HEPATOLOGY
Official Journal of the American Association for the Study of Liver Diseases

High-Throughput T-Cell Receptor Sequencing Across Chronic Liver Diseases Reveals Distinct Disease-Associated Repertoires

Evaggelia Liaskou,[1] Eva Kristine Klemsdal Henriksen,[2,3,4] Kristian Holm,[2,3,4] Fatemeh Kaveh,[5] David Hamm,[6] Janine Fear,[1] Marte K. Viken,[7] Johannes Roksund Hov,[2,3,4,11] Espen Melum,[2,3,4] Harlan Robins,[8] Johanna Olweus,[4,9,10*] Tom H. Karlsen,[2,3,4,11*] and Gideon M. Hirschfield[1*]

Hepatic T-cell infiltrates and a strong genetic human leukocyte antigen association represent characteristic features of various immune-mediated liver diseases. Conceptually the presence of disease-associated antigens is predicted to be reflected in T-cell receptor (TCR) repertoires. Here, we aimed to determine if disease-associated TCRs could be identified in the nonviral chronic liver diseases primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), and alcoholic liver disease (ALD). We performed high-throughput sequencing of the TCRβ chain complementarity-determining region 3 of liver-infiltrating T cells from PSC (n = 20), PBC (n = 10), and ALD (n = 10) patients, alongside genomic human leukocyte antigen typing. The frequency of TCRβ nucleotide sequences was significantly higher in PSC samples (2.53 ± 0.80, mean ± standard error of the mean) compared to PBC samples (1.13 ± 0.17, $P < 0.0001$) and ALD samples (0.62 ± 0.10, $P < 0.0001$). An average clonotype overlap of 0.85% was detected among PSC samples, significantly higher compared to the average overlap of 0.77% seen within the PBC ($P = 0.024$) and ALD groups (0.40%, $P < 0.0001$). From eight to 42 clonotypes were uniquely detected in each of the three disease groups (≥30% of the respective patient samples). Multiple, unique sequences using different variable family genes encoded the same amino acid clonotypes, providing additional support for antigen-driven selection. In PSC and PBC, disease-associated clonotypes were detected among patients with human leukocyte antigen susceptibility alleles. *Conclusion*: We demonstrate liver-infiltrating disease–associated clonotypes in all three diseases evaluated, and evidence for antigen-driven clonal expansions. Our findings indicate that differential TCR signatures, as determined by high-throughput sequencing, may represent an imprint of distinctive antigenic repertoires present in the different chronic liver diseases; this thereby opens up the prospect of studying disease-relevant T cells in order to better understand and treat liver disease. (HEPATOLOGY 2015; 00:000-000)

Marked hepatic mononuclear infiltration is a feature of a number of chronic liver diseases. Primary sclerosing cholangitis (PSC) and primary biliary cirrhosis (PBC) are considered prototypical immune-mediated liver diseases, whereas alcoholic liver disease (ALD) is a prototypical environmentally driven liver disease. In PSC a marked $CD4^+$ and $CD8^+$ T-cell infiltration in portal tracts and around bile ducts is

Abbreviations: ALD, alcoholic liver disease; ANOVA, analysis of variance; CDR3, complementarity-determining region 3; D, diversity; HLA, human leukocyte antigen; J, joining; MHC, major histocompatibility complex; PBC, primary biliary cirrhosis; PCR, polymerase chain reaction; PDC-E2, pyruvate dehydrogenase component E2; PSC, primary sclerosing cholangitis; SEM, standard error of the mean; TCR, T-cell receptor; V, variable.

From the [1]Centre for Liver Research, National Institute for Health Research Birmingham Liver Biomedical Research Unit, University of Birmingham, Birmingham, UK; [2]Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery, and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway; [3]Research Institute of Internal Medicine, Division of Cancer Medicine, Surgery, and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway; [4]K.G. Jebsen Inflammation Research Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [5]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; [6]Adaptive Biotechnologies Corp., Seattle, WA; [7]Institute of Immunology, Oslo University Hospital Rikshospitalet, Oslo, Norway; [8]Fred Hutchinson Cancer Research Center, Seattle, WA; [9]Department of Immunology, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo, Norway; [10]K.G. Jebsen Center for Cancer Immunotherapy, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [11]Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery, and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway

evident.[1,2] T-helper type 17 cells are present, and reduced numbers of T regulatory cells have been reported.[3,4] $CD4^+$ and $CD8^+$ T cells are also involved in the pathogenesis of PBC, presumably a result of environmental insults in genetically susceptible individuals.[5,6] In alcoholic cirrhosis, lymphocytes comprise the dominant infiltrating cell type, with $CD4^+$ and $CD8^+$ T cells being detected in expanded portal tracts and in periseptal areas associated with interface hepatitis and progressive fibrosis.[7] While the predominant T-cell infiltrates suggest that T cells are important for the initiation or propagation of disease in PSC, PBC, and ALD, their antigenic specificities have not been broadly determined and therapies targeting the T-cell infiltration are not in clinical use.

In the autoimmune cholestatic liver diseases PSC and PBC, the strongest genetic associations are within the major histocompatibility complex (MHC),[8] with the most important being the classical human leukocyte antigen (HLA) class I and II molecules.[9] Presentation of the triggering antigen(s) by professional antigen-presenting cells through HLA class I and II molecules results in activation and expansion of the T cells bearing the cognate antigen-specific T-cell receptors (TCRs). Therefore, studies of liver-infiltrating T cells classified on the basis of their TCR, rather than by means of cell surface markers or cytokine profiles, has potential to allow study of disease-relevant T cells, the relationship between their antigenic specificities, and the strong disease associations with the HLA locus. Deep characterization of complex disease phenotypes by stratification of patients based on genetic signatures and clinical features is an important approach needed not only to better understand disease but also develop rational effective therapies.[10]

We hypothesized that in different immune-mediated liver diseases a disease-associated antigenic repertoire drives the recruitment and expansion of distinctive T-cell repertoires within the liver. If so, this would be expected to leave an imprint on the corresponding TCR repertoires. We used high-throughput TCRβ sequencing, a previously inaccessible level of TCR repertoire analysis, and a large number of samples to explore if such disease-associated TCR repertoires could be identified in fresh frozen disease-affected liver tissue from patients with two strongly HLA-associated diseases (PSC and PBC), with ALD as a control.

Materials and Methods

Human Tissue. Freshly explanted liver tissue was available through our transplant program at the Queen Elizabeth Hospital in Birmingham, United Kingdom. Fresh frozen explanted liver tissue from a total of 20 patients with PSC, 10 patients with PBC, and 10 patients with ALD were used. All samples were collected after local research ethics committee approval and patient consent.

