Investigating Primary Biliary Cholangitis for a Human Betaretrovirus **Superantigen** signature

by

Hussain Syed

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Translational Medicine

Department of Medicine

University of Alberta

Abstract

Primary Biliary Cholangitis (PBC) is an idiopathic autoimmune disease characterized by destruction of small hepatic bile ducts and presence of antimitochondrial antibodies (AMA). A Human Betaretrovirus (HBRV) sharing 91-99% sequence similarity with the Mouse Mammary Tumor virus has been identified in PBC patients. The autoimmune biliary disease mouse model for PBC produces spontaneous AMA and is infected by the MMTV suggesting a role for MMTV involvement in disease development. The MMTV also has well documented superantigen activity. Viral superantigen bind the MHC class II receptor on antigen presenting cells to induce clonal expansion of specific TCR-Vβ subsets followed by eventual depletion of cognate T-cells. This skewing of specific TCR-Vβ subsets is the hallmark of superantigen activity. To evaluate a similar role of HBRV infection and superantigen activity in PBC patients, we investigated a transcriptional database from the POISE phase III clinical trials to 1) classify CD4+/CD8+ ratios and Neutrophil to Lymphocyte ratios (NLR, an indicator of lymphopenia) 2) characterize the TCR-Vβ repertoire profile in PBC to find evidence of VB skewing and 3) conduct differential gene expression analysis between PBC patient subsets to identify immunological pathways associated with disease development and progression. Analyzing RNA-seq database, we detected an increased range and higher mean CD4+/CD8+ ratio in PBC. A depleted CD8+ T-cell expression was notably observed in the high CD4+/CD8+ ratios. The CD4+/CD8+ ratio was also identified to be positively correlated with increased lymphopenia and negatively correlated with platelet counts indicating hypersplenism and overall worse prognosis. TCR-Vβ repertoire alteration consistent with superantigen activity was observed. Skewing of multiple Vβ subsets as well as the tendency of these skewed Vβ to skew together in clusters suggests the presence of multiple

HBRV superantigen with varying binding avidities. Finally, pathway expression analysis identified increased anti-viral activity along with interferon responses and TLR activation in PBC samples with high CD4+/CD8+ ratio and high NLR. This suggests ongoing HBRV activity and may predict worse prognosis. Meanwhile, low NLR and CD4+/CD8 ratio samples demonstrated increased T-cell activation, activated innate immunity, NK cell immunity and negative regulation of viral transcription indicating a resolving disease process. We also identified CD3E-204, a transcript for the CD3E lymphocyte marker, which had a strong negative correlation with TCR-Vβ skewing. Further studies into the mechanistic differences between the CD3E transcripts need to be made to evaluate the significance of this finding. In future studies, we hope to investigate serial samples from PBC patients to identify changes in TCR-Vβ skewing over time and changes in CD4+/CD8+ ratios over time in patients with known clinical outcomes. We will also investigate TCR-Vβ skewing in isolated CD4+ and CD8+ T-cells in order to better understand the source of CD8+ T-cell depletion. Finally, we propose to clone HBRV superantigen from PBC patient lymphnode using conserved viral superantigen sequences. This will help highlight superantigen proteins in PBC and shed light on their contribution to PBC pathogenesis.

PREFACE

This thesis is an original document by Hussain Syed. No part of this thesis has been previously published.

This document represents the culmination of efforts put forth by several talented people who assisted me in my research. First, I would like to appreciate the efforts of Dr. Andrew Mason in providing me with the resources, information, guidance, and supervision to successfully conduct investigations described in this document. I would like to thank Tracy Jordan who extracted RNA from whole blood of primary biliary cholangitis (PBC) patients, autoimmune hepatitis (AIH) patients and healthy control and performed all the necessary quality control steps for sequencing. PBC samples were obtained from patients enrolled in the POISE phase III clinical trials studying the efficacy of obeticholic acid (OCA) as a treatment for PBC. All library construction and next-generation sequencing was performed by Novogene Co. Dr. Juan Jovel was instrumental in carrying out RNA-seq alignments against the human genome and determining estimated counts and transcript per million values using Kallisto. Kiandokht Bashiri, Dimple Patel and Emma Zwaigenbaum carried out flow cytometry analysis on PBC patients, liver disease control and healthy control to provide the CD4+/CD8+ ratios. Bishoi Aziz obtained platelet counts for regression analysis against CD4+/CD8+ ratios in PBC patients. I performed data analysis determining TCR-Vβ skewing in PBC, linear regression, hierarchical clustering, differential gene expression analysis, pathway analysis, data interpretation and all statistical analysis.

These studies were approved by the University of Alberta, Human Research Ethics Board (Pro00085859)

ACKNOWLEDGMENTS

It is with great gratitude that I would like to express my heartfelt thanks to everyone who has supported my research during my Master's research. My colleagues, friends, supervisor, committee members and family have all provided me with vital support during my journey. It would not have been possible without your guidance and support.

To begin, I would like to extend my sincerest gratitude to my supervisor, Dr. Andrew Mason for his invaluable advice, guidance, mentorship, and constant support. With his mentorship, I was able to develop my technical ability, critical thinking, and information analysis skills. Learning to methodologically approach the investigation of complex biological phenomenon was a valuable undertaking that has allowed me to develop vital skills as a biological researcher. These lessons will be the steppingstones for future progress in my journey of medical research. I would also like to thank Dr. Juan Jovel, who patiently taught me R and assisted me in developing an operational understanding of the many useful packages developed on the platform. I also express my deepest thanks to members of my research committee, Dr. John Elliott, and Dr. Colin Anderson for providing valuable insight and commentary on my research.

My colleagues and friends at the Mason Lab who created a warm and welcoming work environment and assisted me in my research endeavours deserve all my gratitude. Dr. Steven Willows, Dr. Hiatem Aboufayed, Dr. Kerolous Messeha, Dr. Bishoi Aziz, Dr. Kiandokht Bashiri, Tracy Jordan, Ning Sun and Bruna Dutra all provided support, advice, and constructive criticism throughout our shared work experience.

Most importantly, I would like to express my deepest appreciation towards my family. My parents, Syeda Momina Muhammad and Syed Taqi Muhammad who gave me unconditional love and taught me the value of listening. My brothers Raza Muhammad Syed and Ali Muhammad Syed who were the backbone of my support system and were always there for me. My sister-in-law, Syeda Masooma Muhammad for her constant support. And finally to my young nephew, Syed Hasan Muhammad who always found a way to cheer me up during difficult times.

Finally, I would like to extend my thanks to all those who have helped me on my journey, it is through your support that I have climbed this far.

Table of Contents

List of Tables

List of Figures

List of abbreviations

Abbreviation	Meaning
AIH	Auto-Immune Hepatitis
ALP	Alkaline Phosphatase
AMA	Antimitochondrial Antibodies
BEC	Biliary Epithelial Cells
BLB	Bilirubin
DGE	Differential Gene Expression
FDR	False Discovery Rate
HBRV	Human Betaretrovirus
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
IFI	Interferon Alpha Inducible
IHL	Intrahepatic lymphocyte
MHC	Major Histocompatibility Complex
MMTV	Mouse Mammary Tumor Virus
NK	Natural Killer
NLR	Neutrophil to Lymphocyte Ratio
NOD	Non-obese diabetic
NRAS	Neuroblastoma RAS Viral (V-Ras) Oncogene Homolog
OAS	Oligoadenylate Synthase
OCA	Obeticholic acid
PBC	Primary Biliary Cholangitis
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase Chain Reaction
pcytC	Pigeon cytochrome c peptide
PDC-E2	Pyruvate Dehydrogenase complex E-2
POISE	The PBC OCA International Study of Efficacy
SAg	Superantigen
SD	Standard Deviation
TCR	T-cell receptor
TLR	Toll-Like receptor
TPM	Transcript Per Million
TRIM	Tripartite motif
UDCA	Ursodeoxycholic acid
ULN	Upper Limit of Normal
vSAg	Viral Superantigen

1. Introduction

1.1 Primary Biliary Cholangitis (PBC)

PBC was initially called Primary Biliary Cirrhosis to describe a liver disease characterized by destruction of small intrahepatic bile ducts, and the development of cirrhosis [1]. Since this description, clinicians have found that only 50% of patients develop cirrhosis. In addition to cirrhosis being an inaccurate term to describe the disease, the association between alcohol use and cirrhosis development inadvertently led to stigmatization of patients with PBC [2, 3]. In efforts to negate the stigmatization and employ more descriptive diagnostic criteria, PBC is now referred to as Primary Biliary Cholangitis.

Primary Biliary Cholangitis is an idiopathic autoimmune disease of unknown etiology. It is characterized by the non-suppurative destruction of small interlobular bile ducts. Progressive destruction of the small bile ducts in PBC leads to further complications involving inflammation, fibrosis and cirrhosis which may result in the need for a liver transplant.

Serologically, PBC is currently defined by presence of antimitochondrial antibodies (AMA) as well as elevated levels of IgM in patient sera [4, 5]. The presence of autoreactive plasma cells and lymphocytes has also been identified in PBC patients [2, 3]. Additionally, elevated levels of Alkaline Phosphatase (ALP) and Bilirubin (BLB) in PBC patient serum have been observed to be surrogate indicators of disease outcomes [6].

However, evidence of an autoreactive etiology in PBC is still lacking. About 20% of PBC patients AMA are found to be AMA negative while AMA presence has also been noted in patients with other liver disease such as chronic active hepatitis and cryptogenic cirrhosis [4].

Additionally, AMA titer and presence appear to be unlinked with disease severity [7, 8]. Overall, a role of a antimitochondrial antibody in PBC development has yet to be persuasively established [9]. Meanwhile, IgM isolated from PBC patients has been found to be similar in electrophoretic properties, complexes, clonality and aggregates when compared to patients with hyper IgM expression without liver disease [5]. Finally, the role of plasma cells and lymphocyte self-reactivity in PBC development still lacks direct evidence [10]. Similarly, while ALP levels were found to be effective indicators of disease progression in some patients, the existence of PBC patients with normalized ALP levels represents a challenge for accurate clinical assessment of PBC patients [11].

Overall, it is still unclear what mechanisms trigger the disease process in patients with PBC [2, 3] and the factors contributing to PBC progression are poorly understood and gaps in knowledge of disease mechanism still exist.

Pathologically, the development of PBC has been attributed to destruction of small interlobular bile ducts resulting in liver damage which may progress to cirrhosis. Clinically, majority of patients symptoms include pruritus, fatigue, and sicca syndrome with dry eyes and dry mouth. Patients may also develop hypercholesterolemia, portal hypertension, osteopenia, and osteoporosis [12].

As the etiology of the disease is unclear, various investigations on different mechanisms that may contribute to PBC development are still underway. Several genetic and environmental influences have been attributed as contributors to disease development including genetic predisposition in correlation with MHC class II DR and DQ alleles, association with mucosal infections, contribution of xenobiotic elements, exacerbation by alcohol and cigarette smoke as

well as the role of a transmissible virus in PBC pathogenesis [12]. Here, we investigate the role of a Human Betaretrovirus infection contributing to PBC development.

1.2 PBC Prevalence and Vulnerable Populations

PBC is a rare liver disease that disproportionately targets a specific demographic. Several epidemiological studies have demonstrated that PBC manifests differently across demographics and geographical areas. However, it has been noted to appear more frequently in middle-aged women. Although the female/male ratio does vary, the discrepancy between cases has been observed to be as high as 10/1 in some populations. Additionally, the majority of epidemiological studies indicate an increase in PBC incidence and prevalence globally [13].

Varied distribution of PBC across geographical location has also been observed within Canada. An epidemiological study investigating the incidence and prevalence of PBC in Canada identified a range of 283 cases per million (Ontario) to 465 cases per million (Atlantic Provinces). The transplant rates were also found to be varied across populations. Overall, the study identified an increasing prevalence of PBC in the Canadian population, reporting an increase of 25% in PBC cases from 2013 (255 cases per million) to 2015 (318 cases per million)[14]. Earlier studies have also reported the prevalence of PBC increasing from 100 cases per million in 1996 to 277 cases per million in 2002. The female: male ratio in Canada was observed to be 5:1, with the majority of afflicted women in middle to elderly age group. Despite being classified as a rare disease, PBC patients accounted for 5% of all liver transplants in Canada from 2010-2015. PBC remains a significant contributor to liver disease in Canada, not only imposing a substantial economic burden, but is also the cause of substantial negative affect and decrease in quality of life of afflicted patients and their families [14, 15]. Studies reviewing

the psychological and sociological burdens imposed on patients and their families identified several factors including depression, anxiety, sleep disturbances, social isolation along with negatively impacted mental, physical, and emotional health [16].

1.3 PBC Etiological and Disease Triggers

As previously mentioned, studies investigating PBC development have identified several compounding factors contributing to disease progression. These include genetic predisposition, xenobiotic elements, cellular stress, environmental influences and infectious agents which contribute to loss of immune tolerance [13]. Although these factors have yet to be proven to be solely and conclusively responsible for PBC etiology [3], we discuss investigations into different mechanisms which are thought to contribute towards PBC pathogenesis.

Autoreactivity in PBC is noted to manifest as organ specific damage to biliary epithelial cells leading to bile duct destruction. This may eventually progress to liver fibrosis and cholestasis is some PBC patients which then require liver transplantation. Specifically, autoreactive immune cells are thought to target the E2 subunit of the mitochondrial pyruvate dehydrogenase complex (PDC-E2) typically located on the inner mitochondrial membrane. Aberrant expression of the PDC-E2 on BEC surface may contribute towards bile duct specific damage [3, 17]. However, the direct role of antimitochondrial antibody or self-reactive T-cells in PBC pathogenesis has yet to be convincingly established [9, 10].

Mapping concordance rates of PBC in genetically related individuals identified a concordance rate of 63% for monozygotic twins [18]. While PBC patient's daughters were identified as the group with the highest prevalence of increased risk [19]. In-depth genome analyses have also allowed the discovery of single nucleotide polymorphisms in PBC patients [17] along with

variant alleles at the HLA locus associated with the disease [20, 21]. Xenobiotics such as 2-octynoic acid commonly found in perfumes, lipsticks and common food flavorings were also found to induce high Ig reactivity in PBC patient sera indicating a possible modified, autoreactive immune response to xenobiotic modified PDC-E2 complex [22]. Factors contributing to cellular stress along with abnormal regulation of the endoplasmic reticulum were also found to correlate with increased autophagy in biliary epithelial cells. Additionally, oxidative and nitrosative stress associated with metabolic deregulation were found to impact PBC development and progression. [23, 24]. Other environmental factors such as cigarette smoking, geographical location and bacterial infection have also been implicated as exacerbating factors towards loss of immune tolerance through different mechanisms [17].

Studying infectious agents as contributing factors to PBC development and progression, a betaretrovirus infection was identified to coincide with PBC presentation. The Human Betaretrovirus (HBRV) genome was found in the biliary epithelium and peri-hepatic lymph nodes of PBC patients through ligation mediated PCR and next generation sequencing. However, in liver tissue, the viral titer was noted to be below the limit of detection. HBRV infection was also found to be associated with atypical mitochondrial protein expression on BEC surface [25, 26].

The complex etiology of PBC progression is reflected by the interaction of these many different mechanisms with PBC development and pathogenesis [3].

1.4 PBC Immunology

Only the small intrahepatic bile ducts are subjected to chronic non-suppurative destructive cholangitis in PBC patients. Reactive immune cells working in conjunction with cytokines and chemokines released by the biliary epithelium result in increasing damage to the bile ducts which may develop into ductopenia. Overall, immune dysfunction at site of damage has been recognized as a contributing factor to PBC pathogenesis [3].

Investigation on liver necrosis in PBC patients confirmed presence of CD4+ and CD8+ T-cell at site of biliary epithelium destruction along with increased presentation of MHC-class I and MHC-class II antigens on liver and sinusoidal lining cells. The study noticed an increase in antigen specific cytotoxicity at site of necrosis indicating a topographically distributed deregulation of immune mechanisms [27]. Examinations of the immune cell repertoire at site of inflammatory lesions in PBC patients have identified Ig+ cells, CD20+ B cells, CD4+ and CD8+ T-cells. Notably, CD3+, CD4+ alpha/beta T-cells were the predominant cell type in the infiltrating lymphocytes [28]. The expression of interleukin 2 (IL-2) receptor on the surface of a majority of these CD3+, CD4+, alpha/beta T-cells displaying class II antigens suggests a potential role for these activated T-cells in PBC pathogenesis. Additionally, HLA class II antigens were found to be predominantly displayed on the biliary epithelium which is normally populated by HLA class I [29]. Although it is unclear which of the CD4+ or CD8+ T-cells comprise the majority contribution to PBC development and progression, the pervasiveness of CD4+ T-cells, identification of MHC Class II activity on site of necrosis, association of MHC Class II antigens with prognosis [29, 30] taken in conjunction with the role of CD4+ T-cell dysfunction in development of autoimmune inflammation [31] suggests a CD4+ T-cell associated mechanism contributing to PBC pathogenesis.

Exploring the behavior of CD4+ T-cells, Mayo et al. studied the gene expression of T-cell receptor Variable Beta region (TCR-Vβ) in PBC patients. An increased skew of certain TCR-Vβ subsets was identified in PBC patient peripheral blood compared to normal control. Also, a greater accumulation of skewed Vβ subsets was identified in liver samples from PBC patients compared to disease control. Although a large variety of distribution patterns were noticed in individual patients, mean trends for subset expression held significant across PBC vs healthy comparisons for including upregulation of TCR-Vβ 6-1.3 and TCR-Vβ 3 in PBC patients compared to healthy controls. Although no definitive source of TCR-VB subset skewing was identified in the study, a superantigen effect was tentatively suggested as one possible mechanism that could contribute to TCR- Vβ skewing [32]. The role of superantigen (Sag) activity in the development and progression of autoimmunity has been postulated in a number of diseases including multiple sclerosis, rheumatoid arthritis, sjogren's syndrome, psoriasis, and autoimmune thyroiditis [33].

1.5 Human Betaretrovirus (HBRV) Infection and PBC

Investigation into PBC pathogenesis led to the identification of a novel human betaretrovirus in PBC patient biliary epithelium and perihepatic lymph nodes initially through RT-PCR and later through ligation-mediated PCR and next generation sequencing. Evidence of viral infection was found in ~75% of PBC patients along with autoimmune hepatitis and patients with cryptogenic liver disease, however patients with other liver disease were rarely found be positive for HBRV integration [25, 26]. Interestingly, the majority of HBRV signature was found in patient lymph nodes while lower or undetectable viral titers were observed in most PBC patient liver tissue samples [26].

The betaretroviral signature in PBC patients has been linked with PBC development and progression through co-incidence of viral activity with the atypical display of mitochondrial autoantigen on cell surface following co-cultivation of normal biliary epithelial cells with PBC patient lymph nodes [34].

Alignment studies evaluating the HBRV against the Mouse Mammary Tumor Virus (MMTV) and breast cancer betaretrovirus sequences have identified a 85%-97% sequence homology in the partial *pol* gene, a 93% to 99% homology to the p27 capsid proteins, a 93% to 97% homology to the betaretrovirus envelope proteins and a 76% to 85% homology with the superantigen proteins [35]. Interestingly, MMTV infection was observed to manifest a PBC analogous phenotype in mouse models. In the NOD.c3c4 congenic mouse model, MMTV infected mice developed autoimmune biliary disease (ABD) displaying disease characteristics similar to PBC including infiltration of CD3+, CD4+ and CD8+ T-cells at site of biliary epithelium destruction, spontaneous anti-mitochondrial antibodies in sera and aberrantly expressed PDC-E2 on cell surface. A correlation between anti-MMTV antibody production and AMA production in addition to development of ABD after adoptive transfer of CD4+ T-cells or splenocytes suggests a link between MMTV infection and ABD development mediated by CD4+ T cell dysfunction in NOD.c3c4 mice [36, 37]. Additionally, combination antiretroviral therapy of Nod.c3c4 mice was found to result in cholangitis attenuation, decreased MMTV titer and reduction of serum liver enzymes [38]. The mouse mammary tumor virus is known to produce superantigen proteins which help generate an environment of immune dysfunction conducive to its survival [39].

