-cells of common clonal origin in primary sclerosing cholangitis-inflammatory bowel disease**. Journal of Hepatology 2017;66:116-122.

Paper III

Henriksen EK, Viken MK, Wittig M, Holm K, Folseraas T, Mucha S, Melum E, Hov JR, Lazaridis KN, Juran BD, Chazouillères O, Färkkilä M, Gotthardt DN, Invernizzi P, Carbone M, Hirschfield GM, Rushbrook SM, Goode E, The UK-PSC Consortium, Ponsioen CY, Weersma RK, Eksteen B, Yimam KK, Gordon SC, Goldberg D, Yu L, Bowlus CL, Franke A, Lie BA, Karlsen TH. **HLA haplotypes in primary sclerosing cholangitis patients of admixed and non-European ancestry**. (Manuscript)

4 INTRODUCTION

Primary sclerosing cholangitis (PSC) is a liver disease characterized by inflammation and fibrosis of the bile ducts.[1] The typical patient with PSC is a male in his 30s or 40s with coexistent inflammatory bowel disease (IBD). The etiology and pathogenesis of PSC are largely unknown, but several lines of evidence support the presence of an autoimmune component.[2] To date, there are no effective medical therapies for PSC, and ultimately the disease will progress to liver cirrhosis with the need for liver transplantation. For years, PSC has been among the leading indications for liver transplantation in the Nordic countries.[3] Better understanding of the pathogenic mechanisms in PSC will likely enable the development of medical therapies to treat the underlying cause of disease or delay further disease progression.

Reports of associations between PSC and numerous loci harboring genes that encode proteins affecting T-cell biology (including human leukocyte antigen [HLA] molecules) strongly suggest T cells have a role in PSC pathogenesis.[4] This is further supported by the observation that the majority of infiltrating lymphocytes in PSC-affected livers are $CD4^+$ and $CD8^+$ T cells.[5, 6] The overall focus of the present thesis was therefore to further characterize the T-cell repertoire of patients with PSC (specifically, their T-cell receptors) and investigate whether studying populations of admixed or multi-ethnic ancestry might aid in fine mapping the HLA association in PSC.

4.1 The adaptive immune system and mechanisms of autoimmunity

4.1.1 A brief introduction to the adaptive immune system

The immune system

The immune system protects the body from pathogens and abnormal cells. Traditionally, the immune system is divided into innate and adaptive components with distinct roles and functions (**Figure 1**).[7] The innate immune system confers immediate response to infections. It comprises physical barriers (such as the epithelium of the gastrointestinal tract), innate immune cells (such as dendritic cells and macrophages) and circulating plasma proteins that constitute the complement system. The innate immune cells carry a set of invariant receptors called pattern recognition receptors that recognize conserved molecular patterns derived from microbial pathogens and molecules released by stressed cells.[8] This allows them to non-specifically eliminate the pathogens. Unlike the innate immune system, the adaptive immune system confers a relatively slow response as it relies on the proliferation and differentiation of T and B lymphocytes (also called T and B cells).[7] Once established, the adaptive immune system is highly specific and efficient in clearing the infection. T and B cells can differentiate into long-lived memory cells that will rapidly respond to a second encounter of the same pathogen, creating immunological memory.

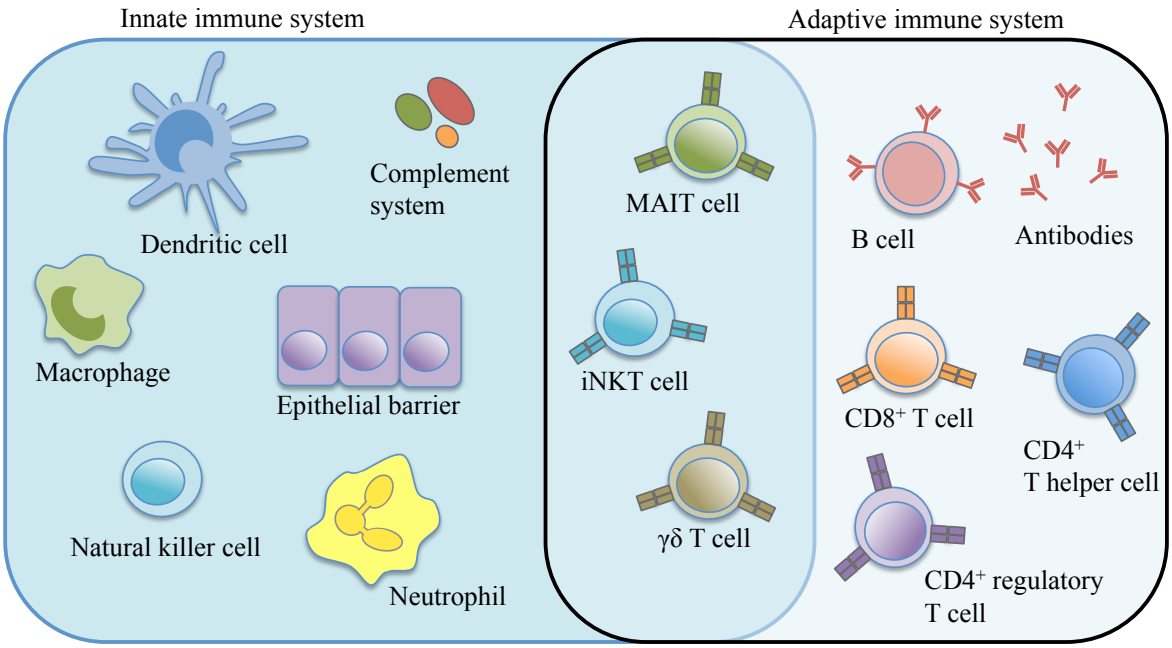

Figure 1. The immune system. The immune system is divided into innate and adaptive components. The innate immune system acts as the first line of defense. It comprises epithelial barriers (such as the skin and in the gastrointestinal tract), a set of plasma proteins constituting the complement system, and various innate immune cells (some of which are depicted in this figure). Unlike the innate immune system, the adaptive immune system confers a relatively slow response, but is highly specific and has memory. The adaptive immune system comprises $CD4^+$ (helper and regulatory) T cells, $CD8^+$ T cells, B cells and antibodies. Innate-like T cells such as the mucosal-associated invariant T (MAIT) cells, invariant natural killer T (iNKT) cells and γδ T cells bridge the innate and adaptive immune system. Inspired by reference[9].

Antigen presentation by major histocompatibility complex molecules

An antigen is a molecule capable of inducing an immune response by binding antibodies or antigen-specific receptors called T-cell receptors (TCRs) and B-cell receptors (BCRs) on the cell surfaces of T and B cells, respectively.[7] The receptors are generated by the random recombination of receptor-coding gene segments, ensuring an enormously diverse TCR and BCR repertoire (in section 4.1.2, the recombination process in T cells is described in detail). An epitope is the specific structure of an antigen that is recognized by the TCR or by the BCR. The TCR differs from the BCR in an important way: it recognizes only short fragments of antigens in complex with molecules on the surfaces of other cells, whereas the BCR recognizes both soluble and membrane-bound intact (*i.e.* non-fragmented) antigens.[10, 11] All nucleated cells express major histocompatibility complex (MHC) class I molecules.[12] These molecules present peptides derived from intracellular proteins to TCRs of $CD8^+$ T cells (**Figure 2**). The TCRs of $CD4^+$ T cells recognize peptides in complex with MHC class II molecules, which are mainly expressed by professional antigen presenting cells (APCs) such as B cells, dendritic cells and macrophages. During inflammation, the expression of MHC class II molecules can also be induced and upregulated in various non-immune cells.[13, 14] The MHC class II molecules present peptides derived from proteins that have been captured by endocytosis from the surroundings.[12] By a mechanism known as cross-presentation, certain APCs are also able to present extracellular-derived peptides on MHC class I molecules to the TCRs of $CD8^+$ T cells.[15]

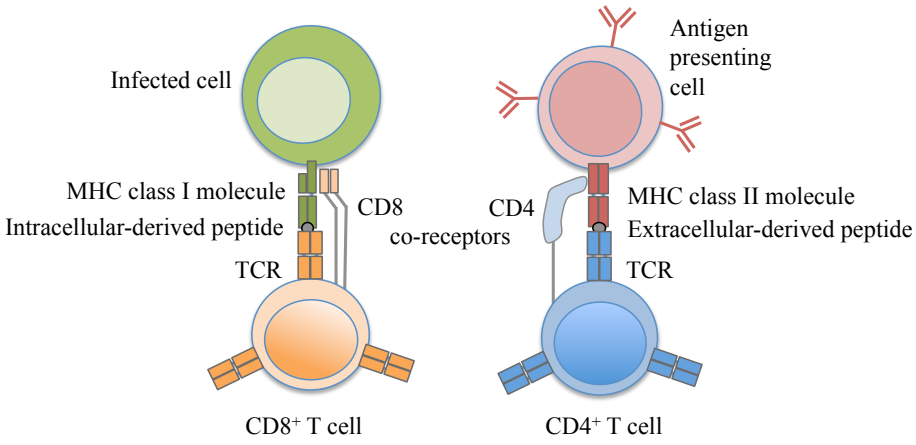

Figure 2. Antigen presentation by major histocompatibility complex (MHC) molecules. MHC class I molecules present intracellular-derived peptides (including viral peptides) to the T-cell receptors (TCRs) of $CD8^+$ T cells. MHC class II molecules present extracellular-derived peptides to the TCRs of $CD4^+$ T cells. All nucleated cells express MHC class I molecules. The MHC class II molecules are expressed by specialized antigen presenting cells (APCs) such as B cells (depicted in this figure), dendritic cells and macrophages.

The T-cell repertoire also comprises several T cells that do not recognize the peptide-MHC complex: mucosal-associated invariant T (MAIT) cells recognize metabolites in complex with the MHC class I-like MHC-related protein 1 (MR1) molecule, and invariant natural killer T (iNKT) cells, germline-encoded, mycolyl lipid-reactive (GEM) T cells and LDN5-like T cells recognize glycolipids presented by MHC class I-like cluster-of-differentiation 1 (CD1) molecules.[16] These 'unconventional' T cells represent a bridge between the innate and the adaptive immune system.

T cells

A mature lymphocyte that has not yet been activated by its corresponding antigen is called a naïve T or B cell. The naïve T cells circulate mainly through the secondary lymphoid tissues and the peripheral blood.[17] Upon an infection, activated dendritic cells that reside in the infected peripheral tissue will engulf cellular debris and pathogens. The dendritic cells will upregulate expression of co-stimulatory molecules and migrate to regional secondary lymphoid tissues where they will present the processed antigens on MHC class I and class II molecules to naïve $CD8^+$ and $CD4^+$ T cells, respectively. Following recognition of a peptide and additional co-stimulatory signals provided by the dendritic cell, the activated naïve T cells will proliferate and differentiate into effector T cells and long-lived memory T cells (**Figure 3**). The collection of daughter cells derived from each naïve T cell is called a T-cell clone, with all cells of that clone carrying identical TCRs (*i.e.* having identical antigenic specificity). Activated T cells will express a distinct set of surface molecules (such as adhesion molecules and chemokine receptors) that will allow the T cells to enter the peripheral tissue where the dendritic cell had encountered the antigen. Upon recognition of their corresponding antigenic peptide in complex with MHC class II molecules on the surface of infected cells, effector $CD8^+$ T cells (also called cytotoxic T cells) will kill the infected cells by binding death receptors on their cell surface and by releasing cytotoxic effector proteins that ultimately will induce apoptosis (programmed cell death) of the target cell.[18] Effector $CD4^+$ T cells will interact with other immune cells and modulate immune responses through direct cell-cell interaction and the release of cytokines.[19] For instance, effector $CD4^+$ T cells activate B cells that respond to the same antigens (see below and **Figure 3**), provide help to establish $CD8^+$ T-cell memory and induce activation of macrophages. Several subtypes of effector $CD4^+$ T cells with various functions have been identified, such as T helper (Th) 1, Th2 and Th17 cells. These $CD4^+$ T-cell subsets are distinguished by the specific transcription factors, surface molecules and cytokines they produce.[20] In addition to the conventional T helper cells, the $CD4^+$ T cells comprise regulatory T cells (Tregs).[19] Using various molecular mechanisms, Tregs suppress T-cell responses against self-antigens (also called autoantigens).[21, 22]

B cells

Naïve B cells internalize antigens that bind to their BCR.[11] Following processing of the antigen, peptides are bound to MHC class II molecules and presented on the cell surface of the B cell. Interaction of the B cell with an antigen-specific $CD4^+$ T cell will promote proliferation and differentiation of the B cell into antibody-producing cells (plasmablasts and plasma cells) or memory B cells, as illustrated in **Figure 3**. During the proliferation process, the gene segments encoding the BCRs will typically accumulate mutations, resulting in a gradual increased affinity for the antigen for the surviving B cells. This is called 'affinity maturation', and is a process unique to the B cells.

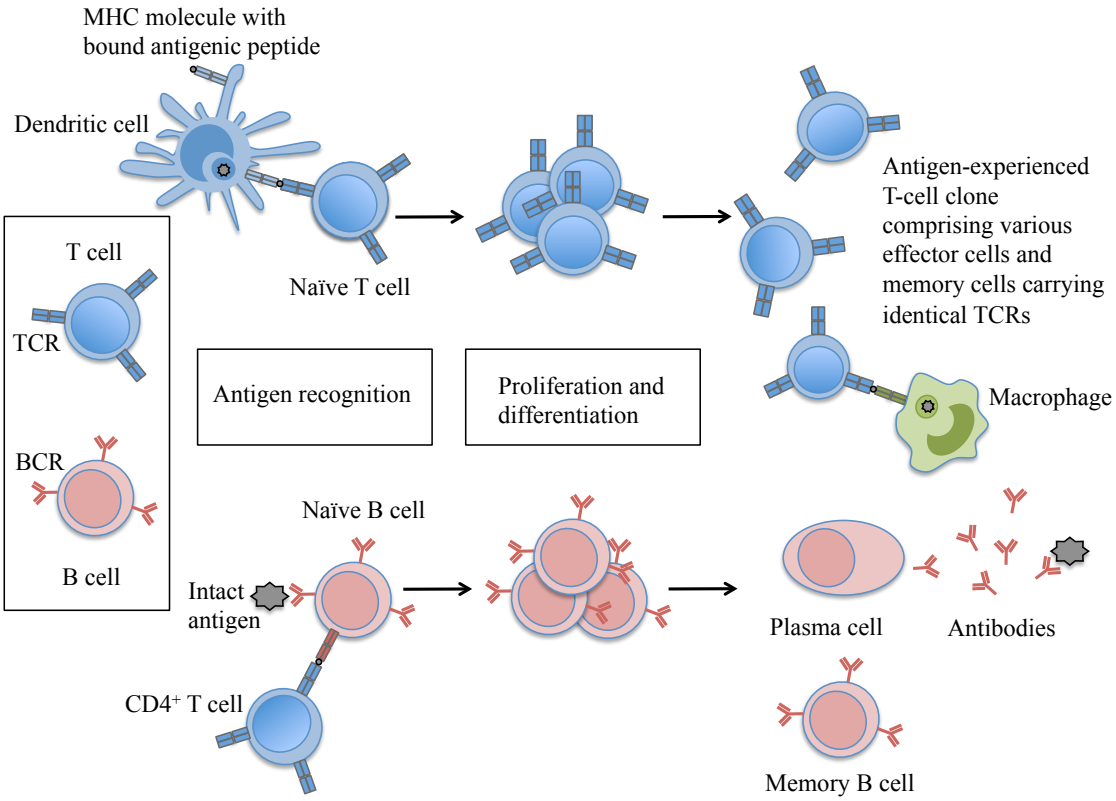

Figure 3. Activation, clonal expansion and differentiation of T and B cells. Activated dendritic cells migrate to secondary lymphoid tissues were they present antigenic peptides on major histocompatibility complex (MHC) class II and class I molecules to naïve $CD4^+$ (shown in this figure) and $CD8^+$ T cells (not shown), respectively. Binding of the T-cell receptor (TCR) to the peptide-MHC complex and the binding of co-stimulatory molecules that are present on the cell surface of the dendritic cell will induce activation of the naïve T cell. The activated T cell will undergo clonal expansion (*i.e.* proliferation) and will differentiate into a collection of daughter cells called a clone that comprises various effector and memory T cells with identical antigenic specificity. Activated T cells will migrate to the peripheral tissue where the dendritic cell had encountered the antigen. Here, the effector $CD8^+$ T cells will kill infected cells that present their corresponding antigenic peptide (not shown). The $CD8^+$ T cells induce apoptosis by binding to death receptors on the cell surface of the infected cell and by releasing cytotoxic proteins. The various subtypes of effector $CD4^+$ T cells will interact with other immune cells (such as $CD8^+$ T cells, B cells [shown] and macrophages [shown]) and modulate immune responses through the release of cytokines and direct cell-cell interactions. B cells recognize both soluble antigens (shown) and intact antigens that are bound to the membranes of macrophages and dendritic cells (not shown). Naïve B cells internalize antigens that bind to their B-cell receptor (BCR). Following processing of the antigen, peptides are bound to MHC class II molecules and presented on their cell surface to $CD4^+$ T cells. Upon recognition of the peptide by an antigen-specific $CD4^+$ T cell, the B cell will proliferate and differentiate into antibody-secreting cells (plasmablasts [not shown] and plasma cells [shown]) and memory B cells (shown).

Antibodies

Antibodies are soluble forms of the BCR that are secreted from plasmablasts and plasma cells.[11] Binding of antibodies to bacterial toxins or viruses will neutralize the pathogens, as they will no longer access host cells.[23] Antibodies binding to antigens on a bacterial surface promote binding of the complement system and phagocytosis of the bacteria by innate immune cells.

Central and peripheral tolerance

The immune system balances the defense against infections and cancers with tolerance to self, food proteins and commensal microbiota. Self-reactive T cells are killed or inactivated by tolerance mechanisms in the thymus (central tolerance) or in the periphery (peripheral tolerance).[24] Precursor T cells (also called thymocytes) that express TCRs with high affinity for self-peptide-MHC complexes are induced to undergo apoptosis in the thymus.[25] Essential to this process, which is known as negative selection, is the expression and presentation of peripheral, tissue-restricted self-antigens on MHC class I and class II molecules by thymic stromal cells. Self-reactive thymocytes can also be induced to differentiate into Tregs. Peripheral tolerance ensure that self-reactive T cells that escape central tolerance are killed or inactivated.[24] Tregs use a number of different mechanisms to suppress peripheral T-cell responses against self, targeting both self-reactive T cells and APCs.[21, 22] For instance, interactions between Tregs and dendritic cells lead to downregulation of the co-stimulatory molecules needed for T-cell activation.[26] Binding of TCRs to peptide-MHC complexes in the absence of co-stimulatory signals will not induce T-cell activation but rather induce T-cell anergy (a state of long-term hyporesponsiveness). If self-reactive T cells escape both central and peripheral tolerance, immune responses against healthy cells and tissues develop, a condition called autoimmunity.

Box 1. Immunology glossary. Definitions of selected immunology terms used in this thesis.

Adaptive immune system	Highly specific component of the immune system comprising T cells, B cells and antibodies. Confers a relatively slow response but creates immunological memory.
Antigen presenting cell (APC)	Cell that expresses MHC class II molecules capable of activating $CD4^+$ T cells. Expresses also MHC class I molecules capable of activating $CD8^+$ T cells.
Autoantibody	Antibodies reactive against self-antigens.
Antigen	Any molecule that can induce immune responses by binding specifically to an antibody, BCR or TCR.
Autoimmune disease	Condition arising from sustained immune responses against the body's own tissues and organs.
B-cell receptor (BCR)	The receptor of B cells responsible for recognizing intact antigens.
Central tolerance	Tolerance mechanisms that occur in the thymus before the release of mature naïve T cells to the periphery.
Human leukocyte antigen (HLA) molecules	The human equivalent to the MHC molecules.
Innate immune system	The first line of defense comprising physical barriers, the compliment system and various innate immune cells.
Lymphocytes	White blood cells, which include T cells and B cells of the adaptive immune system and the natural killer (NK) cells of the innate immune system.
Major histocompatibility complex (MHC) molecules	Molecules responsible for the presentation of antigenic peptides to T cells.
Naïve lymphocyte	A mature T or B cell that has not yet been activated by its corresponding antigen.
Peripheral tolerance	Tolerance mechanisms that occur outside the thymus.
Regulatory T cell (Treg)	T cell that suppresses the functional activity of other (self-reactive) T cells and of APCs.
T-cell clone	Collection of T cells that carry identical TCRs, deriving from a single naïve T cell.
T-cell receptor (TCR)	The receptor of T cells responsible for recognizing fragments of antigens.
Thymocyte	Developing T cell in the thymus.
V(D)J recombination	Process in which precursor T cells and B cells randomly rearrange gene segments that will constitute their TCRs and BCRs.

4.1.2 The T-cell receptor

TCR structure and function

The majority of peripheral T cells express a heterodimeric TCR comprising an α-chain and a β-chain that are linked by a disulfide bond.[10, 27] Such T cells are broadly classified as αβ T cells, and they recognize peptides presented by MHC molecules (**Figure 4**), notably with the exception of the innate-like 'unconventional' MAIT cells, iNKT cells, GEM T cells and LDN5-like T cells.[16] The remaining T cells express a TCR comprising a γ-chain and a δ-chain. The γδ T cells do not commonly recognize peptides presented by MHC molecules but rather recognize various other antigenic targets (such as lipids presented by CD1 molecules), and are therefore grouped as being 'unconventional' T cells.[10, 16, 28] In the remaining pages of this section and throughout this thesis, the focus will be on describing the αβ T cells and their receptor.

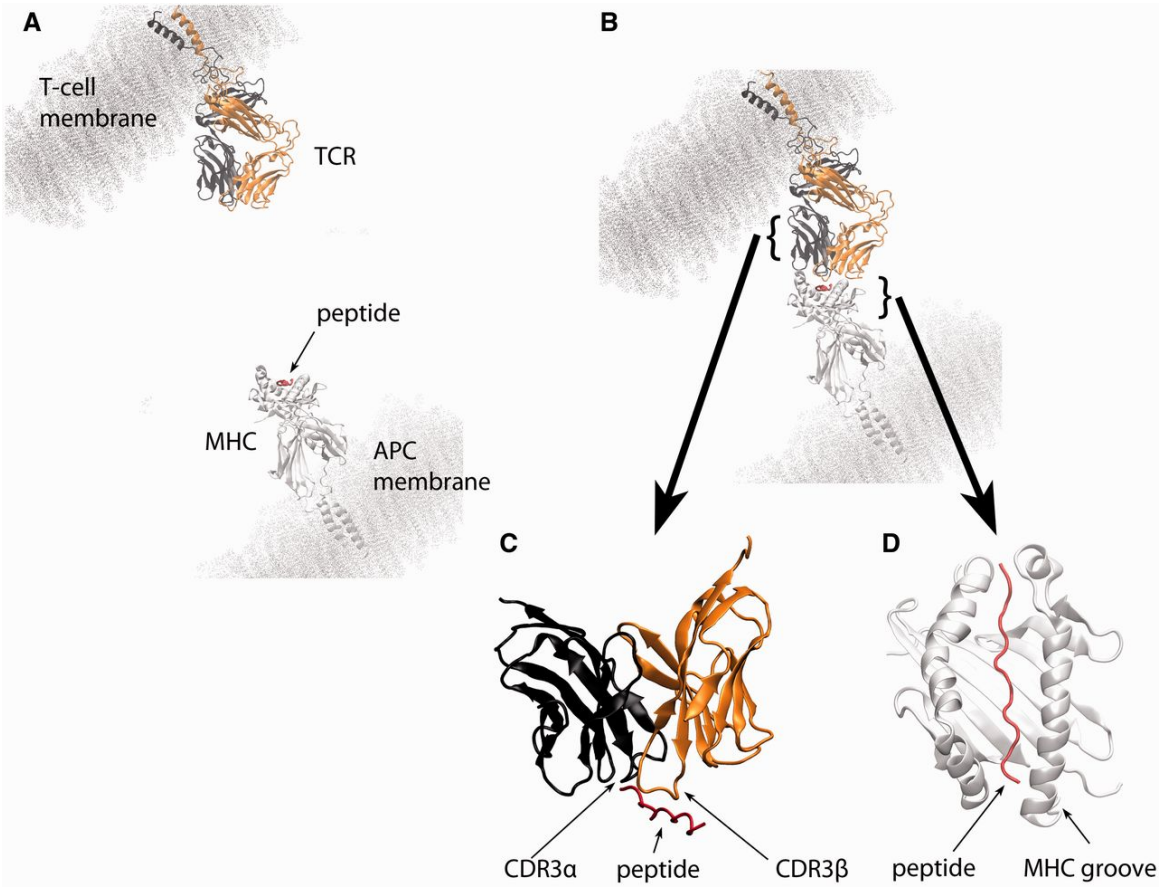

Figure 4. Interaction between an αβ T-cell receptor (TCR), a peptide and a major histocompatibility complex (MHC) class II molecule. Black cartoon: TCR α-chain; orange cartoon: TCR β-chain; white cartoon: MHC; grey dots: cell membrane. (A) The TCR is expressed on the T-cell surface. The MHC class II molecule presents a peptide on the cell surface of an antigen-presenting cell (APC). (B) Bound configuration of the TCR-peptide-MHC complex. (C) Zoomed and rotated view of the TCR variable region. The complementarity-determining region 3 (CDR3) loops of the α-chain and β-chain (labeled) make the main interaction with the peptide. (D) Zoomed and rotated view of the peptide bound to the peptide-binding groove of the MHC class II molecule. Reprinted from reference[29] by permission of Oxford University Press.

Both the α-chain and the β-chain have an extracellular variable region that includes three short hairpin loops called complementarity-determining regions (CDR1–3).[10] Together, the six CDRs constitute the antigen-binding site of the TCR: the CDR1 and CDR2 loops mainly contact the MHC molecule, and the CDR3 loops mainly contact the antigen,[27] as shown in **Figure 4**. The variable region is linked to an extracellular, invariant constant region, which is anchored to the T-cell plasma membrane by a hydrophobic transmembrane region that ends in a short cytoplasmic tail. In the membrane, the TCR associates with invariant accessory proteins (*i.e.* the CD3 complex and ζ chains) in order to make a functional TCR complex,[30] as illustrated in **Figure 5**.

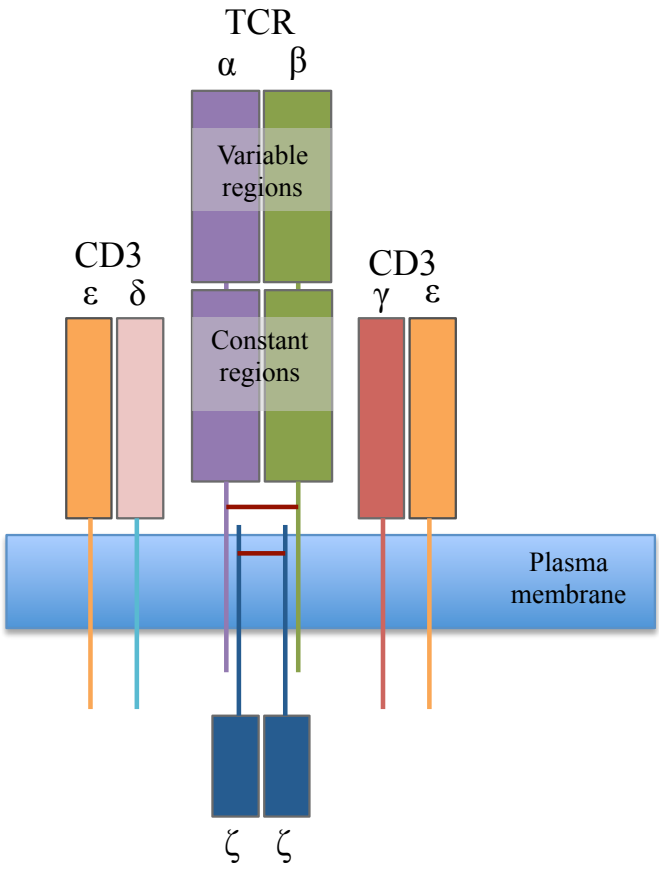

Figure 5. A schematic drawing of the αβ T-cell receptor (TCR) and invariant accessory proteins. The functional TCR complex comprises a TCR heterodimer, the CD3 complex and a ζ homodimer. The TCR heterodimer comprises an α-chain and a β-chain that are linked by a disulfide bond. Both the α-chain and the β-chain have an extracellular variable region, an extracellular constant region, a transmembrane region and a cytoplasmic tail. Together, the variable regions make up the antigen-binding site. The accessory proteins (*i.e.* the CD3 complex and ζ chains) are required for expression of the TCR at the cell surface and for signaling following binding of the TCR to the peptide-MHC complex.