TCRβ Chain Amplification, High-Throughput Sequencing, and Analysis. Genomic DNA from whole fresh frozen human liver tissue (20-30 mg) was extracted using the DNeasy Mini Kit (Qiagen, Valencia, CA). The TCRβ complementarity-determining region 3 (CDR3), which encompasses the V (variable), D (diversity), and J (joining) recombination junctions and encodes the portion of the TCR that directly contacts the peptide-MHC,[11,12] was amplified from genomic DNA using the ImmunoSEQ assay (Adaptive Biotechnologies, Seattle, WA).[13] In this assay a multiplex polymerase chain reaction (PCR) system was used to amplify the rearranged CDR3β sequences from sample DNA. The 87-bp fragment is sufficient to identify the VDJ region spanning each unique CDR3β. Amplicons were sequenced using the Illumina platform. TCRβ V, D, and J gene definitions were provided by the IMGT database (www.imgt.org). The assay is quantitative, having used a complete synthetic repertoire of TCRs to establish an amplification baseline and adjust the assay

Received March 31, 2015; accepted August 5, 2015.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.28116/suppinfo.

Supported by the Norwegian PSC Research Center, the South-Eastern Norway Regional Health Authority grant no. 2015024, and partially by PSC Partners Seeking a Cure. This paper presents independent research funded by the National Institute for Health Research. The views expressed are those of the authors(s) and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

**These authors contributed equally to this work.*

Address reprint requests to: Dr. Gideon M. Hirschfield, Centre for Liver Research, NIHR Birmingham Liver Biomedical Research Unit, University of Birmingham, Birmingham, UK. E-mail: g.hirschfield@bham.ac.uk. Or: Prof. Johanna Olweus, K.G. Jebsen Center for Cancer Immunotherapy, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. E-mail: johanna.olweus@medisin.uio.no.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.28116

Potential conflict of interest: Dr. Robins owns stock, holds intellectual property rights, and is employed by Adaptive. Dr. Hamm owns stock and is employed by Adaptive.

chemistry to correct for primer bias. In addition, barcoded, spiked-in synthetic templates were used to measure the degree of sequencing coverage and residual PCR bias. This information was used for further PCR bias correction and to estimate the abundance of templates in each sample that could be sequenced. The resulting data are filtered and clustered using both the relative frequency ratio between similar clones and a modified nearest-neighbor algorithm, to merge closely related sequences and remove both PCR and sequencing errors. Data were analyzed using in-house developed bioinformatics scripts, the tcR package of R software (R Foundation for Statistical Computing, Vienna, Austria), and the ImmunoSEQ analyzer tool set (http://www.adaptivebiotech.com/immunoseq/analyzer) (analytical flow chart; see Supporting Fig. S1).

HLA Typing. Genomic DNA was extracted from the same human PSC, PBC, and ALD liver tissue using the DNeasy Mini Kit (Qiagen). Amplification of exons 2 and 3 for *HLA-B* and *HLA-DRB1* was performed, followed by sequencing using BigDye chemistry (Applied Biosystems, Foster City, CA) and the ABI3730 genetic analyzer (Applied Biosystems). Allele assignment was done using the AssignSBT software (Conexio Genomics, Fremantle, Australia). In some PBC and ALD liver samples, *HLA-B* and *HLA-DRB1* were amplified and prepared for sequencing using NGSgo workflow (GenDX, Utrecht, The Netherlands). For these samples, sequencing was performed on a MiSeq (Illumina, San Diego, CA) at the Norwegian Sequencing Centre (Oslo, Norway), and the alleles were assigned using the NGSengine software (GenDX).

Assessment of T-Cell Clonality. T-cell clonality, a term used to quantify the diversity of clones in a sample, was provided by Adaptive Biotechnologies. Clonality is calculated as 1 – (Shannon's entropy)/log2(number of productive unique sequences). By dividing the number of unique sequences being measured that are independent of sampling depth this provides a diversity measurement that is strictly a function of the spectrum of clone frequencies in a sample.[14]

Important Terms Used in the Study. The term *productive sequences* describes TCRβ sequences in which the entire TCR gene is expected to translate to a functional protein. The term *productive unique sequences* describes TCRβ sequences comprised of a unique nucleotide sequences. The term *clonotypes* describes TCRβ sequences comprised of unique amino acid sequences. *Clonality* is a measurement of TCRβ diversity, calculated as 1 – (Shannon's entropy)/log2(number of productive unique sequences). *Entropy* is a measurement of the uncertainty associated with the predictability of the identity of a unique TCR in a sample as a function of the frequency spectrum of all TCRs in a sample.

Table 1. Characteristics of the Study Population

Characteristics	PSC (n = 20)	PBC (n = 10)	ALD (n = 10)
Mean age at transplantation (years)	48 ± 13	47 ± 7	45 ± 6
Male sex (%)	85%	20%	80%
Serum ALT (IU/L)	71 ± 35*	152 ± 147	28 ± 14*
Serum AST (IU/L)	102 ± 62	112 ± 7†	27 ± 6†
Serum BIL (μmol/L)	179 ± 204	194 ± 233	63 ± 34
Serum ALP (IU/L)	776 ± 540	436 ± 333	268 ± 287
Serum ALB (g/L)	32 ± 6	38 ± 7	32 ± 7
Antimitochondrial antibody (%)		90%	

Plus or minus values are means ± standard deviation.
*Mean ALT values from two of 20 PSC and five of 10 ALD patients.
†Mean AST values from three of 10 PBC and five of 10 ALD patients.
Abbreviations: ALB, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BIL, bilirubin.

Statistical Analysis. For normally distributed data, one-way analysis of variance (ANOVA) was performed. Group differences were compared using two-way ANOVA. Pairwise differences in continuous data were compared using the nonparametric Wilcoxon rank-sum test. $P < 0.05$ was considered statistically significant.

Results

A Higher T-Cell Diversity Is Detected in Liver Samples From PSC Relative to PBC and ALD Patients. Our study cohort consisted of 20 patients with PSC, 10 patients with PBC, and 10 patients with ALD (Table 1). The total number of sequencing reads achieved for each disease group was 0.46×10^6 to 1.18×10^6 for PSC, 0.65×10^6 to 3.17×10^6 for PBC, and 0.46×10^6 to 2.94×10^6 for ALD; and the number of resulting productive sequences was 0.37×10^6 to 0.95×10^6 for PSC, 0.52×10^6 to 2.61×10^6 for PBC, and 0.39×10^6 to 2.39×10^6 for ALD. A detailed description of the total number of sequence reads, productive sequence counts, number of productive unique sequences, the mass of reaction template (amount of genomic DNA) used per sample, the number of diploid genomes, the number of gene rearrangements and productive gene rearrangements, the percent of T cells (calculated as productive genomic rearrangements/total input DNA [ng] × [6.5 pg DNA/diploid genome]), and richness (a metric showing the average number of unique TCRs per T cell) of all samples is displayed in Supporting Table S1.