1.6 Betaretrovirus & Superantigen

Superantigens (SAg) are a class of viral or bacterial proteins that facilitate a unconventional interaction between the Major Histocompatibility complex class II (MHC II) and the T-cell receptor via its variable beta region [39]. The SAg bind as unprocessed antigens to MHC II in a promiscuous manner, outside of the peptide binding groove. The subsequent interaction of this complex with the Vβ region of CD4+ T-cells stimulate 2-30% of the host T-cell repertoire compared to <0.01% T-cell activation by normal antigens [33].

Over 20 years of investigation on the endogenous and exogenous MMTV has allowed extensive documentation of viral components and virus life cycle leading towards the identification of the MMTV Superantigen. The SAg is noted to drive the MMTV infection by inducing proliferation of naive T-cell in a Vβ specific manner which may be followed by eventual deletion of cognate T-cells (Figure 1). However, the fate of the activated T-cells varies significantly depending on the avidity of the superantigen mediated interaction. Generally, strong responses are often coupled with acute type deletion while weaker responses over a prolonged period of time are associated with slow, chronic depletion [40]. While, MMTV SAG have been known to induce proliferation of naïve T-cells in the presence of env proteins, this proliferation does not correlate with higher levels of SAg gene expression [41]. However, enhanced expression of MHC II receptors may induce stronger SAg activity [40].

Despite advances in our understanding of MMTV superantigen activity, the cell surface expression of SAg proteins remains difficult to detect due to paucity. Instead skewing in the TCR-Vβ repertoire is recognized as the hallmark of MMTV superantigen activity [40]. As SAg sequences sharing 76% to 85% homology to the MMTV SAg have been identified in the HBRV

isolated from PBC patients, there is a possibility that SAg mediated changes to the TCR repertoire may play a crucial role in PBC pathogenesis [35].

Figure 1.

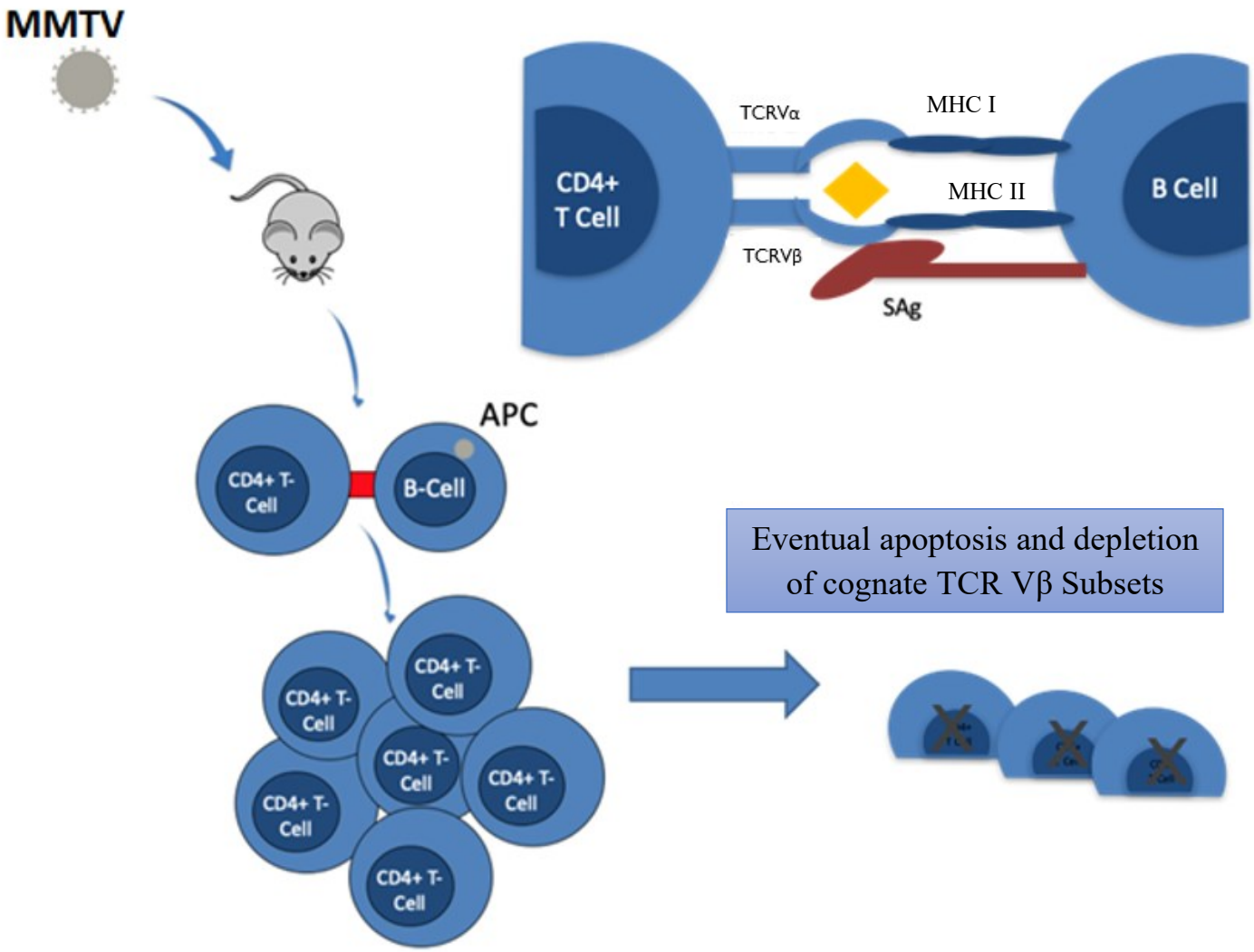


Figure 1. MMTV infected B-cells induce superantigen mediated clonal expansion in CD4+ T-cells which eventually leads to loss of cognate T-cells through apoptosis.

1.7 Superantigen activity and Auto Immunity ?

Although the complexities of the superantigen interaction with MHC class II and TCR-Vβ receptors are still being investigated, several reports have highlighted variable interactions resulting in varying consequences depending on the avidity of the binding interaction, nature of the specific superantigen, specificity of SAg activated T-cells and the overall host immune environment [40].

Despite these variances, superantigens are universally found to induce naïve T-cell proliferation magnitudes higher compared to regular antigens [33], along with having the ability to activate T-cells with various specificities. This, along with the ability of the SAg to drive polyclonal antibody responses lead researchers to investigate the possibility of superantigen involvement in autoimmunity development [42].

Initial studies postulated three main mechanisms of SAg mediated autoimmune development. 1) Localized trauma caused by a viral infection or by a viral SAg related inflammatory response might facilitate exposure of cellular epitopes promoting T-cell self-reactivity 2) Viral SAg mediated expansion of already present T-cells to cross-reactive epitopes. 3) Viral SAg mediated reactivation and expansion of anergic self-reactive T-cells present in the periphery [40, 42].

More recently, researchers investigating bacterial superantigen staphylococcal enterotoxin B (SEB) modeled a chronic SEB SAg exposure. The SEB SAg facilitated high levels of CD4+ activation with a diverse specificity leading to proliferation of self-reactive T-cells. The activated T cells then activated self-reactive B cells resulting in self-reactive auto-antibody production

[43]. Investigators have found evidence implicating both viral and bacterial superantigen involvement in immune mediated disease. Diseases such as multiple sclerosis, psoriasis, rheumatoid arthritis, Kawasaki disease, HIV etc. have displayed some evidence for the role of a superantigen protein contributing to disease induction/exacerbation through SAg mediated polyclonal expansion, T-cell receptor Vβ specific activation and autoreactive T-cells with restricted TCR-Vβ. Animal models for multiple sclerosis have displayed autoreactive T-cell development as well. [44, 45]. Viral antigens such as HIV associated superantigens have been discussed to play a role in HIV associated psoriasis following previous studies linking psoriasis presentation with possible superantigen activity [46, 47]. Most recently, a SARS-CoV-2 superantigen motif has also been noted in the COVID-19 virus alongside the identification of autoantibody presence in up to 50% of COVID-19 patients [48, 49]. Taken together, these findings illustrate the ability of viral and bacterial superantigens to facilitate immunopathology that might contribute to development of autoimmunity.

In addition to the conventionally known superantigen mediated transient T-cell proliferation followed by depletion of cognate Vβ subsets, studies have also investigated the effects of co-stimulation of T-cells with superantigen and a conventional antigen. McCormark et al. showed that upon cotreatment of mice with the staphylococcal enterotoxin A (SEA) SAg with the pigeon cytochrome c peptide (pcytC), T-cells responding to pcytC stimulation as well as the SEA SAg were not deleted and continued to proliferate. Meanwhile, the SEA SAg continued to carry out activation and subsequent deletion of T-cells which did not respond to stimulation from the pcytC antigen [50]. Considering this data, it may be possible that aberrant expression of autoantigens on cell surface coinciding with viral superantigen activity might allow T-cell subsets reactive to the specific autoantigen to resist deletion which could contribute to disease.

1.8 TCR-VB profiles in PBC

Studies by Mayo et al. utilized RT-PCR technique to analyze the T-cell receptor B-chain variable (TCR-VB) gene expression in PBC patient liver tissue and peripheral blood. In 28 PBC patients, it was determined that all 22 Vβ subsets investigated had detectable levels of gene expression. The Vβ profiles of individual PBC patients were found to be highly diverse in both peripheral blood and liver tissue. Using aggregate data, mean Vβ profiles of PBC patients in peripheral blood were found to vary significantly compared to control. Vβ 6.1-3 and Vβ 3 were noted to be higher in PBC samples. Comparing liver tissue with blood, the pattern of bias was noted to be heterogeneous in individual patients. Aggregate means of PBC patient liver Vβ subset expression were found to significantly vary from peripheral blood. Notably, Vβ 6.1-3, Vβ 7 and Vβ 13.1 were observed to be higher while Vβ 4 and 5.2-3 were noted to be lower.

Additionally, observations of Vβ profile expression in the same patient at different time points identified a change in Vβ subset bias overtime. Comparing patients with late-stage disease to early-stage disease indicated changes in Vβ skewing overtime [32]. In another study, Mayo et al. also found that Vβ skewing in PBC patients was a consistent phenomenon regardless of anti-mitochondrial antibody presence [51].

Other researchers have also identified TCR- Vβ bias in PBC patients. One study examining 10 PBC patients reported a disease associated clonotype expansion [52]. Another reported *NRAS* upregulation alongside T-cell activation and TCR Vβ 5-2 and Vβ 10-1 skewing [53]. However, another study reported no abnormality in Vβ repertoire in 8 female PBC patients in early stages of the disease [54].

1.9 Hypothesis and Specific Aims

A) Hypothesis

Based on previous findings of HBRV integration in PBC patients [26] alongside the postulated role of betaretrovirus encoded superantigen facilitating self-reactive T-cell proliferation, as well as the extensive literature discussing MMTV mediated TCR-Vβ skewing in mice [40], we hypothesize that PBC patients experience SAg mediated immune dysfunction (Figure 2). Investigating the immune cell population in PBC patients we expect to find deregulation of immune cell expression, TCR-Vβ skewing, and activation of immunological pathways indicating superantigen induced immunopathology.

Figure 2.

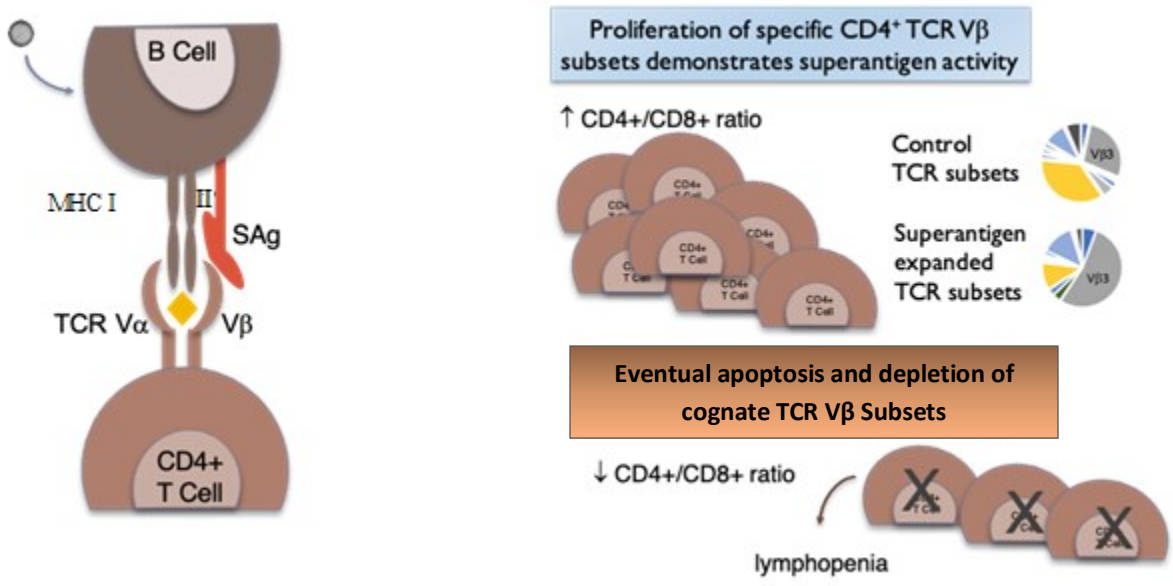


Figure 2. Hypothesized Superantigen mediated clonal expansion of TCR-VB subsets in primary biliary cholangitis patients where cognate T-cells may experience eventual depletion in later stages.

B) Specific Aims:

Specifically, we investigate:

1) CD4+/CD8+ ratio of PBC patients compared to healthy control.
2) Distribution of CD4+ T-cell and CD8+ T-cell gene expression within PBC patients
3) Neutrophil to Lymphocyte ratio of PBC patients and the relationship between NLR and markers of disease progression
4) The distribution of TCR-Vβ subsets in Primary Biliary Cholangitis
5) Differential gene expression of PBC patients subsetted through dichotomization of immune state markers (CD4+/CD8+ ratio and NLR)

2. Methods and Materials

2.1 Sample description and characteristics.

A total of 173 total RNA samples were collected from whole blood of 128 Primary biliary cholangitis (PBC) patients enrolled in the POISE phase III OCA trial [55]. All PBC patients were subject to study recruitment criteria of unsuccessful ursodiol treatment with an alkaline phosphatase level (ALP) of at least 1.67 x ULN or abnormal bilirubin (BLB) of < 2x ULN [55]. Of the 128 PBC patients from which RNA-seq data was collected, 86 had single timepoint samples, 35 had at least one baseline and one treatment sample while 7 had no baseline or no after treatment samples. A total of 173 PBC samples were sequenced. Samples were simply stratified as “Before” signifying baseline sample collection and “After” signifying collection of blood sample at a timepoint after at least a 12-month OCA treatment period. Additionally, 16 samples were sequenced from 11 autoimmune hepatitis patients with 2 patients having samples from two different timepoints and 1 patient with samples from 4 different timepoints. Finally, 16

healthy individuals with no indications of liver disease were also sequenced as healthy controls. Table 1. Lists mean age (along with standard deviation and range), sex distribution and patient biochemistry at baseline.

To supplement our RNA-seq database, we also evaluated immune cell counts from peripheral blood of 108 PBC patients, 26 liver disease patients and 14 healthy controls. Liver disease patient samples included Hepatitis B patients, Hepatitis C patients, Non-alcoholic fatty liver disease patients, Non-alcoholic steatohepatitis patients and Alcoholic liver disease. We also evaluated immune cell counts of intrahepatic lymphocytes extracted from 12 PBC patients and 9 patients with other liver diseases including Autoimmune hepatitis, hepatocellular carcinoma, Non-alcoholic fatty liver disease and Alcoholic liver disease.

Table 1.

Characteristic	PBC	Healthy
Age (Average)	56 (± 10.3)	40 (± 15.5)
Age (Range)	30-86	18-66
Female Sex (%)	93%	93%
Baseline ALP	> 1.67 X ULN	-
Baseline BLB	< 2 x ULN	-

Table 1. Showing mean age, age range and sex ratios of PBC patients and healthy control along with PBC patient biochemistry at baseline.

2.2 RNA-Seq pipeline

All total RNA was extracted from PaxGene tubes. Samples were subjected to nanodrop and Qubit for quality control to ensure that all standards were met before proceeding to sequencing. Libraries were processed by Illumina HiSeq to generate 5GB of data per library. Base quality in FASTQ files were inspected with fastqc, bases with quality scores (Q) lower than 20 were trimmed with fastq-mcf. Pseudo-alignments against the human transcriptome (ensembl Hs_GRCH38.rel79) were done using Kallisto [56], with 100 bootstraps and bias correction. Out files from Kallisto were estimated counts and transcript per million (TPM) normalized values (Figure 3).

Figure 3.

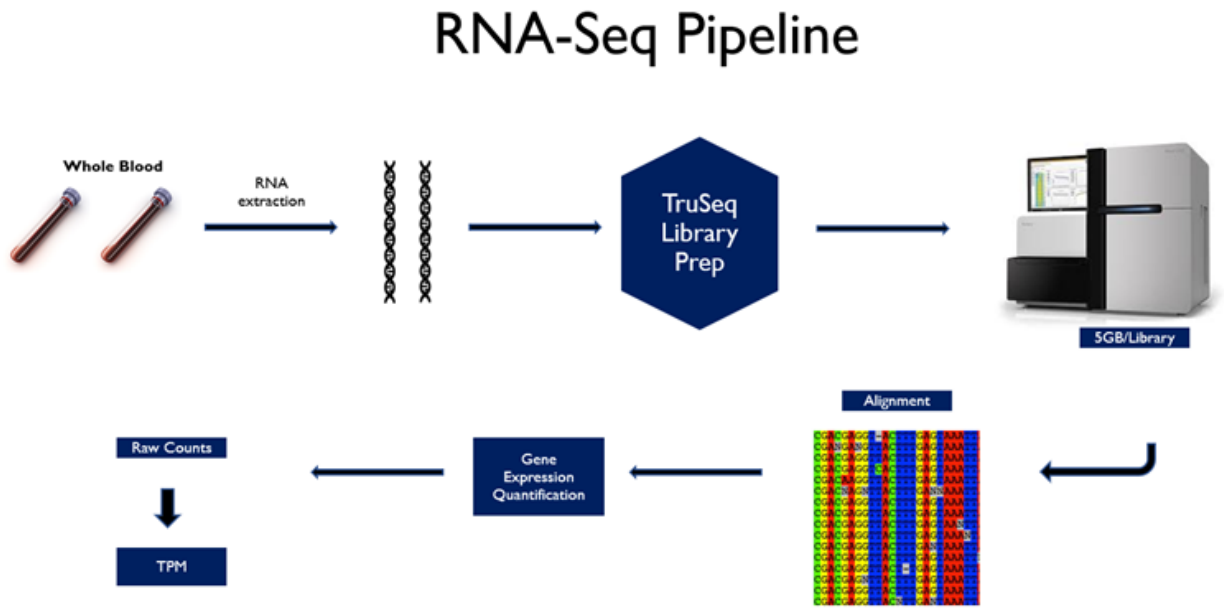


Figure 3. RNA-Seq data collection pipeline highlighting steps taken to arrive at normalized transcript per million values for each transcript expression for all PBC patients and healthy control.

2.3 RNA extraction and PBMC isolation

Blood samples from local patients (not a part of the POISE phase III clinical trials) were collected from patients in Pax Gene tubes (*Becton Dickinson, Franklin Lakes, NK, USA*) and Heparinized tubes. Peripheral blood mononuclear cells (PBMC) were isolated using the Lymphoprep density gradient media and the SepMate tube protocol (*Stemcell, Vancouver, BC, CA*). While RNA was isolated using the E.Z.N.A PX blood RNA kit (*Omega Bio-Tek, Norcross, GA, USA*)

2.4 Calculating Immune gene expression

To ascertain normalized gene expression and determine CD4/CD8 ratios, Neutrophil to Lymphocyte ratios from our RNA-seq data, we used sum TPM expression of all relevant functional transcripts to analyze immune cell signature at the gene level. CD4 transcripts included ENST00000011653, ENST00000539492 and ENST00000541982. CD8a transcripts included ENST00000352580, ENST00000283635, ENST00000409511 and ENST0000040978. CD3e transcripts included ENST00000361763 and ENST00000528600 while CD16b (FCGR3B, neutrophil marker) transcripts included ENST00000367964, ENST00000531221, ENST00000421702, ENST00000294800, ENST00000614870 and ENST00000613418.