TCR genes

The α-chain and β-chain of the human TCR are encoded by the TRA locus on chromosome 14 and TRB locus on chromosome 7, respectively.[31] The TCR variable region is encoded by two or three different gene segments: variable (V), diversity (D) and joining (J). The germline TRA locus contains multiple copies of the Vα and Jα gene segments, but do not contain any D genes. The germline TRB locus contains two copies of the Dβ gene segment in addition to multiple copies of the Vβ and Jβ gene segments. The official IMGT (the international ImMunoGeneTics information system) TCR gene and allele nomenclature[32, 33] is presented in **Figure 6**.

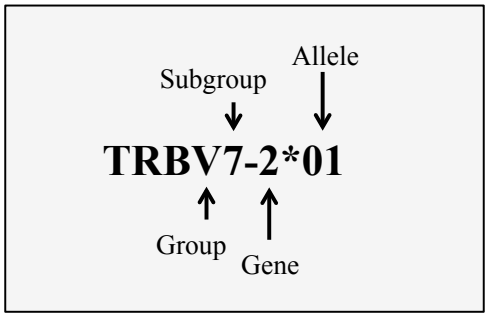

Figure 6. T-cell receptor gene and allele nomenclature. The figure presents the official IMGT (the international ImMunoGeneTics information system) standardized gene and allele nomenclature for T-cell receptors.[32, 33]

V(D)J recombination and allelic exclusion

Stem cells migrate from the bone marrow to the thymus, where they differentiate into mature naïve T cells with highly diverse TCRs.[34] The maturation process is carefully orchestrated to ensure that functional, self-tolerant T cells enter the periphery. Collectively, more than 95% of thymocytes are subject to death by neglect as their TCRs have no or too low affinity for self-peptide-MHC complexes, or undergo apoptosis due to either failure to generate a functional TCR β-chain or high-affinity interactions of their TCRs with self-peptides (as mentioned in section 4.1.1, the latter process is called negative selection).[25, 35, 36]

During T-cell development in the thymus, the TRA and TRB loci are rearranged in a process called V(D)J recombination: one V gene, one D gene for the β-chain only and one J gene are joined to form a contiguous sequence (see **Figure 7**).[31] This process also involves the deletion and insertion of nucleotides in the junctions between the gene segments. The CDR1 and CDR2 loops are encoded within the germline V gene, whereas the CDR3 loop is encoded by the V(D)J junction.[10] The random combination of gene segments encoding the variable regions of each TCR, as well as the random insertion and deletion of nucleotides in the sequence encoding the CDR3 loop, ensures the development of T cells with highly diverse antigenic specificities. The theoretical number of different TCRs that can be produced during V(D)J recombination is estimated to more than 10^{15} in humans.[37-39]

The V(D)J recombination process is carefully coordinated in the thymocyte, with the TRB locus being rearranged before the TRA locus.[34] A Dβ gene rearranges to a Jβ gene on both the maternally- and the paternally-derived chromosome. Following this, a Vβ gene rearranges to DβJβ on just a single chromosome. Most rearrangements result in a non-productive, rearranged TRB locus, *i.e.* the Vβ and Jβ gene segments are not in the same reading frame, a stop codon is created or a pseudogenic gene is used.[40] If the first VβDβJβ recombination results in a productive, rearranged TRB locus, the β-chain is expressed on the cell surface together with an invariant pre-Tα (pTα) chain, the CD3 complex and ζ chains.[34] The complex is a functional pre-TCR, and induces signaling to stop further rearrangement of the TRB locus on the second chromosome. This ensures only one of the two TRB loci is expressed in each T cell, a phenomenon known as allelic exclusion. The pre-TCR induces rapid cell proliferation followed by expression of both co-receptor CD4 and

CD8. This results in several double-positive thymocytes carrying identical pre-TCRs. Once the cells stop dividing, the TRA loci on both the maternally- and the paternally-derived chromosome will rearrange simultaneously and independently in each cell.[41] Due to the high number of Vα and Jα genes, the VαJα recombination can be repeated in several cycles and is almost always successful on at least one chromosome. Once a productive, rearranged TRA locus is generated, the α-chain is co-expressed with the β-chain on the cell surface. VαJα recombination continues on both chromosomes until the thymocyte undergoes positive selection, a process during which thymocytes that express TCRs with low or intermediate affinity for self-peptide-MHC complexes are induced to differentiate into mature T cells, expressing either the CD4 or the CD8 co-receptor.[25, 41] Thymocytes that fail to undergo positive selection will die by neglect. Thymocytes that express TCRs with high affinity for self-peptide-MHC complexes are eliminated by negative selection or are induced to differentiate into Tregs.

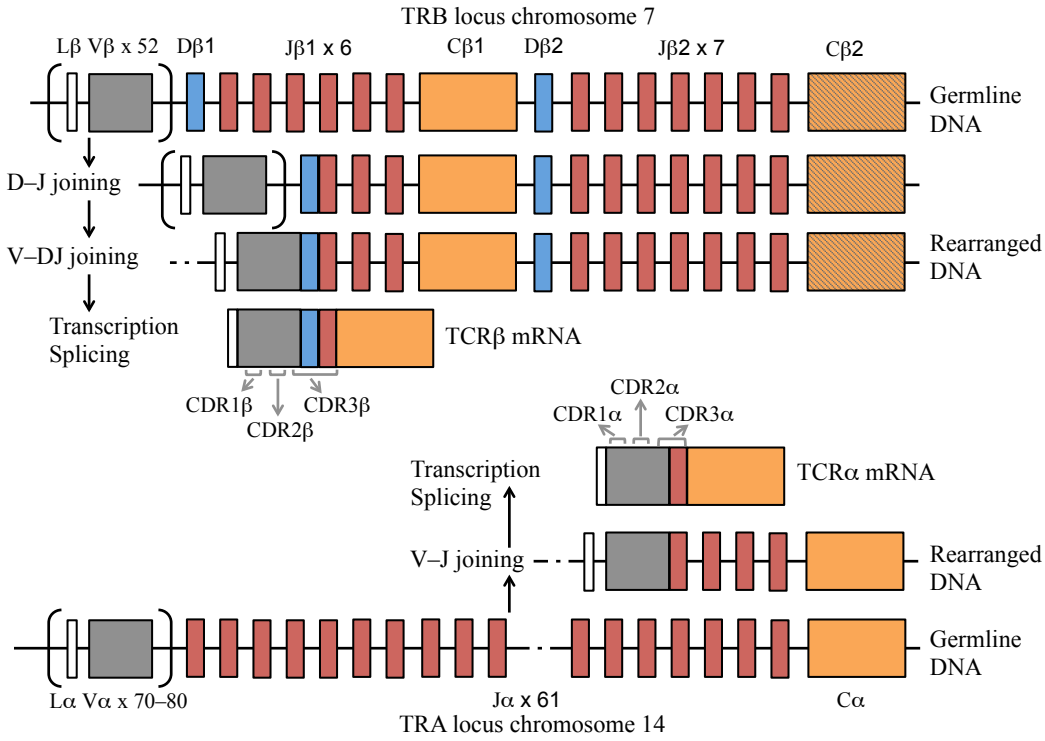

Figure 7. V(D)J recombination. During T-cell development, the loci encoding the T-cell receptor α-chain and β-chain are rearranged in the process called V(D)J recombination. At the TRB locus, one diversity (D) β gene rearranges to a joining (J) β gene, then one variable (V) β gene rearranges to the recombined DβJβ genes. At the TRA locus, one Vα gene rearranges to a Jα gene. An exon encoding a leader (L) sequence is located upstream each V gene. The V(D)J recombination process also involves the deletion and insertion of nucleotides at the Vβ-Dβ, Dβ-Jβ and Vα-Jα junctions (not shown). Following transcription, the sequence between the recombined V(D)J genes and the gene segment encoding the constant region is removed by splicing. While the complementarity-determining region (CDR) 1 and CDR2 loops are encoded within the germline V gene, the CDR3 loop is encoded by the V(D)J junction.

Since there is no allelic exclusion of the TRA locus during VαJα recombination, about 10-20% of all mature T cells express two distinct, productively rearranged α-chains.[42-44] VαJα recombination ceases once positive selection occurs, and most mature T cells carry only a single functional TCR

on their cell surface, existing in numerous copies at the cell surface.[41, 45] Naïve T cells can share the same rearranged β-chain nucleotide sequence but have different α-chain sequences[46] (potentially able to recognize different antigenic triggers) as a result of the cell proliferation that is induced by the pre-TCR expression, followed by independent TRA gene rearrangement in each individual thymocyte[34, 41] (as illustrated in **Figure 8**). In the antigen-experienced T-cell repertoire, the rearranged nucleotide sequence of the β-chain is generally unique to each T-cell clone.[42, 46]

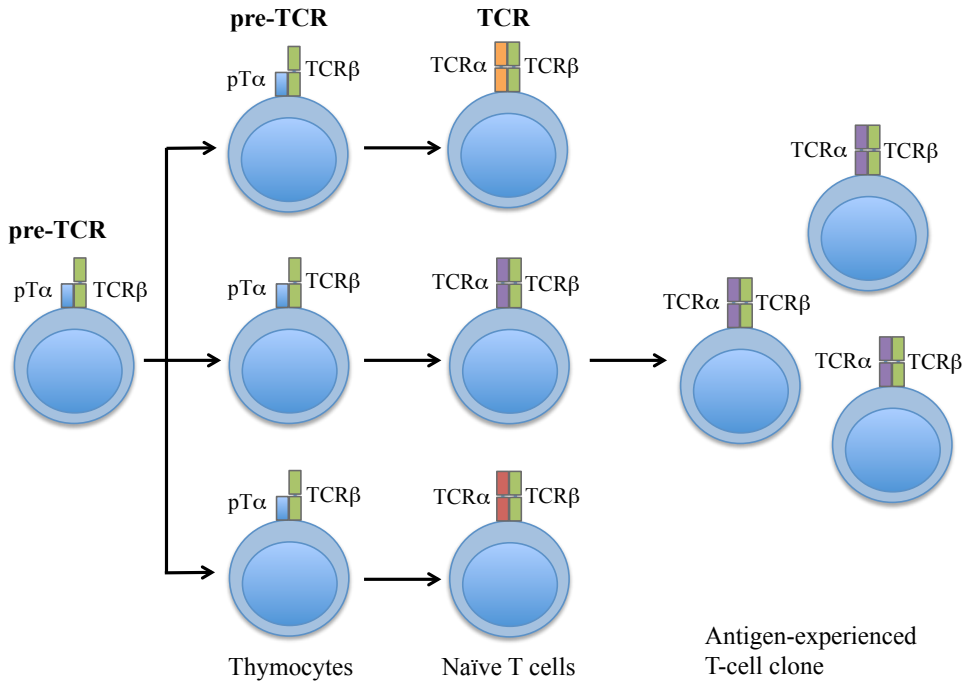

Figure 8. Naïve T cells can share the same T-cell receptor (TCR) β-chain but express different TCR α-chains. Upon successful recombination of a TRB locus, the β-chain is expressed on the cell surface together with the invariant pre-Tα (pTα) chain and accessory proteins to form the pre-TCR. This induces rapid cell proliferation. Once the thymocytes stop dividing, the TRA loci are rearranged independently in each individual cell, thus ultimately resulting in a set of naïve T cells that share the same rearranged TRB nucleotide sequence but have different α-chain sequences (potentially able to recognize different antigenic triggers). Among the antigen-experienced T cells, each uniquely rearranged TRB nucleotide sequence is generally detected in a single T-cell clone and not in other T-cell clones. Hence, the rearranged nucleotide sequence of the β-chain provides a unique molecular tag for each antigen-experienced T-cell clone.

Cross-reactivity

Accumulating evidence indicates that T cells can exhibit cross-reactivity: a given TCR is able to recognize more than one single peptide–HLA complex.[47] For instance, $CD4^+$ memory T cells that recognize antigenic peptides of human immunodeficiency virus type 1 (HIV-1), cytomegalovirus and herpes simplex virus were detected in healthy individuals that had never been infected with these viruses.[48] It has further been shown that HIV-1 reactive T cells are able to recognize bacterial peptides in the gut and soil, other bacterial peptides and peptides from ocean algae and plants. A cross-reactive TCR seems to recognize peptides that are biochemically and structurally related to each other.[49]

4.1.3 Human leukocyte antigens and linkage disequilibrium

Structure of MHC molecules

The MHC class I and II molecules are heterodimeric molecules that are closely related in overall structure, but differ in their subunit compositions,[50] as illustrated in **Figure 9**. The MHC class I molecules consist of an α-chain linked to a smaller chain called the β_2-microglobulin. The MHC class II molecules comprise an α-chain and a β-chain. The peptide-binding groove of MHC class I molecules is formed by two domains of the α-chain, whereas both the α-chain and the β-chain contribute to the peptide-binding groove of MHC class II molecules. In humans, the MHC molecules are called HLA molecules.

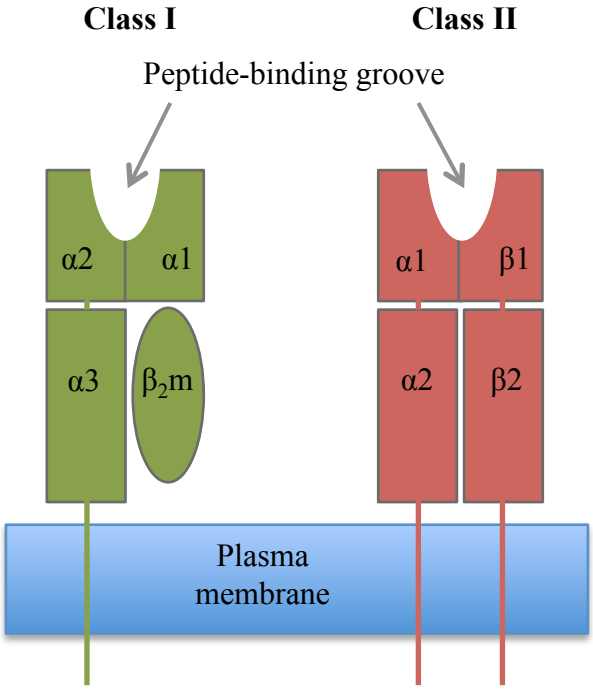

Figure 9. A schematic drawing of major histocompatibility complex (MHC) class I and class II molecules. The MHC class I molecules consist of an α-chain and a β_2-microglobulin (β_2m). Two domains of the α-chain (*i.e.* α1 and α2) form the peptide-binding groove of MHC class I molecules. The MHC class II molecules consist of an α-chain and a β-chain, each comprising a peptide-binding domain (α1 and β1). In humans, the MHC molecules are called HLA molecules.

HLA genes

The genes encoding the HLA class I and II molecules are located on the short arm of chromosome 6 (**Figure 10**), in a region with approximately 260 genes (where about 30% encode proteins of immunological functions) spanning approximately 4 million base pairs.[51] The exception is the β_2-microglobulin, which is encoded on chromosome 15.[50] The α-chains of the HLA class I molecules are encoded by three classical genes, namely *HLA-A*, *HLA-B* and *HLA-C*.[52] There are also three pairs of genes encoding the α-chain and β-chain of the classical HLA class II molecules, called HLA-DR, HLA-DQ and HLA-DP. In some individuals, more than one gene encodes the β-chain of the HLA-DR molecules, namely *HLA-DRB1* (ubiquitously present) and the *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5* genes (variably present). The classical HLA genes are highly polymorphic, with numerous alleles identified for each gene (with the exception of the gene encoding the α-chain

of HLA-DR, *i.e. HLA-DRA*).[51] For instance, *HLA-B* is the most polymorphic gene in the human genome, with more than 4500 known alleles as of December 2016.[53] The α-chain and β-chain of HLA class II molecules can either be encoded in *cis* (on the same chromosome) or in *trans* (on homologous chromosomes),[54] increasing even further the functional diversity of the HLA class II molecules. The current nomenclature for the classical HLA genes and their alleles[55] is depicted in **Figure 11**.

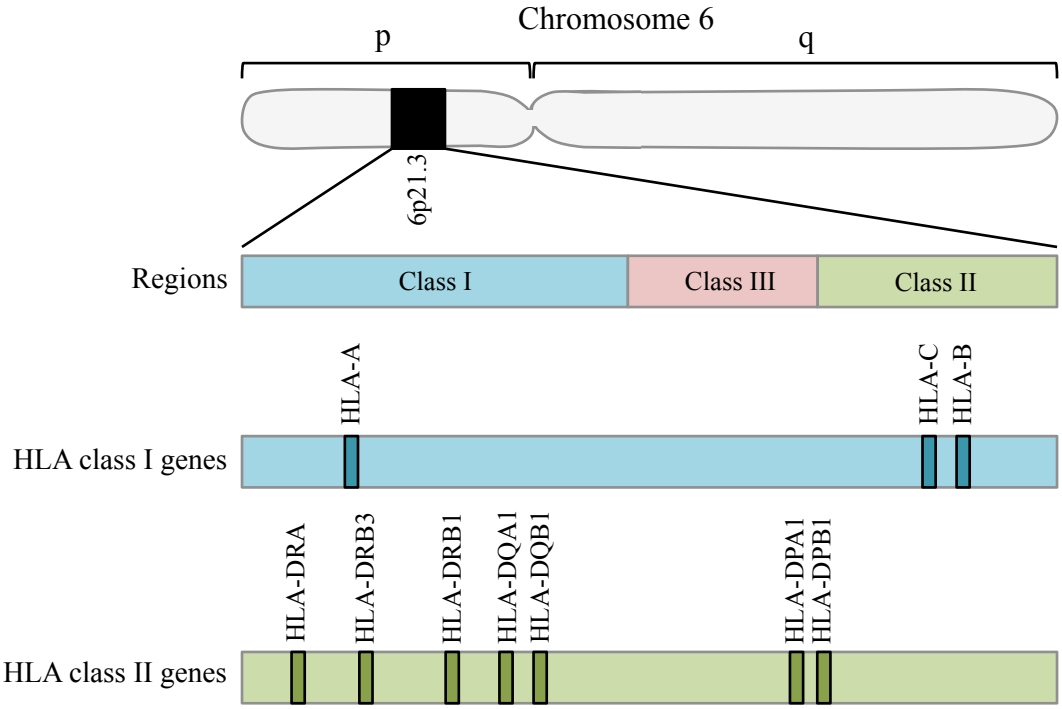

Figure 10. Location and organization of the human leukocyte antigen (HLA) complex. The HLA complex is located on chromosome 6 and is conventionally divided into three regions: class I, class II and class III. Each region contains numerous genes, some of which are depicted in the figure. The class III genes (not shown) differs from the class I and class II genes both structurally and functionally. The figure is inspired by reference[50].

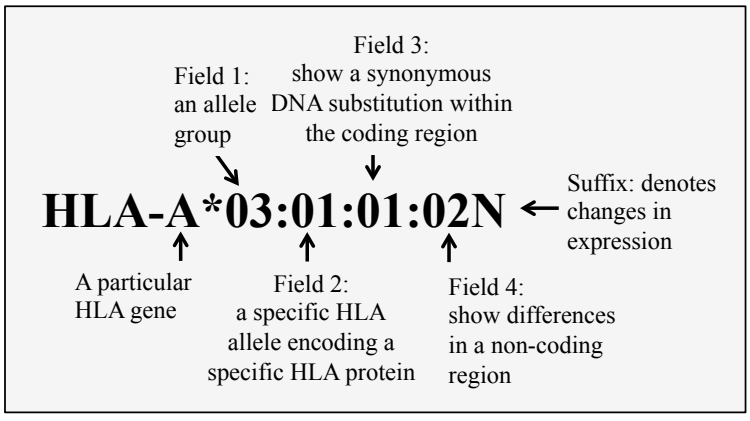

Figure 11. Human leukocyte antigen (HLA) allele nomenclature. The figure is inspired by a figure at the website http://hla.alleles.org/nomenclature/naming.html

Haplotype and linkage disequilibrium

The specific combination of alleles at two or more loci that are present on the same chromosome constitutes a haplotype. Sometimes, certain haplotypes are observed more or less frequently than expected by the individual allele frequencies in the population (**Figure 12**). This non-random association of alleles at different loci is known as linkage disequilibrium (LD).[56] The HLA complex is characterized by a high degree of LD, giving rise to highly conserved HLA haplotypes, some even spanning the HLA complex from the HLA class I genes to the HLA class II genes.[51] For instance, a very common haplotype in populations of Northern-European descent is the 8.1 ancestral haplotype (AH8.1; HLA-A*01:01-C*07:01-B*08:01-DRB3*01:01-DRB1*03:01-DQA1*05:01-DQB1*02:01), typically found at frequencies of about 10%.[57, 58] The AH8.1 is one of the longest haplotypes in the human genome, with 311 annotated loci and spanning 4.7 million nucleotides.[59]

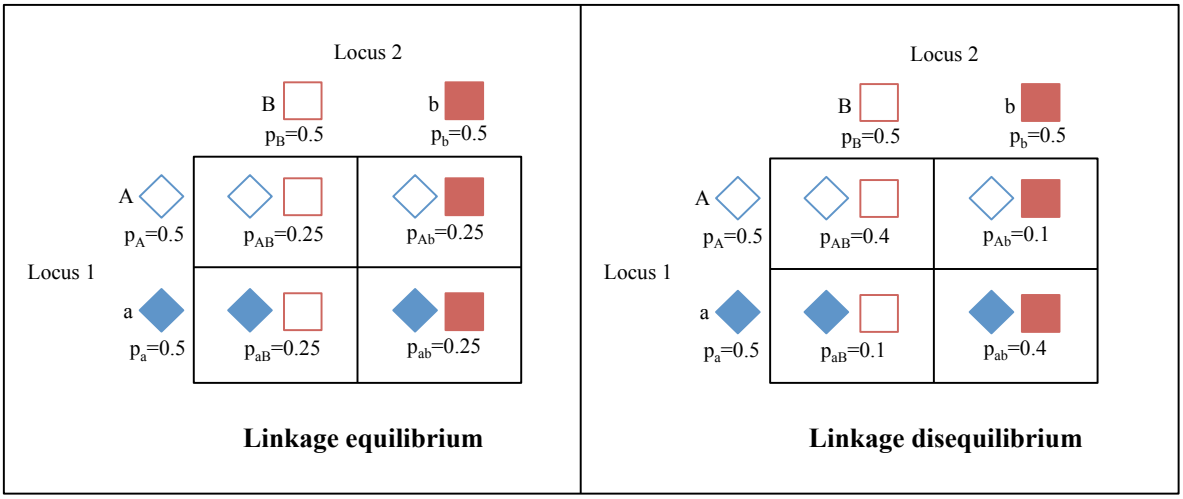

Figure 12. Linkage disequilibrium. The observed population frequencies of alleles at locus 1 (*i.e.* A and a) and at locus 2 (*i.e.* allele B and b) are each 50%. Left panel: if the alleles at locus 1 and 2 are completely independent, the alleles are said to be in linkage equilibrium and the frequency of each of the four haplotypes (*i.e.* AB, aB, Ab and ab) will be p = 0.5 x 0.5 = 0.25. Right panel: if the alleles are in linkage disequilibrium, the various haplotypes are observed more or less often than expected by the individual allele frequencies. The figure is inspired by reference[60].

Box 2. Genetics glossary. Definitions of selected genetic terms used in this thesis.	
Allele	One of two or more variants of a single gene that reside at the same position (i.e. locus) on homologous chromosomes.
Exon	The coding region within a gene.
Gene	A DNA nucleotide sequence that contains biological information, encoding an RNA molecule and/or amino acid sequence (*i.e.* protein).
Gene expression	The process in which a functional gene product is generated from the information encoded by a gene. The first step is transcription. For protein-coding genes, the second step is translation.
Genetic code	The rules that determines which triplet of nucleotides codes for which amino acid during the process of translation.
Genetic variant	An alteration in the DNA nucleotide sequence. Often refers to a single nucleotide polymorphism (SNP).
Genotype	The genetic combination of two alleles at a single locus. Also refers to the entire set of genes in an organism.
Haplotype	The distinct combination of alleles at two or more loci that are present on a single chromosome.
Heterozygous	Carrying two different alleles for a particular gene.
Homologous chromosomes	The two copies of a chromosome in a diploid cell, one inherited from the mother and one inherited from the father.
Homozygous	Carrying two copies of the same allele for a particular gene.
Intron	The non-coding region within a gene. Introns are transcribed together with the exons but are later removed during splicing.
Linkage disequilibrium (LD)	The phenomenon in which alleles at different loci occur together on a single chromosome at a higher frequency in a population than expected by the individual allele frequencies.
Locus (plural loci)	The location on a chromosome where a particular gene resides.
Phenotype	The observable traits of an organism.
Pleiotropy	The phenomenon in which a single gene influences multiple traits.
Polymorphic gene	A gene that exists as more than one variant (*i.e.* allele) in a population.
Polymorphic position	A single nucleotide position occupied by two different nucleotides. Also known as a heterozygous position.
Pseudogene	A segment of DNA that resembles a gene but is usually not transcribed and does not encode a functional protein.
Single nucleotide polymorphism (SNP)	A nucleotide variation at a single nucleotide position in the DNA sequence, occurring at appreciable frequency (often defined to occurrence >1%) in the population.
Splicing	The process in which introns are removed from the mRNA molecules.
Transcription	The process in which an mRNA molecule, which is complementary to the DNA sequence of a gene, is synthesized.
Translation	The process in which an amino acid sequence is produced by the ribosome. The nucleotide sequence of the mRNA molecule determines the amino acid sequence in accordance with the rules of the genetic code.

4.1.4 Escaping tolerance: mechanisms of T-cell mediated autoimmunity

Autoimmune diseases are characterized by sustained immune responses against the body's own tissues or organs. In this section, potential mechanisms for T-cell mediated autoimmunity are introduced.

Escaping central tolerance

Not all self-antigens are available in the thymus for efficient negative selection of self-reactive thymocytes.[25] For instance, certain self-antigens might be expressed at very low levels in the thymic stromal cells.[61] The majority of autoimmune diseases are complex disorders, associated with a number of genetic variants. About 90% of these genetic variants have been located to non-coding regions in the genome, typically at sites that regulate gene expression.[62] Disease-associated genetic variants might exert (subtle) effects on the expression levels of self-antigens in thymic stromal cells,[63] as shown in type 1 diabetes (T1D), myasthenia gravis (MG) and Graves' disease.[64-67] Secondly, certain self-antigens might not be present in the thymus due to tissue-specific alternative mRNA splicing or post-translational modifications dependent on tissue-restricted enzymes.[25] An example is the recently reported hybrid insulin peptides that are recognized by $CD4^+$ T cells isolated from patients with T1D.[68] These are formed by covalent cross-linking of proinsulin peptides to other pancreatic peptides.