The number of productive unique TCRβ sequences, which is clones differing by their TCRβ nucleotide sequence, including different V, D, and J rearrangements, relative to the number of productive sequences

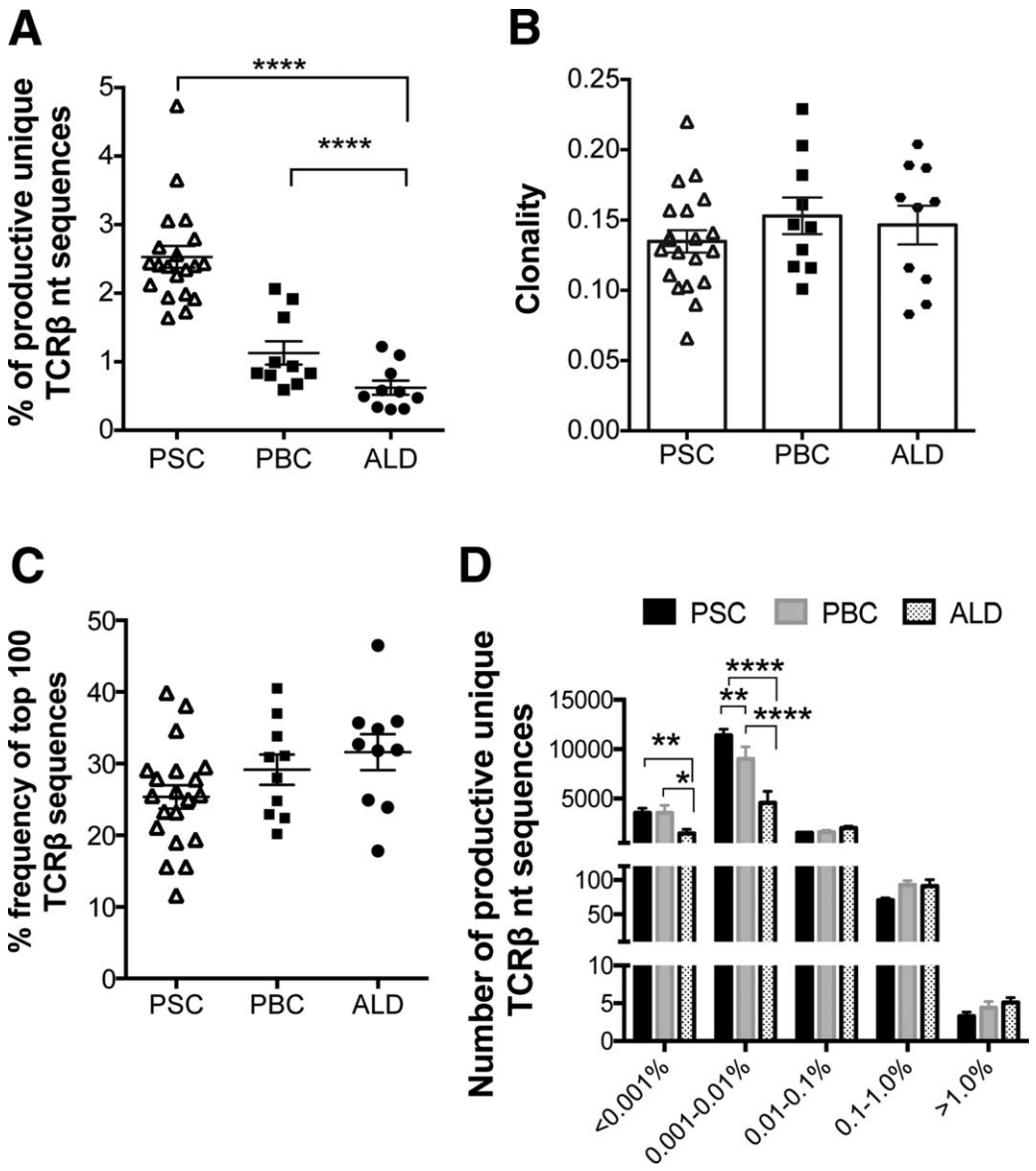

Fig. 1. Clonal distribution of T cells in patients with PSC, PBC, and ALD. (A) Frequency of unique TCRβ sequences identified in the liver of patients with PSC, PBC, and ALD). Data points represent the percentage of unique sequences in the total productive T-cell receptor repertoire of each individual, and bars depict the mean (±SEM) of the groups. Data were normally distributed, and differences between groups were compared using one-way ANOVA. ****$P < 0.0001$. (B) Data show the distribution of T-cell clonal frequencies through the measurement of clonality. A value of 0 indicates a maximally diverse sample, whereas a value of 1 indicates an entirely monoclonal sample. Each dot represents the clonality of each patient's sample, and bars show the mean (±SEM) of the groups. (C) Percentage frequency of top 100 TCRβ nucleotide sequences in PSC, PBC, and ALD. Data points represent the percentage of the top 100 TCRβ sequences in the total repertoire of each individual, and bars depict the mean ± SEM of the percentages of each patient. (D) Frequency distributions of TCRβ nucleotide sequences from PSC (n = 20), PBC (n = 10), and ALD (n = 10) patients. Data are presented as mean ± SEM of distributions of each patient. Data were analyzed using two-way ANOVA with Tukey's post hoc test. Abbreviation: nt, nucleotide.

provides a general assessment of diversity within a sample. In PSC patients this percentage was 2.53 ± 0.80% (mean ± standard error of the mean [SEM]), which was significantly higher compared to PBC, 1.13 ± 0.17% ($P < 0.0001$), and ALD, 0.62 ± 0.10% ($P < 0.0001$) (Fig. 1A). This means there is a higher diversity of T cells in PSC, with T cells in PBC and ALD showing a less polyclonal T-cell repertoire. In line with this, clonality values between the three were highest in PBC (0.153 ± 0.013, mean ± SEM) and ALD (0.146 ± 0.013) and lowest in PSC (0.134 ± 0.007), but these differences were not statistically significant (Fig. 1B). Clonal expansion was further assessed by calculating the cumulative percentage of the repertoire that was constituted by the top 100 TCRβ nucleotide sequences. This showed that in all three disease groups the top 100 TCRβ nucleotide sequences consisted of one-quarter or more of the total T-cell receptor

repertoire. The average fraction of the top 100 TCRβ sequences was 25.3% in PSC, 29.1% in PBC, and 31.6% in ALD, suggesting clonally expanded TCRβ nucleotide sequences (Fig. 1C). The TCRβ nucleotide sequences were similarly diverse in all three disease groups, with the majority of sequences being present in copy numbers within 0.001%-0.01%. In PSC and PBC cases a significantly higher number of sequences (3548 and 3550, respectively) were present at <0.001% compared to ALD (1458). In PSC, the highest number of TCRβ nucleotide sequences were within a range of 0.001%-0.01% (11,395), which was significantly higher compared to the number of sequences within that frequency range in PBC (9008) and ALD (4560) (Fig. 1D). Given that it is the amino acid sequence that defines the structure binding both the epitope and the MHC, independent of the nucleotide sequence, subsequent analysis was done using the amino acid sequence of the TCR, with the unique TCRβ amino acid sequences being called clonotypes.