The combined transcript expression was summed to gene level to determine CD4+ T-cell, CD8+ T-cell, CD3+ T-cell and Neutrophil expression respectively. Flow cytometry analysis was used to corroborate findings in RNA-seq data. Peripheral blood mononuclear cells (PBMC) were isolated from whole blood of donors using Lymphoprep^{TM} density gradient media which is comparable to other PBMC isolation methodology [57]. Intrahepatic lymphocytes (IHLs)were collected by flushing livers of transplants recipients with sterile saline. Platelet counts were obtained from patient medical lab test records.

2.5 Analysis of TCR-VB repertoire

TCR-Vβ subgroup profiles in 128 PBC patients (n = 173 samples) were analyzed as transcript per million (TPM) by using gene level expression of all functional TCR-Vβ genes. We further normalized Vβ expression by dividing each Vβ gene expression in all patients with their respective CD3e gene expression to normalize by total lymphocyte expression. Of the 61 unique TCR-Vβ subsets identified, 19 were found to be non-coding pseudogenes, 36 were functional

coding genes , 2 were non-functional genes and 4 were identified to be functional genes in some individuals and non-functional in others.

To determine abnormal Vβ expression, we established a broad range for normal Vβ expression for each TCR- Vβ subset based on normalized healthy mean TPM Vβ expression + 3*Stdev.S for an upper limit and -2*Stdev.S for the lower limit. Abnormal expression of TCR-Vβ were recorded as frequency of expansions or depletions in the subsets. Mean aggregate comparison between PBC vs control was conducted by an unpaired t-test with fewer assumptions and the Holm-Sidak correction.

2.6 Differential Gene Expression Analysis

All differential gene expression (DGE) analysis were carried out using negative binomial generalized linear models with the DESeq2 R package [58]. Multiple comparisons were conducted within the 173 PBC patient samples by dichotomizing patient subgroups using biological criteria to ascertain the complex immunological pathways involved in changing CD4/CD8 ratios, NLR ratios and TCR-Vβ skewing. All comparisons had an FDR and p-value of at least <0.05.

2.7 Differential Pathway Enrichment

Pathway enrichment analysis on the upregulated and downregulated gene sets derived from DGE output were conducted using the Cytoscape ClueGO application [59] and the GeneOntology Panther tool [60, 61]. For ClueGO, Go term fusion was implemented alongside two-sided hypergeometric test and Bonferroni step down correction in order to reduce pathway redundancy and identify statistically significant dysregulation in biological pathways. Pathway

selection was further narrowed down using % gene involvement with consideration towards enormity of the input gene set. Default settings were utilized when utilizing the GeneOntology software available on the GeneOntology website (http://geneontology.org/).

2.8 Hierarchical clustering and Vβ skewing overlap.

To determine the pattern of Vβ skewing in PBC patients we evaluated co-occurrence of Vβ skewing in individual PBC samples and carried out Hierarchical clustering of Vβ subsets with the R "cluster" package. We evaluated binary input using agglomerative nesting using the manhatten metric and applied the ward method of hierarchical clustering. Furthermore, frequency of each overlap event between two distinct Vβ subsets were recorded in frequency tables for both upregulated skew events as well as downregulated skew events. These data corroborated the findings of the hierarchical clustering method and allowed for mathematical visualization.

3. Results

3.1 CD4+/CD8+ Ratio in Primary Biliary Cholangitis

We quantified lymphocyte expression by calculating the CD4/CD8+ ratio using combined TPM counts of the CD4+ and CD8+ transcripts to evaluate gene level expression from RNA-seq data. A total of 173 samples were compared to 16 disease control and 16 healthy control to determine the range and distribution of CD4+/CD8+ ratios in PBC patients. PBC samples had increased mean and increased range compared to healthy and disease control (Mann Whitney, $p<0.05$) (Figure 4). To validate our findings from RNA-seq data, we analyzed CD4+/CD8+ ratio in PBC patient periphery using flow cytometry. Here, PBC patients (n=108) displayed a similar trend with increased range compared to liver disease (n=26) and healthy control (n=14). (Figure 5). We further investigated the CD4+/CD8+ distribution in PBC intrahepatic lymphocytes. Similar to RNA-seq and peripheral blood, a trend towards higher mean and increased range was observed in PBC compared to liver disease control. (Figure 6.)

Using RNA-seq data, we compared gene expression data for the CD4 and CD8a genes by dichotomizing the patient population into the top 50% (higher) CD4+/CD8+ ratio and the bottom 50% (lower) CD4+/CD8+ ratio. PBC samples with higher CD4+/CD8+ ratio displayed slightly higher CD4+ gene expression ($p<0.05$) along with significantly depleted CD8a gene expression ($p<0.0001$) compared to PBC samples with lower CD4+/CD8+ ratios. This indicates an expanded CD4+ T-cell population and a depleted CD8+ T-cell population contributing to the high CD4+/CD8+ ratio.

Figure 4.

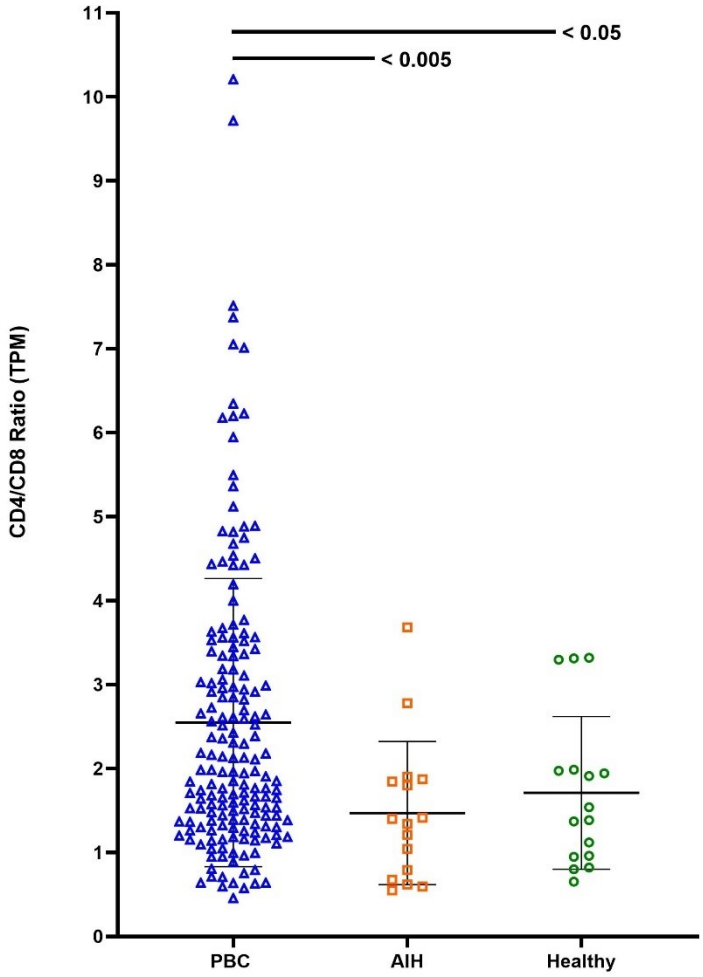


Figure 4. Comparison of CD4+/CD8+ ratios derived from TPM counts between PBC patients, AIH and healthy controls. CD4+/CD8 ratios were calculated using Transcript Per Million (TPM) counts of CD4 and C8a genes obtained from RNA-sequencing of PBC samples (n=173), AIH samples (n=16) and Healthy controls (n=16). Significance was calculated between PBC vs AIH ($p<0.005$) and PBC vs Healthy ($p<0.05$) using Mann-Whitney t-test. PBC samples had a higher mean and range (Mean 2.5, S.D 1.7, Range 0.5-10.2) compared to AIH (Mean 1.5, S.D 0.9, Range 0.5-3.7) and healthy controls (Mean 1.7, S.D 0.9, Range 0.7-3.3). Total RNA was extracted from whole blood PaxGene tubes, cloned into TruSeq libraries, and sequenced by Illumina HiSeq to generate 5GB of data per library. Kallisto was used to generate raw counts and TPM after base reads clean up. Plot was made and statistical analysis was done using PRISM.

Figure 5.

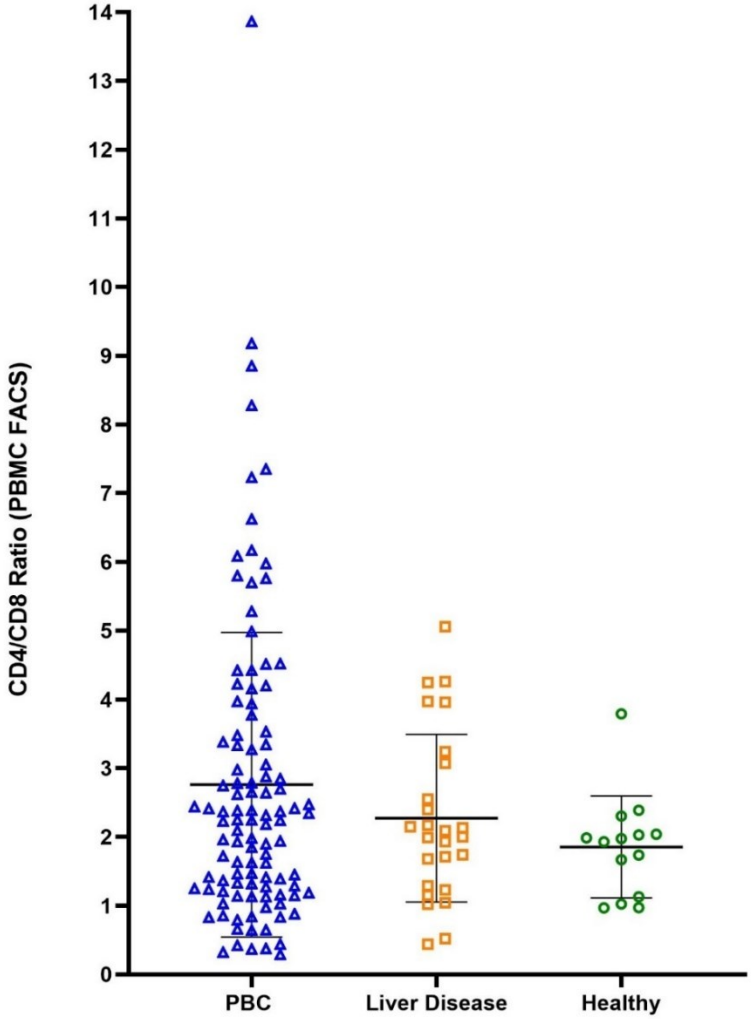


Figure 5. Comparison CD4+/CD8+ ratios derived using flowcytometry of PBMCs from PBC patients, other liver disease patients and healthy controls. CD4+/CD8+ ratios were calculated using proportions of CD4+ and CD8+ T-cells in peripheral blood determined by flowcytometry of peripheral blood mononuclear cells (PBMCs) obtained from PBC patients (n=106), liver disease patients (n=26) and healthy controls (n=14). PBC patient samples showed a trend towards higher mean and range (Mean 2.8, SD 2.2, Range 0.3-13.9) compared to other liver disease patients (Mean 2.3, SD 1.2, Range 0.4-5.1) and healthy controls (Mean 1.9, SD 0.7, Range 1.0-3.8). Other liver disease patient samples included Hepatitis B patients, Hepatitis C patients, Non-alcoholic fatty liver disease patients, Non-alcoholic steatohepatitis patients and Alcoholic liver disease. PBMCs were isolated from whole blood using Lymphoprep density graded media and SepMate lymphocyte separation tubes (*Stemcell, Vancouver, BC, CA*). PBMCs were stained for vitality, CD3+, CD4+ and CD8+ for flowcytometry. Plot was made and statistical analysis was done using PRISM.

Figure 6.

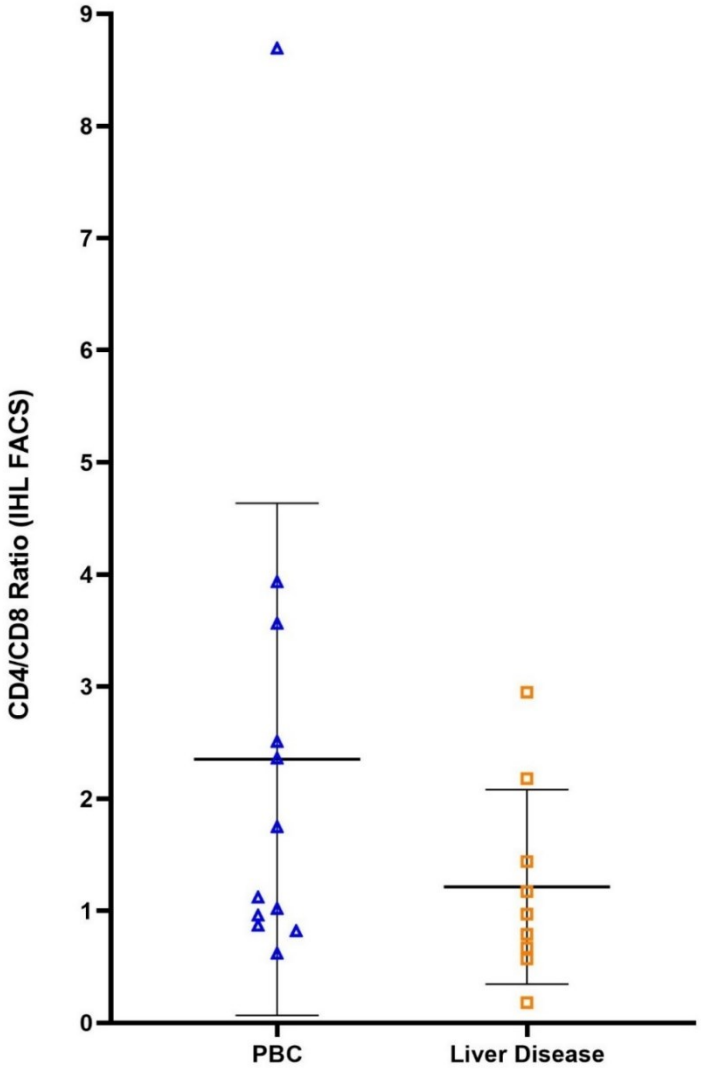


Figure 6. Comparison of CD4+/CD8+ ratios from flowcytometry of intrahepatic lymphocytes between PBC patient samples and liver disease patient samples. CD4+/CD8+ ratios in Intrahepatic lymphocytes (IHL) of PBC (n = 12) and Liver disease (n = 9) were calculated using CD4+ and CD8+ T-cell proportions determined by flow cytometry. PBC patient samples had a trend towards higher mean and range (Mean 2.4, SD 2.3, Range 0.6-8.7) compared to other liver disease patients (Mean 1.2, SD 0.9, Range 0.2-2.8). Other liver disease included patients with Autoimmune hepatitis, hepatocellular carcinoma, Non-alcoholic fatty liver disease and Alcoholic liver disease. IHL's were collected by flushing livers of transplant patients with normal saline. Lymphocytes were extracted from liver flush and stained for flowcytometry as described in Figure 5. Plot was made and statistical analysis was done using PRISM.

Figure 7.

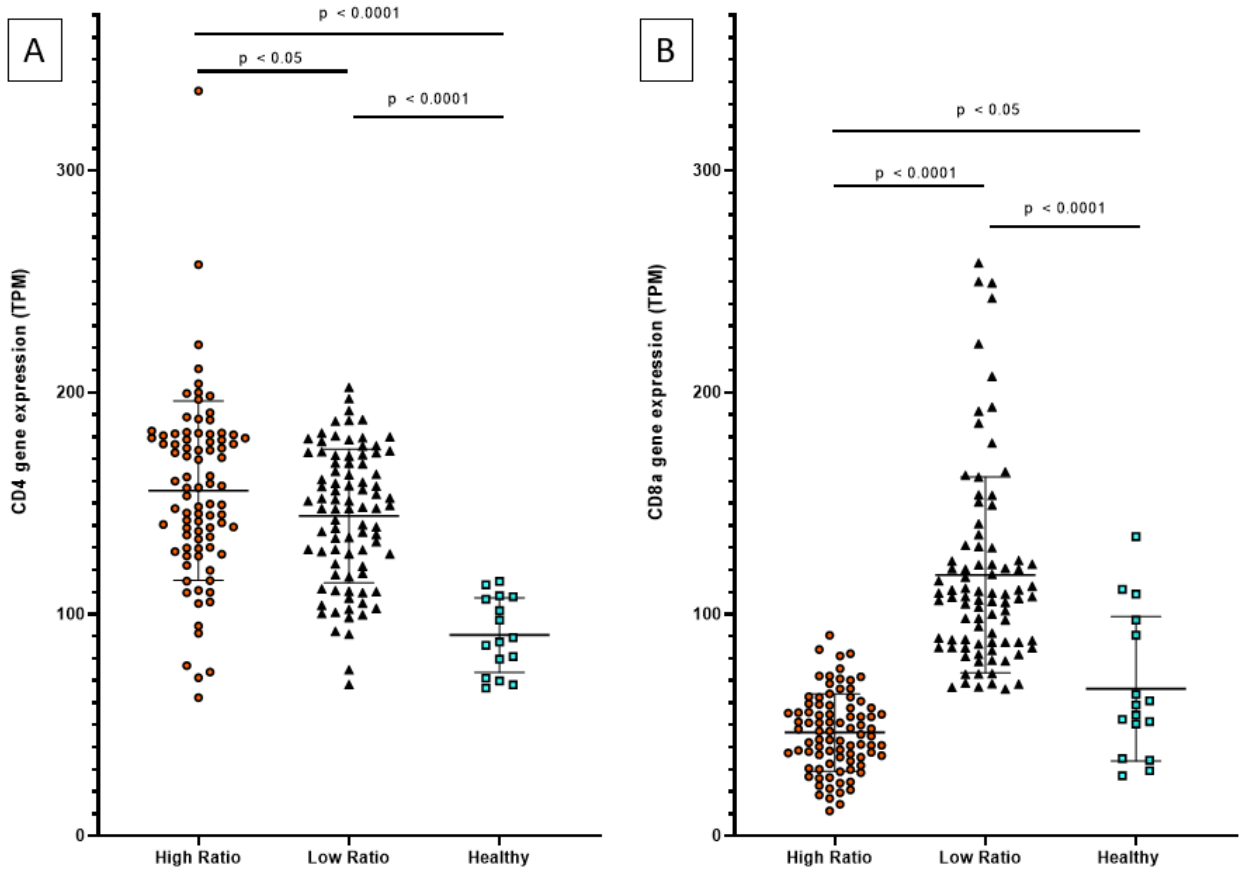


Figure 7. Comparison of CD4 (A) and CD8a (B) gene expression between High CD4+/CD8+ ratio PBC samples, Low CD4+/CD8+ ratio PBC samples and healthy controls. CD4+/CD8+ ratios from 173 PBC patient samples examined in Figure 4 were dichotomized to top 50% "High Ratio" group (n=86) and bottom 50% "Low Ratio" group (n=87). Figure 7 A shows increased CD4 TPM counts in the High ratio (Mean 155.7, SD 40.5, Range 62.4-336.0) compared to Low ratio (Mean 144.3, SD 30.1, Range 68.1-202.2) and healthy controls (Mean 90.6, SD 16.9, Range 66.7-114.7) ($p<0.05$, Mann-Whitney t-test). While Figure 7B shows decreased CD8a TPM counts in the High ratio (Mean 46.7, SD 17.4, Range 11.4-79.2) compared to Low ratio (Mean 117.7, SD 44.1, Range 66.0-258.2) and healthy controls (Mean 66.35, SD 32.6, Range 27.1-135.0) ($p<0.0001$, Mann-Whitney t-test). RNA-sequencing data was obtained as described in Figure 4. Plot was made and statistical analysis was done using PRISM..

3.2 CD4+/CD8+ ratio and the immune repertoire

CD4+/CD8+ ratios calculated using RNA-seq data were plotted against neutrophil to lymphocyte ratios (NLR) also determined from RNA-seq data to conduct a linear regression analysis. While a separate linear regression was carried out between CD4+/CD8+ ratios calculated using flowcytometry and total blood platelet counts obtained from patient charts.