Self-reactive thymocytes might further escape central tolerance by expressing dual TCRs, *i.e.* two distinct, productively rearranged α-chains, at the cell surface.[69] It has been shown that murine T cells with dual TCRs, expressing each TCR at lower levels on the cell surface, required higher concentrations of peptides to induce similar T-cell responses compared to T cells with single TCRs.[70] This suggests that thymocytes with dual TCRs can escape central tolerance as they express less of each TCR, hence are less sensitive to the levels of self-antigen that are present in thymus.

Escaping peripheral tolerance

Defective peripheral tolerance mechanisms have been suggested to play a role in the propagation of autoimmune diseases. As described in section 4.1.1, Tregs are vital in maintaining peripheral tolerance. In several autoimmune diseases (*e.g.* systemic lupus erythematosus and rheumatoid arthritis [RA]), there are conflicting reports on Treg frequency (increased, normal or decreased) and function (normal or impaired).[22] These discrepancies might be explained by the lack of a universal human Treg marker and differences in disease stage and treatment regimen of patients.

The number of potential foreign peptides that T cells might encounter is greater than the number of TCRs available.[38, 71] Cross-reactivity of T cells is believed to be necessary to provide efficient immune coverage with limited numbers of available T cells.[38, 71] If the TCR repertoires were unable to recognize virtually all foreign peptides presented by HLA molecules, pathogens would be expected to rapidly evolve in order to exploit 'holes' in the T-cell repertoire, ultimately overwhelming the human host. On the other hand, cross-reactivity has also been suggested to have a role in initiation of autoimmunity. T cells that are weakly reactive against self-antigens survive central tolerance. These self-reactive T cells can potentially be activated through the cross-recognition of *e.g.* bacterial peptides that are structurally similar to the self-antigen. This phenomenon is known as molecular mimicry. Memory T cells can be stimulated by peptide concentrations that are more than 50-fold lower than the concentrations required to stimulate naïve T cells.[38, 72] Hence, viruses or bacteria could potentially 'trigger' an autoimmune disease by priming self-reactive T cells, leading to tolerance being broken. In line with this, infection with Epstein-Barr virus is linked to significant risk of developing multiple sclerosis (MS).[73]

In celiac disease, it has been hypothesized that gluten-reactive $CD4^+$ T cells provide help to gluten-reactive B cells as well as self-reactive B cells that recognize transglutaminase 2 (TG2), an enzyme that can deamidate dietary gluten.[74] TG2-specific BCRs can internalize complexes of TG2 and deamidated gluten. Following processing of the complex, deamidated gluten peptides will be presented by HLA class II molecules on the cell surface of the B cells to the gluten-reactive $CD4^+$ T cells. Similar mechanisms – in which pathogenic $CD4^+$ T cells that are reactive against exogenous antigens can stimulate self-reactive B cells and the production of autoantibodies – could potentially play a role in propagation of other autoimmune diseases as well.

4.1.5 The role of HLA in autoimmune diseases

The strong association between autoimmune diseases and the HLA complex (as shown in **Figure 13**) supports a role for HLA molecules in initiation and/or propagation of autoimmune diseases.[75, 76] However, the mechanism remains mostly elusive. Disease-associated HLA molecules might exert their function both during thymic education and in the periphery at the site of organ damage. It is estimated that an HLA class I molecule typically binds between 2,000 and 10,000 different peptides, and more than 2,000 different peptides are estimated to bind an HLA class II molecule.[77] The regions encoding the peptide-binding groove are particularly polymorphic.[51] This enables different allelic variants of HLA molecules to bind and present a different range of peptides to T cells. Thus, the most prominent hypothesized molecular mechanism to explain disease-association of certain HLA alleles is their unique ability to bind and present antigens that can trigger or drive disease, as seen in celiac disease[78].

Celiac disease is confined to individuals carrying HLA-DQ2.2 (HLA-DQB1*02:02, HLA-DQA1*02:01), HLA-DQ2.5 (HLA-DQB1*02:01, HLA-DQA1*05:01) and/or HLA-DQ8 (HLA-DQB1*03:02, HLA-DQA1*03:01).[79] The deamidation of gluten by TG2 creates negatively charged gluten-derived peptides that bind with higher affinity to the disease-associated HLA molecules.[78] In line with this, gluten-reactive $CD4^+$ T cells isolated from celiac disease patients preferentially recognize the deamidated gluten peptides when presented by disease-associated HLA molecules, but not in the context of other HLA molecules.[80, 81] In children, autoantibodies against TG2 are exclusively detected in those with the celiac disease-associated *HLA-DQ* alleles.[82]

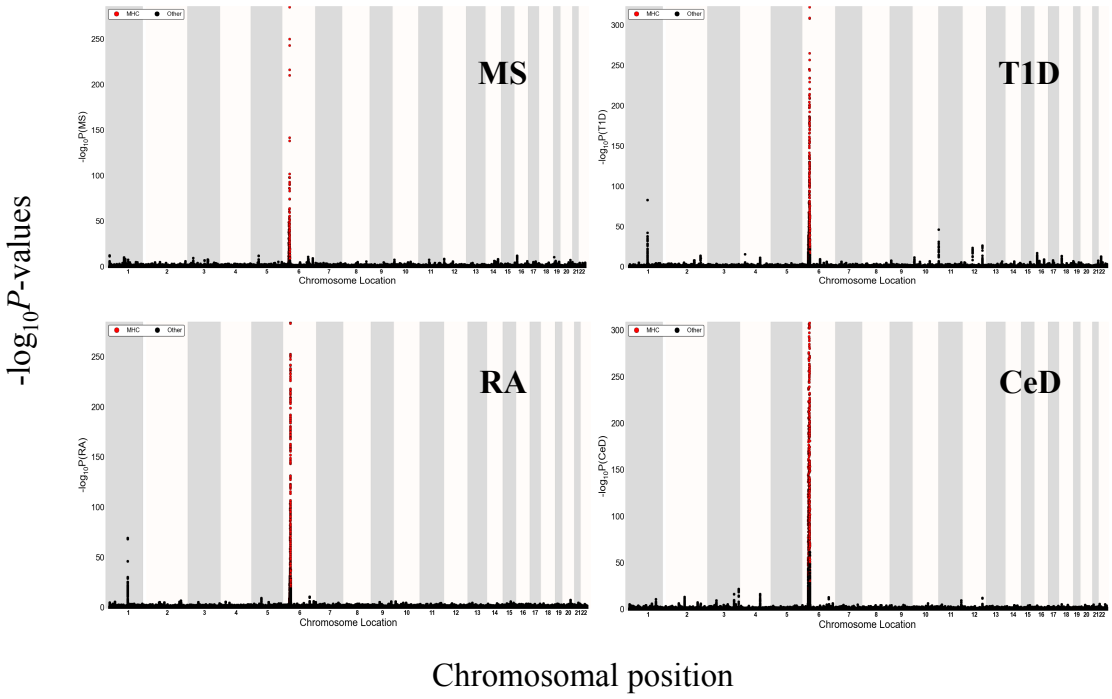

Figure 13. Autoimmune diseases are strongly associated with the human leukocyte antigen (HLA) complex. The figure shows Manhattan plots of genome-wide association study data in various autoimmune diseases. The X axis shows the chromosomal position, and the Y axis shows the strength of the association signal as the negative decadic logarithm of the p-values. HLA data on chromosome 6p21 are plotted in red. The non-HLA data are plotted in black. Data have been plotted using summary statistics in references[83, 84]. Abbreviations: CeD, celiac disease; MS, multiple sclerosis; RA, rheumatoid arthritis; T1D, type 1 diabetes.

In addition to their function in presenting peptide antigens to $CD8^+$ T cells, HLA class I molecules act as ligands for killer immunoglobulin-like receptors (KIRs), which are expressed mainly by natural killer (NK) cells (lymphocytes of the innate immune system, see **Figure 1**).[85] Through this interaction, disease-associated HLA molecules might influence the activation of NK cells. Other possible molecular mechanisms of genetic associations with HLA genes include differential expression of disease-associated HLA alleles and aberrant processing of the peptide–HLA complex.[86]

4.2 The human liver

The liver is the largest internal organ of the human body. It has numerous functions, including production of bile, lipids and proteins (such as components of the complement system), metabolism of nutrients, hormone excretion, storage of glycogen, vitamins and minerals, clearing the blood of harmful substances (*i.e.* detoxification) and immune surveillance.[87-91] Anatomical and physiological features of the human liver that are relevant for the present thesis are described below in sections 4.2.1–4.2.4.

4.2.1 The portal triad and sinusoid

A hepatic lobule is the structural unit of the liver.[87, 92] Portal triads (also called portal tracts) are located at the periphery of the lobule, with each portal triad comprising two blood vessels (branches of the portal vein and the hepatic artery) and a bile duct (**Figure 14**). Nutrient-rich blood from the portal vein and oxygenated blood from the hepatic artery mixes when percolating through capillary-like vessels called the sinusoids within the lobule. Finally, the blood is drained into the central vein, returning to the systemic circulation.

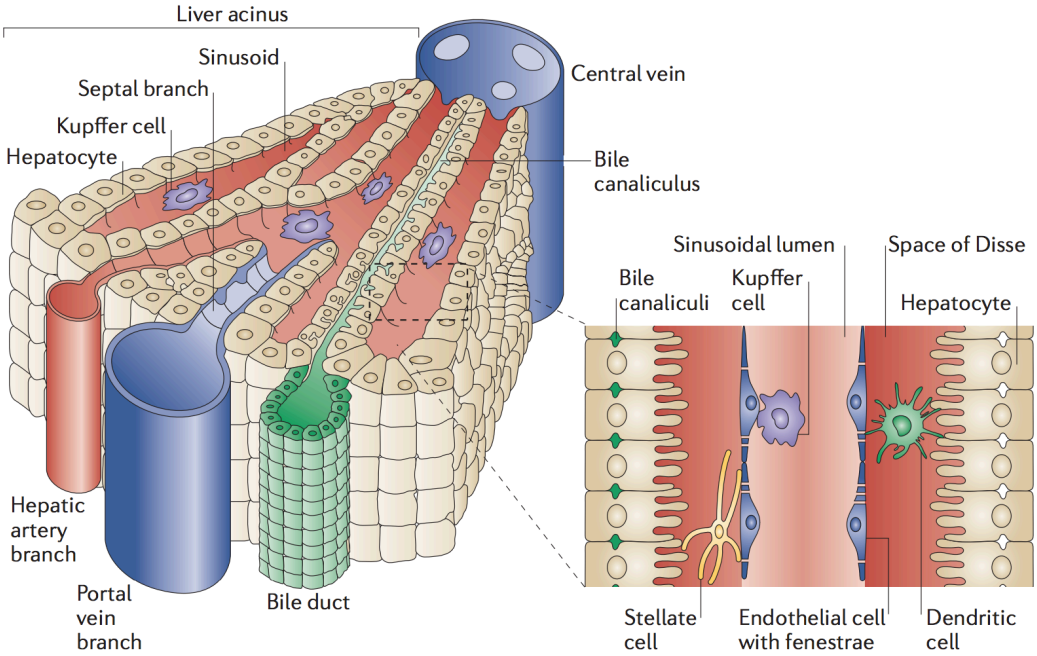

Figure 14. The portal triad and sinusoid. An intrahepatic portal triad comprises three vessels: one bile duct, one branch of the hepatic artery and one branch of the portal vein. The portal triads are located in the corners of hexagonal liver lobules (partially shown in this figure). The arterial and portal venous blood mixes when percolating through the sinusoids and is eventually drained into the central vein, returning to the systemic circulation. The sinusoids (see inset) are lined by unique endothelial cells that are characterized by the absence of tight junctions, the absence of a recognizable basement membrane and the presence of open fenestrae (small holes) that are organized into sieve plates. The sinusoidal endothelium is interspersed with macrophages that are called Kupffer cells. The space of Disse is located between the hepatic epithelial cells (which are called hepatocytes) and the sinusoidal endothelial cells, and contains extracellular matrix proteins and hepatic stellate cells. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology[92], copyright (2006).

4.2.2 The enterohepatic circulation of bile acids

Bile is mainly composed of water, bile acids, bilirubin, phospholipids and cholesterol. The primary bile acids are synthesized from cholesterol in the liver, and secreted through the bile ducts for storage in the gallbladder.[90, 93] The biliary and intestinal epithelia are continuous. Bile is released into the intestine after each meal, where bile acids facilitate the absorption of fats and fat-soluble vitamins.[94] About 95% of the bile acids are reabsorbed in the intestine and returned to the liver via the portal vein.[90, 93] This is known as the enterohepatic circulation of bile acids. Intestinal bacteria can metabolize the primary bile acids into secondary bile acids.[90] In addition to their role in the absorption of fat, bile acids function as signaling molecules and inflammatory agents.

The biliary epithelium initiates several mechanisms to protect itself from the toxic effects of bile acids.[93] This includes continuous secretion of bicarbonate, resulting in alkalinity at the apical surface of the epithelium.[95]

4.2.3 Cell composition in the healthy liver

The liver is predominantly composed of epithelial cells called hepatocytes, comprising 80% of all liver cells.[91] The liver-resident non-hepatocyte cell types comprise biliary epithelial cells (also called cholangiocytes) lining the bile ducts, liver sinusoidal endothelial cells (LSECs) lining the sinusoids, macrophages called Kupffer cells residing in the sinusoids (which constitutes 80–90% of all tissue-resident macrophages in the body), hepatic stellate cells, dendritic cells and lymphocytes (see **Figure 14**).

T cells constitute nearly 50% of the liver-resident lymphocytes, including conventional αβ T cells, innate-like MAIT cells (comprising up to 50% of all T cells in the human liver), iNKT cells (comprising about 1% of human liver-resident T cells) and γδ T cells (comprising about 15% of human liver-resident T cells).[91, 96] The remaining liver-resident lymphocytes comprise NK cells (constituting up to 50% of the liver-resident lymphocytes) and B cells.[97]

4.2.4 Immune surveillance and T-cell tolerance

Naïve T cells are normally activated in secondary lymphoid tissues upon recognition of their corresponding antigen presented by APCs that have migrated from an infected or inflamed non-lymphoid tissue (see section 4.1.1).[17] The liver serves as an exception to this 'rule', as naïve T cells can be retained and activated within this non-lymphoid organ following antigen-presentation by various liver-resident cells.[98, 99] Liver-resident cells that are able to present antigens to T cells include LSECs, Kupffer cells, dendritic cells, hepatocytes, hepatic stellate cells and cholangiocytes.[91, 100, 101]

The blood flows slowly through the sinusoids, facilitating the detection and capture of pathogens by liver-resident cells.[91] Seventy-five percent of the blood entering the liver is supplied from the intestines via the portal vein, thus is rich in nutrients and bacterial-derived molecules.[92] Multiple tolerance mechanisms prevent immune responses against the harmless antigens that are derived from food or the commensal microbiota.[91, 102] Under steady-state conditions, the presentation of antigens to liver-resident and circulating T cells is not accompanied by the co-stimulatory signals that are needed for the activation of naïve T cells. The liver-resident APCs express anti-inflammatory cytokines and co-inhibitory ligands that further induce T-cell tolerance. In response to infections or liver injury, additional signals will break the tolerance and lead to rapid activation of T cells.

Cholangiocytes are the first line of defense against pathogens in the biliary system.[103, 104] Upon activation of their pattern recognition receptors (see section 4.1.1), the cholangiocytes will express adhesion molecules, cytokines and chemokines, which will promote inflammation. It has further been shown that cholangiocytes can present antigens to and activate MAIT cells and iNKT cells.[100, 101] Despite the fact that they constitutively express HLA class II molecules, it has not yet been proven that cholangiocytes can activate conventional T cells.[103, 105, 106]

4.3 Primary sclerosing cholangitis

PSC is a chronic inflammatory liver disease characterized by fibrotic strictures and dilatations of the intrahepatic and extrahepatic bile ducts,[1] as illustrated in **Figure 15**. A mixed inflammatory cell infiltrate consisting of lymphocytes and neutrophils is observed in portal triads, usually more intense around the bile ducts.[107, 108] The majority of infiltrating lymphocytes are $CD4^+$ and $CD8^+$ T cells (**Figure 16**).[5, 6] The strictures lead to destruction and loss of bile ducts and impaired bile flow (*i.e.* cholestasis).[1] In parallel, there is a disorganized proliferation of new bile ducts, disrupting the liver architecture. The disease course is highly variable, but in the majority of patients, the disease will progress to an irreversible scarring of liver tissue (*i.e.* cirrhosis) and liver failure. Ultimately, many patients will need a liver transplantation.

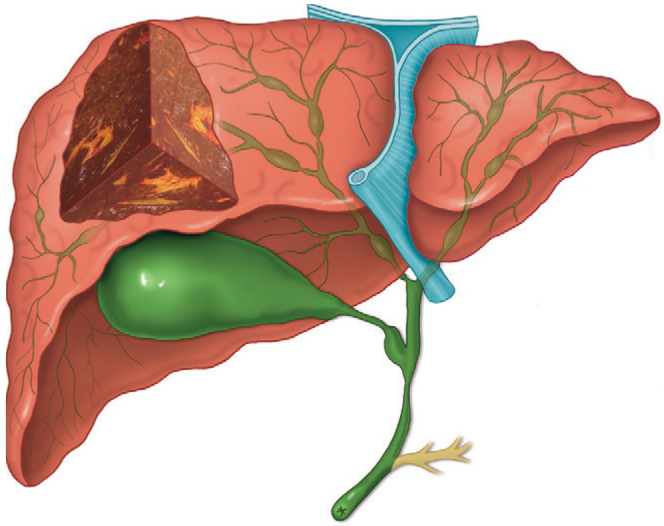

Figure 15. Illustration of a primary sclerosing cholangitis (PSC)-affected liver and its biliary tree. Strictures alternating with dilatations of both the intrahepatic (*i.e.* in the liver) and extrahepatic (*i.e.* outside the liver) bile ducts are commonly observed in patients with PSC. The strictures lead to cholestasis, patchy affection of peribiliary fibrosis and ultimately liver cirrhosis. Reproduced with permission from Kari C. Toverud.

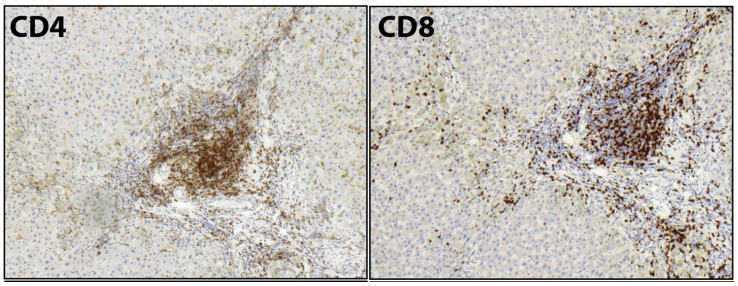

Figure 16. T cells in primary sclerosing cholangitis (PSC) liver. In the PSC-affected liver, $CD4^+$ and $CD8^+$ T cells are located mainly in portal triads, near bile ducts. Single-color immunohistochemistry (magnification: 200x). Reprinted from Liaskou *et al*., Gastroenterology copyright (2014),[6] with permission from Elsevier.

4.3.1 Epidemiology

PSC is a relatively rare disease, with a geographical variation in prevalence and incidence. The prevalence and mean annual incidence of PSC in Norway is 8.5 per 100,000 inhabitants and 1.3 per 100,000 inhabitants, respectively.[109] Similar results have been reported in other Northern-European countries and in the United States of America (USA) and Canada,[110-116] with the highest prevalence reported in a study of Swedish inhabitants (16.2 per 100,000).[110] In Southern Europe and Asia, PSC has been reported to be 10-100 fold less prevalent.[2, 112, 117, 118] PSC typically affects males (a 2:1 male to female ratio) in their third or forth decade of life.[1, 112]

4.3.2 Clinical presentation

The clinical presentation of PSC is variable. Most common symptoms and signs upon diagnosis include fatigue, pruritus, weight loss, abdominal pain, jaundice, enlarged liver and enlarged spleen,[119] but none of these are specific for PSC. About 50% of patients have no symptoms at time of diagnosis.[110, 120] Asymptomatic patients are identified through incidental (or in the case of patients with IBD, selective) testing and discovery of elevated levels of serum liver enzymes (marker of cholestasis).[121]

4.3.3 Diagnosis and clinical entities

A diagnosis of PSC is made in patients with elevated levels of serum liver enzymes that are not otherwise explained, with typical findings of bile duct changes on cholangiography and with no evidence of a secondary (*i.e.* identifiable) cause of sclerosing cholangitis.[122, 123] An elevated serum alkaline phosphatase is the most common biochemical abnormality in PSC. Serum aminotransferase levels are also often elevated, but most patients show normal levels of serum bilirubin at time of diagnosis. A diagnosis of PSC should not be excluded based on normal levels of serum liver enzymes alone, as these fluctuate during disease course. Visualization of the bile ducts using endoscopic retrograde cholangiography (ERC) has been the gold standard in diagnosing PSC. However, due to the non-invasive nature of magnetic resonance cholangiography (MRC) and its reduced risk for complications, MRC has replaced ERC as the first diagnostic modality when PSC is suspected. Finally, secondary causes of sclerosing cholangitis (*e.g.* IgG4-associated cholangitis, toxins, infections and an inadequate blood supply)[124] must be excluded before establishing a diagnosis of PSC.[122, 123]

A liver biopsy is not required for the diagnosis of PSC in patients with typical findings of bile duct irregularities on cholangiography. Yet, for a subgroup of patients who only have affection of the small intrahepatic bile ducts that are not visualized by cholangiography (*i.e.* small duct PSC) a liver biopsy is necessary for diagnosis.[122, 123] Small duct PSC is associated with better long-term prognosis and reduced risk of bile duct cancer (cholangiocarcinoma) compared to 'classical' PSC, also known as large duct PSC.[125-128]

4.3.4 Treatment and prognosis

At present, there is no curative or disease-halting medical therapy for PSC. Several anti-inflammatory and immunosuppressive drugs have been investigated, but none have so far proven to be effective in PSC.[129] A secondary bile acid called ursodeoxycholic acid (UDCA) is currently the main treatment of chronic cholestasis, but the use of UDCA in treating PSC remains controversial. Studies have shown improvements of levels of serum liver enzymes, but there is no proven beneficial effect on symptoms or time to liver transplantation.[129-133] Moreover, high-dose treatment

was associated with higher rates of liver transplantation and death.[134] The European Association for the Study of the Liver (EASL) concluded that the limited data did not yet allow for a specific recommendation for the general use of UDCA in PSC,[122] while the American Association for the Study of Liver Diseases (AASLD) recommends against the use of UDCA in PSC.[123] Novel therapeutic agents are currently in clinical trials and include agents that alter the composition of bile acids and the gut microbiota and agents that target fibrotic processes and immune responses.[129, 135]

Liver transplantation is currently the only therapy that can cure PSC, and is recommended in patients with PSC who progress to end-stage liver disease or who suffer from complications of the disease such as impaired quality of life.[122, 123, 136] PSC is today among the leading indications for liver transplantation in the Nordic Countries, accounting for approximately 15% of all liver transplantations that were performed from 1982 to the end of 2013.[3] The reported median time from diagnosis until liver transplantation or PSC-related death is 10–21 years.[137-139] Patients who undergo liver transplantation have a good prognosis, with a 5-year survival rate of approximately 85%.[3, 140] However, there is a substantial risk of recurrence of PSC after liver transplantation, occurring in at least 20% of transplanted patients.[140, 141] The absence of an inflamed intestine, either due to the absence of concurrent IBD or to colectomy (the surgical removal of the colon) before or during liver transplantation, has a protective effect against recurrent PSC.[142-145]

4.3.5 Comorbidities

Inflammatory bowel disease

PSC is strongly associated with IBD. The reported prevalence of IBD in patients with PSC is 60–80% in populations of Northern European ancestry and 20–54% in Southern Europe and Asia.[2] Conversely, the prevalence of PSC in patients with IBD is 8.1%.[146] PSC can develop in patients with IBD years after colectomy,[147] and IBD can develop for the first time in patients who have undergone liver transplantation for PSC.[147-150] According to commonly accepted criteria, PSC-associated IBD is classified as ulcerative colitis (UC) in 80-90% of the patients, while the remaining 10–20% of patients are diagnosed with Crohn's disease (CD) or indeterminate colitis.[151-154] It has been suggested that PSC-associated IBD represents a distinct disease entity, which is often presented as a mild or asymptomatic inflammation of the colon (*i.e.* colitis) with no rectal bleeding or inflammation.[151-153, 155] Inflammation is more pronounced in the ascending colon.

Autoimmune diseases

Some patients show signs of both PSC and autoimmune hepatitis (AIH) at the same time or sequentially.[156] In addition, up to 25% of patients with PSC are diagnosed with at least one autoimmune disease outside the liver and the colon, such as T1D, thyroid disease, RA and psoriasis.[157]

Cancer

Patients with PSC have an increased risk of cancer, predominantly cholangiocarcinoma (CCA).[137] The 5-year survival rate for patients with CCA is less than 5%,[158] making it one of the most feared complications of PSC. PSC is also associated with an increased risk of colorectal cancer (in patients with colitis) and gallbladder cancer.[137, 159]

4.3.6 PSC etiology and pathogenesis I: Genes and environmental factors

The etiology and pathogenesis of PSC are largely unknown. PSC is considered a complex disease, meaning it results from the interplay between multiple genetic variants and the environment.[160]

There is evidence of a genetic predisposition to PSC, with a 9–39 fold increased risk of disease in siblings of PSC patients compared to the healthy population.[161] Genome-wide association studies (GWAS) have identified 23 loci associated with susceptibility to PSC at genome-wide significance level (p-value < $5x10^{-8}$).[162-169] About half of the risk loci overlap with associations seen in IBD, and many risk loci are shared with autoimmune diseases such as T1D, celiac disease and RA (see **Table 1**).[4] This phenomenon, in which a single gene affects multiple traits, is known as pleiotropy.[170]

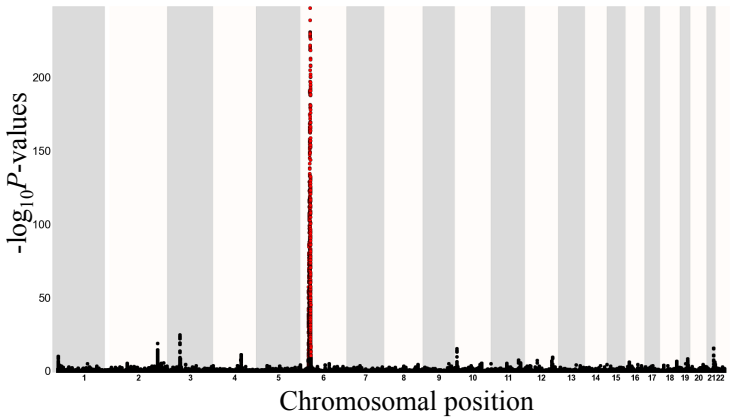

Figure 17. Manhattan plot of genome-wide association study data in primary sclerosing cholangitis. The X axis shows the chromosomal position, and the Y axis shows the strength of the association signal as the negative decadic logarithm of the p-values. Human leukocyte antigen (HLA) data on chromosome 6p21 are plotted in red. The non-HLA data are plotted in black. Data have been plotted using summary statistics in reference[167].