Larger Overlap in TCRß Repertoires Within Than Between Disease Groups. We next investigated the degree of overlap in TCRβ repertoires, both between samples within the same disease and between the three disease groups. In PSC, an average overlap of 0.85% was detected between any two PSC patient samples (range 0.47%-1.41%). This was significantly higher than the average overlap of 0.77% (range 0.39%-1.73%) seen in PBC ($P = 0.024$) and 0.40% (range 0.14%-0.85%) seen in ALD ($P < 0.0001$). The difference in average overlap between the PBC and ALD groups was also statistically significant ($P < 0.0001$) (Fig. 2A). Between any two PSC samples a significantly higher number of clonotypes was observed (8506) to be common compared to PBC (2501, $P < 0.01$) and ALD (964, $P < 0.0001$) (Supporting Fig. S2). On average 0.72% (range 0.37%-1.17%) of clonotypes were common between PSC and PBC liver samples, an overlap significantly higher than that observed between PSC and ALD (0.50% [range 0.1%-1.13%], $P < 0.0001$) and between PBC and ALD samples (0.48% [range 0.08-0.91%], $P < 0.0001$) (Fig. 2B). Only a small proportion of shared clonotypes were among the most frequent ones (>0.01%) within each disease group (Fig. 2C) as well as between disease groups (Fig. 2D). Collectively, these data suggest that there is greater sharing of TCRβ repertoires within the PSC and PBC groups, respectively, compared to the ALD group; that there is a higher degree of sharing TCRβ repertoires between PSC and PBC than between either of these and ALD; and that the majority of shared clonotypes are present at low frequencies.

Presence of Disease-Associated Clonotypes in PSC, PBC, and ALD Patients. To investigate if immune response characteristics specific to each of the three diseases could be detected, we searched for clonotypes present in patients within one disease group and absent from the data obtained from the other two. We identified eight clonotypes that were common between several PSC samples but not found in PBC and ALD liver samples. The clonotype CASSPGQGEGYEQYF and the clonotypes CASSEYSNQPQHF and CASSLGSGANV LTF were found in eight of 20 (40%) and in seven of 20 (35%) PSC patients, respectively, while five clonotypes were detected in 30% of PSC cases (6/20) (Supporting Table S2). The frequencies of these PSC-associated clonotypes were low; in only a few patient samples they were present at 0.01%-0.04% (Supporting Table S2). In PBC patients, 39 clonotypes were seen among 30% of patients and three were found in 40% of the cases (CASSSTGRYGYTF, CASSLSPGLASYEQYF, CATDP PPPNYGYTF) but not in any PSC or ALD samples (Supporting Table S3). In PBC, only a few of these disease-associated clonotypes were frequent (>0.1%), whereas several were found at frequencies of 0.01%-0.05%. Among 30% of ALD samples, eight clonotypes were disease-associated that were completely absent from PSC and PBC samples (Supporting Table S4). Three ALD associated clonotypes were among the most frequent clonotypes in three patients (0.1%, 2.2%, 11.9%). Different V family genes were used to create the same amino acid clonotype (Supporting Tables S2-S4) in all three diseases, supporting antigen-driven selection as the basis of these clonotypes.

Notably, five out of the eight PSC-associated clonotypes were present in a patient (PSC 10) (Supporting Table S5) carrying the PSC-associated HLA-DRB1*03:01-B*08:01 and DRB1*15:01-B*40:01 haplotypes (Supporting Table S2). In 10 more PSC patients 37.5%-50% of the disease-associated clonotypes were detected; all of these patients had at least one *HLA-DRB1* susceptibility allele, DRB1*03:01 or DRB1*13:01 or DRB1*15:01 coupled with or without the HLA-B*08 susceptibility allele (Supporting Table S2). Of note, PSC patients who did not present any of the disease-associated clonotypes had also at least one HLA susceptibility allele (Supporting Table S6). In PBC, the disease-associated clonotypes were detected in patients with the DRB1*08:01 or DRB1*04:04 susceptibility alleles, but disease-associated clonotypes (33.3%-40.5%) were also detected in PBC patients with no *HLA-DRB1* susceptibility alleles (Supporting Table S3). Five ALD patients also presented 37.5%-50% of the disease-associated clonotypes; these patients had a