Figure 8 displays that the NLR was found to have a positive correlation with CD4+/CD8+ ratios ($p=0.0003$). This indicates a progression lymphopenia correlating with a slightly expanding CD4+ T-cell population as well as a depleting CD8+ T-cell population in our PBC patient samples. Meanwhile, Figure 9 demonstrates a negative correlation between CD4+/CD8+ ratios and total blood platelet counts ($p<0.05$), reflecting an association between high CD4+/CD8+ ratios and hypersplenism which is a possible consequence of progressive PBC.

Figure 8.

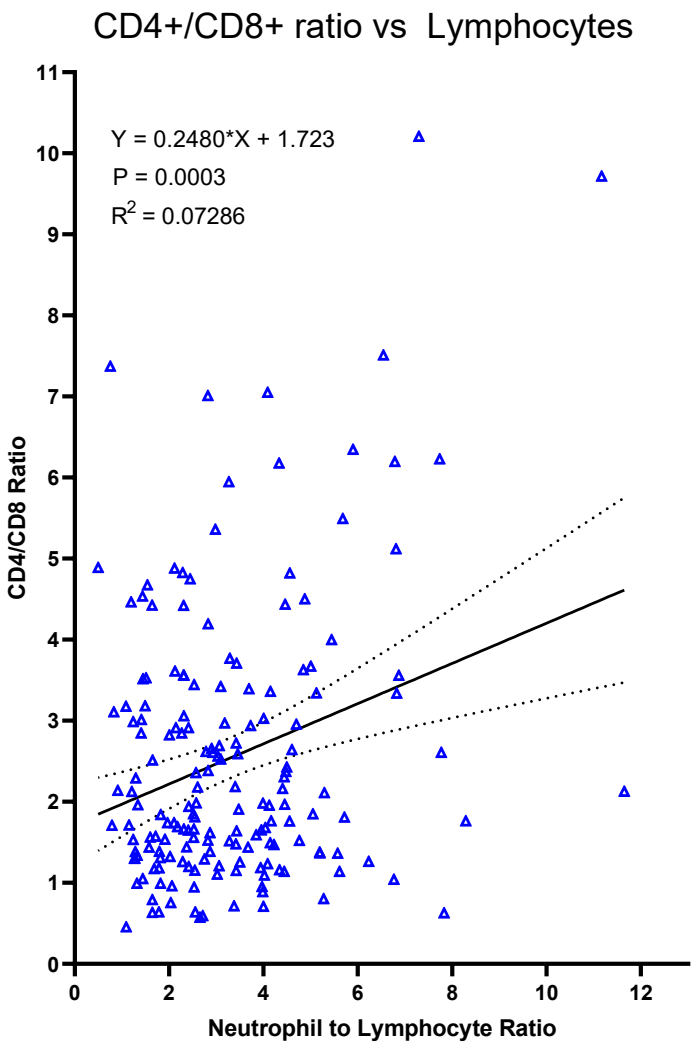


Figure 8. Linear regression of CD4+/CD8+ ratios and Neutrophil to Lymphocyte ratios (NLR) both derived from RNA-seq TPM data of PBC patients. Ratios were determined from TPM gene expression of PBC patient samples (n=173) using CD4 and CD8a genes for CD4+/CD8+ ratios and C3E and FCGR3B (CD16b) genes for NLR. Linear regression shows a positive correlation between CD4+/CD8+ ratio and NLR (R=0.07, p=0.0003) and dotted lines indicate 95% confidence bands of the best fit line. RNA-seq data was analyzed as described in Figure 4. Plot was made and statistical analysis was done using PRISM.

Figure 9.

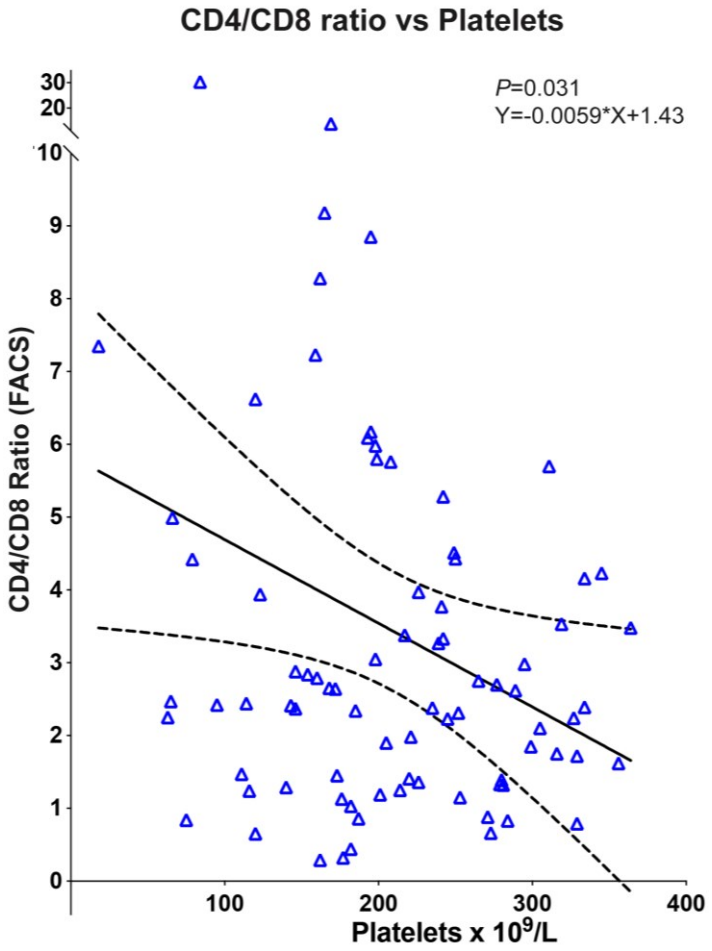


Figure 9. Linear regression of CD4+/CD8+ ratios and platelet counts of PBC patients. CD4+/CD8+ ratios were determined using proportions of CD4+ and CD8+ T-cells in PBMC of PBC patients (n=108) as described in Figure 5, while peripheral blood platelet counts for the same patients were obtained from clinical charts. Linear regression shows a negative correlation between CD4+/CD8+ ratios and platelet counts (p=0.03) while the dotted lines represent the 95% confidence bands of the best fit line. PBMCs isolation and flowcytometry was carried out as described in Figure 5. Plot was made and statistical analysis was done using PRISM.

3.3 TCR-VB repertoire in PBC patients

In our preliminary analysis we compared the TPM expression of all functional TCR-Vβ genes from 5 PBC patients against 5 healthy controls. Figure 10 displays distinct Vβ subsets as slices on a pie-chart to display TCR-Vβ distribution. We identified uniform TCR-Vβ distribution across healthy samples. In comparison, PBC patient samples had varied a Vβ distribution where different Vβ subsets emerged as the dominant Vβ expression.

Figure 10.

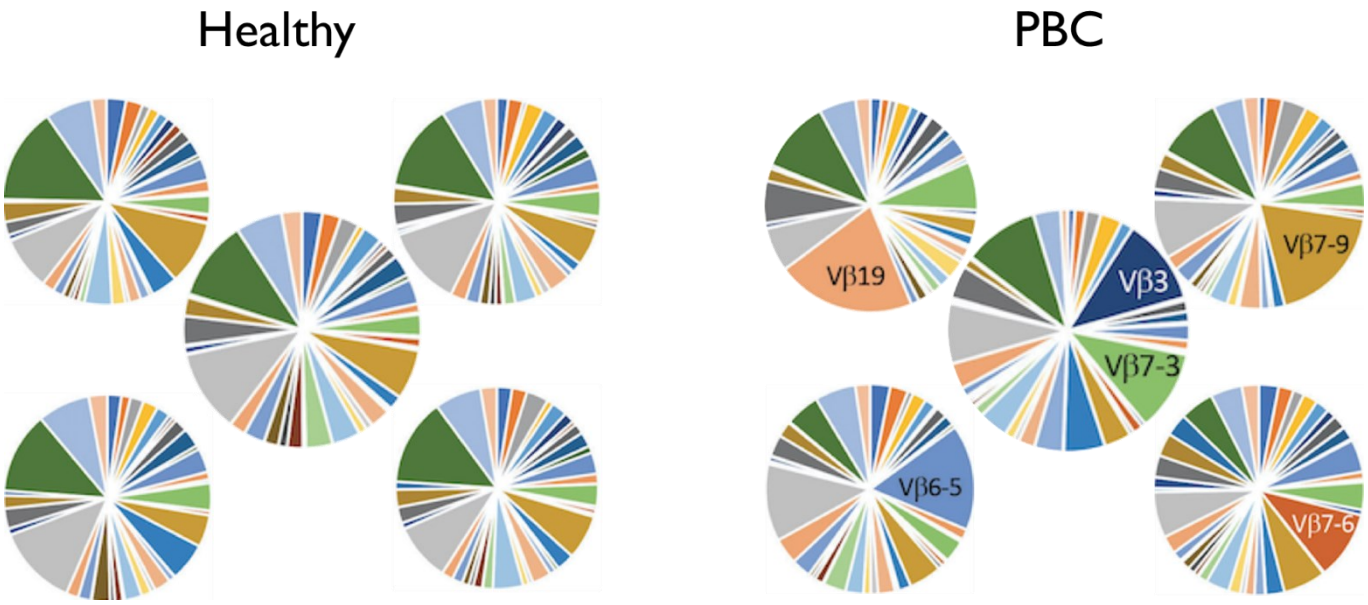


Figure 10. Pie-charts of TCR-Vβ expression of healthy controls and PBC patients displaying TCR-Vβ skewing in PBC samples. Transcript per million (TPM) data was used to determine TCR-Vβ gene expression in a small subset of healthy control (n = 5) and PBC samples (n = 5). TPM expression was obtained from RNA-sequencing data as described in Figure 4. Pir-charts were made using excel.

3.4 TCR-VB mean expression in PBC vs Healthy

To analyze mean expression of TCR-Vβ , we compared the mean expression of each Vβ subset between the PBC samples and the healthy cohort using the Mann-Whitney test using the stringent Holm-Sidak correction. Figure 11 shows mean comparison of 40 Vβ genes. We identified TCR-Vβ 28 ($p<0.005$) and Vβ 5-5 ($p<0.05$) which were downregulated in PBC patients compared to healthy control.

Utilizing an alternate methodology to evaluate TCR-Vβ distribution we also conducted a differential gene expression analysis using the DESeq2 R package [58] comparing 173 PBC patient samples against 15 healthy control. The analysis identified TCR-Vβ 28 and Vβ 6-4 as downregulated in PBC (Supplement Table 1).

Figure 11.

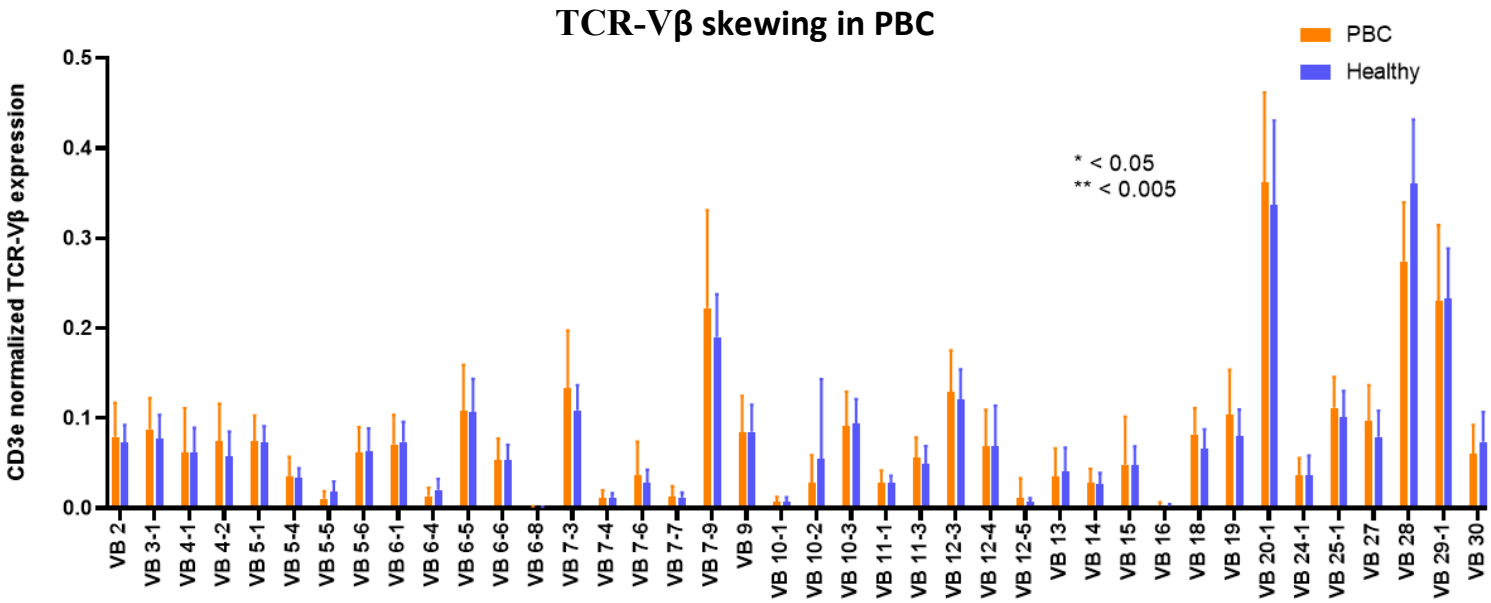


Figure 11. Comparison of mean TCR-Vβ expression between PBC patient samples (n=173) and healthy controls (n=15). TCR-Vβ TPM expression was normalized by dividing TPM counts from each Vβ subset with the respective CD3e TPM expression from that sample. Difference in mean expression was determined by multiple t-test corrected by the Holm-Sidak method. * Indicates a p value of < 0.05 while ** represent a p value of < 0.005. RNA-sequencing data was obtained as described in Figure 4. Plot was made and statistical analysis was done using PRISM.

3.5 Analysis of TCR-Vβ expansion and depletion and evaluation of frequency of Vβ skewing.

To better evaluate TCR-Vβ skewing in PBC patients we recorded the frequency of expansion or depletion of each Vβ subset in individual patients using Vβ expression of healthy controls (n=15) to establish a criterion for normal Vβ expression. Criteria for skewed VB expression was established by defining the range of normal VB expression as healthy mean + 3 x SD (standard deviation) to healthy mean – 2 x SD. Using conditional formatting we visualized positive skewing in Vβ as red cells and negative skewing as yellow cells in order to determine the extent of abnormal Vβ expression in PBC patient samples (n=173) and healthy controls. We identified that healthy controls had almost no indication of skewing, although one control patient was observed to have depletions in multiple Vβ subsets. This patient was identified to be the sister of a PBC patient. Supplement Figure 1 showcases excerpts from PBC patients and healthy control.

A total of 173 samples were collected from 128 PBC patients. Around 80% of all PBC patients displayed either an expansion or depletion in at least one Vβ subset. Overall, 62% of PBC patients had expansions (79/128), 42% had depletions (54/128) and 24% (31/128) had both. Samples from PBC patients were observed to have expansions in variable VB subsets, however some VB subsets were more commonly skewed than others. Figure 12 shows that, from the PBC samples with expansions, Vβ 7-3 was most frequently expanded (22%) followed by Vβ2 (17%), Vβ 12-5 (15%), Vβ16 (14%) and Vβ18 (13%).

Similarly, depletion of Vβ subsets was also found to be varied in PBC patients. Figure 13 shows that of the PBC samples with depletions, Vβ 28 was most frequently depleted (53%) followed by Vβ-30 (20%), Vβ 6-1 (17%), Vβ 29-1 (16%) and Vβ 10-3 (13%).

Figure 12

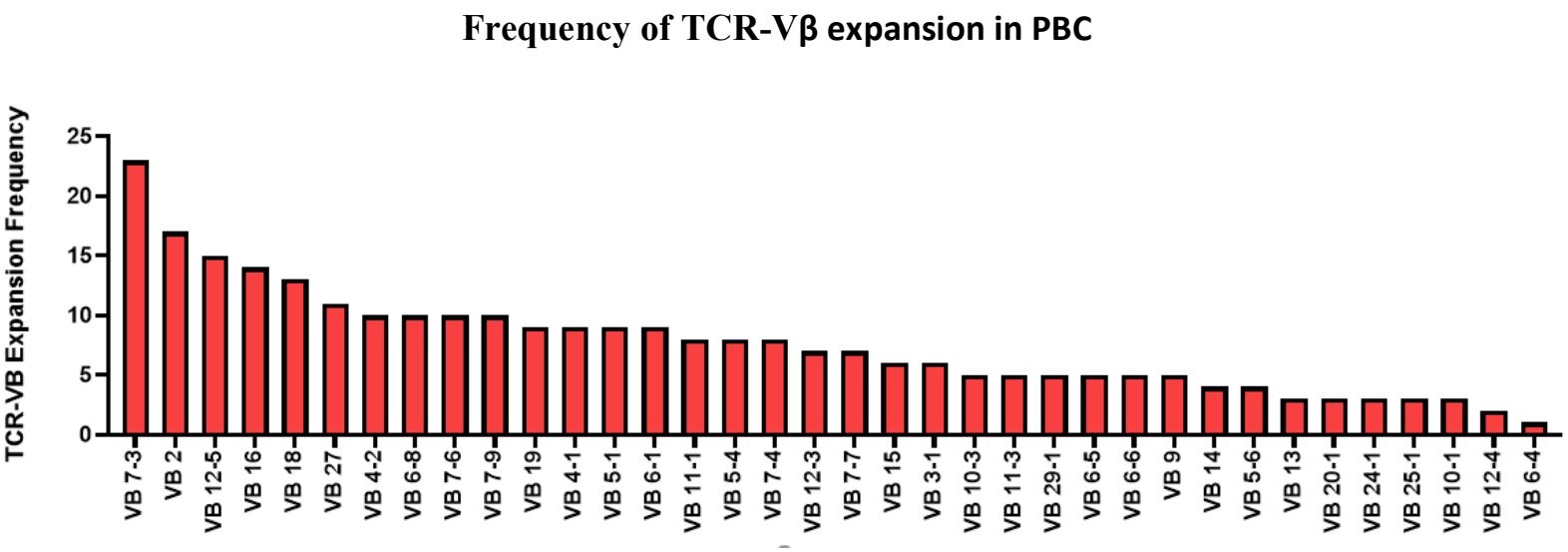


Figure 12. Frequency of TCR-Vβ expansion in PBC patient samples. A total of 173 PBC patient samples (from 128 PBC patients) were investigated for positive skewing as determined by > healthy mean + 3 x SD of normal Vβ expression. 138 out of 173 PBC patient samples displayed events of positive skewing. Vβ TPM was normalized using CD3E TPM expression as described in Figure 11. RNA-sequencing data was obtained as described in Figure 4. Bar plot was made using PRISM.

Figure 13.