Associations with HLA haplotypes

The by far strongest genetic association signal is found within the HLA complex (**Figure 17**).[162, 163, 167] An association between PSC and the HLA class I allele HLA-B*08:01 and HLA class II allele HLA-DRB1*03:01 was first described in the early 1980s.[171, 172] HLA-B*08:01 and HLA-DRB1*03:01 are found on the same extended haplotype: the AH8.1 (which was introduced in section 4.1.3). These associations have later been confirmed by numerous studies, and associations with several other alleles of AH8.1 have further been reported (see **Figure 18**).[162, 173-186] Alleles of the AH8.1 are strongly associated with a large number of immune-mediated diseases.[187] For some diseases, such as PSC, associations have been reported for both HLA class I and class II alleles of the AH8.1. For other diseases, the primary association with AH8.1 is confined to the HLA class I

region, as seen in MG (*HLA-B*),[188] or to the HLA class II region, as seen in T1D (*HLA-DRB1-DQB1*) and celiac disease (*HLA-DQA1-DQB1*).[189, 190]

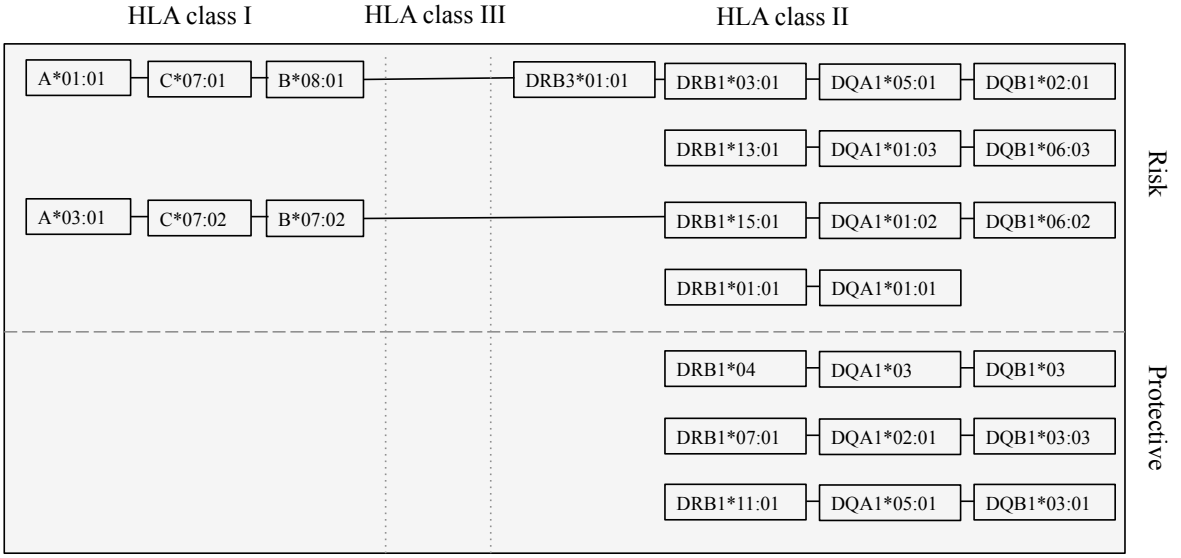

Figure 18. Human leukocyte antigen (HLA) haplotypes associated with susceptibility for primary sclerosing cholangitis (PSC). Four haplotypes are associated with risk of PSC. Three haplotypes are associated with protection.

Associations between PSC and various other HLA haplotypes have also been reported (**Figure 18**).[162, 167, 173-176, 182-186, 191, 192] In addition to AH8.1, a prominent risk haplotype is the HLA-DRB1*13:01-DQA1*01:03-DQB1*06:03 haplotype. It has been suggested that the HLA-DRB1*13:01 haplotype specifically increases the susceptibility of chronic cholangitis since it is associated with both large and small duct PSC, irrespective of IBD status.[193] The HLA-DRB1*13:01 allele is also associated with susceptibility for AIH in South America.[194] The HLA-DRB1*04-DQA1*03-DQB1*03 haplotype is the most consistently reported protective haplotype in PSC.[162, 173-176, 183, 184, 186, 192] The phenotypic heterogeneity in PSC might (partly) explain the association with multiple HLA haplotypes. In line with this, there is evidence of clinical subgroups in PSC whom have different HLA predisposition than the PSC group as a whole.[193, 195]

Due to the extensive LD and multiple polymorphic candidate genes in the HLA complex,[51] it is difficult or even impossible to identify which of the gene(s) within the haplotypes that is/are responsible for these associations. Since allele frequencies and LD patterns differ between different ethnicities,[196] studying non-European or admixed populations might aid in fine mapping the causative HLA gene(s) and alleles in PSC. This was previously demonstrated in MS: to better localize the HLA gene responsible for the association between MS and the HLA-DRB1*15:01-DQB1*06:02 haplotype, the *HLA-DRB1* and *HLA-DQB1* genes of African American MS patients and controls were assessed, showing a selective association between MS and the HLA-DRB1*15:01 allele.[197] In PSC, the low prevalence of disease in populations of non-Northern European ancestry[2, 112, 117, 118] makes it challenging to establish the necessary sample size for such a

study. Genetic studies in PSC have therefore mainly been performed in populations of Northern European origin, and the causative HLA gene(s) and alleles remain unknown.

Associations with non-HLA loci

The non-HLA risk loci appear to be enriched for genes encoding proteins involved in innate or adaptive immune responses, in particular proteins regulating T-cell biology.[4] For instance, CD28 (encoded on the risk loci 2q33) is a co-stimulatory surface molecule necessary for T-cell activation, proliferation and survival.[198] A study reported an accumulation of pro-inflammatory $CD28^-$ T cells in livers of patients with PSC, localized around bile ducts.[6] Other examples are the two cytokines interleukin-2 (IL-2) and IL-21, which are encoded within risk loci 4q27. IL-21 regulates differentiation and function of multiple target cells, including $CD4^+$ and $CD8^+$ T cells and B cells.[199, 200] IL-2 signals induce proliferation and optimize differentiation of $CD8^+$ naïve T cells into effector cells or memory cells.[201] IL-2 further influences the differentiation of $CD4^+$ T cells into various effector T-cell subsets. Specifically, IL-2 signals are crucial for the development, survival and function of Tregs, thus this cytokine is important in maintaining immune tolerance and preventing autoimmunity. The α subunit of the trimeric IL-2 receptor (IL-2Rα, also known as CD25) is encoded within another PSC risk loci (10p15). This suggests an important role for the IL-2 pathway in PSC pathogenesis. In line with this, liver-derived T cells from patients with PSC have been shown to produce significantly lower levels of IL-2 compared to T cells derived from livers of patients with primary biliary cholangitis (PBC, previously known as primary biliary cirrhosis), AIH and healthy controls.[202] Other risk loci harboring genes that encode T-cell regulating proteins are shown in bold in **Table 1**.

Environmental factors

Many individuals who carry PSC-associated HLA haplotypes (such as the AH8.1) never develop disease. Conversely, there are patients with PSC who do not carry any of the risk HLA haplotypes. Genetic variants at the 22 non-HLA risk loci that are identified to date are common variants that are also present to a large extent in healthy individuals.[4, 168, 169] This suggests additional factor(s) must be present to trigger and/or drive development of PSC. Collectively, genes identified so far are estimated to account for less than 10% of the overall disease liability, and likely more than 50% of the susceptibility to PSC is explained by environmental factors.[167, 168, 203] There are only two established environmental factors for PSC, smoking and coffee consumption, both of which are suggested to protect against development of PSC.[204-207] Hormonal factors have further been suggested to influence development of PSC in women: fewer female patients with PSC than healthy controls reported ever using hormonal contraception, and there was a strong correlation between increasing age at diagnosis of PSC and increasing number of children before the diagnosis.[204]

Table 1. Twenty-two genome-wide significant non-human leukocyte antigen (HLA) risk loci that are associated with primary sclerosing cholangitis (PSC). Candidate gene(s) shared with other AIDs refer to the same candidate gene(s) having been reported in both PSC and the listed AID based on mapping of genome-wide significant (p-value < $5x10^{-8}$) single nucleotide polymorphisms to the candidate gene(s) in both diseases. Genes in bold represent genes encoding proteins involved in regulation of T-cell biology.

Locus	Candidate gene(s)	Other AID associated with candidate gene(s)	Putative role of encoded protein(s)
1p36	*MMEL1*, ***TNFRSF14***	CeD, MS, PBC, RA, UC	*MMEL1*: Little is known about its function. *TNFRSF14*: Modulation of T-cell activation and regulation of immune tolerance.
2q13	***BCL2L11***	None	Regulation of T-cell and B-cell apoptotic processes and termination of inflammatory response.
2q33	***CD28***, ***CTLA4***	AA, CeD, GV, MG, RA, T1D	*CD28*: Regulation of T-cell activation, proliferation and survival. *CTLA4*: Major negative regulation of T-cell responses.
2q36	***CCL20***	PBC, IBD, UC	Regulation of antibacterial activity in mucosal lymphoid tissues, including migration of Tregs.
2q37	***GPR35***	AS, IBD, UC	Regulation of IL-4 release from iNKT cells. Potentially involved in mucosal and systemic immune regulation.
3p13	***FOXP1***	None	Regulation of the development and function of Tregs.
3p21	*MST1*	CD, IBD, UC	Regulation of inhibitory functions towards macrophages during inflammation.
4q24	***NFKB1***	PBC, UC	Regulation of multiple biological processes, including inflammation, immunity and apoptosis.
4q27	***IL2***, ***IL21***	AA, CeD, IBD, T1D	*IL2*: Regulation of the differentiation and proliferation of various T-cell subsets, including Tregs. *IL21*: Regulation of differentiation and function of multiple target cells in the innate and adaptive immune response.
6q15	***BACH2***	CD, CeD, IBD, MS, T1D, VT	Regulation of T-cell and B-cell differentiation. Implicated in antiviral innate immune response.
10p15	***IL2RA***	AA, CD, IBD, MS, RA, T1D, VT	Regulation of immune tolerance by modulating Treg activity.
11q13	***CCDC88B***	CD, IBD, PBC, SARC	Regulation of T-cell maturation and inflammatory function.
11q23	***SIK2***	None	Regulation of IL-2 expression in macrophages and leukocyte function.
12q13	***HDAC7***	IBD	Regulation of negative selection of T cells in thymus and development of immune tolerance.
12q23	*RFX4*, *RIC8B*	None	*RFX4*: Potentially involved in regulation of immune- and infectious responses. *RIC8B*: Little is known about its function.
12q24	***SH2B3***, *ATXN2*	AS, CeD, HT, IBD, RA, PBC, SLE, T1D, VT	*SH2B3*: Regulation of cytokine signaling. *ATXN2*: Involved in EGFR trafficking.
16q12	***CLEC16A***, ***SOCS1***	CD, CeD, IBD, MS, PBC, SLE, T1D	*CLEC16A*: Regulation of B-cell function and of thymocyte selection and reactivity. Role in autophagy. *SOCS1*: Regulation of cytokine signaling and thymocyte development.
18q21	***TCF4***	None	Regulation of early B-cell and T-cell development.
18q22	***CD226***	IBD, RA, T1D	Regulation of lymphocyte signaling, cytotoxicity and lymphokine secretion
19q13	***PRKD2***, *STRN4*	T1D	*PRKD2*: Regulation of the negative selection of T cells. *STRN4*: Little is known about its function.
21q22	*PSMG1*	AS, IBD, UC	Potentially implicated in secondary immune response.
21q22	***UBASH3A***	RA, T1D, VT	Regulation of TCR pathways.

Abbreviations: AA, alopecia areata; AID, autoimmune disease; AS, ankylosing spondylitis; CD, Crohn's disease; CeD, celiac disease; GV, Graves' disease; HT, hypothyroidism; IBD, inflammatory bowel disease; iNKT cell, invariant natural killer T cell; MG, myasthenia gravis; MS, multiple sclerosis; PBC, primary biliary cholangitis; RA, rheumatoid arthritis; SARC, sarcoidosis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; TCR, T-cell receptor; Tregs, regulatory T cells; UC, ulcerative colitis; VT, vitiligo.

4.3.5 PSC etiology and pathogenesis II: Pathogenic models

Several pathogenic models in PSC have been proposed over the years.[2, 208] These include immune-driven pathogenic mechanisms and aspects related to bile acid physiology and the integrity of bile duct epithelium. Notably, the proposed pathogenic mechanisms (which are described below) are not mutually exclusive.

The 'leaky gut' hypothesis

The frequent comorbidity with IBD indicates that there is a close relationship between the gut and the liver in patients with PSC. The earliest proposed pathogenic mechanism in PSC was built on a 'leaky gut' concept, in which an increased permeability of the mucosal barrier during ongoing colonic inflammation results in an increased 'leakage' of bacteria and/or bacterial products into the portal vein.[208] In the liver, immune tolerance may be broken due to the increased exposure to the gut-derived molecules or bacteria.[91] In support of this hypothesis, biliary changes resembling PSC have been reported in murine models with exposure to bacterial-derived peptides or intestinal bacterial overgrowth.[209, 210] The observation that not all patients with PSC have IBD, or that IBD can develop after the onset of PSC or even after liver transplantation,[147-150] contradicts that bacterial leakage through an inflamed colon is mandatory for the onset of PSC. However, it does not exclude the fact that bacteria or bacterial products may be important in the pathogenesis of PSC.

The 'gut lymphocyte homing' hypothesis

As mentioned in section 4.1.1, activated lymphocytes express a distinct set of surface molecules that direct migration to the inflamed tissue. Lymphocytes that have been activated by intestinal dendritic cells in the gut-associated lymphoid tissues (GALT) co-express the $\alpha4\beta7$ integrin and CC chemokine receptor 9 (CCR9).[211, 212] These molecules are receptors for the mucosal addressin cell adhesion molecule 1 (MAdCAM-1) and CC chemokine ligand 25 (CCL25), respectively, which in the steady-state condition are restricted to the colon.[213, 214] In PSC, an aberrant expression of MAdCAM-1 and CCL25 has been detected on portal vein- and sinusoid endothelium,[215-217] and approximately 20% of the lymphocytes infiltrating the livers of patients with PSC are effector memory T cells co-expressing $\alpha4\beta7$ and CCR9.[215] These findings suggest an aberrant homing of long-lived gut-derived memory T cells to the liver.[92] Once in the liver, the T cells can promote inflammation upon recognition of the same or structurally similar antigen. In agreement with this hypothesis, murine $CD8^+$ T cells primed in the GALT by ovalbumin migrate to the liver and cause cholangitis when recognizing the same antigen on bile ducts.[218]

Vascular adhesion protein 1 (VAP-1) functions both as an enzyme and as an endothelial adhesion molecule for lymphocytes. VAP-1 is constitutively expressed in the human liver.[219] It has been demonstrated that the deamination of primary amines (such as methylamine, a constituent of food

and cigarette smoke) by VAP-1 can lead to expression of MAdCAM-1 in the presence of the pro-inflammatory cytokine tumor necrosis factor alpha (TNFα) in human primary hepatic endothelial cells.[220] Thus, in the presence of an inflamed gut, increased portal vein levels of dietary amines or amines produced by the gut microbiota may activate hepatic MAdCAM-1 expression, hence promote the recruitment of gut-derived lymphocytes to the liver.

The 'autoimmune' hypothesis

The predominant HLA association in PSC is a feature shared with prototypical autoimmune diseases (see **Figure 13** and **Figure 17**).[76, 162, 163, 167] Patients with PSC frequently have concurrent autoimmune diseases,[156, 157] and the low prevalence of disease in Southern Europe and Asia[2] mirrors the North-South risk gradient of autoimmune diseases[221]. Findings of pleiotropic risk loci[4] and reports of autoantibodies[222, 223] further support an autoimmune component in PSC pathogenesis.

The most prevalent autoantibody in PSC, reported in up to 94% of the patients (although it is not disease-specific), is a particular type of perinuclear anti-neutrophil cytoplasmic antibody (pANCA).[222] The main antigenic target of pANCA in PSC is beta tubulin isotype 5 (TBB-5) in human neutrophils.[224] pANCA is also able to react with the structurally related microbial protein FtsZ, which is expressed by intestinal bacteria during cell division. Hence, through FtsZ, intestinal bacteria might prime self-reactive immune cells and direct an immune response towards human cells; this is an example of molecular mimicry (see section 4.1.4). There has also been reported a high frequency of autoantibodies in PSC that are reactive against cholangiocytes.[225] Upon binding to cholangiocytes, the autoantibodies induce the expression of pattern recognition receptors that recognize bacterial products and viral DNA.[226] In addition to the autoantibodies mentioned above, several other autoantibodies have been detected in patients with PSC, such as anti-nuclear antibodies (ANA) and rheumatoid factor.[222, 223] None of the autoantibodies that are identified so far are PSC-specific.

It has been reported a reduced number of Tregs in patients with PSC compared to patients with PBC or healthy controls.[227] Furthermore, $CD4^+$ T cells (but not $CD8^+$ T cells) isolated from the peripheral blood of patients with PSC exhibit significantly reduced apoptosis compared with healthy controls and patients with PBC.[228] This is associated with impaired upregulation of the proapoptotic protein BIM (which is encoded by *BCL2L11* within the PSC risk loci 2q13) in $CD4^+$ T cells from patients with PSC. These observations might suggest peripheral tolerance mechanisms are defective in patients with PSC, which would indicate these patients are predisposed for autoimmunity.

Despite a probable autoimmune component in PSC, the PSC-specific self-antigen(s) remains unknown, and the patients do not respond to classical immunosuppressant drugs.[129] The latter observation might be due to the timing of therapy, as strictures of the bile ducts are already evident

at time of diagnosis. The strong male predominance in PSC contrasts observations made in most autoimmune diseases[229] (the exception being ankylosing spondylitis[230]). Interestingly, MRC screening of patients with colitis have suggested that women and men have a similar predisposition to PSC, but that women acquire a milder disease phenotype.[146] This is further supported by observations of later onset of disease and reduced risk of liver transplantation and death for female patients with PSC.[138]

The 'toxic bile' hypothesis

An abnormal bile composition and impaired protection mechanisms against bile have been suggested to contribute in the pathogenesis of PSC.[231] The 'toxic bile' hypothesis is based on the findings of bile duct injury resembling human PSC in mice devoid of the phospholipid transporter multidrug resistance protein 2 (Mdr2).[232] This transporter normally mediates secretion of phospholipids in bile, which together with cholesterol form micelles with bile acids to protect the biliary epithelium against the toxic effects of bile acids.[231] No significant association between the gene encoding the human ortholog MDR3 (ATP-binding cassette sub-family B member 4; *ABCB4*) and PSC has so far been identified.

5 AIMS

The overall focus of this thesis was to further characterize components of the immune system in patients with PSC to provide better insight into the disease mechanisms.

The strong association between PSC and the HLA complex as well as associations with numerous loci enriched for genes encoding proteins that directly or indirectly affect T-cell biology strongly suggest that T cells are involved in PSC pathogenesis. This is further supported by the observation that the lymphocytic infiltrates in portal triads predominantly consist of $CD4^+$ and $CD8^+$ T cells. T cells are hypothesized to migrate between the gut and the liver of patients with PSC, promoting inflammation upon recognition of antigen(s). Using high-throughput sequencing of the TCRβ chain, we therefore aimed to:

1. Characterize the liver T-cell repertoire of patients with PSC and determine whether disease-associated TCRβ sequences could be identified in their explanted liver tissue, with PBC patients and alcoholic liver disease (ALD) patients as disease controls. (**Paper I**).

2. Investigate whether gut-infiltrating T cells and liver-infiltrating T cells of patients with concurrent PSC and IBD (*i.e.* PSC-IBD) carry the same TCRs, and further assess whether this is related to PSC pathogenesis by investigating the potential overlap in the TCRβ repertoires of paired tumor-adjacent normal gut and liver tissue sampled from patients with colon cancer and metastasis to the liver (**Paper II**).

The identity of the PSC-specific antigenic trigger(s) has remained unknown. Dissecting the HLA association might provide knowledge about the antigen(s) and type(s) of T-cell response(s) involved in PSC pathogenesis. However, the high degree of LD extending across large parts of the HLA complex has made it difficult to identify the HLA gene(s) and alleles responsible for the strong association with PSC. Given LD patterns, haplotype frequencies and allele frequencies differ across ethnicities, the final aim of this thesis was to:

3. Evaluate whether studying populations of admixed and non-European ancestry could aid in pinpointing the causative alleles of PSC-associated HLA haplotypes (**Paper III**).

6 METHODOLOGICAL CONSIDERATIONS

6.1 Study populations and starting material

In every study, we included patients with PSC. The diagnosis of PSC was based on accepted criteria with typical findings of bile duct irregularities on cholangiography.[122, 123] The studies that were described in paper I and II had a case-control study design. Herein, we further included patients with the liver diseases PBC and ALD (paper I) and patients with colon cancer and liver metastasis (paper II). Paper III depicts a descriptive, hypothesis-generating study. We did not include any healthy controls in our studies and the reasons for this are described in sections 6.1.2–6.1.4.

The studies were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all study participants. Ethical approval was obtained by the Regional Committee for Research Ethics at the South-Eastern Norwegian Health Authority (2012-286, 2015-744, 2016-1540 and 2012-341) and from research ethics committees at each collaborating center abroad.

6.1.1 Considerations regarding starting material in paper I and paper II

In PSC, as in most autoimmune conditions, the access and use of diseased human tissue is limited and challenging. Explanted livers were the only source of diseased liver tissue available for study in paper I and paper II. About a hundred liver transplantations are performed each year in Norway, of which nearly 20% are due to PSC.[3] A similar number of PSC patients (*i.e.* about 20 patients) are transplanted each year at the Queen Elizabeth Hospital in Birmingham, United Kingdom (UK). Tissue is only sampled from the explanted livers of the patients whom have given their written consents to participate in health research prior to transplantation. Ultimately, the scarce number of patients with PSC and the rare access to their diseased liver tissue limits the conceivable sample size of these studies.

At time of liver transplantation, patients might have progressed to end-stage liver disease, and thus it is impossible to know whether our findings are representative for mechanisms that take place in the early phase of inflammation or represent secondary toxic or inflammatory insults due to *e.g.* procedure-related colonization, cirrhosis and cholangitis. Most patients with PSC do not undergo regular biopsies since these are rarely diagnostic (except when suspecting a diagnosis of small duct PSC[122, 123]). Hence, it is impossible to study the TCR repertoires at an earlier disease stage.

The livers of patients with PSC show a pronounced patchy distribution of inflammation. The relatively small samples that were used in paper I and II might therefore represent a small proportion of the diseased microenvironment. Nevertheless, the use of inflamed human tissue rather

than animal models or *in vitro* studies ensures our results are relevant when aiming to increase our knowledge of complex human disease mechanisms.

The availability of either fresh-frozen or formalin-fixed, paraffin-embedded (FFPE) tissue in paper I and paper II did not allow for isolation and sorting of T-cell subsets and/or single T cells prior to subsequent analyses. This limited our choices for TCR sequencing strategies and ultimately constrained our options for downstream analysis, as discussed further in section 6.2.1. Extracting genomic DNA (gDNA) of high molecular weight from FFPE tissue is challenging as it becomes fragmented due to the fixation process conditions.[233] The use of gDNA extracted from FFPE tissue might therefore affect the amount of sequencing data retrieved. However, fragmentation should be unbiased and will not specifically affect particular rearranged TCRβ sequences.

6.1.2 Study populations in paper I

In paper I, we included fresh-frozen explanted whole liver tissue from patients with PSC and from disease controls, namely patients with PBC and patients with ALD since predominant T-cell infiltrates are also evident in the liver tissue of these patients.[234, 235] PBC is an immune-mediated liver disease characterized by progressive destruction of the small intrahepatic bile ducts.[236] Disease-specific anti-mitochondrial antibodies (AMA) are detected in about 95% of patients with PBC. ALD is a liver disease related to alcohol consumption.[237] We did not include liver tissue from healthy controls since the aim and focus of the study was to define the TCR repertoire in PSC in comparison to other liver diseases in order to identify disease-associated TCR sequences. Statistical frameworks for power calculations in TCRβ sequencing studies are currently lacking. We ultimately set the sample size to n=20 patients with PSC and n=20 disease controls (*i.e.* n=10 patients with PBC and n=10 patients with ALD), which is comparable to or greater than the sample size of similar studies.[238-240] The liver samples were acquired through the transplant program at the Queen Elizabeth Hospital, Birmingham, UK.

The patients with PSC, patients with PBC and patients with ALD were of similar age at liver transplantation (on average, 48, 47 and 45 years, respectively). Gender differences between disease groups (PSC: 80% males, PBC: 20% males, ALD: 80% males) reflect the reported male-to-female ratios in these diseases.[112, 241, 242] HLA-matching of study participants is generally recommended in TCR studies given the potential effect of HLA alleles on the TCR repertoire. However, this was impossible in paper I due to the scarcity of available explanted liver tissue from these patients. Besides, considering that certain HLA haplotypes are strongly associated with susceptibility for PSC or PBC[243] and further that we aimed to identify disease-associated TCR repertoires, it was not in our interest to intentionally match the patients and disease controls with regards to HLA type. Most of the patients that were included in the study (including patients with ALD) did carry HLA alleles associated with susceptibility to PSC or PBC.

6.1.3 Study populations in paper II

In paper II, we included every Norwegian patient with a concurrent diagnosis of PSC and IBD who had fresh-frozen explanted whole liver tissue, fresh-frozen biopsies from the ascending colon and EDTA anti-coagulated peripheral (whole) blood available at Oslo University Hospital (OUH) Rikshospitalet, Oslo, Norway. Of the ten patients who fit these criteria, nine were diagnosed with UC and one was diagnosed with CD. This largely resembles the prevalence of UC versus CD in patients with concurrent PSC and IBD.[151-154] The diagnosis and classification of IBD was based on commonly accepted clinical, endoscopic and histopathologic criteria.[244] The colonic biopsies and peripheral blood samples had been collected during routine colonoscopy at OUH Rikshospitalet.[152] The time from colonoscopy until liver transplantation ranged from 0.6 years to 7.3 years (median 3.5 years). Three patients were using immunosuppressant drugs at time of sampling, including one patient using Cyclosporin A prior to liver transplantation. Immunosuppressive drugs are generally expected to have 'global' inhibitory effects on T-cell proliferation,[245] however a study reported that the immunosuppressive agents Cyclosporin A and Tacrolimus might affect the expression of specific TCRβ sequences.[246] Given the limited access to matched material from patients with PSC, we decided to continue with all ten patients in subsequent analyses.

We further included paired tumor-adjacent normal tissue that had been sampled during colectomy and liver resection from ten patients with colon cancer and liver metastasis. The normal tissue was available as FFPE material at Human Biomaterials Resource Center, University of Birmingham, UK. With the limited amount of gDNA extracted from the tumor-adjacent normal FFPE tissue, we chose to prioritize sequencing the TCRβ repertoire and thus did not perform HLA typing of the colon cancer patients. Six colon cancer patients (including every patient that was ultimately included in the TCRβ overlap analyses) had received chemotherapy prior to sample collection. The patients completed their chemotherapy treatments one month to two years prior to colectomy and/or liver resection. Of note, increased number of Tregs has been reported in peripheral blood, tumor-draining lymph nodes and at tumor sites of colorectal patients.[247] Reduced number of Tregs has been reported in peripheral blood seven days after chemotherapy for colorectal cancer patients who had high levels of Tregs prior to chemotherapy.[248] However, the total number of lymphocytes and the percentages of $CD4^+$ and $CD8^+$ T cells were reported to be unchanged following chemotherapy.