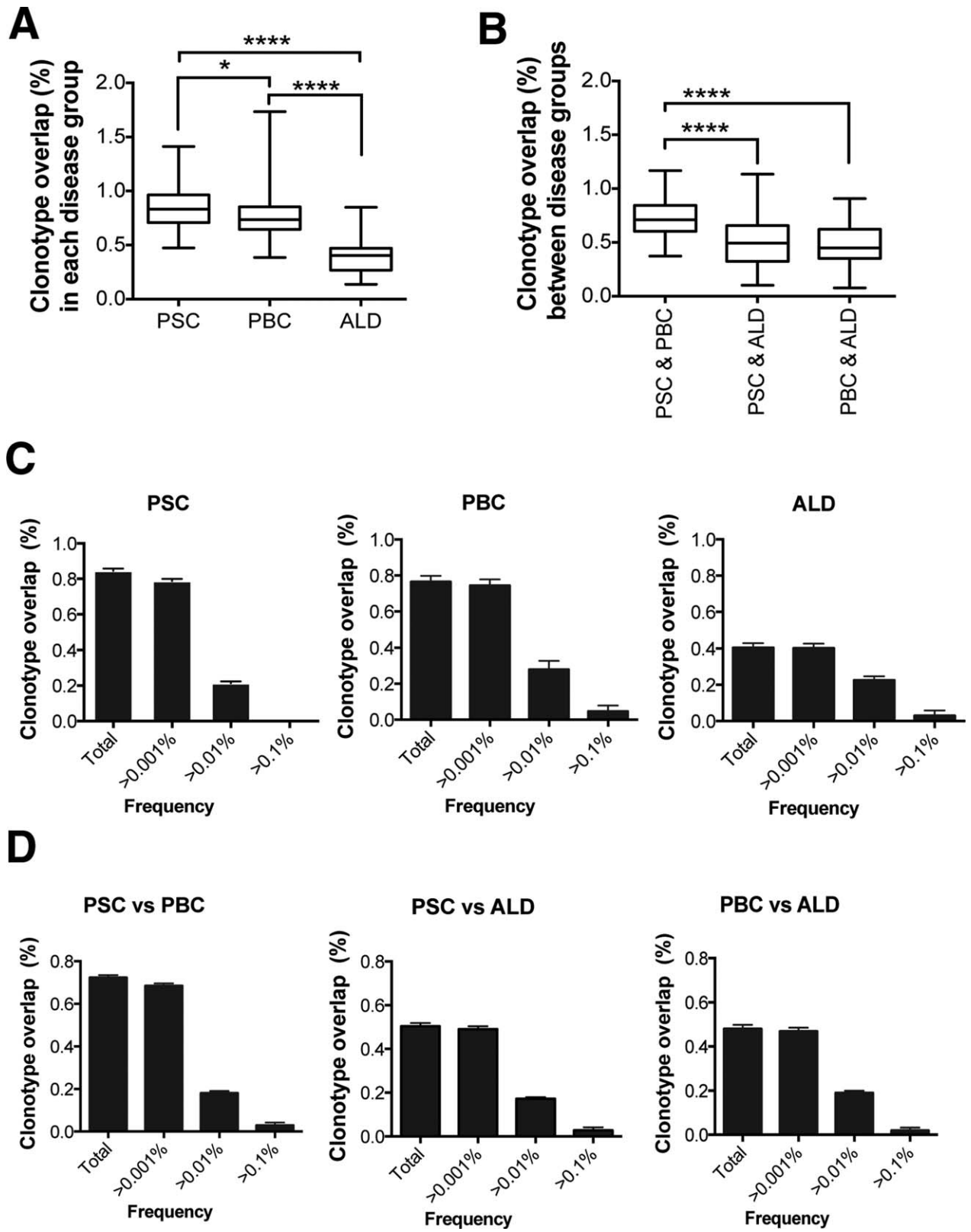

Fig. 2. Overlap in T-cell specificity in each group and between disease groups. A TCRβ overlap was calculated based on the number of common amino acid clonotypes in two samples– (number of common amino acid clonotypes in the two samples $\times$ 2)/(total number of amino acid clonotypes in sample 1 + total number of amino acid clonotypes in sample 2) $\times$ 100 –and the average of all samples in each group was reported. (A) Data show the overlap of clonotypes in each disease group, PSC, PBC, and ALD. (B) Data show the overlap of clonotypes between disease groups, PSC and PBC, PSC and ALD, and PBC and ALD. Data were normally distributed, and differences between groups were compared using one-way ANOVA. *$P = 0.024$, ****$P < 0.0001$. (C) Degree of overlap between clonotypes in PSC, PBC, and ALD patients according to their frequency. (D) Degree of overlap between clonotypes shared between PSC and PBC, PSC and ALD, and PBC and ALD patients. Data are presented as mean $\pm$ SEM of each patient combination.

mixture of PSC susceptibility alleles (DRB1*15:01, DRB1*03:01) and PSC and PBC protective alleles (DRB1*07:01, DRB1*11:01) (Supporting Table S4). Collectively, between eight and 42 TCRs were associated with each of the three disease groups and were common between a large fraction of patients within each group.

Disease-Associated Clonotypes Share Common Amino Acid Characteristics. The presence of one hydrophobic residue glycine (G) among two polar residues threonine (T) and/or serine (S) was revealed in the CDR3 region of 62.5% of the PSC-associated clonotypes (T-G-T, TS-GG-T, GG-T, TS-GG-T, S-G-T) (Fig. 3A). An aspartic acid (D) or glutamic acid (E) was present in 87.5% of PSC disease–associated clonotypes. In PBC, serine (S), glycine (G), and tyrosine (Y) were common in two out of the three disease-associated clonotypes that were found in 40% of cases (Fig. 3B). In ALD, the presence of charged amino acid residues in the CDR3 region of the disease-associated clonotypes was more evident, with five of eight clonotypes sharing a glutamic acid (E) and three of eight clonotypes sharing an arginine (R). The presence of G-N (glycine-asparagine) was evident in 62.5% of the ALD-associated clonotypes (Fig. 3C). Taken together these data suggest that the disease-associated clonotypes share similar TCRβ protein characteristics.

T Cells From PSC, PBC, and ALD Patients Show a Differential Vβ and Jβ Gene Usage. The TCR repertoire is shaped by the genetic background of the individual and the response to (self)antigens; therefore, antigen-driven stimulation in the context of similar HLA molecules will lead to oligoclonal expansion of T cells using common V-D-J segments.[15] Thus, it is important to delineate the Vβ and Jβ gene distribution pattern among T cells in the different liver diseases. In PSC a significantly higher percentage of clonotypes used V01-01 (0.07%), V05-07 (0.02%), V06-04 (1.14%), V06-09 (0.03%), V07-01 (0.007%), V11-02 (2.21%), and V22-01 (0.0035%) compared to PBC: V01-01 (0.052%, $P = 0.04$), V05-07 (0.009%, $P = 0.008$), V06-04 (0.87%, $P = 0.03$), V06-09 (0.02%, $P = 0.02$), V07-01 (0.001%, $P = 0.003$), V11-02 (1.98%, $P = 0.04$), V22-01 (0%, $P = 0.04$). In PBC, significantly higher percentages of clonotypes used V06-05 (4.3%), V07-08 (1.73%), and V07-09 (4.03%) compared to PSC: V06-05 (3.36%, $P = 0.001$), V07-08 (1.53%, $P = 0.02$), and V07-09 (3.40%, $P = 0.001$). A significantly higher percentage of clonotypes in PSC used V06-08 (0.03%) and V29-01 (2.71%) compared to ALD (V06-08 [0.01%, $P = 0.005$] and V29-01 [2.42%, $P = 0.04$]), whereas in ALD a significantly higher percentage of clonotypes used V06-05 (4.23%) compared to PSC (3.36%, $P = 0.0002$). No significant differences were found in the usage of V genes between PBC and ALD liver T-cell clones, except the V07-09 gene, which was significantly more frequent in PBC (4.04%) compared to ALD (3.60%, $P = 0.02$) (Fig. 4A). No significant differences were detected for J gene usage between PSC and PBC T-cell clones, but in PSC and PBC a significantly lower percentage of clonotypes used J02-06 clonotypes (1.84% and 1.78%, respectively) compared to ALD (2.05%, $P = 0.03$) (Fig. 4B). The percentage frequency of all V and J genes in all three diseases are shown in Supporting Figs. S3-S6. We further performed principal component analysis of the V/J gene usage of all clonotypes, the top 1000, and the