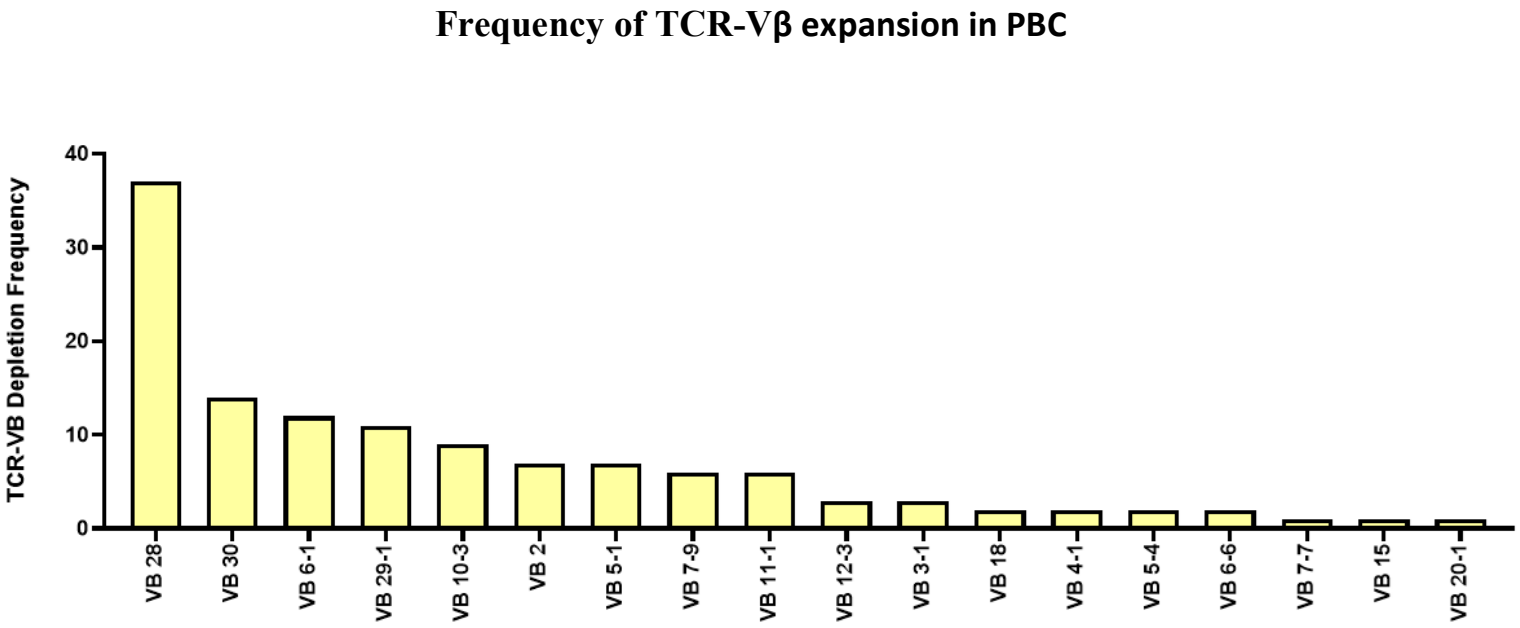


Figure 13. Frequency of TCR-Vβ depletion in PBC patient samples. A total of 173 PBC patient samples (from 128 PBC patients) were investigated for negative skewing as determined by < healthy mean - 2 x SD of normal Vβ expression. 70 out of 173 PBC patient samples displayed events of positive skewing. Vβ TPM was normalized using CD3E TPM expression as described in Figure 11. RNA-sequencing data was obtained as described in Figure 4. Bar plot was made using PRISM.

3.6 Clustering of Skewed TCR-Vβ subset**.**

Recognizing the variance of skewed TCR-Vβ subsets in PBC patients, we evaluated the pattern of Vβ skewing in PBC patients. We conducted hierarchical clustering analyses on events of positively and negatively skewed Vβ. Additionally, we created heatmaps of frequency tables marking cooccurrence of expanded Vβ subsets in individual PBC samples.

We recorded the tendency of frequently skewed Vβ subsets to experience concurrent skewing events. Utilizing dendrograms of positively and negatively skewed subsets (Figure 14 and Figure 15) along with their respective frequency table heatmaps (Supplement Figure 2 and Supplement Figure 3) we characterized specific subgroups of Vβ subsets which demonstrated frequent overlap. For the 40 Vβ groups investigated in our study, out of numerous possible combinations we identified 8 unique combinations for positively skewed Vβ subsets (Table 1.) and 5 possible combinations for negatively skewed Vβ subsets (Table 2.). For positively skewed subsets, Vβ7-3, Vβ 12-5 and Vβ 2 comprised the most prominent group while the most prominent group for negatively grouped subsets was comprised of Vβ 28, Vβ6-1, Vβ29-1, and Vβ5-1.

Figure 14.

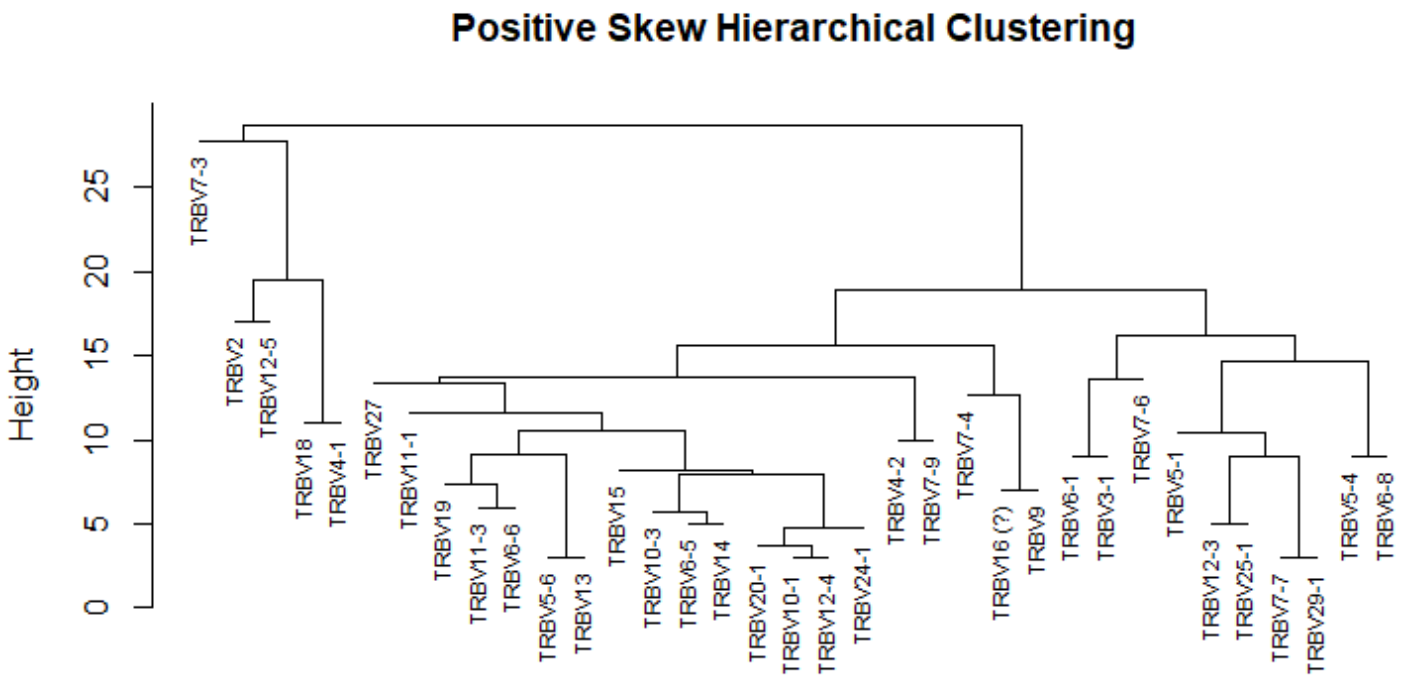


Figure 14. Hierarchical clustering of positive TCR-Vβ skewing in PBC patients. Clustering was conducted by converting skewing frequency to binary data using 0 to indicate Vβ TPM expression within the normal range (healthy mean + 3 x SD to healthy mean – 2 x SD) and 1 to indicate positive skewing events in Vβ subsets (> healthy mean + 3 x SD). Clustering was done in Agglomerative Nesting manner using the "agnes" command in the "cluster" R package using the ward method and the manhatten metric. Dendrogram was constructed using the dendrogram of agnes specification in the pltree function.

Figure 15.

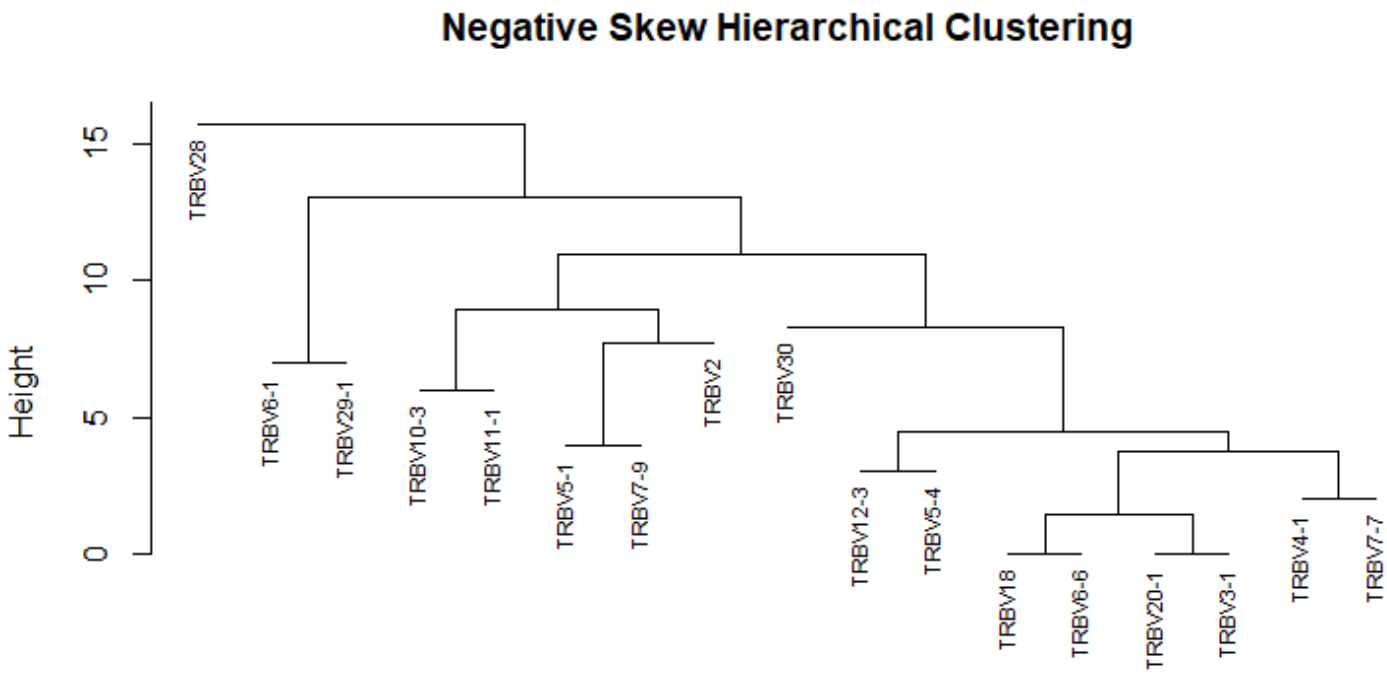


Figure 14. Hierarchical clustering of negative TCR-Vβ skewing in PBC patients. Clustering was conducted by converting skewing frequency to binary data using 0 to indicate Vβ TPM expression within the normal range (healthy mean + 3 x SD to healthy mean – 2 x SD) and 1 to indicate negative skewing events in Vβ subsets (< healthy mean – 2 x SD). Clustering was done in Agglomerative Nesting manner using the "agnes" command in the "cluster" R package using the ward method and the manhatten metric. Dendrogram was constructed using the dendrogram of agnes specification in the pltree function.

Table 1.

TCR-Vβ Positively Skewed Groups
Vβ 7-3, Vβ 12-5, Vβ 2
Vβ 7-3, Vβ 27, Vβ 2
Vβ 29-1, Vβ 7-7, Vβ 5-1
Vβ 18, Vβ 2, Vβ 4-1
Vβ 12-5, Vβ 2, Vβ 4-1
Vβ 29-1, Vβ 7-7, Vβ 5-1
Vβ 2, Vβ6-1
Vβ 18, Vβ 7-9

Table 1. **Possible TCR-Vβ groupings of positively skewed Vβ subsets based on hierarchical clustering and frequency of Vβ subset overlap.** Hierarchical clustering of positively skewed Vβ subsets displayed in Figure 14 was used in conjunction with the observed frequency of overlap between positively skewed TCR-Vβ subsets (Supplement Figure 2) to create predictive groups for superantigen TCR-Vβ specificity.

Table 2.

TCR-Vβ Negatively Skewed Groups
Vβ 28, Vβ 6-1, Vβ 29-1, Vβ 5-1
Vβ 28, Vβ 5-1, Vβ 7-9
Vβ 2, Vβ 6-1, Vβ5-1
Vβ 11-1, Vβ 6-1, Vβ 10-3
Vβ 29-1, Vβ 6-1

Table 2. **Possible TCR-Vβ groupings of negatively skewed Vβ subsets based on hierarchical clustering and frequency of Vβ subset overlap.** Hierarchical clustering of negatively skewed Vβ subsets displayed in Figure 15 was used in conjunction with the observed frequency of overlap between negatively skewed TCR-Vβ subsets (Supplement Figure 3) to create predictive groups for superantigen TCR-Vβ specificity.

3.7 Differential Gene Expression and Pathway analysis of High vs Low CD4+/CD8+

To identify biological pathways involved in T-cell dysregulation, we dichotomized patient population using the top and bottom 10% of CD4+/CD8+ ratios. Differential gene expression analysis identified 1257 genes to be upregulated in the high CD4+/CD8+ ratio group compared to the low ratio group. While 755 genes were found to be downregulated in the low CD4+/CD8+ ratio group compared to the high ratio group. Conducting pathway analysis on these deregulated genes using Cytoscape ClueGO [59], we were able to identify relevant immunological pathways that provide insight into PBC patient immune dysfunction.

In the High CD4+/CD8+ ratio group, we identified upregulation of anti-viral pathways, Type 1 interferon signaling, Interleukin 8 and Interleukin 12 production and activation of leukocyte differentiation.

Antiviral pathways included the Oligoadenylate Synthase family (OAS1,OAS2,OAS3) which are activators of RNase L [62], HERC5 a modulator of antiviral response [63], EIF2AK2 which regulates transcription factors and viral immune response [64] and TRIM family proteins (TRIM 34, TRIM 5) which are involved in anti-retroviral activity [65]. Type 1 interferon response included Interferon Alpha Inducible (IFI) Protein family genes (IFI27, IFI6, IFIT1,IFIT2,IFIT3), Immune signal transducer MYD88 [66], and MX1 an antagonist of DNA/RNA virus replication [67]. We also identified upregulation in SP100, a gene involved in antiviral immunity, tumorigenesis, and a potential target for autoimmunity [68, 69]. Interleukin 12 and Interleukin-8 production pathway upregulation was observed by activation of TLR1, TLR2, TLR3, TLR4, TLR7 and TLR8 and IRF5 (interferon regulatory factor 5) genes. DDX58 gene which encodes RIG-I, an innate immune receptor of cytoplasmic viral nucleic acids and activator of Type 1

interferon was also found to be upregulated [70]. Finally, leukocyte differentiation pathways were also upregulated including ZBTB7B, which encodes a protein involved in lineage specification of T-cells [71] and HCLS1 which is involved in antigen receptor signaling for clonal expansion and deletion [72].Table 3. Showcases some of the identified pathways along with the associated genes found to be upregulated in the High CD4+/CD8+ ratio group.

In contrast, the Low CD4+/CD8+ ratio group highlighted pathways involved in natural killer cell mediated immunity, T-cell differentiation in thymus, Alpha-Beta T-cell activation and Negative T-cell selection (Table 4.) Upregulation of natural killer immunity involved KLRD1, CRATM and LAG3. While T-cell differentiation and Alpha-Beta T-cell activation involved the ATG5, CD28, WNT1, DNAJA3 and ZAP70 genes. Interestingly LAG 3 is also involved in negative regulation of T-cell activation [73], DNAJA3 is also involved in tumor suppression [74] while Zap70 also regulates B-cell activation [75]. Entrez Gene from NCBI and UniProt were used to identify gene names from gene symbols and identify proteins encoded by the identified genes [76, 77].

Table 3.

Differential Gene Enrichment in PBC samples with High CD4/CD8

Upregulated in High CD4/CD8	Associated Genes	% Associated Genes	Corrected PValue
Type I interferon signaling pathway	ADAR, EGR1, FADD, IFI27, IFI6, IFIT1, IFIT2, IFIT3, IRF5, ISG15, MX1, MX2, MYD88, OAS1, OAS2, OAS3, OASL, PTPN6, RSAD2, SAMHD1, SP100, TREX1, USP18	23.96	3.04E-16
Defense response to virus	ADAR, APOBEC3A, ATG7, BCL2L1, CARD9, DDX58, DHX58, EIF2AK2, FADD, HERC5, IFI16, IFI27, IFI44L,IFI6, IFIT1, IFIT1B, IFIT2, IFIT3, IFNGR2, IL10RB, IL1B, IRF5, ISG15, MX1, MX2, NLRP3, NLRX1, OAS1,OAS2, OAS3, OASL, PCBP2, PLSCR1, RNASE6, RNF135, RSAD2, SAMHD1, TLR7, TLR8, TREX1, TRIM34, TRIM5, UNC93B1	16.48	3.04E-16
Interleukin-12 production	ARRB2, CD36, IL16, IRAK3, IRF5, LGALS9, LILRA5, MAPK14, MEFV, NFKB1, PRKCD, SCIMP, SYK, THBS1, TLR2, TLR4, TLR8	26.56	1.08E-30
Positive regulation of leukocyte differentiation	AP3B1, ATP6AP1, BTK, CCR1, CD4, CD86, CTNNBIP1, EVI2B, FADD, FES, HCLS1, IL1B, INPP5D, LGALS9, LILRB2, LILRB4, MMP14, NCKAP1L, NLRP3, NOTCH2, PF4, RHOA, SYK, TESC, VSIR, ZBTB7B	16.25	1.08E-30
Positive regulation of interleukin-8 production	CD14, DDX58, FADD, FCN1, HSPA1A, IL1B, LGALS9, MYD88, SYK, TLR1, TLR2, TLR4, TLR5, TLR7, TLR8	25.86	1.08E-30

Table 3. **Biological pathways identified as upregulated in the <u>High</u> CD4+/CD8+ ratio PBC samples compared to the low CD4+/CD8+ ratio samples.** Differential gene expression analysis was carried out between RNA-seq data from PBC patients with the highest CD4+/CD8+ ratio samples (top 10%) vs the lowest CD4+/CD8+ ratio samples (bottom 10%) subsetted from the total PBC samples displayed in Figure 4. “Upregulated in High CD4/CD8” highlights upregulated pathways, “Associated genes” lists the specific genes from the relevant pathways identified as upregulated, “% associated genes” provides the proportion of genes in each specific pathway that were found to be upregulated. “Corrected pValue” provides the pValue after correcting for multiple comparisons. A total of 1257 genes were found to be differentially upregulated in the high CD4/CD8 group. Differential gene expression was conducted using DESeq2 R package and pathway analysis was done using Cytoscape ClueGo and GeneOntology.

Table 4.

Differential Gene Enrichment in PBC samples with Low CD4/CD8

Upregulated in Low CD4/CD8	Associated Genes	% Associated Genes	Corrected P Value
Natural killer cell mediated immunity	CD160, CD226, CD96, CRTAM, KLRD1, KLRK1, LAG3, RASGRP1, SH2D1A, SLAMF6, SLAMF7	16.67	9.88E-05
T cell differentiation in thymus	ATG5, CD28, CD3D, CD3E, CD3G, CDKN2A, DNAJA3, IL7R, RABL3, RASGRP1, SOD1, TOX, WNT1, ZAP70	17.72	2.38E-03
Alpha-Beta T-cell activation	BATF, CD160, CD28, CD3E, CRTAM, EOMES, GNAO1, GPR18, GPR183, IFNG, LY9, PRKCQ, PTPN22, RORA, RUNX3, SLAMF6, TBX21, TOX, ZAP70, ZNF683	13.16	2.38E-03

Table 4. **Biological pathways identified as upregulated in the Low CD4+/CD8+ ratio PBC samples compared to the high CD4+/CD8+ ratio samples.** Differential gene expression analysis was carried out between RNA-seq data from PBC patients with the highest CD4+/CD8+ ratio samples (top 10%) vs the lowest CD4+/CD8+ ratio samples (bottom 10%) subsetted from the total PBC samples displayed in Figure 4. “Upregulated in High CD4/CD8”, “Associated genes”, “% associated genes”, and “Corrected pValue” describe pathway enrichment data as listed in Table 3. A total of 755 genes were found to be differentially upregulated in the low CD4/CD8 group. Differential gene expression was conducted using DESeq2 R package and pathway analysis was done using Cytoscape ClueGo and GeneOntology.