The average age at liver transplantation for the patients with PSC-IBD was 47 years (range: 36–63 years), while the colon cancer patients were nearly 20 years older, with the average age at liver resection being 65 years (range: 53–79 years). It has been reported a reduction in TCR repertoire diversity in peripheral blood with advancing age.[249] The diversity correlated with percentage of naïve T cells (which decreased linearly with advancing age), but there were no significant age-related changes in the diversity of non-rare TCRβ sequences (those with more than one T cell per million). During subsequent TCRβ analyses in paper II, we normalized our sequencing data as

described in section 6.2.3, and therefore assume our results reflect age-independent differences between paired PSC-IBD affected tissue and paired tumor-adjacent normal tissue.

There is a pronounced clinical heterogeneity in PSC. This has led to speculations that PSC could represent a 'mixed bag' of different conditions with various undefined etiologies, however the major subgroup of patients who have coexistent IBD likely represent one disease entity with common pathogenic mechanisms.[2] We did not include matched material from the rare subgroup of PSC patients who do not have IBD. It is very difficult to delineate these patients from the overall PSC population since IBD can occur even after liver transplantation.[147-150] Hence, it is possible that findings in paper II reflect a PSC-IBD specific feature. We were further unable to include paired tumor-adjacent normal gut and liver tissue that had been sampled from Norwegians, thus it is possible that results reflect differences across populations from different geographical origins. Finally, we were unable to include disease controls in paper II. Sampling of matched gut tissue from patients with other liver diseases such as PBC and ALD (or from healthy liver donors) is ethically impossible, as is sampling of matched liver tissue from patients with bowel disease without clinical features of liver dysfunction.

6.1.4 Study populations in paper III

In paper III, we included every PSC patient who had sufficient amount of gDNA available in Oslo, the Netherlands or Kiel (Germany) among those whose non-European genetic ancestry had been revealed in the Immunochip-based PSC study[167] (n=67). The patients had originally been recruited in European and North-American countries (**Table 2**). The gDNA had been extracted from whole blood or liver tissue. Using single nucleotide polymorphism (SNP) genotyping data from the Immunochip-based PSC study[167], we estimated the genetic ancestry of these patients in a multidimensional scaling analysis.[250] We additionally included PSC patients of self-reported admixed or non-European ancestry sampled in the USA (n=21, African Americans) and Canada (n=4, one Iranian, one Pakistani/Indian, one admixed Canadian Caucasian/Iranian and one admixed Canadian Caucasian/African Canadian). Occasionally, self-reported ancestry is inaccurate in defining an individual's genetic ancestry. However, it is reasonable to assume that patients of self-reported admixed or non-European ancestry are truly not of 'unmixed' European ancestry.

Table 2. Overview of the primary sclerosing cholangitis (PSC) patients of non-European ancestry who were included in paper III. [a]Patients who were originally recruited for the Immunochip-based PSC study[167].

Country	Number of patients		
	of confirmed non-European genetic ancestry[a]	of self-reported admixed or non-European ancestry	Total
Finland	1		1
The United Kingdom	26		26
Germany	2		2
The Netherlands	3		3
France	6		6
Italy	1		1
The United States of America	6	21	27
Canada	22	4	26
Total	67	25	92

We further included 135 Norwegian and 15 Swedish PSC patients. They were selected from a previously described Scandinavian PSC population.[182, 251] The gDNA had been extracted from whole blood. Since we aimed to assess the previously reported PSC-associated HLA haplotypes, the Scandinavian patients were randomly selected from among patients carrying *HLA-DRB1* alleles that are found on these haplotypes (*i.e.* HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*07:01, HLA-DRB1*11:01, HLA-DRB1*13:01 and HLA-DRB1*15:01).[243] With a frequency of approximately 10%,[57, 58] AH8.1 is a very common haplotype in Northern-European populations. To avoid that patients carrying AH8.1 constituted the majority of the Scandinavian study population, we selected only ten patients from among those who were homozygous for both HLA-B*08 and HLA-DRB1*03:01 alleles. The selection of specific Scandinavian patients rendered us unable to appropriately compare the strength of LD that was measured in the Scandinavian study population and the admixed/non-European study population, having excluded any Scandinavian patient with *e.g.* HLA-DQB1*06:03 on non-PSC associated haplotypes. In our subsequent analyses, we therefore included additional data, which had been generated and described previously, from the overall Scandinavian PSC population.[182, 251]

Allele frequencies differ between populations of different geographical origins.[196] In genetic association analyses, it is important that healthy controls are comparable with patients with regards to their ethnic origin so that differences in allele frequencies do not reflect different ancestry. Due to the selection process of the patients of admixed or non-European ancestry, it would have been impossible to recruit an adequate healthy control population for genetic association analyses in paper III.

6.2 High-throughput sequencing and analyses of the TCRβ chain

In paper I and paper II, we sequenced the TCR repertoires using the commercially available high-throughput sequencing assay 'immunoSEQ' at Adaptive Biotechnologies (Seattle, WA, USA).[252, 253] Adaptive Biotechnologies is a well-recognized company offering immune repertoire sequencing, with a great track record of publications that have used their technology.[238, 239, 254-256] For reasons elaborated below, we chose to sequence the β-chain only, using gDNA as starting template.

6.2.1 Considerations concerning high-throughput TCR sequencing strategy

Sequencing target: the α-chain and/or β-chain

Most high-throughput TCR sequencing methods target single subunits, *i.e.* the α-chain or the β-chain. The α-chain is difficult to characterize due to more V and J genes present on the TRA locus compared to the TRB locus (see **Figure 7**) and the fact that about 10–20% of T cells express two productively rearranged α-chains[42-44], which complicates the interpretation of TCRα sequencing data. In contrast, most T cells express only one productively rearranged β-chain as a result of allelic exclusion mechanisms, and knowledge of the nucleotide sequence of the β-chain is considered sufficient to identify a given antigen-experienced T-cell clone.[42, 44, 46] The latter statement was recently confirmed in a study where 98.2% of the TCRβ sequences that were detected in both tumor tissue and matched samples of peripheral blood mononuclear cells were paired with the same TCRα sequences across sample types, indicating the tumor-infiltrating T cells and peripheral blood T cells were members of the same T-cell clones.[42] For these reasons, we decided to focus our TCR repertoire studies exclusively on sequencing the β-chain.

In recent years, technologies have made possible the paired sequencing of the α-chain and β-chain,[42, 44, 257] which allows subsequent functional analyses of the TCRs. Single-chain sequencing of previously uncharacterized TCR repertoires renders the functional analysis of the TCRs impossible, since either the α-chain or the β-chain remains unknown. Paired sequencing methods are often technically challenging and require the use of single-sorted cells, thus are generally limited in throughput. An exception is the pairSEQ technology developed by Adaptive Technologies that uses bioinformatics to facilitate pairing of TCRα and TCRβ sequences at high-throughput without the need for single cells.[42] Nevertheless, the pairSEQ assay still requires isolated T cells as starting material, thus was not applicable in the studies included in this thesis.

Starting template: genomic DNA or messenger RNA

A dilemma during planning of these studies was whether to use gDNA or messenger RNA (mRNA). In both gDNA and mRNA sequencing strategies, the target sequences must be enriched

using polymerase chain reaction (PCR) strategies in order to efficiently sequence the enormously diverse TCR repertoire. When starting with gDNA, a multiplex PCR strategy is employed, using a mixture of different primers that together amplify every V and J gene that constitute the rearranged TCRα or TCRβ sequences.[252, 258] However, some primers amplify their targets more efficiently than others, introducing bias. Different strategies have been utilized in attempts to overcome the amplification bias within multiplexed PCR.[258] While developing their immunoSEQ assay, Adaptive Biotechnologies utilized a complete synthetic TCR repertoire to assess and correct for PCR amplification bias within the multiplex PCR, adjusting the relative concentrations of each PCR primer.[253] With this strategy, the differences in amplification efficiency were minimized, and 'all' residual bias is computationally removed after sequencing. In spite of this, a few PCR and sequencing errors might remain, but these are difficult to distinguish from biological variation.

Simplified PCR strategies are often employed when mRNA is the starting template due to the close proximity of the gene segments encoding the variable region and the constant region following splicing (see **Figure 7**), ultimately minimizing PCR bias.[258, 259] However, mRNA is highly susceptible to fragmentation, and differences in the efficiency of reverse transcription (*i.e.* the conversion of mRNA to complementary DNA [cDNA]) might affect yield and introduce sequence errors prior to the PCR amplification.

When using gDNA as starting template, every non-productively rearranged TCR sequence will be sequenced along with the productively rearranged, functional TCR sequences. In contrast, only expressed TCR sequences will be sequenced if mRNA is the starting template, thus reducing the amount of data from non-productive TCR sequences. Moreover, strategies have been developed for the sequencing and analysis of full-length TCR mRNA sequences, allowing the exact identification of the V gene used by each rearranged TCR sequence.[260, 261] In comparison, strategies using gDNA as starting template sequence only a minor portion of the V gene, and as a consequence investigators will occasionally be unable to determine which V gene was used in the rearranged TCR sequence. Another advantage of using mRNA as starting template is the possibility to sequence mRNA transcripts of additional genes at the same time. Several investigators have employed this strategy in order to simultaneously investigate the TCR and the function of single-sorted T cells.[43, 44]

The major disadvantage of using mRNA as starting template is that information about the number of T cells in a sample is lost: a given amount of mRNA might represent a large amount of mRNA from a few cells or a small amount of mRNA from many cells. Moreover, different T cells might harbour different number of TCR transcripts.[262, 263] In contrast, their rearranged gDNA template exists as only one copy per T cell. Thus, using gDNA as starting template enables the investigator to accurately determine the quantity of each T-cell clone: the fraction of sequencing reads that

represent a specific TCRβ sequence within a sample should be directly proportional to the fraction of T cells that carry the particular TCR within the sample, *i.e.* should be directly proportional to the size of the T-cell clone.

Ultimately, we decided to use gDNA as starting template for our sequencing studies since we wanted to accurately determine the size of each T-cell clone. For future studies, strategies using mRNA as starting template should be considered if it is possible to isolate and sort single T cells prior to sequencing. This will allow for paired sequencing of full-length TCRα and TCRβ sequences and subsequent functional analyses.

Sequencing depth

One additional factor for consideration was the sequencing depth to be used. Fundamentally, if a given TCR sequence is not observed, it might be because it was not present in the investigated T-cell repertoire or it might be missing from the sequencing data due to insufficient biological or technical sampling. Adaptive Biotechnologies offers several different sequencing depths in their immunoSEQ assay. We sequenced the gut and liver TCRβ repertoires using survey resolution and the peripheral blood TCRβ repertoires using deep resolution, as recommended by Adaptive Biotechnologies for the assessment of non-lymphoid tissues and whole blood, respectively. Sequencing at a higher sequencing depth will increase the probability of detecting sequences from naïve and low-frequent T-cell clones. Nevertheless, the estimation of relative clone frequencies is generally regarded independent of sampling depth: for every additional low-frequent clone that is detected using a higher sequencing depth, additional T cells of the higher-frequent clones are sampled and sequenced as well.

6.2.2 High-throughput TCRβ sequencing

The gDNA was extracted from each tissue sample and peripheral blood sample using various kit-based protocols (Qiagen, Valencia, CA, USA) and a modified salting-out technique[264], respectively. The gDNA was sent to Adaptive Biotechnologies, who performed library preparation, high-throughput TCRβ sequencing and data processing. The immunoSEQ amplification strategy uses a two-step biased-controlled PCR assay (**Figure 19**). First, a multiplex PCR strategy targets the CDR3s using gene-specific primers with universal adapter sequences. The second PCR adds barcodes and Illumina adapters to each sequence to allow for multiplex sequencing. The immunoSEQ assay is compatible with gDNA extracted from FFPE samples.

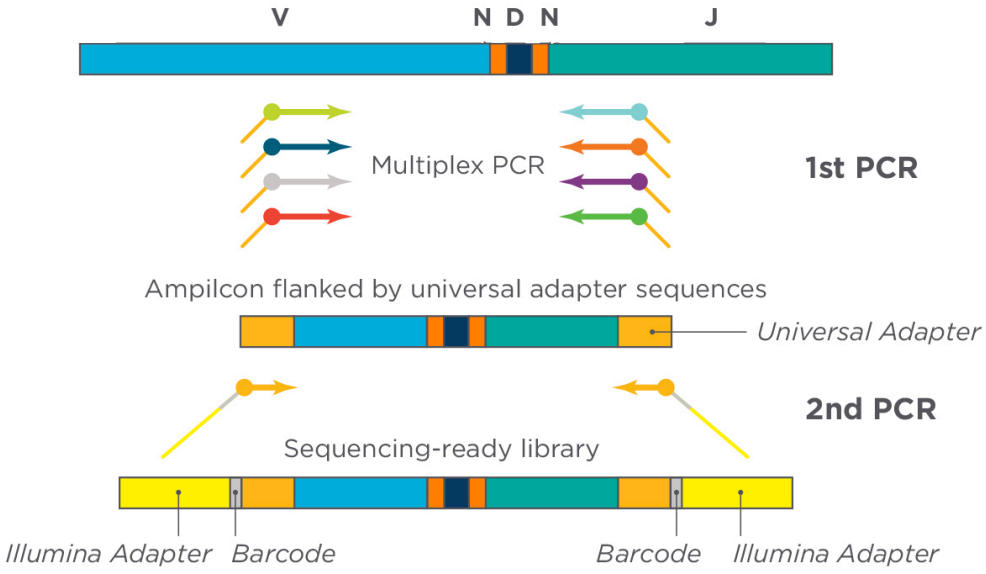

Figure 19. The immunoSEQ assay. A multiplex polymerase chain reaction (PCR) targets the complementarity-determining region 3 (CDR3) of rearranged TCR sequences. On the end of each gene-specific primer is a universal adapter sequence. The adapter sequences allow the addition of barcodes and Illumina adapters in the second PCR. Figure is reproduced and modified with permission from Adaptive Biotechnologies.

Following sequencing and initial data processing, the encoded amino acid sequence representing the CDR3 was determined for each unique TCRβ nucleotide sequence, and the Vβ, Dβ and Jβ genes (and when possible: the alleles) were annotated. The data was further sorted to exclude any non-productively rearranged sequences (*i.e.* sequences with an out-of-frame rearrangement or a stop codon introduced in the CDR3 region). Finally, the frequency of each remaining productively rearranged TCRβ unique nucleotide sequence was determined.

Adaptive Biotechnologies add barcoded synthetic templates to every sample before the PCRs. By assessing these, the amplification efficiency (also called the amplification factor) of each sample can be estimated, allowing for the assessment of the absolute number of T cells in a given sample.

6.2.3 Nomenclature and analyses of TCRβ sequencing data

In this thesis, we performed data analysis using the immunoSEQ Analyzer (http://www.adaptive-biotech.com/immunoseq/analyzer), R statistical software environment (https://www.r-project.org/) and general bioinformatics strategies. Numerous bioinformatics and statistical methods and tools have been developed in recent years, aiming to maximize the information extracted from TCR sequencing data.[265-270] However, no standard pipeline for data processing and analysis has been established yet, thus comparing results across studies is challenging.

Nomenclature

A standardized nomenclature for sequencing data is missing from the field. In paper I, we designated the productively rearranged unique TCRβ nucleotide sequences as 'productive unique sequences' and the unique CDR3 amino acid sequences as 'clonotypes'. However in paper II, we redesignated these sequences as 'nucleotide clonotypes' and 'amino acid clonotypes', respectively. While each nucleotide clonotype generally represents a single T-cell clone, an amino acid clonotype might represent several T-cell clones since different nucleotide clonotypes can encode the same CDR3 amino acid sequence. The nomenclature as defined in paper II will be used in the remainder of this thesis.

In paper I, we searched for amino acid clonotypes present in samples from several patients with a particular disease but absent in the TCRβ sequencing data from patients with either of the two other diseases. We called such amino acid clonotypes 'disease-associated clonotypes'.

In paper I and paper II, we designated the estimated number of T cells as 'number of productive gene rearrangements', as initially suggested by Adaptive Biotechnologies and their immunoSEQ Analyzer. In hindsight, a more precise nomenclature is rather 'number of template molecules' or simply 'number of T cells'.

T-cell diversity

'Clonality' provides a measurement of T-cell diversity that is strictly a function of the spectrum of T-cell clone frequencies in a sample, independent of sampling depth. In a purely diverse T-cell repertoire – with an even distribution of T-cell clones – the clonality score will be 0. In contrast, a clonality score of 1 indicates a monoclonal T-cell repertoire.

In paper I, we further described T-cell diversity by calculating the number of nucleotide clonotypes relative to the number of sequencing reads. Although we normalized by dividing by the total number of (productive) sequencing reads retrieved for each sample, we have later learned that this assessment might be biased by differences in PCR amplification across samples. 'Richness', calculated as the number of nucleotide clonotypes relative to the number of productive gene

rearrangements, might be a better measurement of T-cell diversity since it takes into account the amplification factor of each sample as permitted by the immunoSEQ assay, thus indicating the true clone-to-cell ratio.

We further investigated the clonal distribution of T cells by assessing the cumulative percentages of the 100 most abundant nucleotide clonotypes in each sample, the cumulative percentage distributions of nucleotide clonotypes at various frequencies and the frequency distributions of the nucleotide clonotypes from various diseases and tissues.

Clonotype overlap

In paper I, we investigated whether liver TCRβ repertoires are shared between different diseased individuals by assessing the overlap in amino acid clonotypes. In paper II, we investigated whether T-cell clones are shared between paired gut and liver tissue (within the same individual), and therefore assessed the overlap in nucleotide clonotypes. In both papers, we used an index reporting the proportion of clonotypes that are shared between the two investigated TCRβ repertoires. We further calculated two related overlap indices in paper II: the Jaccard index and the overlap coefficient (also known as the Szymkiewicz-Simpson coefficient).

As described in section 4.2.4, naïve T cells can be retained and activated within the liver.[98, 99] Accumulating evidence indicate naïve T cells also migrate through other non-lymphoid tissues, such as the gut mucosa.[271-273] Naïve T cells can share the same TCRβ nucleotide clonotype but carry different TCRα nucleotide clonotypes (see **Figure 8**).[46] However, in the antigen-experienced T-cell repertoire, the TCRβ nucleotide clonotypes are generally unique to each T-cell clone.[42, 46] It is therefore of importance to omit data from naïve T cells to ensure that detecting the same nucleotide clonotype across tissues indicates T cells of common clonal origin are present in both tissues, and not the presence of naïve T cells with shared TCR β-chain. As sorting of antigen-experienced T cells prior to TCRβ sequencing was impossible in our studies, we decided to set a frequency cut-off when assessing clonotype overlap in paper II in order to exclude data from the least expanded T cells, assuming that naïve T cells where among them. It has been shown that naïve T cells are present in PSC-IBD affected tissues: naïve T cells constitute approximately 30% of the gut T cells of newly diagnosed IBD patients[274] and about 10-20% of the liver T cells in patients with PSC (Liaskou E, unpublished data). Following assessment of the cumulative percentage distribution of nucleotide clonotypes detected at various frequencies, we chose 0.01% as the cut-off for our overlap analysis, hence including the clonotypes of the 63.6% and 60.2% most expanded T cells in the liver and gut samples, respectively. We designated the clonotypes at frequencies higher than 0.01% as 'memory TCRβ repertoire' to emphasize that clonotypes from naïve T cells were not likely present in this fraction. In hindsight, 'expanded TCRβ repertoire' would have been a more accurate term. Among the sequencing reads representing a single clonotype, there might be data

from both effector and memory T cells that originated from the same naïve T cell. Moreover, choosing this rather strict cut-off, we possibly excluded clonotypes originating from the many unexpanded memory T-cell clones.[275]

An important source of error in our studies is the possible presence of peripheral blood T cells in explanted or resected liver tissue and in gut tissue samples due to remaining blood residues in tissues after sampling. In an attempt to avoid 'contaminating' data from such circulating T cells when investigating the overlap of clonotypes in paired PSC-IBD affected tissues, we decided to further exclude any liver or gut nucleotide clonotype in paper II that was observed at frequencies higher than 0.01% in the matched blood sample. Unfortunately, we did not have access to matched peripheral blood samples of the colon cancer patients, nor did we have TCRβ sequencing data from matched blood samples of patients included in paper I.

Overlap analyses are highly sensitive to differences in sample size. In fact, the degree of observed overlap has been shown to be directly related to the sequencing depth.[276] In paper I, fewer nucleotide clonotypes were observed in the TCRβ sequencing data from ALD samples compared to PSC and PBC samples, yet the average number of sequencing reads retrieved were similar across diseases. In paper II, substantially fewer sequencing reads were retrieved from samples of colon cancer patients compared to samples of PSC-IBD patients (the use of gDNA extracted from FFPE tissue and the fact that fewer number of T cells infiltrate non-diseased tissue might partly explain this). We therefore normalized by downsampling the PSC-IBD affected tissue samples and the tumour-adjacent normal tissue samples to the same number of sequencing reads. Following exclusion of five colon cancer patients due to low read count (less than 10,000 reads) of their gut or liver sample, 20,264 reads (which is the size of the smallest remaining sample) were randomly picked from the total repertoire of each PSC-IBD affected tissue sample and remaining tumour-adjacent normal tissue sample. Despite having the same number of sequencing reads in each tissue sample, approximately ten times as many nucleotide clonotypes were identified in PSC-IBD affected tissue samples compared to tumour-adjacent normal tissue samples. Therefore, we further normalized by assessing the overlap only amongst the 100 most abundant nucleotide clonotypes from each sample. We decided to repeat the downsampling procedure and overlap measurements for a total of five times.

Downsampling to the same number of sequencing reads is a well-known strategy for normalization.[277] However, PCR amplification efficiencies might differ between samples. Hence, normalization by same number of sequencing reads in each sample might not necessarily reflect that data from the same number of T cells are analysed from each sample, and might therefore not fully resolve the issue of different sample sizes. In future studies, downsampling to the same

number of productive gene rearrangements could be considered as an alternative normalization strategy, provided that TCR sequencing is performed using the immunoSEQ assay.

V and J gene usage

Finally, we reported the relative frequency of each Vβ and Jβ gene used by various nucleotide clonotypes (*i.e.* T-cell clones). Using this strategy, the frequency of each nucleotide clonotype is ignored, so a skewed (also called biased) usage of certain Vβ or Jβ genes would indicate that a skewed number of T-cell clones carry TCRs encoded by that/those gene(s). An alternative and complimentary strategy would have been to report the frequency of each Vβ and Jβ gene on the T-cell level (*i.e.* sequencing reads), thus taking into account the differences in clonal size. Using the latter strategy, a biased usage of Vβ or Jβ genes in the T-cell repertoire could indicate clonal expansion.

6.2.4 Statistical methods used in paper I and paper II

For normally distributed data in paper I, we used one-way analysis of variance (ANOVA) and two-way ANOVA to compare the three disease groups. In paper II, we compared the clonality scores of the three sample types from PSC-IBD patients and the tumor-adjacent normal tissue samples using the non-parametric Mann-Whitney *U* test (also called the Wilcoxon rank-sum test).

While investigating whether there were any differences in clonal Vβ and Jβ gene usage across tissues and between PSC patients and (disease) controls, we compared the frequency of each Vβ and Jβ gene using the Mann-Whitney *U* test. We used the non-parametric Wilcoxon signed rank test for related samples when investigating whether the overlapping gut-liver clonotypes had a different Vβ and Jβ gene usage compared to the nucleotide clonotypes that were found uniquely in either the gut or the liver sample. In paper I, we did not correct for multiple testing, hence there is an increased risk of type I error (*i.e.* 'false positive' findings). In paper II, we adjusted the p values for the number of Vβ and Jβ genes tested using Benjamini-Hochberg false discovery rate-controlling procedure to reduce the risk of type I error. A p-value less than 0.05 was regarded as statistically significant.

In paper I, we further investigated the V/J gene usage of various sets of nucleotide clonotypes by performing a principal component analysis (PCA) using the 'tcR' package in R.[268] We have later learned that the method that is available in the 'tcR' package is highly sensitive to differences in the total number of nucleotide clonotypes. When investigating the V/J gene usage of 'all clonotypes', different numbers of nucleotide clonotypes had been retrieved from the samples. The PCAs of the V/J gene usage of the 500 and 1000 most abundant clonotypes likely represent true differences in clonal Vβ and Jβ gene usage.

6.3 HLA typing and analyses

6.3.1 Sequencing approaches

Sanger sequencing

The current gold standard in HLA typing is Sanger sequencing, and the minimum requirement is sequencing the exons that encode the peptide-binding groove, *i.e.* exon 2 and 3 of HLA class I genes and exon 2 of HLA class II genes[53] (**Figure 20**). In paper I, HLA typing was partly performed by collaborators using Sanger sequencing protocols: following amplification of exon 2 and 3 of *HLA-B* and *HLA-DRB1*, the amplicons were sequenced and the possible HLA allele combinations were assigned using AssignSBT software (Conexio Genomics, Fremantle, Australia) that compared sequencing results with the IMGT/HLA reference database.[53] Using Sanger technology, amplicons derived from the maternally- and the paternally-inherited chromosomes are sequenced simultaneously and a single consensus sequence is generated. If the genotyped individual is heterozygous for a particular HLA gene (*i.e.* has two different alleles of the particular HLA gene), the polymorphic position(s) within the sequence will appear on the chromatogram as two peaks of different colors at a single nucleotide position, as shown in **Figure 21**. Sometimes, it is impossible to decipher which were the two alleles as two or more different allele combinations could produce identical consensus sequences; this is known as ambiguous allele combinations.[278] There are different approaches to deal with ambiguities, *e.g.* consecutive rounds of sequencing using allele-specific primers, down-scale results to two-digit (one-field) resolution, or select the most probable allele combination based on how frequent these alleles are in the ethnic group matching the study population. However, the latter approach might lead to underestimation of rare HLA alleles and is generally not applicable when assigning the alleles of individuals of admixed or unknown genetic ancestry, as was the case in paper III.

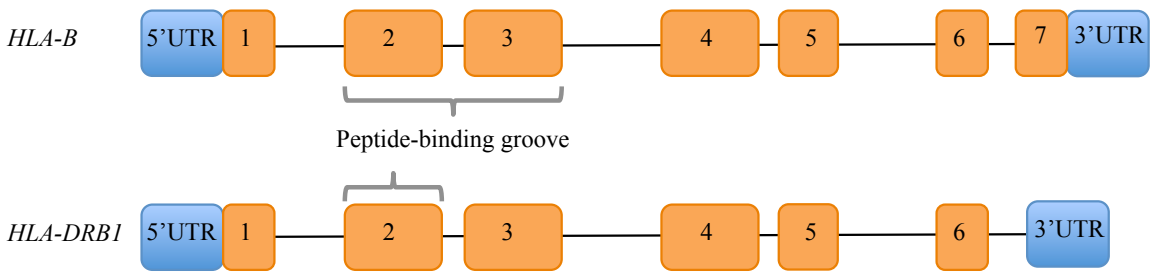

Figure 20. Schematic structure of *HLA-B* and *HLA-DRB1*. Schematic representation of the *HLA-B* and *HLA-DRB1* genes, which encode the α-chain of a human leukocyte antigen (HLA) class I molecule and the β-chain of a HLA class II molecule, respectively. The peptide-binding groove of HLA molecules are encoded by exon 2 and 3 of HLA class I genes (*e.g. HLA-B*) and exon 2 of HLA class II genes (*e.g. HLA-DRB1*). Abbreviation: UTR: untranslated region.