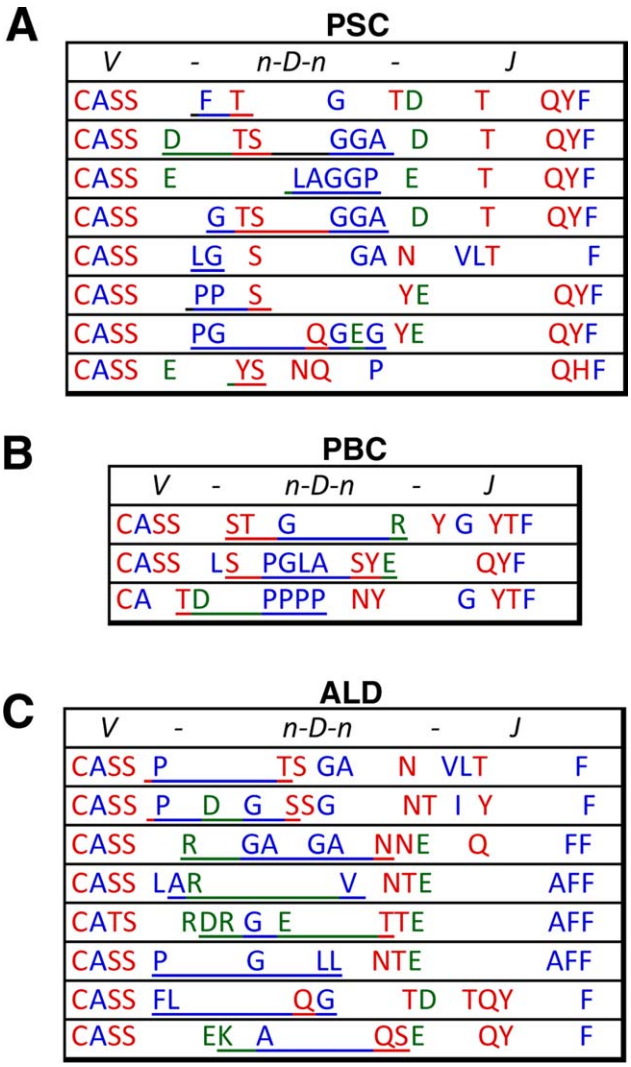

Fig. 3. Amino acid characteristics of disease-associated clonotypes. Data show the CDR3 of (A) PSC, (B) PBC, and (C) ALD disease-associated clonotypes. The CDR3 region consisted of the V, D, and J segments with nontemplated nucleotides inserted or deleted in the junctional sites. Polar amino acids are shown in red, hydrophobic amino acids are shown in blue, and charged amino acids are shown in green. The n-D-n region of each clonotype is underlined. Abbreviation: n, nucleotide.

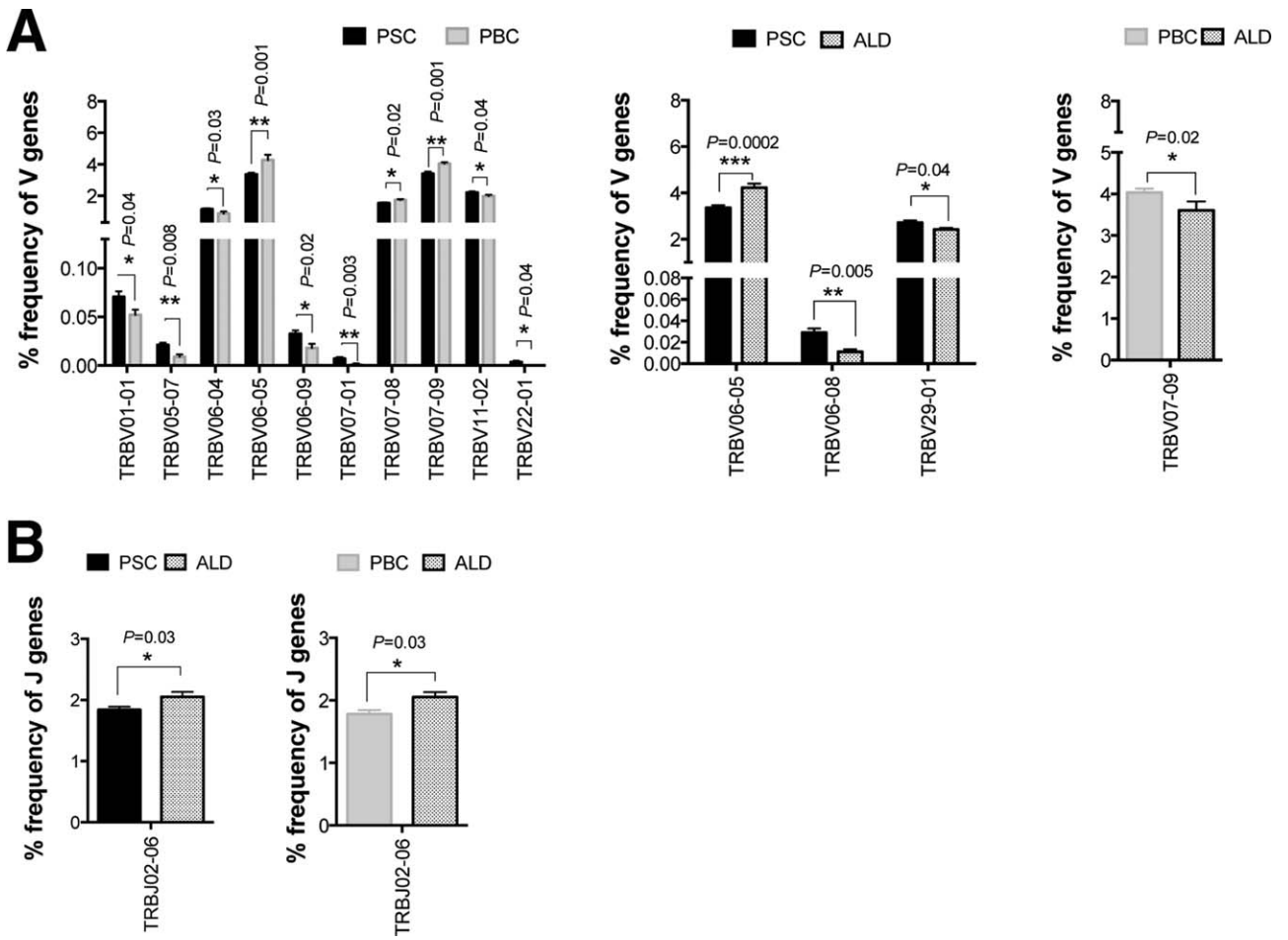

Fig. 4. Vβ gene and Jβ gene usage of clonotypes in the liver of PSC, PBC, and ALD patients. Data show the percentage frequency of (A) V genes and (B) J genes used by clonotypes in PSC, PBC, and ALD. Data show mean ± SEM frequency of each patient. Data were compared using Wilcoxon rank-sum test. $*P < 0.05$, $**P < 0.01$, $***P < 0.0001$.

top 500 most frequent clonotypes in order to evaluate their diversity in the three disease groups. When looking at all clonotypes a clustering of ALD clonotypes along PC1 is evident, but no clear differences appear between PSC and PBC clonotypes (Supporting Fig. S7A). When selecting the top 1000 and top 500 clonotypes, based on their frequency, a clustering of PSC clonotypes along PC2 (Supporting Fig. S7B,C) and a clustering of PBC clonotypes along PC1 (Supporting Fig. S7C) appear. The present data thus demonstrate a preferential usage of Vβ and Jβ genes from T-cell clones infiltrating the different liver diseases.