3.8 Differential Gene Expression and Pathway analysis of High vs Low NLR.

We carried out differential gene expression analysis on the PBC patient samples with the highest (top 10%) NLR values against the PBC patient samples with the lowest (bottom 10%) NLR values. We identified 3868 upregulated genes in the high NLR group (consequently these genes were downregulated in the low NLR group). While 2902 genes were downregulated genes in the high NLR group (these were upregulated in the low NLR group) (p-adjusted <0.05). Pathway analysis identified activation of Type 1 interferon, Cytokine response, TLR signaling and defense response in patients with high NLR (Table 5). While regulation of viral transcription, innate immunity and myeloid cell activation was upregulated in patients with low NLR (Table 6).

Table 5.

Differential Gene Enrichment in PBC samples with High NLR

Upregulated in High NLR	Associated Genes	% Associated Genes	Corrected PValue
Type I interferon signaling pathway	ADAR, CNOT7, FADD, GBP2, HFE, HLA-E, IFI27, IFI35, IFI6, IFIT1, IFIT2, IFIT3, IFITM1, IFITM2, IFITM3, IFNAR1, IFNAR2, IRF1, IRF2, IRF9, ISG15, JAK1, LSM14A, MX1, MX2, MYD88, OAS1, OAS2, OAS3, OASL, PTPN1, PTPN11, PTPN2, RNASEL, RSAD2, SAMHD1, SP100, STAT1, STAT2, TBK1, TRIM6, UBE2K, USP18, XAF1, YTHDF2, YTHDF3, ZBP1	48.96	1.27E-08
Defense response to virus	ABCC9, ADAR, AIM2, ANKRD17, APOBEC3A, APOBEC3B, ATG7, ATL1, BCL2L1, BECN1, BIRC2, BNIP3L, CGAS, CNOT7, CRCP, CXCL10, DDX17, DDX21, DDX58, DDX60, DHX36, DHX58, DHX9, DNAJC3, DTX3L, EIF2AK2, EIF2AK4, ELMOD2, ERCC6, EXOC1, F2RL1, FADD, FAM111A, FGL2, GBP1, HERC5, IFI16, IFI27, IFI44L, IFI6, IFIH1, IFIT1, IFIT1B, IFIT2, IFIT3, IFIT5, IFITM1, IFITM2, IFITM3, IFNAR2, IFNGR2, IL10RB, IL15, IL1B, ILRUN, IRF1, IRF2, IRF9, ISG15, ITCH, ITGAX, LSM14A, LYST, MICB, MX1, MX2, NCBP3, NLRP3, NT5C3A, OAS1, OAS2, OAS3, OASL, PARP9, PCBP2, PLSCR1, PMAIP1, PPM1B, PTPN22, PTPRC, PUM1, PUM2, RIOK3, RNASEL, RNF135, RSAD2, RTP4, SAMHD1, SEC14L1, SERINC3, SETD2, SLFN11, STAT1, STAT2, TBK1, TLR3, TLR7, TLR8, TNFAIP3, TRIM22, TRIM25, TRIM34, TRIM38, TRIM5, TRIM6, USP15, ZC3HAV1, ZMPSTE24	41.38	4.15E-11
MyD88-dependent toll-like receptor signaling pathway	BTK, CD300A, CD36, IRAK3, IRAK4, IRF1, LY96, MAP3K1, MAP3K7, MYD88, TAB2, TAB3, TIRAP, TLR1, TLR2, TLR4, TLR5, TLR6, TLR7, TLR8, TNIP1, TRAF6	59.46	4.66E-17
Positive regulation of response to cytokine stimulus	ADAM17, CASP1, CASP4, CREBRF, CXCR4, DDX58, DHX9, FADD, HIF1A, IFIH1, IL1R1, LSM14A, MBD4, MED1, PAFAH1B1, PARP14, PARP9, RIPK1, RIPK2, TANK, TBK1, TICAM2, TLR2, TLR4, TMED7-TICAM2, TRIM6, UBE2K, ZBP1	45.16	2.74E-15
Positive regulation of toll-like receptor signaling pathway	DDX3X, F2RL1, FLOT1, PELI1, PJA2, PTPN22, RSAD2, TICAM2, TIRAP, TLR1, TLR2, TLR3, TLR5, TMED7-TICAM2, TREML4, WDFY1	59.26	4.66E-17

Table 5. Biological pathways identified as upregulated in PBC samples with <u>High</u> NLR compared to PBC samples with low NLR. Differential gene expression analysis was conducted on RNA-seq data from PBC patients with the highest NLR values (top 10%) vs the lowest NLR values (bottom 10%). These PBC patient samples were subsetted from the total PBC patient samples (n=173) for which we obtained RNA-sequencing data as described in Figure 4. "Upregulated in High CD4/CD8", "Associated genes", "% associated genes", and "Corrected pValue" describe pathway enrichment data as listed in Table 3. A total of 3868 genes were found to be differentially upregulated in the high NLR group. Differential gene expression was conducted using DESeq2 R package and pathway analysis was done using Cytoscape ClueGo and GeneOntology.

Table 6.

Differential Gene Enrichment in PBC samples with Low NLR

Upregulated in Low NLR	Associated Genes	% Associated Genes	Corrected pValue
Regulation of Viral transcription	AAAS, CDK9, MRPL23, MRPS12, NELFA, NELFB, NUP188, NUP210, NUP85, POLR2E, POLR2F, POLR2H, POLR2I, POLR2L, PSMC3, RPL10, RPL10A, RPL13, RPL13A, RPL14, RPL18, RPL18A, RPL23A, RPL28, RPL29, RPL3, RPL32, RPL35A, RPL37A, RPL4, RPL5, RPL7A, RPL8, RPLP0, RPLP1, RPLP2, RPS10, RPS11, RPS15, RPS15A, RPS16, RPS18, RPS19, RPS2, RPS21, RPS28, RPS3, RPS4X, RPS5, RPSA, SMARCB1, SUPT5H, TARBP2, TFAP4, TRIM11, TRIM32, USF2	29.84	1.75E-06
Innate immune response	ADAMTS13, ADGRB1, AKAP8, AMH, APOE, C1QBP, C1S, C8G, CACTIN, CARD11, CD6, CDC37, CHID1, COCH, DAPK3, DDX56, EPRS1, EVL, FAM3A, GATA3, GZMM, HLA-DPB1, HLA-DQB2, HLA-G, HMGB3, HPX, HRAS, ICAM2, IGHA1, IGHA2, IGHD, IGHE, IGHG1, IGHG2, IGHM, IGHV1OR15-1, IGHV1OR15-9, IGHV3-13, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-72, IGHV3-73, IGHV3-74, IGHV3OR16-9, IGHV4-28, IGHV5-51, IGLC2, IKBKE, IRAK1, IRF3, IRF4, LAG3, LY9, MAP4K2, METTL3, MIF, MST1R, MUC1, MUC20, MYC, MYO1C, NCR3, PKN1, PNKD, POLR3D, POLR3E, POLR3H, PPP1R14B, PQBP1, PSMB11, PSMC3, PSMC5, PSMD3, PSMD8, RAB7B, RBM14, RBM14-RBM4, RPL13A, RPS19, RPSA, SARM1, SFPQ, SHMT2, STING1, TAB1, TICAM1, TKFC, TP53, TRBC2, TRGV4, TRIM11, TRIM28, TRIM32, TRIM35, TRIM41, VSIG4, WRNIP1, ZDHHC11	9.81	1.07E-07
Myeloid cell activation involved in immune response	A1BG, ACTR1B, AIFM2, ALDOC, AMH, AP2A2, APEH, APRT, ARHGAP45, ATAD3B, C1orf35, DDOST, DPP7, EEF2, EGFL7, FOXF1, GHDC, GPI, GSTP1, HMOX2, IMPDH2, ITM2C, KCNAB2, LAT, MIF, MMP28, NME2, ORMDL3, PFKL, PLPPR3, PPIE, PRKCSH, PSMC3, PSMD3, PTGES2, PYGB, RHOF, RPSA, SPHK2, SPNS1, STING1, SYNGR1, TBC1D10C, TICAM1, TMC6, TMEM63A, VAT1	7.77	2.15E-05

Table 6. Biological pathways identified as upregulated in PBC samples with <u>Low</u> NLR compared to PBC samples with low NLR. Differential gene expression analysis was conducted on RNA-seq data from PBC patients with the highest NLR values (top 10%) vs the lowest NLR values (bottom 10%). These PBC patient samples were subsetted from the total PBC patient samples (n=173) for which we obtained RNA-sequencing data as described in Figure 4. "Upregulated in High CD4/CD8", "Associated genes", "% associated genes", and "Corrected pValue" describe pathway enrichment data as listed in Table 3. A total of 2902 genes were found to be differentially upregulated in the low NLR group. Differential gene expression was conducted using DESeq2 R package and pathway analysis was done using Cytoscape ClueGo and GeneOntology.

3.9 Transcriptional marker for TCR-Vβ skewing in PBC patients.

Investigating potential genetic markers for TCR-Vβ skewing in PBC patients, we were able to identify CD3E-204 transcript expression to have a strong inverse correlate with Vβ skewing.

CD3E is a marker for CD3+ lymphocytes [78]. Alignment of RNA-seq against the human genome allowed us to identify six different transcripts for the CD3E gene including CD3E-201, CD3E-202, CD3E-203, CD3E-204, CD3E-205, CD3E-206. Using the ensembl database we were able to identify CD3E-201 and CD3E-204 as the two functional transcripts that encode the CD3E protein [79]. Aside from their listing in the ensembl database as well as the nucleotide sequence of both transcripts, no further information was available on the functional differences between the two transcripts.

To determine the relationship between TCR-Vβ skewing and CD3E-204 transcript expression, we had to evaluate changes in TCR-Vβ skewing as well as changes in CD3E-204 expression across different timepoints in the same PBC patients. For this investigation, we identified PBC patients for which we had RNA-sequencing data at two different timepoints. A "Before" indicator was used to describe the baseline timepoint sample (before treatment in the OCA trial) while a "After" indicator was used to describe a later timepoint sample (collected after at least 12-month treatment with OCA). A total of 35 PBC patients were identified which had a RNA-seq samples available at both the "Before" timepoint and the "After" timepoint. 5 of these patients did not display any changes in TCR-Vβ skewing between the timepoints, 20 exhibited expansions in TCR-Vβ subset skewing from the "Before" to the "After" timepoint while 11 showed contractions in TCR-Vβ subset skewing from the "Before" to the "After"

timepoint. Figure 16 shows changes in A) CD3E gene expression, B) CDE3-201 transcript expression and C) CD3E-204 transcript expression in PBC patients experiencing Expansions or Contractions of TCR-Vβ skewing across two different timepoints.

The CD3E gene expression was noticed to be decreased in PBC patient samples with higher TCR-Vβ skewing as identified in Figure 16A (i.e., lower CD3E expression after expansion of TCR-Vβ skewing and before contraction of TCR-Vβ skewing). On the other hand, Figure 16B displays an increase in CD3E-201 transcript expression correlating with higher TCR-Vβ skewing (i.e., higher CD3E-201 expression after expansion and before contraction). Finally, Figure 16C shows a drastic reduction in CD3E-204 transcript expression in correlation with higher TCR-Vβ skewing (i.e., significant reduction in CD3E-204 transcript expression after expansion and recovery of CDE-204 transcript expression after contraction).

Figure 16.

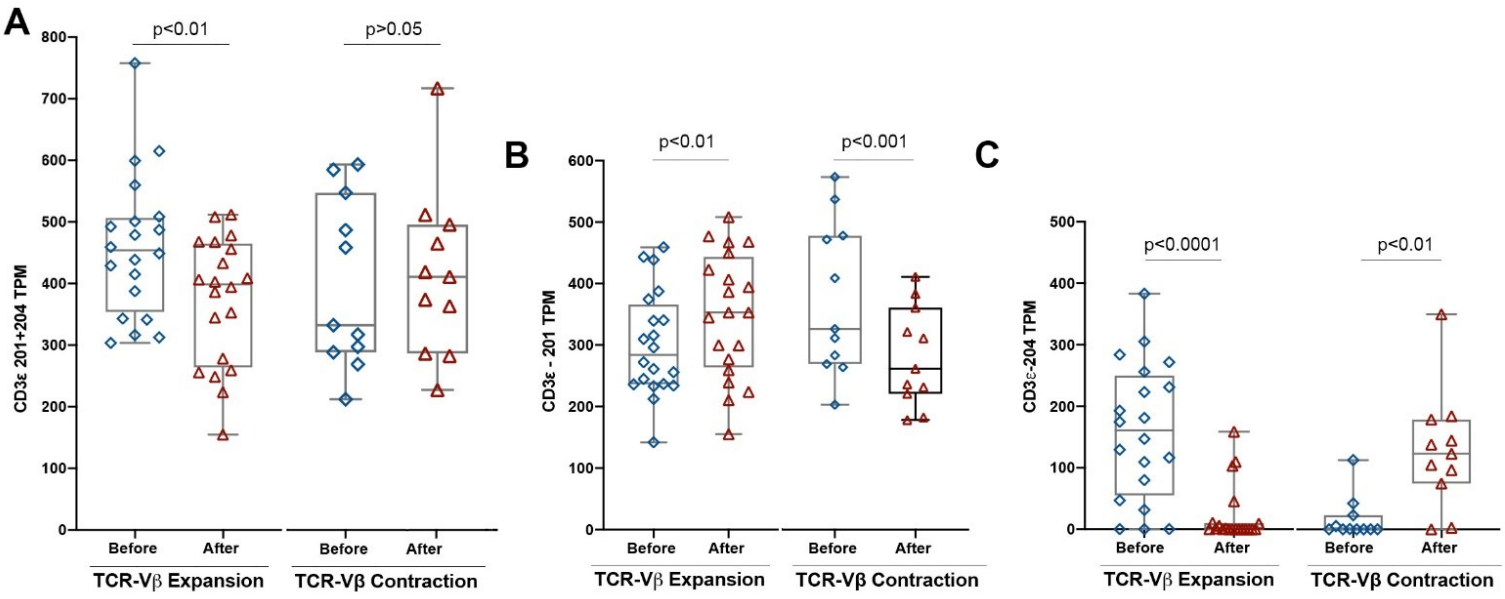


Figure 16. Changes in CD3E gene expression (A), CD3E-201 transcript expression (B) and CD3E-204 transcript expression (C) in PBC patient samples obtained at baseline "Before" treatment and "After" treatment with OCA. A total of 35 PBC patients with both a "Before" and an "After" timepoint sample were identified. 20 patients exhibited Expansion in TCR-Vβ skewing from Before to After, 11 showed Contraction in TCR-Vβ skewing from Before to After while the remaining 5 displayed no change in TCR-Vβ skewing across the two timepoints. Figure 16 A, B and C show the correlation between change in CD3E gene/functional transcript expression and increased or decreased TCR-Vβ skewing across time (marked as expansion or contraction). CD3E gene expression was determined by combining expression of the functional transcripts CD3E-201 and CD3E-204. TPM expression of CD3E-201 and CD3E-204 was determined from RNA-seq data which was obtained from PBC patients as described in Figure 4. Plot construction and statistical analysis was conducted using PRISM.

4. Discussion

4.1 Introduction

Primary Biliary Cholangitis is a chronic liver disease characterized by autoimmune mediated destruction of small hepatic bile ducts [1]. Currently, treatments such as Ursodeoxycholic acid and Obeticholic acid are useful in normalizing PBC patients, however 50% of patients still require additional therapy with modulating drugs [80], and PBC patients still make up around 10% of all liver transplants in North America and Europe. Furthermore, recurrence of PBC after transplant is a common phenomenon ranging from 10% to 35% of all PBC transplants depending on external risk factors, age at transplant and time since transplant [81]. Treatments to control PBC progression and reduce risks of recurrent PBC after transplant have shown variable levels of success [80]. As PBC is a multifactorial disease with strong autoimmune, genetic, and infectious elements [13, 25], further understanding of the mechanisms surrounding disease initiation and development is needed to develop appropriate therapies.

Previous investigations on the immunological state of Primary Biliary Cholangitis patients have individually recorded snapshots of CD4+ T-cell and CD8+ T-cell populations, abnormal neutrophil to lymphocyte ratio in as well as T-cell receptor Variable β region gene bias in PBC.

Researchers studying T-cell population in 25 PBC patients identified normal CD4+ T-cell counts while documenting lower levels of CD8+ T-cells. The absolute numbers of CD8+ T-cells was also inversely correlated with total serum bilirubin levels indicating higher CD4+/CD8+ ratios in patients with worse disease [82]. A study aiming to determine the prognostic value of neutrophil to lymphocyte ratio in 88 retrospective and 63 prospective patients (followed for 1-year) reported higher NLR scores to correlate with higher short-term mortality rates in PBC [83]. Another study

investigating the immune repertoire in 28 PBC patients described TCR-Vβ skewing in PBC patients. The study identified a bias towards specific Vβ in peripheral blood compared to healthy control (notably Vβ 6.1) while also describing preferential intrahepatic accumulation of specific Vβ subsets in the liver (Notably Vβ7 and Vβ-13) [32]. While these findings provide valuable contributions to our current understanding of the immunological environment of PBC patients, they also indicated a need for more comprehensive analysis to investigate the mechanisms driving immune system dysfunction in PBC.

Herein, we attempt to comprehensively characterize the immunological environment of PBC patients by investigating a RNA-seq database of 173 PBC patient samples from 128 PBC patients supplemented by flowcytometry analysis of peripheral blood from 108 PBC patients and intrahepatic lymphocytes from 12 PBC patients.

4.2 Immune environment in Primary Biliary Cholangitis.

A) CD4+/CD8+ ratio and distribution of CD4+T-cells and CD8+ T-cells gene expression in PBC.

A comprehensive analysis of the CD4+ and CD8+ TPM gene expression in peripheral blood identified a higher mean and increased range of the CD4+/CD8+ ratio in PBC patients compared to autoimmune hepatis and healthy controls. These findings were corroborated with increased range observed in PBMC and Intrahepatic lymphocytes of PBC patients using flowcytometry analysis.

Taking previous reports of CD4+/CD8+ ratio and negative correlation of CD8+ T-cells with bilirubin values into consideration [82] we dichotomized the CD4+/CD8+ ratio values derived

from TPM counts to the top 50% and the bottom 50% to better illustrate the T-cell population. On plotting the individual CD4 and CD8a gene expression values for the high CD4+/CD8+ ratio cohort, the low CD4+/CD8+ ratio cohort and the healthy control we identified a severe depletion in CD8a expression in the high ratio cohort accompanied by an increase in the CD4 gene expression. While the CD4+ T-cell expansion is in line with our expectations of superantigen – MHC Class II mediated CD4+ T-cell activation, the CD8 depletion does not necessarily fit with the classical MMTV like superantigen activity, there might be a form of viral superantigen that interacts with the Vβ subset of CD8+ T-cell cells.

To further investigate the relationship between CD4+/CD8+ ratio and PBC progression, we plotted the CD4+/CD8+ ratio against the Neutrophil to Lymphocyte ratio (NLR) as higher values for NLR are correlated with increased short-term mortality in PBC [83]. Linear regression showed a positive correlation between the CD4+/CD8+ ratio and NLR which suggests a link between lymphopenia (driven by a reduction in CD8+ T-cells despite an expansion of CD4+ T-cells) and increased NLR. Additionally, we plotted CD4+/CD8+ ratio of PBC patients against total platelet counts which showed a negative correlation suggesting a relationship between increased CD4+/CD8+ ratio and thrombocytopenia which is a consequence of cirrhosis in liver disease [84]. Cirrhosis develops in late stage PBC patients and is even utilized as a prognostic indicator for UDCA treatment failure in PBC [85].

Taken together, these findings suggest that higher CD4+/CD8+ ratios are characterized by slightly increased CD4+ T-cell expression, markedly decreased CD8+ T-cell expression, overall lymphopenia and increased NLR all contributing to a worse prognosis for PBC patients.