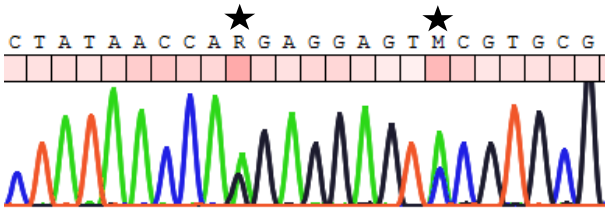

Figure 21. Sanger sequencing chromatogram. Using Sanger technology, the maternally- and the paternally-inherited chromosomes are sequenced simultaneously, generating a single consensus sequence. If the genotyped individual has two different alleles, the nucleotides that differ between the alleles will appear on the chromatogram as two peaks of different colors at a single nucleotide position. In this figure, a star points to the polymorphic positions (also known as heterozygous positions): R denotes an adenine (A) and guanine (G); M denotes an A and cytosine (C). The alleles could be either –A—A– and –G—C– or –A—C– and –G—A–.

High-throughput sequencing

In high-throughput sequencing technologies, the maternally-inherited chromosome and the paternally-inherited chromosome produce separate sequencing reads. This enables phasing of the polymorphic positions (*i.e.* deducing which nucleotides belong to the same allele), hence reducing the number of ambiguities. Two approaches for high-throughput HLA sequencing have been used in this thesis: a PCR-based method developed by GenDx (Utrecht, the Netherlands) was employed on the samples that failed Sanger sequencing in paper I as well as on samples from patients with PSC-IBD in paper II, whereas in paper III, collaborators carried out a capture-based method[279]. The PCR-based method utilized a long-range PCR to enrich the HLA genes of interest. For the enrichment of *HLA-B*, the whole gene was amplified. For the enrichment of the *HLA-DRB1* gene, exon 2–exon 4 (including introns) were amplified. The capture-based method involved the targeted enrichment of HLA genes using sequence-specific biotinylated RNA probes.[279] The probes were hybridized to the DNA library and collected using streptavidin magnetic beads and a magnet. In both approaches, HLA-typing software programs aligned sequencing reads to known allele sequences that are present in the IMGT/HLA reference database,[53] and the best matching alleles were selected using various alignment statistics.[279] Several alleles were annotated at full resolution (eight-digit), yet we chose to report the alleles at four-digit (two-field) resolution for subsequent analyses since the first four digits of an HLA allele name specify the amino acid sequence (see **Figure 11**).[55] Higher resolution of alleles will generally reduce the statistical power of a study.

With the advent of high-throughput sequencing technologies, the number of known HLA alleles has increased vastly (16,251 alleles reported in the IMGT/HLA reference database as of December 2016,[53] compared to approximately 1250 alleles known in 2001[280]) and the number is still increasing. In fact, the technology surpasses the IMGT/HLA reference database and for many alleles, the intron sequence and sequences of exons other than exon 2 and 3 are still missing in the reference database.[53]

6.3.2 Post-genotyping quality control

In order to ensure high quality data sets, it is important to perform a quality control of genotype data. Genotyping success rate might be influenced by several factors, such as DNA quality and appropriate probe and primer design.[281] The HLA typing success rate in paper III, excluding samples with unsuccessful genotyping results, was ≥98.9% for all HLA genes.

6.3.3 Analyses of HLA data

The terms 'haplotype' and 'LD' were introduced in section 4.1.3. In paper III, HLA haplotypes were predicted using the software PHASE v2.1.1.[282-284] A number of factors influence the accuracy of computational haplotype phasing, such as sample size and allele frequency.[285] Increasing the sample size will improve the haplotype phasing accuracy. An allele must be observed several times within its haplotype context in order for the software to confidently phase the haplotypes. As a result, rare alleles are inherently difficult to phase computationally.

In paper III, LD between pairs of HLA alleles were estimated using the software Unphased v.3.0.10.[286] The two most common measures of LD is D' and r^2, with absolute values between 0 and 1.[56] Consider two loci on the same chromosome with alleles A/a and B/b, respectively. A value of D'=1 states that either allele A or allele B always occur on the same haplotype as the other, *e.g.* that the haplotypes detected in the population are AB, aB and ab (not Ab). In contrast to D', the measure of r^2 takes into account the individual allele frequencies. A value of r^2=1 denotes that the alleles A and B always occur together on the same chromosome, *i.e.* that allele A and allele B have the same allele frequencies and the only observed haplotypes are *e.g.* AB and ab. In the example above (with the observed haplotypes AB, aB and ab), allele B is more frequent than allele A and the value of r^2 will be <1. Hence, the value of r^2 indicates to what extent an allele can predict the presence of the other. In contrast, the relative magnitude of D'<1 has no clear interpretation and is difficult to compare between samples. For these reasons, r^2 was reported in paper III.

7 MAIN RESULTS

Paper I: High-throughput T-cell receptor sequencing across chronic liver diseases reveals distinct disease-associated repertoires

In this study, we aimed to characterize the liver T-cell repertoire of patients with PSC using high-throughput TCRβ sequencing. We further sought to assess whether disease-associated TCRβ sequences are evident in the explanted livers of these patients, with patients with PBC and patients with ALD as disease controls. Liver tissue from patients with PSC had on average a higher number of unique TCRβ nucleotide sequences (*i.e.* nucleotide clonotypes) relative to the total number of sequencing reads (2.53%) compared to liver tissue from patients with PBC (1.13%, $p<0.0001$) and ALD (0.62%, $p<0.0001$). On average, PSC samples had a clonality score of 0.134, whereas PBC and ALD samples had clonality scores of 0.153 and 0.146, respectively. The 100 most abundant nucleotide clonotypes constituted on average 25.3% of the TCRβ repertoire in PSC, 29.1% of the TCRβ repertoire in PBC and 31.6% of the TCRβ repertoire in ALD. We detected an average overlap of 0.85% amino acid sequences (*i.e.* amino acid clonotypes) between any two PSC samples, which was significantly higher than the average clonotype overlap detected between any two PBC samples (0.77%, $p=0.024$) and any two ALD samples (0.40%, $p<0.0001$). There was also greater sharing of TCRβ repertoires between any two PSC and PBC samples (0.72%) than between either of them with ALD samples (between PSC and ALD: 0.50%, $p<0.0001$; between PBC and ALD: 0.48%, $p<0.0001$). We identified eight amino acid clonotypes that were common between the liver TCRβ repertoire of at least 30% of patients with PSC but absent from the liver TCRβ repertoires of patients with PBC or ALD. We further identified 42 amino acid clonotypes that were common between the liver TCRβ repertoire of at least 30% of patients with PBC but were not detected in the liver TCRβ repertoires of patients with PSC or ALD, and eight amino acid clonotypes that were common between the liver TCRβ repertoire of at least 30% of patients with ALD but were not detected in the liver TCRβ repertoires of patients with PSC or PBC. Within the TCRβ repertoire of some patients, multiple nucleotide clonotypes encoded the same disease-associated amino acid clonotype. In conclusion, our findings suggest a diverse, polyclonal T-cell repertoire infiltrating the livers of patients with PSC. We identified between eight and 42 potential disease-associated amino acid clonotypes in every disease investigated, with evidence of antigen-driven clonal expansions.

Paper II: Gut and liver T-cells of common clonal origin in primary sclerosing cholangitis-inflammatory bowel disease.

The aim of this study was to investigate whether T cells that infiltrate the gut and the liver of patients with PSC-IBD carry the same TCRs (*i.e.* are clonally related), and further assess whether clonally related T cells are present in paired tumor-adjacent normal gut and liver tissue sampled from patients with colon cancer and liver metastasis. An average of 9.7% (range: 4.7–19.9%) of the nucleotide clonotypes were shared between the paired gut and liver 'memory' TCRβ repertoires of PSC-IBD patients. T cells carrying overlapping gut-liver clonotypes constituted about 15% of the liver and gut 'memory' T cells. A significantly higher clonotype overlap was observed between paired PSC-IBD affected samples compared to paired tumor-adjacent normal gut and liver samples after downsampling to the same number of sequencing reads (8.7% versus 3.6%, p=0.0007) and further when assessing only the 100 most abundant nucleotide clonotypes in each sample (16.6% versus 3.8%, p=0.0117). In summary, we provide the first human-based evidence of gut and liver T cells of common clonal origin. Our data indicate that a high proportion of the memory T cell repertoire in the gut and liver of PSC-IBD patients is capable of recognizing the same or structurally related antigen(s). Our data further suggest this is related to PSC-IBD pathogenesis.

Paper III: HLA haplotypes in primary sclerosing cholangitis patients of admixed and non-European ancestry.

In this study, we sought to investigate to what extent studying non-European or admixed populations might aid in identifying the causative HLA alleles in PSC. We focused our assessment on the three PSC-associated haplotypes that carry the risk alleles HLA-DRB1*13:01 and HLA-DRB1*03:01 and the protective HLA-DRB1*04 alleles. Thirty-five percent of the HLA-DRB1*13:01 haplotypes in the admixed/non-European study population did not carry the HLA-DQB1*06:03 allele. In comparison, every HLA-DRB1*13:01 haplotype in the Scandinavian study population carried this allele. The LD between HLA-B*08:01 and HLA-DRB1*03:01 alleles was weak in the admixed/non-European study population (r^2=0.17) compared to the previously described Scandinavian PSC population (r^2=0.65). On the other hand, every HLA-DRB1*03:01 haplotype carried the HLA-DQB1*02:01 allele, irrespective of ancestry. Every HLA-DRB1*04:01 and HLA-DRB1*04:04 haplotype in the admixed/non-European study population carried the HLA-DQB1*03:01 and HLA-DQB1*03:02 alleles, respectively. These findings suggest that studying admixed or multi-ethnic populations could help in pinpointing the allele responsible for the HLA-DRB1*13:01-DQA1*01:03-DQB1*06:03 association in PSC, and in fine mapping the AH8.1 association to the HLA class I and/or HLA class II region.

8 GENERAL DISCUSSION

8.1 The PSC T-cell repertoire

In PSC, the majority of liver-infiltrating lymphocytes are T cells that are localized to portal triads in proximity to bile ducts (see **Figure 16**).[5, 6] In paper I, we observed a diverse, polyclonal repertoire of T cells in the PSC-affected liver tissue. This contrasts a study that suggests the presence of oligoclonal T-cell repertoires in PSC livers.[287] The previous study was constrained by a limited number and limited specificities of monoclonal anti-TRBV antibodies used in immunohistochemistry, and therefore likely assessed only a fraction of the complete T-cell repertoire.

As briefly discussed in section 6.1.1, it is impossible to determine whether polyclonality is also representative of early inflammatory responses in PSC. We analyzed the T-cell repertoires of explanted liver tissue, thus secondary insults might contribute to the observed polyclonality in our studies. In a study characterizing the synovial T-cell repertoires of patients with RA, investigators observed an oligoclonal T-cell repertoire in early RA, dominated by a few highly expanded T-cell clones.[240] However, in the T-cell repertoires of patients with established RA, a more even distribution of T-cell clones was observed. This supports the hypothesized autoimmune mechanism called 'epitope spreading', in which chronic inflammation leads to the emergence of new self-epitopes that can trigger additional self-reactive T-cell clones.[288] In line with this, oligoclonality rather than polyclonality could potentially characterize the initial T-cell response in PSC.

The immune response towards a specific peptide can include sets of T-cell clones sharing identical or highly similar TCRs.[289] Most antigen-specific T-cell responses involve TCR repertoires that are distinct between individuals. Other antigen-specific T-cell responses comprise TCRs that are shared between multiple individuals; this is known as 'public' T-cell responses. Several mechanisms have been proposed in the establishment of public T-cell responses, including convergent recombination, HLA-restricted selection in thymus and antigen-driven selection. Formerly believed to be a rare phenomenon, public T cells are now frequently being observed in various TCR studies. In paper I, we reported that any two patients with PSC shared on average 0.85% of their liver amino acid clonotypes. Due to the redundancy of the genetic code, different nucleotide clonotypes could encode the same amino acid clonotype across patients. We further reported an average overlap of 0.72% of amino acid clonotypes between any two PSC and PBC patients, which was significantly higher than the sharing detected between either of these patients and ALD patients. Whether this indicates the presence of shared antigens in response to or driving bile duct injury, or simply reflects the higher number of amino acid clonotypes detected in samples from patients with PSC and PBC compared to patients with ALD, could only be speculated.

8.1.1 PSC-associated T-cell clonotypes

In paper I, we identified eight amino acid clonotypes that were detected in the liver TCRβ repertoires of several British PSC patients but not in the liver TCRβ repertoires of patients with PBC or ALD (**Table 3**). In paper II, we further observed the PSC-associated amino acid clonotypes in the liver, gut and blood TCRβ repertoires of Norwegian PSC-IBD patients. Within the TCRβ repertoire of some patients, multiple different nucleotide clonotypes encoded the same disease-associated amino acid clonotype, thus indicating antigen-driven selection of T-cell clones with the PSC-associated amino acid clonotypes.

Table 3. Primary sclerosing cholangitis (PSC)-associated amino acid clonotypes. The table shows the number of samples in which each of the PSC-associated amino acid clonotypes was observed in paper I and paper II.

PSC-associated amino acid clonotypes	Number of British PSC	Number of Norwegian PSC-IBD			Number of normal	
	liver samples, n=20	liver samples, n=10	gut samples, n=10	blood samples, n=10	liver samples, n=10	gut samples, n=10
CASSDTSGGADTQYF	6	1	0	2	1	0
CASSELAGGPETQYF	6	1	4	3	0	0
CASSEYSNQPQHF	7	1	1	3	0	0
CASSFTGTDTQYF	6	0	1	3	0	0
CASSGTSGGADTQYF	6	3	1	2	0	0
CASSLGSGANVLTF	7	5	2	9	0	1
CASSPGQGEGYEQYF	8	2	1	3	1	0
CASSPPSYEQYF	6	1	2	8	0	0

Our findings could possibly reflect the presence of disease-specific pathogenic T cells in the livers of patients with PSC. However, the absence of the eight PSC-associated amino acid clonotypes in PBC or ALD liver samples could also be explained by insufficient biological or technical sampling or by power constraints, thus should be validated in larger cohorts. In paper II, we further detected two of the PSC-associated amino acid clonotypes in tumor-adjacent normal liver tissue sampled from a British colon cancer patient. This might suggest only a few (or possibly none) of the eight PSC-associated amino acid clonotypes are truly unique for PSC-affected liver tissue. Alternatively, the amino acid clonotypes might be enriched in (rather than unique to) PSC-affected liver tissue compared to the livers of healthy individuals or disease controls, which would reflect the presence of disease-related T cells.

Two of the PSC-associated amino acid clonotypes differ only by a single amino acid: CASS**D**TSGGADTQYF and CASS**G**TSGGADTQYF. It could be speculated that the T cells expressing either of these recognize the same antigenic peptide(s). In line with this, studies in celiac disease have identified several T cells that express a shared motif in their CDRβ amino acid clonotype amongst the T cells that are reactive against one specific gluten peptide.[290, 291] Studies

have further revealed that T-cell responses towards dietary gluten peptides are typically dominated by a few clonally expanded T-cell clones with preferred pairing and usage of certain TRAV and TRBV genes across patients, yet also comprise T cells carrying highly diverse amino acid clonotypes.[290-293] Hence, it is possible that the remaining PSC-associated amino acid clonotypes also recognize the same antigenic peptide(s). Alternatively, they might recognize different peptides derived from the same antigen.

8.1.2 Gut and liver T cells of common clonal origin

In paper II, our data indicated that gut and liver T cells of common clonal origin are present in patients with concurrent PSC and IBD: approximately 10% of the nucleotide clonotypes (*i.e.* the T-cell clones) were shared between the gut and liver 'memory' TCRβ repertoires of PSC-IBD patients. The clonotype overlap detected in paired PSC-IBD affected gut and liver tissue was significantly higher than the clonotype overlap detected in paired tumor-adjacent normal gut and liver tissue, as illustrated in **Figure 22**. This suggests that our findings are related to PSC pathogenesis.

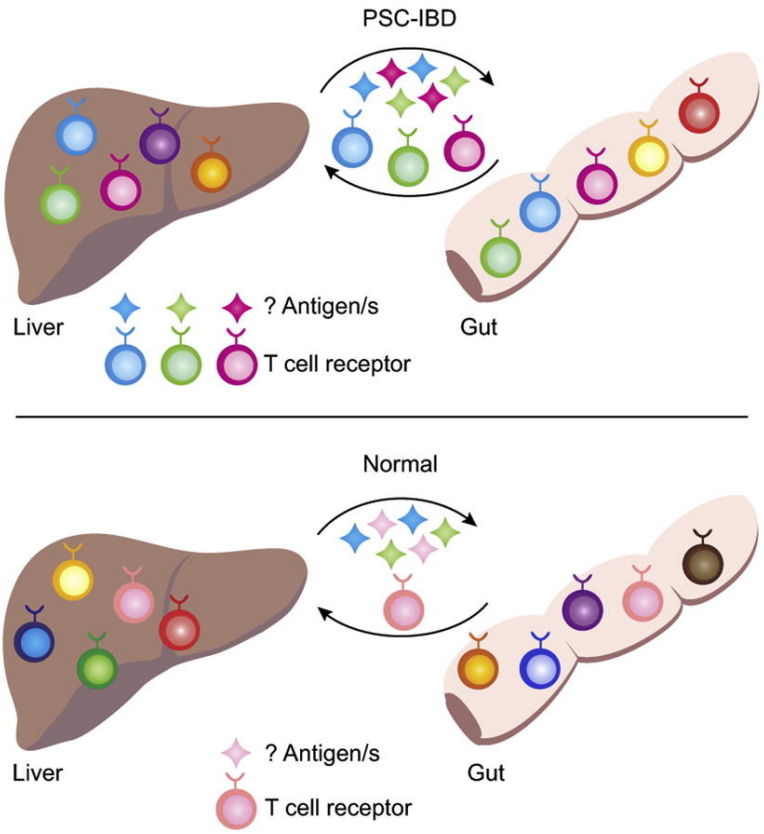

Figure 22. Gut and liver T cells of common clonal origin. Figure illustrates findings in paper II: for the first time it was demonstrated that a proportion of T cells in the human gut and liver originates from the same naïve T-cell clones, hence are able to recognize the same and/or structurally similar antigenic triggers. This proportion is particularly high in patients with concurrent primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD). Reprinted from Henriksen *et al*., Journal of Hepatology copyright (2017),[294] with permission from Elsevier.

Our data are compatible with the proposed 'gut lymphocyte homing' hypothesis, in which an enterohepatic migration of T cells contributes to the pathogenesis of PSC.[92] These findings support efforts to therapeutically target T-cell recruitment in patients with PSC. Vedolizumab is a monoclonal antibody that targets the α4β7 integrin. Vedolizumab is now licensed for the treatment of IBD,[295, 296] and a phase III clinical trial investigating its effect in PSC is currently ongoing (ClinicalTrials.gov identifier NCT03035058).

In our study, we are unable to confirm whether the clonally related T cells were initially activated in the gut or elsewhere (*e.g.* the liver), as this would have required the sampling of tissues at multiple time points, ideally tracking T-cell clonotypes before onset of disease (which is clearly impossible). As described in section 4.3.5, several lines of evidence strongly support a migration of gut-derived T cells to the liver: an aberrant expression of the gut-specific molecules MAdCAM-1 and CCL25 and the presence of $\alpha4\beta7^{+}CCR9^{+}$ effector memory T cells has been detected in the inflamed livers of patients with PSC.[215-217] In further agreement, studies in mice show that $CD8^{+}$ T cells primed in the GALT can migrate to the liver and cause cholangitis when recognizing the same antigen on bile ducts.[218, 297] Murine $CD8^{+}$ T cells that are primed in the liver, on the other hand, are incapable of migrating to the gut.[297] Conversely, a study reported that murine LSECs are able to induce expression of α4β7 and CCR9 on $CD4^{+}$ T cells *in vitro* in a retinoic acid-dependent manner,[298] hence supporting migration of liver-derived $CD4^{+}$ T cells to the gut. The intestinal expression of VAP-1 is greatly increased in patients with IBD,[219] suggesting liver-derived lymphocytes can further enter the inflamed gut by using VAP-1.

Importantly, we do not provide any evidence in paper II that T cells actually migrate between the gut and the liver. The human naïve T-cell pool can persist for decades due to peripheral homeostatic proliferation, generating naïve T-cell clones that each comprise a small number of identical naïve T cells.[256, 299] Hence, the shared gut and liver T-cell repertoire could rather be the result of shared naïve T-cell clones and (structurally similar) antigens occurring at both sites, leading to activation and expansion of the same naïve T-cell clones in the gut and the liver.

The presence of tissue-specific antigens (perhaps a result of epitope spreading) might explain why the remaining nucleotide clonotypes were uniquely detected in either the gut or liver tissue samples. As emphasized earlier, insufficient biological or technical sampling might also explain the absence of a clonotype from sequenced TCRβ repertoires. It is inherently challenging to identify overlapping clonotypes due to the patchy nature of PSC and the use of relatively small tissue samples. In preparation for the study that was described in paper II, we sequenced the gut TCRβ repertoires of two PSC-IBD patients to ensure a considerable number of T-cell clonotypes could be detected in the gDNA extracted from tiny colonic biopsies (1-5mg). Additional colonic biopsies

sampled from the two PSC-IBD patients were later included when sequencing for the main study. We used the immunoSEQ assay for sequencing the TCRβ chain in both the pilot study and the main study. When assessing the overlap of nucleotide clonotypes between the pilot and main study gut TCRβ repertoire for the two patients, we observed an overlap of 17% and 19% nucleotide clonotypes, respectively (data not published). When further assessing the overlap amongst only the 100 most abundant nucleotide clonotypes in each sample, we observed an overlap of 57% and 67% nucleotide clonotypes, respectively (data not published). This suggests sampling variability is substantial, however less considerable for the most abundant nucleotide clonotypes than for the low-frequent nucleotide clonotypes. It can therefore be speculated that the true overlap of nucleotide clonotypes (*i.e.* of T-cell clones) between the gut and the liver of a patient with PSC might be even greater than we reported in paper II.

8.1.3 T-cell subsets

Both $CD4^+$ and $CD8^+$ T cells infiltrate the PSC-affected livers; the CD4/CD8 ratio is 1:1.4.[6] An impaired $CD4^+$ T cell homeostasis has been reported in patients with PSC, with reduced Treg numbers and the presence of the pro-inflammatory Th17 cells in PSC-affected liver tissue.[227, 300] An accumulation of $\alpha4\beta7^+CCR9^+$ effector memory T cells and pro-inflammatory $CD28^-$ T cells in PSC-affected livers has also been reported.[6, 215]

In our studies, we were unable to sort various T-cell subsets prior to TCRβ sequencing, hence the phenotype(s) and function(s) of the sequenced T cells are unknown. Sorting and sequencing separate T-cell subsets might have been informative, as we could have determined whether liver-infiltrating T cells that co-express α4β7 and CCR9 carry overlapping nucleotide clonotypes. However, it has been reported murine LSECs can promote expression of α4β7 and CCR9 on $CD4^+$ T cells,[298] thus it would have been impossible to confirm that the clonally related gut and liver T cells were initially activated in the gut.

The T cells that are described in paper I and paper II could be any type of αβ T cell, including conventional $CD4^+$ and $CD8^+$ T cells, Tregs or even unconventional αβ T cells such as MAIT cells and iNKT cells. The TCRs of MAIT cells and iNKT cells primarily use TRBV6 or TRBV20[301] and TRBV25[302], respectively. Yet, these Vβ genes can also be used by the TCRs of other T cells.[301] Hence, knowledge of the Vβ gene only is not sufficient to identify the unconventional T cells. It has been repeatedly shown that T cells originating from the same naïve T-cell clone can have diverse phenotypes and functions (known as the 'one cell, multiple fates' principle).[43, 254, 303-305] Thus, there might be *e.g.* both Th1 and Th17 cells amongst the T cells carrying identical TCRs.

8.2 The unknown antigen(s) in PSC

Our findings in paper II suggest that the PSC-IBD affected gut and liver tissue share a substantial number of antigen-experienced T-cell clones, thus indicate that the same or structurally similar antigenic trigger(s) are present in the gut and the liver of patients with concurrent PSC and IBD. This strongly supports the notion that exogenous antigen(s) from the gut might trigger T-cell responses in PSC. Viral or bacterial antigen(s) or possibly dietary antigen(s) could break peripheral T-cell tolerance and trigger autoimmunity in PSC by priming self-reactive T cells through cross-recognition (as explained in section 4.1.4). Alternatively, pathogenic $CD4^+$ T cells that are not self-reactive but rather reactive against an exogenous antigen could stimulate the production of autoantibodies by promoting proliferation and differentiation of self-reactive B cells, as seen in celiac disease[74]. In this disease, a continuous exposure to dietary gluten is required for disease progression. If a similar pathogenic mechanism is evident in PSC, this could explain why immunosuppressant drugs appear to not have an effect in PSC. Elimination of the exogenous antigen might then cure the disease in patients with PSC, as seen in patients with celiac disease.

The strong association with IBD and the fact that patients that undergo colon removal at the same time or before liver transplantation have less chance of developing recurrent PSC[142, 144] give further support for the notion of an exogenous trigger in PSC. Studies have in recent years reported the presence of an altered composition of the gut microbiota in patients with PSC, characterized by a reduced diversity and alterations of several bacterial taxa.[306] Whether this finding is relevant for the pathogenesis in PSC is currently unknown.

It is reasonable to presume that the antigenic trigger(s) of the T-cell response in PSC resides within the spectrum of peptides that is presented by PSC-associated HLA molecules, as seen in celiac disease[78]. Hence, defining the ligand characteristics of causative HLA alleles could potentially contribute with relevant information for research aimed at identifying the antigenic trigger(s) in PSC. However, it has not yet been possible to pinpoint the alleles responsible for the HLA association in PSC with certainty because of the high degree of LD extending across large parts of the HLA complex.

8.3 How to identify the causative HLA gene(s) and alleles in PSC

Our findings in paper III suggest that studying further the PSC-associated HLA-DRB1*13:01-DQA1*01:03-DQB1*06:03 haplotype in admixed or multi-ethnic populations of PSC patients and ethnicity-matched healthy controls could aid in defining the causative HLA allele of this haplotype. About one-third of the HLA-DRB1*13:01 haplotypes in the admixed/non-European study population did not carry the HLA-DQB1*06:03 allele. This finding was not unexpected, as we had included several patients of African origin and both HLA-DRB1*13:01-DQB1*06:03 and HLA-DRB1*13:01-DQB1*05:01 are reported common haplotypes in African Americans.[307] Amongst the nine HLA-DRB1*13:01 haplotypes carrying non-DQB1*06:03 alleles in the admixed/non-European study population, three did indeed carry the HLA-DQB1*05:01 allele. In contrast, every HLA-DRB1*13:01 haplotype in the Scandinavian study population carried the HLA-DQB1*06:03 allele, hence illustrating the difficulty in defining the causative HLA gene in this population.

Our findings further suggest that studying admixed or multi-ethnic populations could aid in pinpointing the AH8.1 association in PSC to the HLA class I and/or class II region as we measured a very weak LD between HLA-B*08:01 and HLA-DRB1*03:01 in the admixed/non-European study population compared to the previously described Scandinavian PSC population[251]. However, pinpointing the potential causative allele within the HLA class II region might remain a challenge as every HLA-DRB1*03:01 haplotype carried the HLA-DQB1*02:01 allele, irrespective of ancestry.