Discussion

Using high-throughput sequencing technology on a large collection of fresh frozen human diseased liver tissue samples, we report for the first time a comprehensive characterization of the TCRβ repertoire of hepatic infiltrating T cells in three chronic nonviral liver diseases: PSC, PBC, and ALD. Our findings suggest that (1) among the three liver diseases studied PSC shows the highest T-cell diversity, (2) PBC and ALD T cells are less polyclonal and diverse, and (3) disease-associated clonotypes were demonstrated in all three liver disorders with signs of antigen-driven clonal selection based on differences in the number of unique nucleotide sequences coding for a given amino acid sequence but also signs of antigen-driven clonal expansion as seen in the high fraction of the total T-cell receptor repertoire comprised of the 100 most frequent nucleotide sequences. A key feature of our study is that we characterize the receptor signatures of infiltrating T cells from diseased liver tissue without any *in vitro* manipulation such as cell culture that may bias the results.

In contrast to our study of total T cells, the presence of clonally expanded T cells in the liver of PBC patients has been reported.[16] MHC class II–restricted autoreactive $CD4^+$ T cells specific for the E2 component of pyruvate dehydrogenase (PDC-E2) have been identified

in peripheral blood and liver of PBC patients[17] as well as MHC class I–restricted $CD8^+$ T cells reacting to PDC-E2.[18,19] Notably, the average frequency of PDC-$E2_{159\text{-}167}$-specific $CD8^+$ T cells in expanded cultures of liver-infiltrating lymphocytes from end-stage PBC patients has been calculated as $4.14 \pm 0.95 \times 10^{-4}$, but undetectable levels of PDC-$E2_{159\text{-}167}$-specific $CD8^+$ T cells have been reported in uncultured liver-infiltrating lymphocyte samples,[19] which partly agrees with our findings and the low frequencies of the disease-associated clonotypes detected in our PBC and PSC end-stage patients. In addition, we report herein 42 disease-associated clonotypes found in PBC patients but not in PSC and ALD liver patient samples. Notably, several of these PBC clonotypes were detected in patients with HLA-susceptible alleles. Further characterization of the reactivity of these clonotypes will reveal their relationship to the PDC-E2 antigen, as well as potentially novel antigenic drivers in PBC.

The presence of oligoclonal T cells in PSC livers has also been previously suggested.[20,21] Importantly, these prior studies were limited by technological constraints because they were based on a limited number of antibodies against Vβ segments used in immunohistochemistry. Our data revealed the presence of high T-cell diversity and polyclonality but in addition eight PSC-associated clonotypes. The use of tissue from patients with end-stage disease means that procedure (i.e., endoscopic retrograde cholangiography)–related colonization and secondary insults due to cirrhosis and cholangitis may be contributing to the background polyclonality of the T-cell repertoire in PSC. However, in several patients with clonotypes that appeared associated with PSC, different V family genes were used to create the same amino acid clone, further supporting an antigen-driven selection and a potentially primary role in PSC pathogenesis for the disease-associated clonotypes. Of note, most of the shared clonotypes were found in PSC patients carrying at least one susceptible HLA allele, further supporting the hypothesis of a strong HLA association resulting in a restricted antigenic repertoire.

Surprisingly, our data demonstrate for the first time that ALD is not as non-immune-mediated as sometimes considered. Our findings revealed a less polyclonal nature of ALD liver-infiltrating T cells, with signs of clonal expansions and the least T-cell diversity. Clinical observations of ALD, such as elevation in serum immunoglobulin levels, antibodies to unique hepatic proteins, alterations in the $CD8^+/CD4^+$ T-cell subsets, and changes in the normal cytokine balance do suggest an altered immune regulation with an increased activity toward normal self-proteins.[22] Studies report that the oxidation of ethanol results in the formation of a variety of reactive molecules capable of binding covalently to proteins in order to form adducts. Such adduct formation could impair protein function but could also generate neoantigens capable of eliciting a potentially injurious immune response. In the livers of ALD patients adducts have been identified,[23] and it is believed that such adducts could be presented by antigen-presenting cells to $CD4^+$ T cells, thus inducing clonal T-cell expansions. The less polyclonal nature of liver-infiltrating T cells observed in our settings may indeed be reflective of such neo-antigen-specific T-cell responses and opens the possibility for specific assessments of the antigenic reactivity of detected T-cell clones.

Studies in celiac disease, in which gluten-reactive $CD4^+$ T cells have been identified, have shown that DQ (A1*05:01, B1*02:01)–restricted gluten-specific T cells carried diverse TCRs.[24] This opens the possibility that the different disease-associated TCRs detected in our liver disease samples may indeed be representative of the same antigenic trigger(s). Of note, three of the PSC and four of the PBC disease-associated clonotypes were reported by others as common public clonotypes.[25] Again, paralleling observations with those of celiac disease is useful because the presence of gliadin-specific T cells has been also reported in healthy individuals.[26] Our data revealed a greater TCR overlap among PSC and PBC patients but also between PSC and PBC patients. To what extent this is related to the fact that bile duct injury is common to both conditions, compared with ALD, can only be speculated, i.e., whether this means that there are shared antigenic drivers representative of the local immune environment in PSC and PBC or the results merely reflect the higher number of clonotypes detected in these conditions.

There are several technological and methodological constraints that make the task of TCR characterization challenging. The high-throughput sequencing approach is fairly new, meaning analysis tools and interpretation strategies are still under development. In addition, sampling variability may be relevant to the data generated, particularly so for PSC in which a patchy affection is most pronounced. Furthermore, use of end-stage explanted liver from liver transplantation may render the identification of antigen-specific, disease-relevant T-cell populations involved in the primary insult difficult because several secondary inflammatory complications are common at this stage that could cause infiltration by bystander T cells. Nevertheless, our data provide important insights characterizing the signatures of liver-infiltrating T cells from a large number of diseased chronic liver tissues without any *in vitro* manipulation;

and on the insights gained, we strongly advocate that the specificity phenotype of hepatic T cells needs to be accounted for in the study of their functional properties and role in the disease processes.