B) TCR-Vβ repertoire distribution in Primary Biliary Cholangitis

Studies attempting to determine expression of the MMTV endogenous retrovirus in mice noted the rarity of the SAg encoding viral mRNA. Detectable levels of sag mRNa were initially only reported confidently in lactating mammary glands [86]. Later, optimized Northern analysis improved this detection range to some other organs such as salivary glands and lymphoid tissue and even detected viral sag mRNA in some B cells. However the detected level of sag mRNA was 500 fold lower in these samples when compared to mammary tissue and the viral mRNA level was below the limit of detection for other hematopoietic cell types [87]. Other studies attempting to generate anti-SAg monoclonal antibodies have also been less than successful in detecting SAg proteins in resting lymphoid cells or hematopoietic cell types other than activated B-cells [88, 89]. With TCR-Vβ skewing identified as a hallmark of viral SAg activity [90], and the availability of TCR-Vβ quantification through monoclonal Vβ antibodies or RNA-seq gene expression technology, we examine the TCR-Vβ repertoire for indications of superantigen activity. In the face of assay limitations, low levels of virus and poor antibody response, we examine the TCR-Vβ repertoire in PBC patients to identify mass expansion of certain Vβ subsets as indication of superantigen activity. Identification of targeted Vβ will help in the identification of conserved sequences of HBRV viral SAg.

The population from which we derived RNA-seq data was a subset of patients recruited for the phase 3 trial of The PBC OCA International Study of Efficacy (POISE) [55]. Due to the specific study recruitment criteria, we expected the genetic expression of our samples to be more homogenous compared to a completely randomized sample of PBC patients. Nevertheless, calculating the CD4+/CD8+ ratio allowed us to visualize a range of immune states despite the standardized sampling. Taking the variation in gene expression within our sample population in

tandem with previous studies which have described a high variation in TCR-Vβ skewing from patient to patient [32], we expected to find variable TCR-Vβ skewing in individual PBC patients. However, based on our hypothesis of superantigen mediated clonal expansion of specific Vβ subsets, we also expected to highlight patterns of Vβ skewing which would indicate the effect of multiple HBRV SAg with variable binding efficiencies.

Our initial attempts at mapping the TCR-Vβ repertoire in 5 PBC patients against 5 healthy controls identified an evidently homogenous Vβ gene expression across all healthy controls whereas PBC samples showed clear indications of expanded Vβ subsets. Investigating the TCR-Vβ repertoire further, we compared the overall expression of individual Vβ subsets between PBC patients and healthy control using two different approaches.

In the first approach, we standardized the TCR-Vβ of individual patients by dividing each Vβ subset using the respective patients CD3E gene expression thereby normalizing Vβ expression with overall lymphocyte expression. Calculating the mean, standard deviation, and total number of samples for PBC samples and healthy control we conducted a Mann Whitney u-test with fewer assumptions and the Holm-Sidak correction. This resulted in a stringent analysis considering variations in lymphocyte expression across patients as well as accounting for multiple comparisons across the tested Vβ subsets. The results indicate a depletion of the Vβ 28 and Vβ 5-5 subsets in the entire PBC population. In the second approach we simply used the DESeq2 R package [58] to compare all PBC samples against healthy control to identify TCR-Vβ 28 and Vβ 6-4 depletion in the PBC patient population on average. These findings might suggest that late-stage liver disease patients with a greater need for transplant might experience depletions of specific Vβ more often than those in earlier stages of the disease.

Although the evaluation of TCR-Vβ expression comparing the entire PBC population against control allowed us to identify skewed Vβ subsets that are possible common targets for varied SAg strains, we needed to adopt an alternate methodology to thoroughly analyze the extent of Vβ subset bias in PBC. Using the CD3E normalized TPM counts of individual Vβ subsets, we established a range of healthy mean + 3*Stdev.S to healthy mean – 2*Stdev.S to define a broad set of values as normal Vβ expression. This range identified most Vβ subsets in all healthy population as normally expressed (except for a Vβ 5-6 outlier in one control) and even allowed the identification of one particular healthy control with abnormal Vβ distribution which was later identified as a genetic relation to a PBC patient and removed from healthy control.

Rather than simply comparing mean Vβ expression between PBC patients and healthy individuals, recording the frequency of expanded and depleted TCR-Vβ subsets for each individual patients allowed us better visualize the extent of Vβ skewing in PBC. The 80% of PBC patients that were identified to possess a positive or negative bias in Vβ expression shared overlap in skewing for some TCR-Vβ subsets. These T-cell subsets showing indications of superantigen mediated skewing might not only highlight potential targets for therapy development but will also prove useful in identifying TCR-Vβ specificities for particular superantigen sequences.

Charting out frequency overlaps, we conducted a hierarchical clustering to group commonly skewed Vβ subsets into distinct groups. Using dendrograms and frequency heatmaps, we identified 13 likely groups (8 positively skewed groups and 5 negatively skewed groups) of TCR-Vβ out of a total of 40! (40 factorial) possible groups. Preferential clustering of these frequently skewed TCR-Vβ subsets allows us to propose preferential Vβ binding specificities for

HBRV SAg families. These groupings suggest superantigen activity of multiple HBRV SAg with varied binding avidities.

Overall, the identification of abnormal mean expression of specific Vβ subsets in the PBC population compared to healthy control paired with the identification of significant expansions and depletion of Vβ subsets in individual PBC patients and clustered skewing of Vβ subsets provides evidence for superantigen activity in PBC patients.

C) Differential gene expression of PBC patient subsets dichotomized using the CD4+/CD8+ and NLR score Immune state markers.

To elucidate the biological pathways involved in disease progression we evaluated the gene expression of PBC patients at different stages of the disease development as indicated by the CD4+/CD8+ and Neutrophil to Lymphocyte ratio immunological markers.

As higher CD4+/CD8+ ratios were found to be indicative of worse disease, we conducted a differential gene expression analysis between the top 10% PBC samples with the highest ratios against the bottom 10% of PBC samples with the lowest ratios. The high CD4+/CD8+ cohort had anti-viral response upregulation including the OAS family genes, EIF2AK2 and TRIM family genes which all conventionally display anti-viral activity [62, 64, 65]. However, the activation of this anti-viral response alongside interferon and interleukin activation pathways might suggest an ongoing immune challenge and viral activity. We also observed an upregulation of genes such as ZBTB7B which is involved in lineage specification of T-cells [71] and HCLS1 which is involved in clonal expansion and deletion signaling [72]. Activation of anti-viral activity, interferon activation, lineage commitment, and clonal expansion and/or depletion of might suggest active HBRV infection, increased HBRV SAg activity and worse

prognosis. In contrast, findings of NK cell activation coupled with activation of pathways involved in Alpha-Beta T-cell activation and T-cell differentiation in thymus might suggest an improved immune state and better prognosis.

Similarly, as increased NLR scores were found to correlate with higher CD4+/CD8+ scores and are associated with overall worse outcomes for PBC patients [83], we observed the differential gene expression between the top 10% highest NLR samples against the bottom 10% lowest NLR samples. We noticed an activation of anti-retroviral activity in the high NLR group coupled with type 1 interferon response, TLR activation and response to cytokine stimulus. Similar to the High CD4+/CD8+ ratio differential gene expression, this may suggest active HBRV infection and HBRV SAg activity. While the low NLR group displayed an upregulation of innate immunity, myeloid cell activation and activation of genes involved in negative regulation of viral transcription such as TFAP4, TARBP2 and TRIM32 [91-93]. Activation of innate immunity, myeloid cell activation and anti-viral activity without interferon activation or upregulation in cytokine production pathways might suggest a resolving disease process.

Findings from DGE of high vs low CD4+/CD8+ ratio and high vs low NLR corroborate with each other. They suggest retroviral activity, SAg mediated clonal expansion/depletion and overall worse prognosis in PBC patients with high CD4+/CD8+ ratio and high NLR score. While the innate immune response, myeloid activation, normal T-cell activation, and regulation of viral transcription might suggest a more controlled immune environment.

D) Genetic marker of TCR-Vβ skewing in Primary Biliary Cholangitis

To further understand the pattern of TCR-Vβ skewing in PBC patients and identify potential markers for Vβ skewing, we characterized change in TCR-Vβ across different

timepoints by investigating serial samples of PBC patients. We identified 35 PBC patients for whom we had RNA-seq data at both the baseline timepoint in the study as well as a later timepoint where these patients had received OCA treatment for at least a 12-month period. The baseline timepoints were marked as “Before” treatment samples while the after treatment timepoint were marked as “After” treatment samples. We noticed that sample collection timepoint did not necessarily dictate the same direction for TCR-Vβ skewing for all patients. Of the 35 patients, 20 patients had an expansion in the number of skewed TCR-Vβ subsets from the Before timepoint to the After timepoint while 11 patients had a contraction in the number of skewed TCR-Vβ subsets from the Before to After timepoint. 4 patients displayed no change in skewing across the two timepoints. This variability in TCR-Vβ skewing across PBC patients further suggests the activity of multiple HBRV SAgs with varied binding affinities.

Further investigating TCR-Vβ skewing across timepoints, we looked for genetic markers that might correlate with the direction of TCR-Vβ skewing

As passage of time proved to be an insufficient indicator for the direction of TCR-Vβ skewing in all PBC patients, we instead looked for genetic markers that might correlate with the direction of Vβ skewing. An identification of such a marker could potentially help decipher the mechanism influencing the direction of change in TCR-Vβ skewing across time in PBC patients. We were successful in identifying CD3E-204 transcript expression to be strongly negatively correlated with TCR-Vβ skewing. CD3E-201 and CD3E-204 were idented as the only two functional transcripts of the CD3E lymphocyte marker [76, 78]. As CD3E-204 displayed such a strong relationship with TCR-Vβ skewing, we evaluated overall CD3E expression as well as individual CD3E-201 and CD3E-204 transcript expression to

better understand their relationship with TCR-Vβ skewing. To evaluate the relationship between the CD3E gene and its transcripts with Vβ skewing, we mapped changes in their expression across timepoints as shown in Figure 16. Interestingly, Figure 16B and Figure 16C show that the CD3E-201 and CD3E-204 behave in an opposite manner in regard to TCR-Vβ skewing. CD3E-201 displayed higher expression in PBC patients that had expanded TCR-Vβ skewing at the after timepoint and lower expression in PBC patients that had a contraction of overall TCR-Vβ skewing at the after timepoint. This indicates that CD3E-201 is positively correlated with increased skewing of Vβ subsets in PBC patients. On the other hand, CD3E-204 displayed an extremely noticeable reduction in expression in PBC patients that had expanded TCR-Vβ subsets and seemed to have recovered expression in PBC patients that experienced contraction in TCR-Vβ skewing. This indicates that there may be a mechanistic interaction between TCR-Vβ skewing and CD3E-204 which may prompt CD3E-204 transcript expression shutoff when TCR-Vβ subsets are being expanded by HBRV superantigen activity. Figure 16A shows that overall CD3E gene expression also negatively correlated with TCR-Vβ skewing. A possible explanation for this effect could be the difference in severity of correlation between the two CD3E transcripts. While CD3E-204 displayed a strong negative correlation with TCR-Vβ skewing and a drastic reduction in its expression in response to increased Vβ skewing, CD3E-201 was observed to have a much milder positive correlation. This difference may allow CD3E-204 to drive the correlation of overall CD3E gene expression.

Looking for explanations regarding the significance of this correlation, we searched for mechanistic differences between the CD3E-204 and CD3E-201 transcripts. We were unable to find any studies exploring the difference between the two transcripts or even investigating

the individual CD3E transcripts. Additional inquires need to be made to determine the differences between CD3E-201 and CD3E-204 transcripts to further explore the biological implications of a negative correlation between CD3E-204 expression and TCR-Vβ skewing and how they might be linked mechanistically. Indeed, the correlation between increased Vβ skewing and diminished CD3E-204 expression warrants further investigation as a marker of differential TCR complex signaling via viral SAg activation.

4.4 **Future Directions**

To further model TCR-Vβ skewing in Primary Biliary Cholangitis, we propose to investigate serial samples of diagnosed patients who either responded positively to treatment and were subsequently normalized compared to patients who reached end stage liver disease and required a liver transplant. This will help map changes in Vβ expression across disease progression across longer timepoints. Where possible, supplementing our peripheral blood RNA-seq data with Intrahepatic lymphocyte samples from the same patients will help us characterize differences in TCR-Vβ expression between peripheral blood and liver lymphocytes.

We also suggest separate analysis of CD4+ and CD8+ T-cells to identify potential Vβ skewing influencing CD8+ T-cell depletion in PBC. Recording changes in overall CD4+ T-cell and CD8+ T-cell expression will also help us quantify changes in CD4+/CD8+ ratio across disease progression in patients with known clinical outcomes.

Using the clustered groups of Vβ subsets with evidence of concurrent skewing along with overlaps in amino acid sequences of the CDR2, FR3 and HV4 SAg binding regions, we hope to distinguish conserved sequences in the viral superantigen. Using these conserved sequences, we will attempt to PCR-clone the HBRV SAg sequences from transplant patient's perihepatic lymphnode samples.

4.5 **Conclusion**

Overall, the significant TCR-Vβ bias observed in the mean PBC expression, the heterogeneity of TCR- Vβ skewing in PBC patients, the clustering of skewed Vβ along with increased CD4+/CD8+ ratio correlating with lymphopenia (high NLR) and thrombocytopenia are all consistent with viral Superantigen activity. Furthermore, the upregulation of anti-viral response, increase in interferon and interleukin response, TLR response and leukocyte differentiation in PBC samples with high CD4+/CD8+ ratio and high NLR may indicate increased viral activity, viral proliferation, and progressive disease in patients with high CD4+ T-cells, low CD8+ T-cells and overall lymphopenia. Meanwhile, NK and T-cell activation coupled with upregulation of innate immunity, myeloid activation and regulation of viral transcription may be indicative of a more robust patient immunity contributing to disease resolution. Finally, the correlation between increased Vβ skewing and diminished CD3E-204 expression needs to be further investigated as a potential marker of TCR-Vβ skewing and superantigen activity.

Bibliography

1. Dauphinee, J.A. and J.C. Sinclair, *Primary biliary cirrhosis.* Can Med Assoc J, 1949. **61**(1): p. 1-6.
2. Carey, E.J., A.H. Ali, and K.D. Lindor, *Primary biliary cirrhosis.* The Lancet, 2015. **386**(10003): p. 1565-1575.
3. Tsuneyama, K., et al., *Primary Biliary Cholangitis: Its Pathological Characteristics and Immunopathological Mechanisms.* The Journal of Medical Investigation, 2017. **64**(1.2): p. 7-13.
4. Klatskin, G., *Mitochondrial Antibody in Primary Biliary Cirrhosis and Other Diseases.* Annals of Internal Medicine, 1972. **77**(4): p. 533.
5. Lindgren, S. and S. Eriksson, *IgM in primary biliary cirrhosis. Physicochemical and complement activating properties.* The Journal of Laboratory and Clinical Medicine, 1982. **99**(5): p. 636-645.
6. Lammers, W.J., et al., *Levels of Alkaline Phosphatase and Bilirubin Are Surrogate End Points of Outcomes of Patients With Primary Biliary Cirrhosis: An International Follow-up Study.* Gastroenterology, 2014. **147**(6): p. 1338-1349.e5.
7. Zhang, F.-K., J.-D. Jia, and B.-E. Wang, *Clinical evaluation of serum antimitochondrial antibody-negative primary biliary cirrhosis.* Hepatobiliary & pancreatic diseases international: HBPD INT, 2004. **3**(2): p. 288-291.
8. Joshi, S., et al., *Antimitochondrial antibody profiles: are they valid prognostic indicators in primary biliary cirrhosis?* The American Journal of Gastroenterology, 2002. **97**(4): p. 999-1002.
9. Sutton, I. and J. Neuberger, *Primary biliary cirrhosis: seeking the silent partner of autoimmunity.* Gut, 2002. **50**(6): p. 743-6.
10. Bjorkland, A. and T.H. Totterman, *Is primary biliary cirrhosis an autoimmune disease?* Scand J Gastroenterol Suppl, 1994. **204**: p. 32-9.
11. Terziroli Beretta-Piccoli, B., et al., *Primary biliary cholangitis with normal alkaline phosphatase: A neglected clinical entity challenging current guidelines.* Journal of Autoimmunity, 2021. **116**: p. 102578.
12. Poupon, R., *Primary biliary cirrhosis: A 2010 update.* Journal of Hepatology, 2010. **52**(5): p. 745-758.
13. Lleo, A., et al., *Primary biliary cholangitis: a comprehensive overview.* Hepatology International, 2017. **11**(6): p. 485-499.
14. Myers, R.P., et al., *Epidemiology and natural history of primary biliary cirrhosis in a Canadian health region: A population-based study.* Hepatology, 2009. **50**(6): p. 1884-1892.
15. Yoshida, E.M., et al., *Epidemiology and liver transplantation burden of primary biliary cholangitis: a retrospective cohort study.* CMAJ Open, 2018. **6**(4): p. E664-E670.
16. Sivakumar, T. and K.V. Kowdley, *Anxiety and Depression in Patients with Primary Biliary Cholangitis: Current Insights and Impact on Quality of Life.* Hepatic Medicine: Evidence and Research, 2021. **Volume 13**: p. 83-92.
17. Gulamhusein, A.F. and G.M. Hirschfield, *Primary biliary cholangitis: pathogenesis and therapeutic opportunities.* Nature Reviews Gastroenterology & Hepatology, 2020. **17**(2): p. 93-110.

18. Selmi, C., et al., *Primary biliary cirrhosis in monozygotic and dizygotic twins: Genetics, epigenetics, and environment.* Gastroenterology, 2004. **127**(2): p. 485-492.
19. Jones, D.E.J., et al., *Familial primary biliary cirrhosis reassessed: a geographically-based population study.* Journal of Hepatology, 1999. **30**(3): p. 402-407.
20. Gulamhusein, A., B. Juran, and K. Lazaridis, *Genome-Wide Association Studies in Primary Biliary Cirrhosis.* Seminars in Liver Disease, 2015. **35**(04): p. 392-401.
21. Selmi, C., et al., *Infectious agents and xenobiotics in the etiology of primary biliary cirrhosis.* Dis Markers, 2010. **29**(6): p. 287-99.
22. Amano, K., et al., *Chemical xenobiotics and mitochondrial autoantigens in primary biliary cirrhosis: identification of antibodies against a common environmental, cosmetic, and food additive, 2-octynoic acid.* J Immunol, 2005. **174**(9): p. 5874-83.
23. Sasaki, M., et al., *A possible involvement of endoplasmic reticulum stress in biliary epithelial autophagy and senescence in primary biliary cirrhosis.* J Gastroenterol, 2015. **50**(9): p. 984-95.
24. Grattagliano, I., et al., *Pathogenic role of oxidative and nitrosative stress in primary biliary cirrhosis.* World J Gastroenterol, 2014. **20**(19): p. 5746-59.
25. Mason, A.L. and G. Zhang, *Linking human beta retrovirus infection with primary biliary cirrhosis.* Gastroenterol Clin Biol, 2010. **34**(6-7): p. 359-66.
26. Wang, W., et al., *Frequent proviral integration of the human betaretrovirus in biliary epithelium of patients with autoimmune and idiopathic liver disease.* Aliment Pharmacol Ther, 2015. **41**(4): p. 393-405.
27. van den Oord, J.J., et al., *Immunohistochemical characterization of inflammatory infiltrates in primary biliary cirrhosis.* Liver, 1984. **4**(4): p. 264-74.
28. Krams, S.M., et al., *Analysis of hepatic T lymphocyte and immunoglobulin deposits in patients with primary biliary cirrhosis.* Hepatology, 1990. **12**(2): p. 306-13.
29. Colucci, G., F. Schaffner, and F. Paronetto, *In situ characterization of the cell-surface antigens of the mononuclear cell infiltrate and bile duct epithelium in primary biliary cirrhosis.* Clinical Immunology and Immunopathology, 1986. **41**(1): p. 35-42.
30. Gores, G.J., et al., *Primary biliary cirrhosis: associations with class II major histocompatibility complex antigens.* Hepatology, 1987. **7**(5): p. 889-92.
31. Skapenko, A., et al., *The role of the T cell in autoimmune inflammation.* Arthritis Res Ther, 2005. **7 Suppl 2**: p. S4-14.
32. Mayo, M.J., B. Combes, and R.N. Jenkins, *T-cell receptor Vbeta gene utilization in primary biliary cirrhosis.* Hepatology, 1996. **24**(5): p. 1148-55.
33. Delves Peter J, R.I.M., *Encyclopedia of immunology*. 2nd ed ed. 1998

San Diego: Academic Press.
34. Sadamoto, T., et al., *Expression of pyruvate-dehydrogenase complex PDC-E2 on biliary epithelial cells induced by lymph nodes from primary biliary cirrhosis.* The Lancet, 1998. **352**(9140): p. 1595-1596.
35. Xu, L., et al., *Cloning the human betaretrovirus proviral genome from patients with primary biliary cirrhosis.* Hepatology, 2004. **39**(1): p. 151-6.
36. Zhang, G., et al., *Mouse mammary tumor virus in anti-mitochondrial antibody producing mouse models.* J Hepatol, 2011. **55**(4): p. 876-84.
37. Irie, J., et al., *NOD.c3c4 congenic mice develop autoimmune biliary disease that serologically and pathogenetically models human primary biliary cirrhosis.* J Exp Med, 2006. **203**(5): p. 1209-19.