The HLA-DRB1*04:01 and HLA-DRB1*04:04 alleles were rarely observed in the admixed/non-European study population. This is not unexpected given their overall low frequency in non-European populations[57, 58] and further their suspected protective role in PSC[186]. In the admixed/non-European study population, every HLA-DRB1*04:01 and HLA-DRB1*04:04 haplotype carried the HLA-DQB1*03:01 and HLA-DQB1*03:02 alleles, respectively. This suggests that studying admixed or multi-ethnic populations might not aid in defining the causative allele in the protective HLA-DRB1*04-DQA1*03-DQB1*03 haplotype.

Taken together, our data support the further assessments of certain PSC-associated HLA haplotypes in admixed or multi-ethnic populations to identify causative HLA gene(s) and alleles in PSC. Importantly, this assumes the same causative alleles across ethnicities. Large-scale trans-ancestry studies have recently reported that the majority of associated risk loci in complex diseases such as IBD and RA were shared across populations of different origins.[308, 309] In ankylosing spondylitis, an association with HLA-B27 is observed in most ethnic groups studied, independently of particular HLA haplotypes.[310] If different HLA alleles are associated across ethnicities, their encoded

molecules might share (similar) peptide-binding motif. Three-dimensional modeling has shown that both HLA-DRB1*13:01 and HLA-DRB1*03:01 encode Asparagine at residue 37, which is a residue reported to be associated with PSC.[185] This induces a positive charge in pocket P9 of the peptide-binding groove of the HLA-DR molecule, hence supporting the notion that disease-associated HLA molecules bind and present a different range of peptides than non-associated molecules.

Our study was not the first assessing HLA haplotypes in African American PSC patients.[311] However, the previous study was constrained by low resolution of the assessed HLA alleles and insufficient HLA class II coverage (only *HLA-A*, *HLA-B* and *HLA-DRB1* genotypes were obtained from the United Network for Organ Sharing [UNOS] database). Nevertheless, the investigators reported an association with HLA-DR13, but it is not known whether this was due to the HLA-DRB1*13:01 allele or other HLA-DR13 subtypes that are frequently found in African Americans.[307] Similar to paper III, the investigators further reported a dissociation of the LD between HLA-B8 and HLA-DR3. When comparing to African American ALD patients, they detected an association with HLA-B8 and not with HLA-DR3 in the African American PSC population. A primary role for HLA-B*08:01 and HLA-DRB1*13:01 in PSC is further supported by HLA SNP association plots (which peak in or near *HLA-B*)[162, 163, 167] and by unconditional and conditional logistic regression analyses in a Scandinavian PSC population.[186] However, it is possible that the true causative HLA alleles are rather alleles in strong LD with HLA-B*08:01 and HLA-DRB1*13:01. Our findings in paper III strongly suggest assessments of HLA in admixed or multi-ethnic populations might provide answers.

9 CONCLUSIONS AND FUTURE PERSPECTIVES

The focus of the present thesis was to characterize the T-cell repertoire of patients with PSC using high-throughput sequencing of the TCRβ chain and further investigate whether studying admixed or multi-ethnic populations might aid in fine mapping the HLA association in PSC. We observed a diverse, polyclonal T-cell repertoire in PSC-affected livers, and detected eight PSC-associated amino acid clonotypes with signs of antigen-driven clonal selection. This may reflect the presence of disease-related pathogenic T cells in PSC-affected livers. We confirmed the presence of gut and liver T cells of common clonal origin in patients with concurrent PSC and IBD, which further indicate that the same or structurally similar antigenic trigger(s) are present in the gut and the liver of these patients. Our findings therefore strongly support the notion that an exogenous antigen(s) from the gut trigger T-cell responses in PSC.

In future studies, the main focus will be to identify the antigenic trigger(s) in PSC. To address this, several parallel approaches are needed. It is reasonable to presume that such antigen(s) resides within the spectrum of peptides that is presented by the PSC-associated HLA molecules. At present, the alleles responsible for the strong HLA association in PSC are unknown. Our data support efforts to systematically collect samples from PSC patients of admixed or non-European ancestry for the purpose of pinpointing the causative HLA alleles in PSC in genetic association analyses. Once the causative HLA alleles have been identified, risk HLA molecules will be extracted from liver tissue of patients using monoclonal antibodies, followed by elution and sequencing of the bound peptides by mass spectrometry. Secondly, soluble and biotinylated risk HLA molecules will be used to enrich particular peptides from digests of proteins isolated from the gut and liver of patients with PSC.[312] By investigating whether these complexes can activate T cells isolated from PSC-affected livers, candidate antigen(s) can potentially be identified.

In parallel, single T cells will be extracted from fresh livers of patients with PSC. This will allow for paired sequencing of full-length TCRα and TCRβ sequences and subsequent functional and structural analyses of *e.g.* the PSC-associated TCRs. For instance, knowledge of both α-chain and β-chain of PSC-associated TCRs will allow the creation of cell lines expressing the specific TCR, which will further render possible the screening of various endogenous and exogenous (*e.g.* bacterial) antigenic peptides for their ability to induce T-cell responses. Notably, the identification of both risk HLA molecules and candidate antigenic peptide(s) will enable the development of HLA-peptide tetramers for the purification and subsequent analyses of disease-relevant pathogenic T cells.[313]

The results from paper II are compatible with the proposed 'gut lymphocyte homing' pathogenic mechanism. However, we were unable to confirm that the clonally related T cells were initially activated in the gut. Following antigen-recognition, the gene segments that encode the BCRs will accumulate mutations during the B-cell proliferation process.[11] Hence, a B-cell clone comprises various antigen-experienced cells with different BCR sequences. BCR-sequencing allows for the construction of lineage trees and enables tracking of individual B-cell clones in various tissues, with B cells that have fewest mutations representing the recently activated B-cell member of the clone.[314] BCR-sequencing of B cells from PSC-IBD affected gut and liver tissue may therefore provide answers to whether lymphocytes are able to migrate between the two tissues in human PSC, and further to the direction of migration (from gut to liver and/or liver to gut).

Collectively, these efforts will hopefully increase our understanding of the pathogenic mechanisms in PSC and enable the development of better therapeutic options for patients with PSC.

10 REFERENCES

1. Hirschfield GM, Karlsen TH, Lindor KD, et al. Primary sclerosing cholangitis. Lancet 2013;382:1587-99.
2. Karlsen TH, Schrumpf E, Boberg KM. Update on primary sclerosing cholangitis. Dig Liver Dis 2010;42:390-400.
3. Fosby B, Melum E, Bjoro K, et al. Liver transplantation in the Nordic countries - An intention to treat and post-transplant analysis from The Nordic Liver Transplant Registry 1982-2013. Scand J Gastroenterol 2015;50:797-808.
4. Folseraas T, Liaskou E, Anderson CA, et al. Genetics in PSC: what do the "risk genes" teach us? Clin Rev Allergy Immunol 2015;48:154-64.
5. Ponsioen CY, Kuiper H, Ten Kate FJ, et al. Immunohistochemical analysis of inflammation in primary sclerosing cholangitis. Eur J Gastroenterol Hepatol 1999;11:769-74.
6. Liaskou E, Jeffery LE, Trivedi PJ, et al. Loss of CD28 expression by liver-infiltrating T cells contributes to pathogenesis of primary sclerosing cholangitis. Gastroenterology 2014;147:221-232.e7.
7. Parkin J, Cohen B. An overview of the immune system. Lancet 2001;357:1777-89.
8. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. Int Rev Immunol 2011;30:16-34.
9. Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. Nat Rev Cancer 2004;4:11-22.
10. Attaf M, Legut M, Cole DK, et al. The T cell antigen receptor: the Swiss army knife of the immune system. Clin Exp Immunol 2015;181:1-18.
11. Batista FD, Harwood NE. The who, how and where of antigen presentation to B cells. Nat Rev Immunol 2009;9:15-27.
12. Neefjes J, Jongsma ML, Paul P, et al. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 2011;11:823-36.
13. Momburg F, Koretz K, Von Herbay A, et al. Nonimmune human cells can express MHC class II antigens in the absence of invariant chain--an immunohistological study on normal and chronically inflamed small intestine. Clin Exp Immunol 1988;72:367-372.
14. Ayres RC, Neuberger JM, Shaw J, et al. Intercellular adhesion molecule-1 and MHC antigens on human intrahepatic bile duct cells: effect of pro-inflammatory cytokines. Gut 1993;34:1245-1249.
15. Joffre OP, Segura E, Savina A, et al. Cross-presentation by dendritic cells. Nat Rev Immunol 2012;12:557-69.
16. Godfrey DI, Uldrich AP, McCluskey J, et al. The burgeoning family of unconventional T cells. Nat Immunol 2015;16:1114-23.
17. Pennock ND, White JT, Cross EW, et al. T cell responses: naïve to memory and everything in between. Adv Physiol Educ 2013;37:273-283.
18. Harty JT, Tvinnereim AR, White DW. CD8+ T cell effector mechanisms in resistance to infection. Annu Rev Immunol 2000;18:275-308.
19. Zhu J, Paul WE. CD4 T cells: fates, functions, and faults. Blood 2008;112:1557-69.
20. Raphael I, Nalawade S, Eagar TN, et al. T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. Cytokine 2015;74:5-17.
21. Shevach EM. Mechanisms of Foxp3+ T Regulatory Cell-Mediated Suppression. Immunity 2009;30:636-645.
22. Grant CR, Liberal R, Mieli-Vergani G, et al. Regulatory T-cells in autoimmune diseases: challenges, controversies and--yet--unanswered questions. Autoimmun Rev 2015;14:105-16.
23. Schroeder HW, Jr., Cavacini L. Structure and function of immunoglobulins. J Allergy Clin Immunol 2010;125:S41-52.
24. Xing Y, Hogquist KA. T-Cell Tolerance: Central and Peripheral. Cold Spring Harb Perspect Biol 2012;4:a006957.
25. Klein L, Kyewski B, Allen PM, et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat Rev Immunol 2014;14:377-91.

26. Onishi Y, Fehervari Z, Yamaguchi T, et al. Foxp3+ natural regulatory T cells preferentially form aggregates on dendritic cells in vitro and actively inhibit their maturation. Proc Natl Acad Sci U S A 2008;105:10113-8.
27. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol 2006;24:419-66.
28. Champagne E. gammadelta T cell receptor ligands and modes of antigen recognition. Arch Immunol Ther Exp (Warsz) 2011;59:117-37.
29. Knapp B, Demharter S, Esmaielbeiki R, et al. Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations. Brief Bioinform 2015;16:1035-1044.
30. Clevers H, Alarcon B, Wileman T, et al. The T cell receptor/CD3 complex: a dynamic protein ensemble. Annu Rev Immunol 1988;6:629-62.
31. Murphy K, Travers P, Walport M, et al. Janeway's Immunobiology. Garland Science 2012.
32. Lefranc MP. Nomenclature of the human T cell receptor genes. Curr Protoc Immunol 2001;40:A.10.1-A.10.23.
33. Lefranc MP. From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). Cold Spring Harb Protoc 2011;2011:627-32.
34. Michie AM, Zuniga-Pflucker JC. Regulation of thymocyte differentiation: pre-TCR signals and beta-selection. Semin Immunol 2002;14:311-23.
35. Stritesky GL, Xing Y, Erickson JR, et al. Murine thymic selection quantified using a unique method to capture deleted T cells. Proc Natl Acad Sci U S A 2013;110:4679-84.
36. Kyewski B, Klein L. A central role for central tolerance. Annu Rev Immunol 2006;24:571-606.
37. Zarnitsyna VI, Evavold BD, Schoettle LN, et al. Estimating the Diversity, Completeness, and Cross-Reactivity of the T Cell Repertoire. Front Immunol 2013;4:485.
38. Sewell AK. Why must T cells be cross-reactive? Nat Rev Immunol 2012;12:669-77.
39. Murugan A, Mora T, Walczak AM, et al. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc Natl Acad Sci U S A 2012;109:16161-16166.
40. Brady BL, Steinel NC, Bassing CH. Antigen receptor allelic exclusion: an update and reappraisal. J Immunol 2010;185:3801-8.
41. Gascoigne NR, Alam SM. Allelic exclusion of the T cell receptor alpha-chain: developmental regulation of a post-translational event. Semin Immunol 1999;11:337-47.
42. Howie B, Sherwood AM, Berkebile AD, et al. High-throughput pairing of T cell receptor alpha and beta sequences. Sci Transl Med 2015;7:301ra131.
43. Stubbington MJ, Lonnberg T, Proserpio V, et al. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods 2016;13:329-32.
44. Han A, Glanville J, Hansmann L, et al. Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat Biotechnol 2014;32:684-92.
45. Niederberger N, Holmberg K, Alam SM, et al. Allelic exclusion of the TCR alpha-chain is an active process requiring TCR-mediated signaling and c-Cbl. J Immunol 2003;170:4557-63.
46. Arstila TP, Casrouge A, Baron V, et al. A direct estimate of the human alphabeta T cell receptor diversity. Science 1999;286:958-61.
47. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. J Biol Chem 2012;287:1168-77.
48. Su LF, Kidd BA, Han A, et al. Virus-specific CD4+ memory-phenotype T cells are abundant in unexposed adults. Immunity 2013;38:373-83.
49. Frankild S, de Boer RJ, Lund O, et al. Amino Acid Similarity Accounts for T Cell Cross-Reactivity and for "Holes" in the T Cell Repertoire. PLoS ONE 2008;3:e1831.
50. Klein J, Sato A. The HLA System. First of two parts. N Engl J Med 2000;343:702-709.
51. Trowsdale J, Knight JC. Major Histocompatibility Complex Genomics and Human Disease. Annu Rev Genomics Hum Genet 2013;14:301-323.
52. Shiina T, Hosomichi K, Inoko H, et al. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 2009;54:15-39.

53. Robinson J, Halliwell JA, Hayhurst JD, et al. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 2015;43:D423-31.
54. Megiorni F, Pizzuti A. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. J Biomed Sci 2012;19:88-88.
55. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens 2010;75:291-455.
56. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002;3:299-309.
57. Gonzalez-Galarza FF, Christmas S, Middleton D, et al. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 2011;39:D913-9.
58. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res 2015;43:D784-8.
59. Horton R, Gibson R, Coggill P, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 2008;60:1-18.
60. Naess S. The major histocompatibility complex association in primary sclerosing cholangitis. Akademika Publishing 2015.
61. Gammon G, Sercarz E. How some T cells escape tolerance induction. Nature 1989;342:183-185.
62. Farh KK-H, Marson A, Zhu J, et al. Genetic and Epigenetic Fine-Mapping of Causal Autoimmune Disease Variants. Nature 2015;518:337-343.
63. Gabrielsen IS, Amundsen SS, Helgeland H, et al. Genetic risk variants for autoimmune diseases that influence gene expression in thymus. Hum Mol Genet 2016;25:3117-3124.
64. Colobran R, Armengol Mdel P, Faner R, et al. Association of an SNP with intrathymic transcription of TSHR and Graves' disease: a role for defective thymic tolerance. Hum Mol Genet 2011;20:3415-23.
65. Giraud M, Taubert R, Vandiedonck C, et al. An IRF8-binding promoter variant and AIRE control CHRNA1 promiscuous expression in thymus. Nature 2007;448:934-7.
66. Vafiadis P, Bennett ST, Todd JA, et al. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 1997;15:289-92.
67. Pugliese A, Zeller M, Fernandez A, Jr., et al. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 1997;15:293-7.
68. Delong T, Wiles TA, Baker RL, et al. Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. Science 2016;351:711-4.
69. Cusick MF, Libbey JE, Fujinami RS. Molecular Mimicry as a Mechanism of Autoimmune Disease. Clin Rev Allergy Immunol 2012;42:102-111.
70. Blichfeldt E, Munthe LA, Rotnes JS, et al. Dual T cell receptor T cells have a decreased sensitivity to physiological ligands due to reduced density of each T cell receptor. Eur J Immunol 1996;26:2876-84.
71. Degauque N, Brouard S, Soulillou J-P. Cross-Reactivity of TCR Repertoire: Current Concepts, Challenges, and Implication for Allotransplantation. Front Immunol 2016;7:89.
72. Curtsinger JM, Lins DC, Mescher MF. CD8+ memory T cells (CD44high, Ly-6C+) are more sensitive than naive cells to (CD44low, Ly-6C-) to TCR/CD8 signaling in response to antigen. J Immunol 1998;160:3236-43.
73. Tselis A. Epstein-Barr virus cause of multiple sclerosis. Curr Opin Rheumatol 2012;24:424-8.
74. Sollid LM, Jabri B. Triggers and drivers of autoimmunity: lessons from coeliac disease. Nat Rev Immunol 2013;13:294-302.
75. Sollid LM, Pos W, Wucherpfennig KW. Molecular mechanisms for contribution of MHC molecules to autoimmune diseases. Curr Opin Immunol 2014;31:24-30.
76. Miyadera H, Tokunaga K. Associations of human leukocyte antigens with autoimmune diseases: challenges in identifying the mechanism. J Hum Genet 2015;60:697-702.
77. Marsh SGE, Parham P, Barber LD. The HLA factsbook. Academic Press 2000.

78. Molberg O, McAdam SN, Korner R, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. Nat Med 1998;4:713-7.
79. Mubarak A, Spierings E, Wolters V, et al. Human leukocyte antigen DQ2.2 and celiac disease. J Pediatr Gastroenterol Nutr 2013;56:428-30.
80. Lundin KE, Scott H, Hansen T, et al. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med 1993;178:187-96.
81. Lundin KE, Scott H, Fausa O, et al. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. Hum Immunol 1994;41:285-91.
82. Bjorck S, Brundin C, Lorinc E, et al. Screening detects a high proportion of celiac disease in young HLA-genotyped children. J Pediatr Gastroenterol Nutr 2010;50:49-53.
83. Andreassen OA, Desikan RS, Wang Y, et al. Abundant genetic overlap between blood lipids and immune-mediated diseases indicates shared molecular genetic mechanisms. PLoS One 2015;10:e0123057.
84. Wang Y, Bos SD, Harbo HF, et al. Genetic overlap between multiple sclerosis and several cardiovascular disease risk factors. Mult Scler 2016;22:1783-1793.
85. Thielens A, Vivier E, Romagne F. NK cell MHC class I specific receptors (KIR): from biology to clinical intervention. Curr Opin Immunol 2012;24:239-45.
86. Evans DM, Spencer CC, Pointon JJ, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 2011;43:761-7.
87. Juza RM, Pauli EM. Clinical and surgical anatomy of the liver: a review for clinicians. Clin Anat 2014;27:764-9.
88. Qin X, Gao B. The complement system in liver diseases. Cell Mol Immunol 2006;3:333-40.
89. Ponziani FR, Pecere S, Gasbarrini A, et al. Physiology and pathophysiology of liver lipid metabolism. Expert Rev Gastroenterol Hepatol 2015;9:1055-67.
90. Chiang JY. Bile acid metabolism and signaling. Compr Physiol 2013;3:1191-212.
91. Jenne CN, Kubes P. Immune surveillance by the liver. Nat Immunol 2013;14:996-1006.
92. Adams DH, Eksteen B. Aberrant homing of mucosal T cells and extra-intestinal manifestations of inflammatory bowel disease. Nat Rev Immunol 2006;6:244-51.
93. Tabibian JH, Masyuk AI, Masyuk TV, et al. Physiology of cholangiocytes. Compr Physiol 2013;3:541-65.
94. Maldonado-Valderrama J, Wilde P, Macierzanka A, et al. The role of bile salts in digestion. Adv Colloid Interface Sci 2011;165:36-46.
95. Hohenester S, Wenniger LM, Paulusma CC, et al. A biliary HCO3- umbrella constitutes a protective mechanism against bile acid-induced injury in human cholangiocytes. Hepatology 2012;55:173-83.
96. Kurioka A, Walker LJ, Klenerman P, et al. MAIT cells: new guardians of the liver. Clin Transl Immunology 2016;5:e98.
97. Fasbender F, Widera A, Hengstler JG, et al. Natural Killer Cells and Liver Fibrosis. Front Immunol 2016;7:19.
98. Tay SS, Wong YC, Roediger B, et al. Intrahepatic activation of naive CD4+ T cells by liver-resident phagocytic cells. J Immunol 2014;193:2087-95.
99. Wuensch SA, Pierce RH, Crispe IN. Local intrahepatic CD8+ T cell activation by a non-self-antigen results in full functional differentiation. J Immunol 2006;177:1689-97.
100. Jeffery HC, van Wilgenburg B, Kurioka A, et al. Biliary epithelium and liver B cells exposed to bacteria activate intrahepatic MAIT cells through MR1. J Hepatol 2016;64:1118-27.
101. Schrumpf E, Tan C, Karlsen TH, et al. The biliary epithelium presents antigens to and activates natural killer T cells. Hepatology 2015;62:1249-59.
102. Crispe IN. Immune tolerance in liver disease. Hepatology 2014;60:2109-17.
103. Syal G, Fausther M, Dranoff JA. Advances in cholangiocyte immunobiology. Am J Physiol Gastrointest Liver Physiol 2012;303:G1077-86.

104. Chen X-M, O'Hara SP, LaRusso NF. The immunobiology of cholangiocytes. Immunol Cell Biol 2008;86:497-505.
105. Leon MP, Bassendine MF, Wilson JL, et al. Immunogenicity of biliary epithelium: investigation of antigen presentation to CD4+ T cells. Hepatology 1996;24:561-7.
106. Barnes BH, Tucker RM, Wehrmann F, et al. Cholangiocytes as immune modulators in rotavirus-induced murine biliary atresia. Liver Int 2009;29:1253-61.
107. Chapman RW, Arborgh BA, Rhodes JM, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. Gut 1980;21:870-7.
108. Portmann B, Zen Y. Inflammatory disease of the bile ducts-cholangiopathies: liver biopsy challenge and clinicopathological correlation. Histopathology 2012;60:236-48.
109. Boberg KM, Aadland E, Jahnsen J, et al. Incidence and prevalence of primary biliary cirrhosis, primary sclerosing cholangitis, and autoimmune hepatitis in a Norwegian population. Scand J Gastroenterol 1998;33:99-103.
110. Lindkvist B, Benito de Valle M, Gullberg B, et al. Incidence and prevalence of primary sclerosing cholangitis in a defined adult population in Sweden. Hepatology 2010;52:571-7.
111. Schrumpf E, Boberg KM, Karlsen TH. Primary sclerosing cholangitis - the Norwegian experience. Scand J Gastroenterol 2015;50:781-96.
112. Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review. J Hepatol 2012;56:1181-8.
113. Kaplan GG, Laupland KB, Butzner D, et al. The burden of large and small duct primary sclerosing cholangitis in adults and children: a population-based analysis. Am J Gastroenterol 2007;102:1042-9.
114. Kingham JG, Kochar N, Gravenor MB. Incidence, clinical patterns, and outcomes of primary sclerosing cholangitis in South Wales, United Kingdom. Gastroenterology 2004;126:1929-30.
115. Bambha K, Kim WR, Talwalkar J, et al. Incidence, clinical spectrum, and outcomes of primary sclerosing cholangitis in a United States community. Gastroenterology 2003;125:1364-9.
116. Card TR, Solaymani-Dodaran M, West J. Incidence and mortality of primary sclerosing cholangitis in the UK: a population-based cohort study. J Hepatol 2008;48:939-44.
117. Ang TL, Fock KM, Ng TM, et al. Clinical profile of primary sclerosing cholangitis in Singapore. J Gastroenterol Hepatol 2002;17:908-13.
118. Escorsell A, Pares A, Rodes J, et al. Epidemiology of primary sclerosing cholangitis in Spain. Spanish Association for the Study of the Liver. J Hepatol 1994;21:787-91.
119. Yimam KK, Bowlus CL. Diagnosis and classification of primary sclerosing cholangitis. Autoimmun Rev 2014;13:445-50.
120. Broome U, Olsson R, Loof L, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut 1996;38:610-5.
121. Weismuller TJ, Wedemeyer J, Kubicka S, et al. The challenges in primary sclerosing cholangitis--aetiopathogenesis, autoimmunity, management and malignancy. J Hepatol 2008;48 Suppl 1:S38-57.
122. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J Hepatol 2009;51:237-67.
123. Chapman R, Fevery J, Kalloo A, et al. Diagnosis and management of primary sclerosing cholangitis. Hepatology 2010;51:660-78.
124. Abdalian R, Heathcote EJ. Sclerosing cholangitis: a focus on secondary causes. Hepatology 2006;44:1063-74.
125. Broome U, Glaumann H, Lindstom E, et al. Natural history and outcome in 32 Swedish patients with small duct primary sclerosing cholangitis (PSC). J Hepatol 2002;36:586-9.
126. Angulo P, Maor-Kendler Y, Lindor KD. Small-duct primary sclerosing cholangitis: a long-term follow-up study. Hepatology 2002;35:1494-500.
127. Bjornsson E, Olsson R, Bergquist A, et al. The natural history of small-duct primary sclerosing cholangitis. Gastroenterology 2008;134:975-80.
128. Björnsson E, Boberg KM, Cullen S, et al. Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut 2002;51:731-735.

129. Goode EC, Rushbrook SM. A review of the medical treatment of primary sclerosing cholangitis in the 21st century. Ther Adv Chronic Dis 2016;7:68-85.
130. Lindor KD. Ursodiol for primary sclerosing cholangitis. Mayo Primary Sclerosing Cholangitis-Ursodeoxycholic Acid Study Group. N Engl J Med 1997;336:691-5.
131. Olsson R, Boberg KM, de Muckadell OS, et al. High-dose ursodeoxycholic acid in primary sclerosing cholangitis: a 5-year multicenter, randomized, controlled study. Gastroenterology 2005;129:1464-72.
132. Triantos CK, Koukias NM, Nikolopoulou VN, et al. Meta-analysis: ursodeoxycholic acid for primary sclerosing cholangitis. Aliment Pharmacol Ther 2011;34:901-10.
133. Lindstrom L, Hultcrantz R, Boberg KM, et al. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2013;11:841-6.
134. Lindor KD, Kowdley KV, Luketic VAC, et al. High Dose Ursodeoxycholic Acid for the Treatment of Primary Sclerosing Cholangitis. Hepatology 2009;50:808-814.
135. Ali AH, Carey EJ, Lindor KD. Current research on the treatment of primary sclerosing cholangitis. Intractable Rare Dis Res 2015;4:1-6.
136. Andersen IM, Fosby B, Boberg KM, et al. Indications and Outcomes in Liver Transplantation in Patients With Primary Sclerosing Cholangitis in Norway. Transplant Direct 2015;1:e39.
137. Boonstra K, Weersma RK, van Erpecum KJ, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045-55.
138. Weismuller TJ, Hansen BE, Trivedi PJ, et al. Gender and IBD phenotype are independent predictors of death or transplantation and of malignancy in Primary Sclerosing Cholangitis – a multicenter retrospective study of the International PSC Study Group (IPSCSG). Hepatology 2015;62:246A.
139. Tischendorf JJ, Hecker H, Kruger M, et al. Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: A single center study. Am J Gastroenterol 2007;102:107-14.
140. Hildebrand T, Pannicke N, Dechene A, et al. Biliary strictures and recurrence after liver transplantation for primary sclerosing cholangitis: A retrospective multicenter analysis. Liver Transpl 2016;22:42-52.
141. Fosby B, Karlsen TH, Melum E. Recurrence and rejection in liver transplantation for primary sclerosing cholangitis. World J Gastroenterol 2012;18:1-15.
142. Alabraba E, Nightingale P, Gunson B, et al. A re-evaluation of the risk factors for the recurrence of primary sclerosing cholangitis in liver allografts. Liver Transpl 2009;15:330-40.
143. Cholongitas E, Shusang V, Papatheodoridis GV, et al. Risk factors for recurrence of primary sclerosing cholangitis after liver transplantation. Liver Transpl 2008;14:138-43.
144. Vera A, Moledina S, Gunson B, et al. Risk factors for recurrence of primary sclerosing cholangitis of liver allograft. Lancet 2002;360:1943-4.
145. Ravikumar R, Tsochatzis E, Jose S, et al. Risk factors for recurrent primary sclerosing cholangitis after liver transplantation. J Hepatol;63:1139-1146.
146. Lunder AK, Hov JR, Borthne A, et al. Prevalence of Sclerosing Cholangitis Detected by Magnetic Resonance Cholangiography in Patients With Long-term Inflammatory Bowel Disease. Gastroenterology 2016;151:660-669.e4.
147. Grant AJ, Lalor PF, Salmi M, et al. Homing of mucosal lymphocytes to the liver in the pathogenesis of hepatic complications of inflammatory bowel disease. Lancet 2002;359:150-7.
148. Riley TR, Schoen RE, Lee RG, et al. A case series of transplant recipients who despite immunosuppression developed inflammatory bowel disease. Am J Gastroenterol 1997;92:279-82.
149. Befeler AS, Lissoos TW, Schiano TD, et al. Clinical course and management of inflammatory bowel disease after liver transplantation. Transplantation 1998;65:393-6.