In summary, using high-throughput sequencing, we provide a comprehensive characterization of the immune profile of liver-infiltrating T cells from patients with PSC, PBC, and ALD. Our data reveal the presence of disease-associated clonotypes in all livers investigated. In PSC, these features of antigen-driven clonal expansion occur alongside T-cell diversity and polyclonality. In ALD, surprisingly, a pronounced oligoclonal nature of liver-infiltrating T cells was observed, along with the detection of ALD-associated clonotypes, possibly supporting the presence of neoantigen-specific T-cell responses in the otherwise environmentally driven condition. The presence of disease-associated clonotypes in PSC and PBC patients bearing HLA-susceptible haplotypes supports a link between HLA and antigenic repertoires. Further appreciation of antigenic triggers responsible for the observation made has potential to help develop targeted therapy and allow for functional assessments of T cells to focus on pathogenic subpopulations, rather than the overall selections based on surface marker phenotyping.

References

1. Ludwig J, Barham SS, LaRusso NF, Elveback LR, Wiesner RH, McCall JT. Morphologic features of chronic hepatitis associated with primary sclerosing cholangitis and chronic ulcerative colitis. Hepatology 1981;1:632-640.
2. Liaskou E, Jeffery LE, Trivedi PJ, Reynolds GM, Suresh S, Bruns T, et al. Loss of CD28 expression by liver-infiltrating T cells contributes to pathogenesis of primary sclerosing cholangitis. Gastroenterology 2014;147:221-232.
3. Katt J, Schwinge D, Schoknecht T, Quaas A, Sobottka I, Burandt E, et al. Increased T helper type 17 response to pathogen stimulation in patients with primary sclerosing cholangitis. Hepatology 2013;58:1084-1093.
4. **Sebode M, Peiseler M,** Franke B, Schwinge D, Schoknecht T, Wortmann F, et al. Reduced FOXP3$^+$ regulatory T cells in patients with primary sclerosing cholangitis are associated with IL2RA gene polymorphisms. J Hepatol 2014;60:1010-1016.
5. Lindor KD, Gershwin ME, Poupon R, Kaplan M, Bergasa NV, Heathcote EJ; American Association for Study of Liver Diseases. Primary biliary cirrhosis. Hepatology 2009;50:291-308.
6. Lleo A, Selmi C, Invernizzi P, Podda M, Coppel RL, Mackay IR, et al. Apotopes and the biliary specificity of primary biliary cirrhosis. Hepatology 2009;49:871-879.
7. Haydon G, Lalor PF, Hubscher SG, Adams DH. Lymphocyte recruitment to the liver in alcoholic liver disease. Alcohol 2002;27:29-36.
8. **Hov JR, Kosmoliaptsis V,** Traherne JA, Olsson M, Boberg KM, Bergquist A, et al. Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. Hepatology 2011;53:1967-1976.
9. Naess S, Lie BA, Melum E, Olsson M, Hov JR, Croucher PJ, et al. Refinement of the MHC risk map in a scandinavian primary sclerosing cholangitis population. PLoS One 2014;9:e114486.
10. Berntsen NL, Klingenberg O, Juran BD, Benito de Valle M, Lindkvist B, Lazaridis KN, et al. Association between HLA haplotypes and increased serum levels of IgG4 in patients with primary sclerosing cholangitis. Gastroenterology 2015;148:924–927.
11. Pannetier C, Even J, Kourilsky P. T-cell repertoire diversity and clonal expansions in normal and clinical samples. Immunol Today 1995;16:176-181.
12. Plasilova M, Risitano A, Maciejewski JP. Application of the molecular analysis of the T-cell receptor repertoire in the study of immune-mediated hematologic diseases. Hematology 2003;8:173-181.
13. **Robins HS, Campregher PV, Srivastava SK**, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood 2009;114:4099-4107.
14. Cooper ZA, Frederick DT, Juneja VR, Sullivan RJ, Lawrence DP, Piris A, et al. BRAF inhibition is associated with increased clonality in tumor-infiltrating lymphocytes. Oncoimmunology 2013;2:e26615.
15. VanderBorght A, Geusens P, Vandevyver C, Raus J, Stinissen P. Skewed T-cell receptor variable gene usage in the synovium of early and chronic rheumatoid arthritis patients and persistence of clonally expanded T cells in a chronic patient. Rheumatology (Oxford) 2000;39:1189-1201.
16. Moebius U, Manns M, Hess G, Kober G, Meyer zum Buschenfelde KH, Meuer SC. T cell receptor gene rearrangements of T lymphocytes infiltrating the liver in chronic active hepatitis B and primary biliary cirrhosis (PBC): oligoclonality of PBC-derived T cell clones. Eur J Immunol 1990;20:889-896.
17. Van de Water J, Ansari A, Prindiville T, Coppel RL, Ricalton N, Kotzin BL, et al. Heterogeneity of autoreactive T cell clones specific for the E2 component of the pyruvate dehydrogenase complex in primary biliary cirrhosis. J Exp Med 1995;181:723-733.
18. Kita H, Lian ZX, Van de Water J, He XS, Matsumura S, Kaplan M, et al. Identification of HLA-A2-restricted CD8$^+$ cytotoxic T cell responses in primary biliary cirrhosis: T cell activation is augmented by immune complexes cross-presented by dendritic cells. J Exp Med 2002;195:113-123.
19. Kita H, Matsumura S, He XS, Ansari AA, Lian ZX, Van de Water J, et al. Quantitative and functional analysis of PDC-E2-specific autoreactive cytotoxic T lymphocytes in primary biliary cirrhosis. J Clin Invest 2002;109:1231-1240.
20. Broome U, Grunewald J, Scheynius A, Olerup O, Hultcrantz R. Preferential V beta3 usage by hepatic T lymphocytes in patients with primary sclerosing cholangitis. J Hepatol 1997;26:527-534.
21. Probert CS, Christ AD, Saubermann LJ, Turner JR, Chott A, Carr-Locke D, et al. Analysis of human common bile duct–associated T cells: evidence for oligoclonality, T cell clonal persistence, and epithelial cell recognition. J Immunol 1997;158:1941-1948.
22. Duryee MJ, Willis MS, Freeman TL, Kuszynski CA, Tuma DJ, Klassen LW, et al. Mechanisms of alcohol liver damage: aldehydes, scavenger receptors, and autoimmunity. Front Biosci 2004;9:3145-3155.
23. Stewart S, Jones D, Day CP. Alcoholic liver disease: new insights into mechanisms and preventative strategies. Trends Mol Med 2001;7:408-413.
24. Lundin KE, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, et al. Gliadin-specific, HLA-DQ(alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med 1993;178:187-196.
25. **Britanova OV, Putintseva EV**, Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. J Immunol 2014;192:2689-2698.
26. Christophersen A, Raki M, Bergseng E, Lundin KE, Jahnsen J, Sollid LM, et al. Tetramer-visualized gluten-specific CD4$^+$ T cells in blood as a potential diagnostic marker for coeliac disease without oral gluten challenge. United European Gastroenterol J 2014;2:268-278.

Author names in bold denote shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.28116/suppinfo.