38. Sharon, D., et al., *Impact of combination antiretroviral therapy in the NOD.c3c4 mouse model of autoimmune biliary disease.* Liver Int, 2015. **35**(4): p. 1442-50.
39. Acha-Orbea, H. and H.R. MacDonald, *Superantigens of mouse mammary tumor virus.* Annu Rev Immunol, 1995. **13**: p. 459-86.
40. Tomonari, K., *Viral superantigens*. 1997, Boca raton: CRC Press. 114, 245.
41. Golovkina, T.V., et al., *The mouse mammary tumor virus envelope gene product is required for superantigen presentation to T cells.* J Exp Med, 1994. **179**(2): p. 439-46.
42. Acha-Orbea, H., *Bacterial and viral superantigens: roles in autoimmunity?* Ann Rheum Dis, 1993. **52 Suppl 1**: p. S6-16.
43. Chowdhary, V.R., et al., *Chronic exposure to staphylococcal superantigen elicits a systemic inflammatory disease mimicking lupus.* J Immunol, 2012. **189**(4): p. 2054-62.
44. Torres, B.A. and H.M. Johnson, *Modulation of disease by superantigens.* Current Opinion in Immunology, 1998. **10**(4): p. 465-470.
45. Schlievert, P.M., *Role of superantigens in human disease.* J Infect Dis, 1993. **167**(5): p. 997-1002.
46. Morar, N., et al., *HIV-associated psoriasis: pathogenesis, clinical features, and management.* The Lancet Infectious Diseases, 2010. **10**(7): p. 470-478.
47. Leung, D.Y., et al., *A potential role for superantigens in the pathogenesis of psoriasis.* J Invest Dermatol, 1993. **100**(3): p. 225-8.
48. Cheng, M.H., et al., *Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation.* Proc Natl Acad Sci U S A, 2020. **117**(41): p. 25254-25262.
49. Chang, S.E., et al., *New-onset IgG autoantibodies in hospitalized patients with COVID-19.* Nat Commun, 2021. **12**(1): p. 5417.
50. McCormack, J.E., J. Kappler, and P. Marrack, *Stimulation with specific antigen can block superantigen-mediated deletion of T cells in vivo.* Proc Natl Acad Sci U S A, 1994. **91**(6): p. 2086-90.
51. Mayo, M.J., et al., *Similar T-cell oligoclonality in antimitochondrial antibody-positive and -negative primary biliary cirrhosis.* Dig Dis Sci, 2001. **46**(2): p. 345-51.
52. Liaskou, E., et al., *High-throughput T-cell receptor sequencing across chronic liver diseases reveals distinct disease-associated repertoires.* Hepatology, 2016. **63**(5): p. 1608-19.
53. Nakagawa, R., et al., *CD4(+) T cells from patients with primary biliary cholangitis show T cell activation and differentially expressed T-cell receptor repertoires.* Hepatol Res, 2019. **49**(6): p. 653-662.
54. Hou, X., et al., *No difference in TCRbeta repertoire of CD4+ naive T cell between patients with primary biliary cholangitis and healthy control subjects.* Mol Immunol, 2019. **116**: p. 167-173.
55. Nevens, F., et al., *A Placebo-Controlled Trial of Obeticholic Acid in Primary Biliary Cholangitis.* N Engl J Med, 2016. **375**(7): p. 631-43.
56. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.
57. Chen, H., et al., *Functional comparison of PBMCs isolated by Cell Preparation Tubes (CPT) vs. Lymphoprep Tubes.* BMC Immunol, 2020. **21**(1): p. 15.
58. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

59. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.
60. Gene Ontology, C., *The Gene Ontology resource: enriching a GOld mine.* Nucleic Acids Res, 2021. **49**(D1): p. D325-D334.
61. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
62. Kristiansen, H., et al., *The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities.* J Interferon Cytokine Res, 2011. **31**(1): p. 41-7.
63. Mathieu, N.A., et al., *HERC5 and the ISGylation Pathway: Critical Modulators of the Antiviral Immune Response.* Viruses, 2021. **13**(6).
64. Ge, L., et al., *EIF2AK2 selectively regulates the gene transcription in immune response and histones associated with systemic lupus erythematosus.* Mol Immunol, 2021. **132**: p. 132-141.
65. Giraldo, M.I., et al., *TRIM Proteins in Host Defense and Viral Pathogenesis.* Curr Clin Microbiol Rep, 2020: p. 1-14.
66. Zhou, S., et al., *MyD88 is critical for the development of innate and adaptive immunity during acute lymphocytic choriomeningitis virus infection.* Eur J Immunol, 2005. **35**(3): p. 822-30.
67. Verhelst, J., et al., *Interferon-inducible protein Mx1 inhibits influenza virus by interfering with functional viral ribonucleoprotein complex assembly.* J Virol, 2012. **86**(24): p. 13445-55.
68. Stepp, W.H., et al., *Sp100 colocalizes with HPV replication foci and restricts the productive stage of the infectious cycle.* PLoS Pathog, 2017. **13**(10): p. e1006660.
69. Bluthner, M., et al., *Identification of major linear epitopes on the sp100 nuclear PBC autoantigen by the gene-fragment phage-display technology.* Autoimmunity, 1999. **29**(1): p. 33-42.
70. Brisse, M. and H. Ly, *Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5.* Front Immunol, 2019. **10**: p. 1586.
71. Hedrick, S.M., *Thymus lineage commitment: a single switch.* Immunity, 2008. **28**(3): p. 297-9.
72. Taniuchi, I., et al., *Antigen-receptor induced clonal expansion and deletion of lymphocytes are impaired in mice lacking HS1 protein, a substrate of the antigen-receptor-coupled tyrosine kinases.* The EMBO Journal, 1995. **14**(15): p. 3664-3678.
73. Ye, B., et al., *Increasing LAG-3 expression suppresses T-cell function in chronic hepatitis B: A balance between immunity strength and liver injury extent.* Medicine (Baltimore), 2017. **96**(1): p. e5275.
74. Traicoff, J.L., S.M. Hewitt, and J.Y. Chung, *DNAJA3 (DnaJ (Hsp40) homolog, subfamily A, member 3).* Atlas of Genetics and Cytogenetics in Oncology and Haematology, 2012(3).
75. Fallah-Arani, F., et al., *Redundant role for Zap70 in B cell development and activation.* Eur J Immunol, 2008. **38**(6): p. 1721-33.
76. Maglott, D., et al., *Entrez Gene: gene-centered information at NCBI.* Nucleic Acids Res, 2007. **35**(Database issue): p. D26-31.
77. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.

78. de la Hera, A., et al., *Structure of the T cell antigen receptor (TCR): two CD3 epsilon subunits in a functional TCR/CD3 complex.* J Exp Med, 1991. **173**(1): p. 7-17.
79. Howe, K.L., et al., *Ensembl 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D884-D891.
80. Shah, R.A. and K.V. Kowdley, *Current and potential treatments for primary biliary cholangitis.* The Lancet Gastroenterology & Hepatology, 2020. **5**(3): p. 306-315.
81. Montano-Loza, A.J. and A.L. Mason, *Recurrence of primary biliary cholangitis after liver transplantation: A Japanese perspective.* Hepatol Commun, 2017. **1**(5): p. 391-393.
82. Moreno-Otero, R., et al., *Reduced numbers of CD8+ T cells and B cell-expression of Leu-8 antigen in peripheral blood of patients with primary biliary cirrhosis.* Hepatogastroenterology, 1994. **41**(3): p. 239-43.
83. Lin, L., et al., *Does neutrophil-to-lymphocyte ratio predict 1-year mortality in patients with primary biliary cholangitis? Results from a retrospective study with validation cohort.* BMJ Open, 2017. **7**(7): p. e015304.
84. Aster, R.H., *Pooling of platelets in the spleen: role in the pathogenesis of "hypersplenic" thrombocytopenia.* J Clin Invest, 1966. **45**(5): p. 645-57.
85. van Hoogstraten, H.J.F., et al., *Prognostic factors and long-term effects of ursodeoxycholic acid on liver biochemical parameters in patients with primary biliary cirrhosis.* Journal of Hepatology, 1999. **31**(2): p. 256-262.
86. Ross, S.R. and D. Solter, *Glucocorticoid regulation of mouse mammary tumor virus sequences in transgenic mice.* Proc Natl Acad Sci U S A, 1985. **82**(17): p. 5880-4.
87. Henrard, D. and S.R. Ross, *Endogenous mouse mammary tumor virus is expressed in several organs in addition to the lactating mammary gland.* J Virol, 1988. **62**(8): p. 3046-9.
88. Mohan, N., et al., *Production and characterization of an Mls-1-specific monoclonal antibody.* J Exp Med, 1993. **177**(2): p. 351-8.
89. Winslow, G.M., et al., *Detection and biochemical characterization of the mouse mammary tumor virus 7 superantigen (Mls-1a).* Cell, 1992. **71**(5): p. 719-730.
90. Acha-Orbea, H. and E. Palmer, *Mls — a retrovirus exploits the immune system.* Immunology Today, 1991. **12**(10): p. 356-361.
91. Imai, K. and T. Okamoto, *Transcriptional repression of human immunodeficiency virus type 1 by AP-4.* J Biol Chem, 2006. **281**(18): p. 12495-505.
92. Ling, T., et al., *TARBP2 negatively regulates IFN-beta production and innate antiviral response by targeting MAVS.* Mol Immunol, 2018. **104**: p. 1-10.
93. Fu, B., et al., *TRIM32 Senses and Restricts Influenza A Virus by Ubiquitination of PB1 Polymerase.* PLoS Pathog, 2015. **11**(6): p. e1004960.

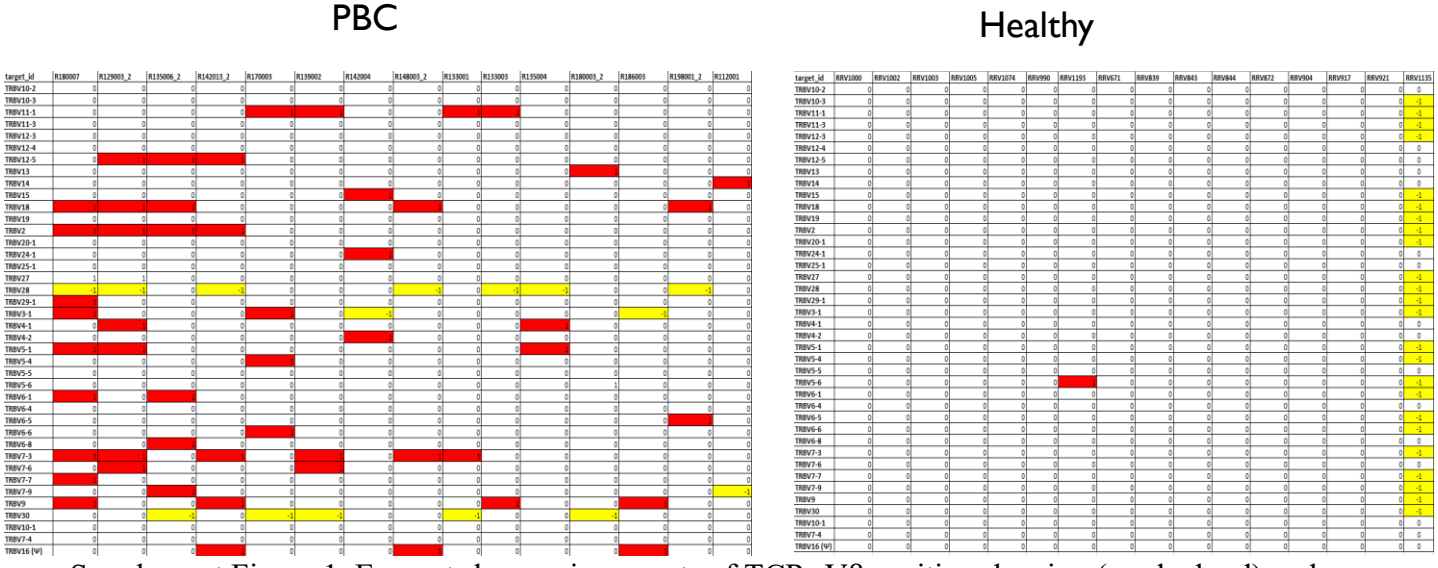
Appendix

Supplement Table 1.

Gene Name	Base Mean	Log2FoldChange	p-adj
TRBV28	1096.0	-0.41	0.007
TRBV6-4	10.3	-0.69	0.040

Supplement Table 1. Differential gene expression from DESeq2 analysis of 173 PBC samples against 15 healthy control showed upregulation in TRBV 23-1 and TRBV 7-1 genes while TRBV28 and TRBV 6-4 were downregulated.

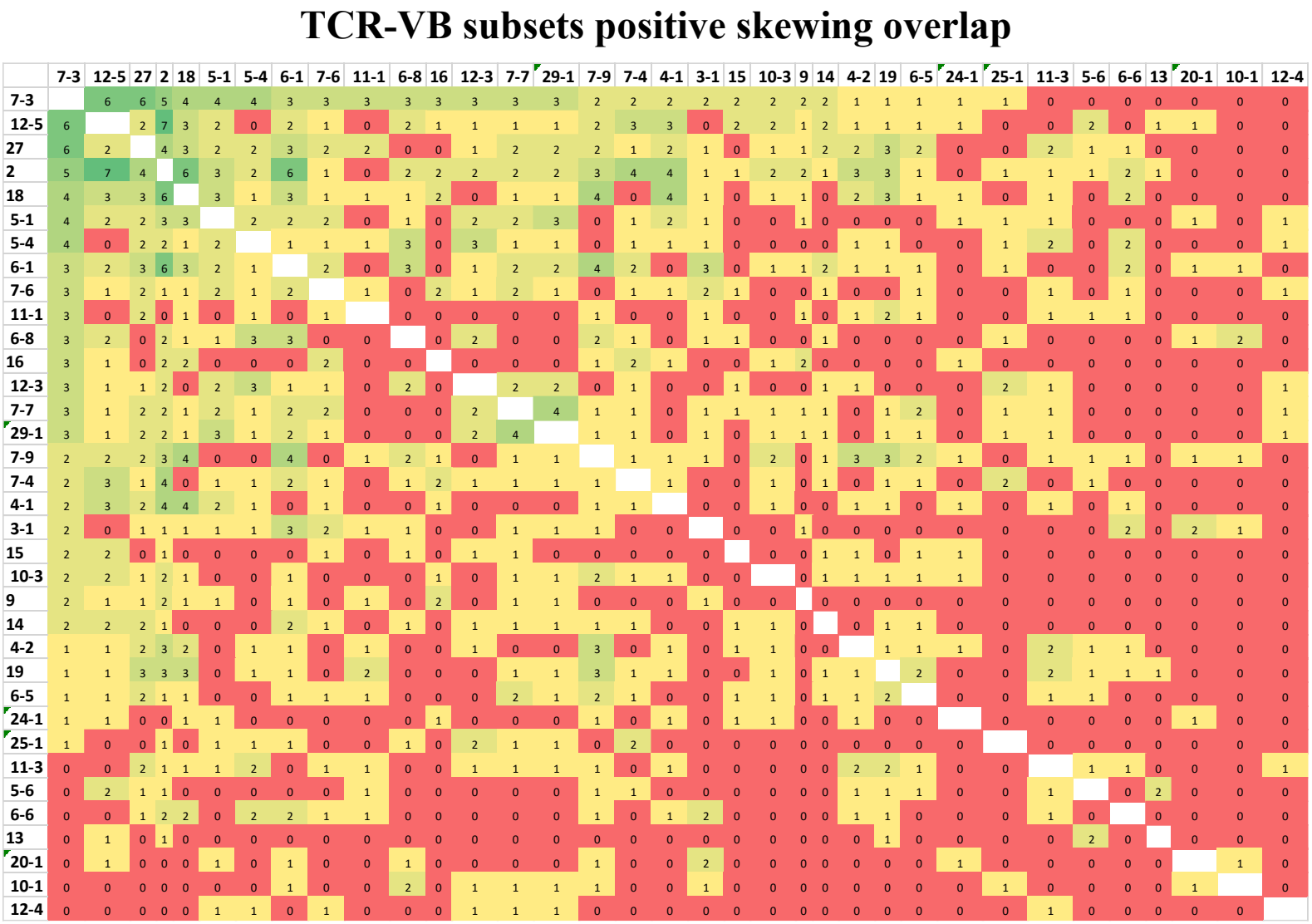
Supplement Figure 1.



Supplement Figure 1. Excerpt showcasing events of TCR- Vβ positive skewing (marked red) and negative skewing (marked yellow) in PBC population and healthy control. Only one event of borderline positive skewing was recorded in healthy control while one healthy individual's

sample was marked as an outlier for displaying extensive negative skewing in multiple TCR- Vβ subsets.

Supplement Figure 2.



Supplement Figure 2. Similarity chart displays the number of PBC patients which have co-occurrence expansion of specific TCR-Vβ subsets. Cells with hues closer to green indicate high level of TCR-Vβ skewing overlap between two TCR-Vβ subsets while hues closer to red indicate less overlap in TCR-Vβ skewing. Numbers listed inside cells record the number of overlaps between two groups. Legends for the rows and columns list the different TCR-Vβ subsets while the values in the chart represent the number of patients which have the respective paired TCR-Vβ subsets skewed together. Counts for the number of patients were obtained by counting each TCR-Vβ positively skewed in each PBC patient sample.

Supplement Figure 3.

	28	6-1	5-1	7-9	10-3	29-1	30	12-3	2	11-1	5-4	18	6-6	20-1	3-1	4-1	7-7
28		6	5	5	4	3	3	3	2	2	2	1	1	1	1	0	0
6-1	6		5	3		5	0	2	3	4	1	2	2	1	1	1	0
5-1	5	5		4	4	3	0	3	3	2	1	1	1	1	1	0	0
7-9	5	3	4		3	2	0	2	2	1	0	1	1	1	1	0	0
10-3	4	4	4	3		2	1	3	2	3	1	2	2	1	1	0	0
29-1	3	5	3	2	2		1	1	2	1	0	2	2	1	1	0	1
30	3	0	0	0	1	1		0	0	0	0	0	0	0	0	0	1
12-3	3	2	3	2	3	1	0		1	1	1	1	1	1	1	0	0
2	2	3	3	2	2	2	0	1		2	0	1	1	1	1	1	0
11-1	2	4	2	1	3	1	0	1	2		1	1	1	0	0	1	0
5-4	2	1	1	0	1	0	0	1	0	1		0	0	0	0	0	0
18	1	2	1	1	2	2	0	1	1	1	0		2	1	1	0	0
6-6	1	2	1	1	2	2	0	1	1	1	0	2		1	1	0	0
20-1	1	1	1	1	1	1	0	1	1	0	0	1	1		1	0	0
3-1	1	1	1	1	1	1	0	1	1	0	0	1	1	1		0	0
4-1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0		0
7-7	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	

Supplement Figure 3. Similarity chart displays the number of PBC patients which have co-occurrence expansion of specific TCR-Vβ subsets. Cells with hues closer to green indicate high level of TCR-Vβ skewing overlap between two TCR-Vβ subsets while hues closer to red indicate less overlap in TCR-Vβ skewing. Numbers listed inside cells record the number of overlaps between two groups. Legends for the rows and columns list the different TCR-VB subsets while the values in the chart represent the number of patients which have the respective paired TCR-Vβ subsets skewed together. Counts for the number of patients were obtained by counting each TCR-Vβ positively skewed in each PBC patient sample.