150. Verdonk RC, Dijkstra G, Haagsma EB, et al. Inflammatory bowel disease after liver transplantation: risk factors for recurrence and de novo disease. Am J Transplant 2006;6:1422-9.
151. Loftus EV, Jr., Harewood GC, Loftus CG, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. Gut 2005;54:91-6.
152. Jorgensen KK, Grzyb K, Lundin KE, et al. Inflammatory bowel disease in patients with primary sclerosing cholangitis: clinical characterization in liver transplanted and nontransplanted patients. Inflamm Bowel Dis 2012;18:536-45.
153. Sinakos E, Samuel S, Enders F, et al. Inflammatory bowel disease in primary sclerosing cholangitis: a robust yet changing relationship. Inflamm Bowel Dis 2013;19:1004-9.
154. Aadland E, Schrumpf E, Fausa O, et al. Primary sclerosing cholangitis: a long-term follow-up study. Scand J Gastroenterol 1987;22:655-64.
155. Lundqvist K, Broome U. Differences in colonic disease activity in patients with ulcerative colitis with and without primary sclerosing cholangitis: a case control study. Dis Colon Rectum 1997;40:451-6.
156. Boberg KM, Chapman RW, Hirschfield GM, et al. Overlap syndromes: the International Autoimmune Hepatitis Group (IAIHG) position statement on a controversial issue. J Hepatol 2011;54:374-85.
157. Saarinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. Am J Gastroenterol 2000;95:3195-9.
158. Ghouri YA, Mian I, Blechacz B. Cancer review: Cholangiocarcinoma. J Carcinog 2015;14:1.
159. Bergquist A, Ekbom A, Olsson R, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. J Hepatol 2002;36:321-7.
160. Rioux JD, Abbas AK. Paths to understanding the genetic basis of autoimmune disease. Nature 2005;435:584-589.
161. Bergquist A, Montgomery SM, Bahmanyar S, et al. Increased risk of primary sclerosing cholangitis and ulcerative colitis in first-degree relatives of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2008;6:939-43.
162. Karlsen TH, Franke A, Melum E, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology 2010;138:1102-11.
163. Melum E, Franke A, Schramm C, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet 2011;43:17-9.
164. Srivastava B, Mells GF, Cordell HJ, et al. Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis. Scand J Gastroenterol 2012;47:820-6.
165. Folseraas T, Melum E, Rausch P, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol 2012;57:366-75.
166. Ellinghaus D, Folseraas T, Holm K, et al. Genome-wide association analysis in primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. Hepatology 2013;58:1074-83.
167. Liu JZ, Hov JR, Folseraas T, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet 2013;45:670-5.
168. Ellinghaus D, Jostins L, Spain SL. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48:510-8.
169. Ji SG, Juran BD, Mucha S, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet 2017;49:269-273.
170. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. Genome Med 2016;8:78.
171. Chapman RW, Varghese Z, Gaul R, et al. Association of primary sclerosing cholangitis with HLA-B8. Gut 1983;24:38-41.
172. Schrumpf E, Fausa O, Forre O, et al. HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. Scand J Gastroenterol 1982;17:187-91.

173. Donaldson PT, Farrant JM, Wilkinson ML, et al. Dual association of HLA DR2 and DR3 with primary sclerosing cholangitis. Hepatology 1991;13:129-33.
174. Farrant JM, Doherty DG, Donaldson PT, et al. Amino acid substitutions at position 38 of the DRβ polypeptide confer susceptibility to and protection from primary sclerosing cholangitis. Hepatology 1992;16:390-395.
175. Olerup O, Olsson R, Hultcrantz R, et al. HLA-DR and HLA-DQ are not markers for rapid disease progression in primary sclerosing cholangitis. Gastroenterology 1995;108:870-8.
176. Mehal WZ, Lo YM, Wordsworth BP, et al. HLA DR4 is a marker for rapid disease progression in primary sclerosing cholangitis. Gastroenterology 1994;106:160-7.
177. Moloney MM, Thomson LJ, Strettell MJ, et al. Human leukocyte antigen-C genes and susceptibility to primary sclerosing cholangitis. Hepatology 1998;28:660-2.
178. Bernal W, Moloney M, Underhill J, et al. Association of tumor necrosis factor polymorphism with primary sclerosing cholangitis. J Hepatol 1999;30:237-41.
179. Mitchell SA, Grove J, Spurkland A, et al. Association of the tumour necrosis factor alpha -308 but not the interleukin 10 -627 promoter polymorphism with genetic susceptibility to primary sclerosing cholangitis. Gut 2001;49:288-94.
180. Norris S, Kondeatis E, Collins R, et al. Mapping MHC-encoded susceptibility and resistance in primary sclerosing cholangitis: the role of MICA polymorphism. Gastroenterology 2001;120:1475-82.
181. Wiencke K, Spurkland A, Schrumpf E, et al. Primary sclerosing cholangitis is associated to an extended B8-DR3 haplotype including particular MICA and MICB alleles. Hepatology 2001;34:625-30.
182. Karlsen TH, Boberg KM, Vatn M, et al. Different HLA class II associations in ulcerative colitis patients with and without primary sclerosing cholangitis. Genes Immun 2007;8:275-278.
183. Wiencke K, Karlsen TH, Boberg KM, et al. Primary sclerosing cholangitis is associated with extended HLA-DR3 and HLA-DR6 haplotypes. Tissue Antigens 2007;69:161-9.
184. Donaldson PT, Norris S. Evaluation of the role of MHC class II alleles, haplotypes and selected amino acid sequences in primary sclerosing cholangitis. Autoimmunity 2002;35:555-64.
185. Hov JR, Kosmoliaptsis V, Traherne JA, et al. Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. Hepatology 2011;53:1967-76.
186. Naess S, Lie BA, Melum E, et al. Refinement of the MHC risk map in a scandinavian primary sclerosing cholangitis population. PLoS One 2014;9:e114486.
187. Candore G, Lio D, Colonna Romano G, et al. Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. Autoimmun Rev 2002;1:29-35.
188. Gregersen PK, Kosoy R, Lee AT, et al. Risk for myasthenia gravis maps to a 151Pro→Ala change in TNIP1 and to human leukocyte antigen-B*08. Ann Neurol 2012;72:927-935.
189. Noble JA, Erlich HA. Genetics of Type 1 Diabetes. Cold Spring Harb Perspect Med 2012;2:a007732.
190. Sollid LM. Molecular basis of celiac disease. Annu Rev Immunol 2000;18:53-81.
191. Hov JR, Lleo A, Selmi C, et al. Genetic associations in Italian primary sclerosing cholangitis: heterogeneity across Europe defines a critical role for HLA-C. J Hepatol 2010;52:712-7.
192. Spurkland A, Saarinen S, Boberg KM, et al. HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. Tissue Antigens 1999;53:459-69.
193. Naess S, Bjornsson E, Anmarkrud JA, et al. Small duct primary sclerosing cholangitis without inflammatory bowel disease is genetically different from large duct disease. Liver Int 2014;34:1488-95.
194. Duarte-Rey C, Pardo AL, Rodriguez-Velosa Y, et al. HLA class II association with autoimmune hepatitis in Latin America: a meta-analysis. Autoimmun Rev 2009;8:325-31.
195. Hov JR, Boberg KM, Taraldsrud E, et al. Antineutrophil antibodies define clinical and genetic subgroups in primary sclerosing cholangitis. Liver Int 2017;37:458-465.

196. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet 2003;361:598-604.
197. Oksenberg JR, Barcellos LF, Cree BA, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet 2004;74:160-7.
198. Sharpe AH, Freeman GJ. The B7-CD28 superfamily. Nat Rev Immunol 2002;2:116-26.
199. Tian Y, Zajac AJ. IL-21 and T Cell Differentiation: Consider the Context. Trends Immunol 2016;37:557-68.
200. Kuchen S, Robbins R, Sims GP, et al. Essential role of IL-21 in B cell activation, expansion, and plasma cell generation during CD4+ T cell-B cell collaboration. J Immunol 2007;179:5886-96.
201. Boyman O, Sprent J. The role of interleukin-2 during homeostasis and activation of the immune system. Nat Rev Immunol 2012;12:180-190.
202. Bo X, Broome U, Remberger M, et al. Tumour necrosis factor alpha impairs function of liver derived T lymphocytes and natural killer cells in patients with primary sclerosing cholangitis. Gut 2001;49:131-41.
203. Henriksen EK, Melum E, Karlsen TH. Update on primary sclerosing cholangitis genetics. Curr Opin Gastroenterol 2014;30:310-9.
204. Andersen IM, Tengesdal G, Lie BA, et al. Effects of coffee consumption, smoking, and hormones on risk for primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2014;12:1019-28.
205. Loftus EV, Jr., Sandborn WJ, Tremaine WJ, et al. Primary sclerosing cholangitis is associated with nonsmoking: a case-control study. Gastroenterology 1996;110:1496-502.
206. van Erpecum KJ, Smits SJ, van de Meeberg PC, et al. Risk of primary sclerosing cholangitis is associated with nonsmoking behavior. Gastroenterology 1996;110:1503-6.
207. Mitchell SA, Thyssen M, Orchard TR, et al. Cigarette smoking, appendectomy, and tonsillectomy as risk factors for the development of primary sclerosing cholangitis: a case control study. Gut 2002;51:567-73.
208. Pollheimer MJ, Halilbasic E, Fickert P, et al. Pathogenesis of primary sclerosing cholangitis. Best Pract Res Clin Gastroenterol 2011;25:727-739.
209. Yamada S, Ishii M, Liang LS, et al. Small duct cholangitis induced by N-formyl L-methionine L-leucine L-tyrosine in rats. J Gastroenterol 1994;29:631-6.
210. Lichtman SN, Keku J, Clark RL, et al. Biliary tract disease in rats with experimental small bowel bacterial overgrowth. Hepatology 1991;13:766-72.
211. Schweighoffer T, Tanaka Y, Tidswell M, et al. Selective expression of integrin alpha 4 beta 7 on a subset of human CD4+ memory T cells with Hallmarks of gut-trophism. J Immunol 1993;151:717-29.
212. Mora JR, Bono MR, Manjunath N, et al. Selective imprinting of gut-homing T cells by Peyer's patch dendritic cells. Nature 2003;424:88-93.
213. Berlin C, Berg EL, Briskin MJ, et al. Alpha 4 beta 7 integrin mediates lymphocyte binding to the mucosal vascular addressin MAdCAM-1. Cell 1993;74:185-95.
214. Zabel BA, Agace WW, Campbell JJ, et al. Human G protein-coupled receptor GPR-9-6/CC chemokine receptor 9 is selectively expressed on intestinal homing T lymphocytes, mucosal lymphocytes, and thymocytes and is required for thymus-expressed chemokine-mediated chemotaxis. J Exp Med 1999;190:1241-56.
215. Eksteen B, Grant AJ, Miles A, et al. Hepatic endothelial CCL25 mediates the recruitment of CCR9+ gut-homing lymphocytes to the liver in primary sclerosing cholangitis. J Exp Med 2004;200:1511-7.
216. Grant AJ, Lalor PF, Hubscher SG, et al. MAdCAM-1 expressed in chronic inflammatory liver disease supports mucosal lymphocyte adhesion to hepatic endothelium (MAdCAM-1 in chronic inflammatory liver disease). Hepatology 2001;33:1065-72.
217. Hillan KJ, Hagler KE, MacSween RN, et al. Expression of the mucosal vascular addressin, MAdCAM-1, in inflammatory liver disease. Liver 1999;19:509-18.
218. Seidel D, Eickmeier I, Kuhl AA, et al. CD8 T cells primed in the gut-associated lymphoid tissue induce immune-mediated cholangitis in mice. Hepatology 2014;59:601-11.
219. Salmi M, Kalimo K, Jalkanen S. Induction and function of vascular adhesion protein-1 at sites of inflammation. J Exp Med 1993;178:2255-60.

220. Liaskou E, Karikoski M, Reynolds GM, et al. Regulation of mucosal addressin cell adhesion molecule 1 expression in human and mice by vascular adhesion protein 1 amine oxidase activity. Hepatology 2011;53:661-72.
221. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med 2002;347:911-20.
222. Hov JR, Boberg KM, Karlsen TH. Autoantibodies in primary sclerosing cholangitis. World J Gastroenterol 2008;14:3781-3791.
223. Chung BK, Guevel BT, Reynolds GM, et al. Phenotyping and auto-antibody production by liver-infiltrating B cells in primary sclerosing cholangitis and primary biliary cholangitis. J Autoimmun 2017;77:45-54.
224. Terjung B, Sohne J, Lechtenberg B, et al. p-ANCAs in autoimmune liver disorders recognise human beta-tubulin isotype 5 and cross-react with microbial protein FtsZ. Gut 2010;59:808-16.
225. Xu B, Broome U, Ericzon BG, et al. High frequency of autoantibodies in patients with primary sclerosing cholangitis that bind biliary epithelial cells and induce expression of CD44 and production of interleukin 6. Gut 2002;51:120-7.
226. Karrar A, Broome U, Sodergren T, et al. Biliary epithelial cell antibodies link adaptive and innate immune responses in primary sclerosing cholangitis. Gastroenterology 2007;132:1504-14.
227. Sebode M, Peiseler M, Franke B, et al. Reduced FOXP3+ regulatory T cells in patients with primary sclerosing cholangitis are associated with IL2RA gene polymorphisms. J Hepatol 2014;60:1010-6.
228. Schoknecht T, Schwinge D, Stein S, et al. CD4+ T cells from patients with primary sclerosing cholangitis exhibit reduced apoptosis and down-regulation of proapoptotic Bim in peripheral blood. J Leukoc Biol 2017;101:589-597.
229. Whitacre CC. Sex differences in autoimmune disease. Nat Immunol 2001;2:777-80.
230. Lee W, Reveille JD, Davis JC, Jr., et al. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. Ann Rheum Dis 2007;66:633-8.
231. Chazouilleres O. Primary sclerosing cholangitis and bile acids. Clin Res Hepatol Gastroenterol 2012;36 Suppl 1:S21-5.
232. Fickert P, Fuchsbichler A, Wagner M, et al. Regurgitation of bile acids from leaky bile ducts causes sclerosing cholangitis in Mdr2 (Abcb4) knockout mice. Gastroenterology 2004;127:261-74.
233. Gilbert MT, Haselkorn T, Bunce M, et al. The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? PLoS One 2007;2:e537.
234. Hashimoto E, Lindor KD, Homburger HA, et al. Immunohistochemical characterization of hepatic lymphocytes in primary biliary cirrhosis in comparison with primary sclerosing cholangitis and autoimmune chronic active hepatitis. Mayo Clin Proc 1993;68:1049-55.
235. Haydon G, Lalor PF, Hubscher SG, et al. Lymphocyte recruitment to the liver in alcoholic liver disease. Alcohol 2002;27:29-36.
236. Carey EJ, Ali AH, Lindor KD. Primary biliary cirrhosis. The Lancet;386:1565-1575.
237. Gao B, Bataller R. Alcoholic liver disease: pathogenesis and new therapeutic targets. Gastroenterology 2011;141:1572-85.
238. Sherwood AM, Emerson RO, Scherer D, et al. Tumor-infiltrating lymphocytes in colorectal tumors display a diversity of T cell receptor sequences that differ from the T cells in adjacent mucosal tissue. Cancer Immunol Immunother 2013;62:1453-61.
239. Lossius A, Johansen JN, Vartdal F, et al. High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T cells. Eur J Immunol 2014;44:3439-52.
240. Klarenbeek PL, de Hair MJ, Doorenspleet ME, et al. Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. Ann Rheum Dis 2012;71:1088-93.
241. Smyk DS, Rigopoulou EI, Pares A, et al. Sex Differences Associated with Primary Biliary Cirrhosis. Clin Dev Immunol 2012;2012:610504.
242. Yang AL, Vadhavkar S, Singh G, et al. Epidemiology of alcohol-related liver and pancreatic disease in the United States. Arch Intern Med 2008;168:649-56.

243. Mells GF, Kaser A, Karlsen TH. Novel insights into autoimmune liver diseases provided by genome-wide association studies. J Autoimmun 2013;46:41-54.
244. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989;170:2-6; discussion 16-9.
245. Heidt S, Roelen DL, Eijsink C, et al. Calcineurin inhibitors affect B cell antibody responses indirectly by interfering with T cell help. Clin Exp Immunol 2010;159:199-207.
246. Miller JL, Ericson SG. Cyclosporin a and tacrolimus (FK506) differentially alter T-cell receptor expression in vivo. Immunopharmacol Immunotoxicol 2007;29:105-18.
247. Zhang X, Kelaria S, Kerstetter J, et al. The functional and prognostic implications of regulatory T cells in colorectal carcinoma. J Gastrointest Oncol 2015;6:307-313.
248. Maeda K, Hazama S, Tokuno K, et al. Impact of chemotherapy for colorectal cancer on regulatory T-cells and tumor immunity. Anticancer Res 2011;31:4569-74.
249. Britanova OV, Putintseva EV, Shugay M, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. J Immunol 2014;192:2689-98.
250. Liu Y, Nyunoya T, Leng S, et al. Softwares and methods for estimating genetic ancestry in human populations. Hum Genomics 2013;7:1-1.
251. Karlsen TH, Boberg KM, Olsson M, et al. Particular genetic variants of ligands for natural killer cell receptors may contribute to the HLA associated risk of primary sclerosing cholangitis. J Hepatol 2007;46:899-906.
252. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood 2009;114:4099-107.
253. Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun 2013;4:2680.
254. Gaide O, Emerson RO, Jiang X, et al. Common clonal origin of central and resident memory T cells following skin immunization. Nat Med 2015;21:647-53.
255. Faham M, Carlton V, Moorhead M, et al. Discovery of T-Cell Receptor Beta Motifs Specific to HLA-B27+ Ankylosing Spondylitis by Deep Repertoire Sequence Analysis. Arthritis Rheumatol 2016; (*in press*). doi: 10.1002/art.40028
256. Thome JJC, Grinshpun B, Kumar BV, et al. Long-term maintenance of human naïve T cells through in situ homeostasis in lymphoid tissue sites. Sci Immunol 2016;1:eaah6506.
257. Turchaninova MA, Britanova OV, Bolotin DA, et al. Pairing of T-cell receptor chains via emulsion PCR. Eur J Immunol 2013;43:2507-15.
258. Robins H. Immunosequencing: applications of immune repertoire deep sequencing. Curr Opin Immunol 2013;25:646-52.
259. Mamedov IZ, Britanova OV, Zvyagin IV, et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. Front Immunol 2013;4:456.
260. Quigley MF, Almeida JR, Price DA, et al. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. Curr Protoc Immunol 2011;Chapter 10:Unit10.33.
261. Redmond D, Poran A, Elemento O. Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. Genome Med 2016;8:80.
262. Paillard F, Sterkers G, Bismuth G, et al. Lymphokine mRNA and T cell multireceptor mRNA of the Ig super gene family are reciprocally modulated during human T cell activation. Eur J Immunol 1988;18:1643-6.
263. Paillard F, Sterkers G, Vaquero C. Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal human T lymphocytes. EMBO J 1990;9:1867-1872.
264. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
265. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12:380-1.

266. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb Protoc 2011;2011:695-715.
267. Kuchenbecker L, Nienen M, Hecht J, et al. IMSEQ--a fast and error aware approach to immunogenetic sequence analysis. Bioinformatics 2015;31:2963-71.
268. Nazarov VI, Pogorelyy MV, Komech EA, et al. tcR: an R package for T cell receptor repertoire advanced data analysis. BMC Bioinformatics 2015;16:175.
269. Vander Heiden JA, Yaari G, Uduman M, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 2014;30:1930-2.
270. Bagaev DV, Zvyagin IV, Putintseva EV, et al. VDJviz: a versatile browser for immunogenomics data. BMC Genomics 2016;17:453.
271. Lewis M, Tarlton JF, Cose S. Memory versus naive T-cell migration. Immunol Cell Biol 2008;86:226-31.
272. Sathaliyawala T, Kubota M, Yudanin N, et al. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. Immunity 2013;38:187-97.
273. Thome JJ, Yudanin N, Ohmura Y, et al. Spatial map of human T cell compartmentalization and maintenance over decades of life. Cell 2014;159:814-28.
274. Horjus Talabur Horje CS, Middendorp S, van Koolwijk E, et al. Naive T cells in the gut of newly diagnosed, untreated adult patients with inflammatory bowel disease. Inflamm Bowel Dis 2014;20:1902-9.
275. Klarenbeek PL, Tak PP, van Schaik BD, et al. Human T-cell memory consists mainly of unexpanded clones. Immunol Lett 2010;133:42-8.
276. Shugay M, Bolotin DA, Putintseva EV, et al. Huge Overlap of Individual TCR Beta Repertoires. Front Immunol 2013;4:466.
277. Woodsworth DJ, Castellarin M, Holt RA. Sequence analysis of T-cell repertoires in health and disease. Genome Med 2013;5:98.
278. Adams SD, Barracchini KC, Chen D, et al. Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification. J Transl Med 2004;2:30.
279. Wittig M, Anmarkrud JA, Kassens JC, et al. Development of a high-resolution NGS-based HLA-typing and analysis pipeline. Nucleic Acids Res 2015;43:e70.
280. Gerlach JA. Human lymphocyte antigen molecular typing: how to identify the 1250+ alleles out there. Arch Pathol Lab Med 2002;126:281-4.
281. Pompanon F, Bonin A, Bellemain E, et al. Genotyping errors: causes, consequences and solutions. Nat Rev Genet 2005;6:847-59.
282. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978-89.
283. Stephens M, Scheet P. Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. Am J Hum Genet 2005;76:449-462.
284. Bettencourt BF, Santos MR, Fialho RN, et al. Evaluation of two methods for computational HLA haplotypes inference using a real dataset. BMC Bioinformatics 2008;9:68.
285. Browning SR, Browning BL. Haplotype phasing: Existing methods and new developments. Nature Reviews. Genetics 2011;12:703-714.
286. Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered 2008;66:87-98.
287. Broome U, Grunewald J, Scheynius A, et al. Preferential V beta3 usage by hepatic T lymphocytes in patients with primary sclerosing cholangitis. J Hepatol 1997;26:527-34.
288. Vanderlugt CL, Miller SD. Epitope spreading in immune-mediated diseases: implications for immunotherapy. Nat Rev Immunol 2002;2:85-95.
289. Li H, Ye C, Ji G, et al. Determinants of public T cell responses. Cell Res 2012;22:33-42.
290. Qiao SW, Christophersen A, Lundin KE, et al. Biased usage and preferred pairing of alpha- and beta-chains of TCRs specific for an immunodominant gluten epitope in coeliac disease. Int Immunol 2014;26:13-9.
291. Qiao SW, Raki M, Gunnarsen KS, et al. Posttranslational modification of gluten shapes TCR usage in celiac disease. J Immunol 2011;187:3064-71.

292. Dahal-Koirala S, Risnes LF, Christophersen A, et al. TCR sequencing of single cells reactive to DQ2.5-glia-alpha2 and DQ2.5-glia-omega2 reveals clonal expansion and epitope-specific V-gene usage. Mucosal Immunol 2016;9:587-96.
293. Broughton SE, Petersen J, Theodossis A, et al. Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. Immunity 2012;37:611-21.
294. Henriksen EK, Jorgensen KK, Kaveh F, et al. Gut and liver T-cells of common clonal origin in primary sclerosing cholangitis-inflammatory bowel disease. J Hepatol 2017;66:116-122.
295. Jin Y, Lin Y, Lin LJ, et al. Meta-analysis of the effectiveness and safety of vedolizumab for ulcerative colitis. World J Gastroenterol 2015;21:6352-60.
296. Stallmach A, Langbein C, Atreya R, et al. Vedolizumab provides clinical benefit over 1 year in patients with active inflammatory bowel disease - a prospective multicenter observational study. Aliment Pharmacol Ther 2016;44:1199-1212.
297. Eickmeier I, Seidel D, Grun JR, et al. Influence of CD8 T cell priming in liver and gut on the enterohepatic circulation. J Hepatol 2014;60:1143-50.
298. Neumann K, Kruse N, Szilagyi B, et al. Connecting liver and gut: murine liver sinusoidal endothelium induces gut tropism of CD4+ T cells via retinoic acid. Hepatology 2012;55:1976-84.
299. den Braber I, Mugwagwa T, Vrisekoop N, et al. Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. Immunity 2012;36:288-97.
300. Katt J, Schwinge D, Schoknecht T, et al. Increased T helper type 17 response to pathogen stimulation in patients with primary sclerosing cholangitis. Hepatology 2013;58:1084-93.
301. Lepore M, Kalinichenko A, Colone A, et al. Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRbeta repertoire. Nat Commun 2014;5:3866.
302. Cianferoni A. Invariant Natural Killer T Cells. Antibodies 2014;3:16-36.
303. Gerlach C, van Heijst JWJ, Swart E, et al. One naive T cell, multiple fates in CD8+ T cell differentiation. J Exp Med 2010;207:1235-1246.
304. Wolf KJ, Emerson RO, Pingel J, et al. Conventional and Regulatory CD4+ T Cells That Share Identical TCRs Are Derived from Common Clones. PLoS One 2016;11:e0153705.
305. Becattini S, Latorre D, Mele F, et al. T cell immunity. Functional heterogeneity of human memory CD4+ T cell clones primed by pathogens or vaccines. Science 2015;347:400-6.
306. Hov JR, Kummen M. Intestinal microbiota in primary sclerosing cholangitis. Curr Opin Gastroenterol 2017;33:85-92.
307. Zachary AA, Bias WB, Johnson A, et al. Antigen, allele, and haplotype frequencies report of the ASHI minority antigens workshops: part 1, African-Americans. Hum Immunol 2001;62:1127-36.
308. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979-86.
309. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376-81.
310. Ramos M, Lopez de Castro JA. HLA-B27 and the pathogenesis of spondyloarthritis. Tissue Antigens 2002;60:191-205.
311. Bowlus CL, Li CS, Karlsen TH, et al. Primary sclerosing cholangitis in genetically diverse populations listed for liver transplantation: unique clinical and human leukocyte antigen associations. Liver Transpl 2010;16:1324-30.
312. Dorum S, Bodd M, Fallang LE, et al. HLA-DQ molecules as affinity matrix for identification of gluten T cell epitopes. J Immunol 2014;193:4497-506.
313. Sims S, Willberg C, Klenerman P. MHC-peptide tetramers for the analysis of antigen-specific T cells. Expert Rev Vaccines 2010;9:765-74.
314. Stern JN, Yaari G, Vander Heiden JA, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. Sci Transl Med 2014;6:248ra107.