Title: Novel autoantibody targets identified in patients with autoimmune hepatitis (AIH) by PhIP-Seq

Arielle Klepper[1], Andrew Kung[1,2], Sara Vazquez[1,2], Anthea Mitchell[1,2], Sabrina Mann[1,2], Isaac Avila-Vargas[1], Swathi Kari[1], Melawit Tekeste [1], Javier Castro[1], Briton Lee[1], Maria Duarte[1], Mandana Khalili [1,4], Michael Wilson[1], Monica Yang[1], Paul Wolters[1], Jennifer Price[1,4], Emily Perito[1,4], Sandy Feng[1,4], Jacquelyn Maher[1,4], Jennifer Lai[1,4], Christina Weiler-Normann[3], Ansgar W Lohse[3], Joseph DeRisi[1,2*],Michele Tana[1,4*]

[1]University of California, San Francisco, USA
[2]Chan Zuckerberg Biohub; San Francisco, CA, USA
[3]Medical Department, University Medical Center Hamburg-Eppendorf, Hamburg, DE
[4]UCSF Liver Center

Abstract

While the identification of autoantigens remains a critical challenge in understanding and treating autoimmune diseases, the role of autoantibodies in the pathophysiology of autoimmune hepatitis (AIH) is uncertain. To better characterize the antigenic landscape of AIH, we employed Phage Immunoprecipitation-Sequencing (PhIP-Seq) to identify novel autoantibodies specific to patients. Using these results, a logistic regression classifier was able to predict which patients had AIH, relative to healthy controls indicating the presence of a distinct humoral immune signature. To further investigate the autoantibodies most specific to AIH, significant peptides were identified with >98% specificity relative to a broad array of controls (298 patients with NAFLD, PBV, or healthy controls). Top ranked autoreactive targets included anti-SLA, a well-recognized autoantibody in AIH, and a second hit to disco interacting protein 2 homolog A (DIP2A). The autoreactive fragment of DIP2A shares a nearly identical 8 amino acid stretch with the U27 protein of HHV-6, a virus known to reside in the liver. Finally, peptides derived from the LRRNT domain of the relaxin family peptide receptor 1 (RXFP1) were highly enriched and specific to AIH. The enriched peptides map to a motif adjacent to the receptor binding domain, required for RXFP1 signaling. RXFP1 is a G-protein coupled receptor that binds relaxin-2, an anti-fibrogenic molecule shown to reduce the myofibroblastic phenotype of hepatic stellate cells. Eight of nine patients with antibodies to RXFP1 had evidence of advanced fibrosis (F3 or greater). Furthermore, serum from AIH patients positive for anti-RFXP1 antibody was able to significantly inhibit relaxin-2 signaling in THP1 cells relative to anti-RXFP1 negative control serum, while depletion of IgG abrogated this effect. These data demonstrate that PhIP-Seq is a powerful discovery tool, capable of identifying known and novel autoantibody targets in AIH, lending additional supporting evidence that HHV6 may play a role in the development of AIH, and pointing to a potential pathogenic role for anti-RXFP1 IgG in some patients. Identification of anti-RXFP1 in patient serum may enable risk stratification of AIH patients for fibrosis progression and lead to the development of novel strategies for disease intervention.

Introduction

Autoimmune hepatitis (AIH) is a chronic, severe liver disease identified in the 1950s, impacting all ages and races, rising in incidence[1], and disproportionately impacting people of color[2,3]. Treatment of AIH frequently requires lifelong therapy with immunosuppressive medications, which have potential side effects. However, despite the long-standing clinical burden of AIH, little is known about the etiopathogenesis of disease. Clinically, AIH onset is often marked by an episode of acute hepatitis and initial work-up includes assessment of total IgG levels (commonly elevated in patients with AIH), as well as evaluation for characteristic autoantibodies, particular anti-nuclear

antibodies (ANA), anti-smooth muscle antibodies (SMA), and liver-kidney microsomes type 1 (anti-LKM-1), none of which are specific to AIH or to the liver itself, as well as anti-soluble liver antigen and liver pancreas (SLA/LP), a liver antigen highly specific to AIH, present in ~20% of AIH patients[4–6]. Determination of this serologic profile is essential to the diagnosis of AIH, and allows for discrimination between the two types of autoimmune hepatitis, AIH-1 (primarily impacting adults) and AIH-2 (primarily impacting children). However, the significance of autoantibodies play in determining prognosis, or their role in disease pathogenesis, is debated.

The pathogenesis of AIH is believed to result from a combination of genetic, immunologic, and environmental factors. Regarding genetic predisposition, genome-wide association studies (GWAS) of patients from Europe and North America have shown a significant association with Type I AIH and HLA alleles DRB1*0301 and DRB1*0401[7]. Several studies have further implicated a break-down in self-tolerance as a core immunologic mechanism of disease[8,9]. Numerous environmental triggers have also been associated with AIH, such as medications and viruses, for example minocycline, nitrofurantoin, hepatitis viruses, and human herpes viruses[10,11]. However, a driving, central autoantigen in Type 1 AIH has not been identified, and the contribution of molecular mimicry as a pathogenic mechanism, while proposed, remains controversial. The role antibodies play in driving the pathogenesis of AIH is also debated.

While autoantibodies play a central role in AIH diagnosis clinically, further analysis of the pathogenic role of B cells and antibodies in the etiopathogenesis of AIH is needed in order to advance our understanding of disease, and has been cited as a core goal of the AIH research agenda by professional societies[12]. In order to perform an unbiased, high-throughput assessment of autoantibody profiles in patients with autoimmune hepatitis, we applied phage display immunoprecipitation sequencing (PhIP-Seq) to characterize the serum or plasma from 115 AIH patients obtained from multicenter, international collaborative cohort of patients, and compared these results to a robust series of 298 control serum or plasma samples from patients with non-alcoholic fatty liver disease (NAFLD), primary biliary cholangitis (PBC), or healthy controls to further define novel disease and tissue specific autoantibody targets with the power to inform our understanding of AIH pathogenesis.

Results

Autoantibody testing was performed on broad, multi-center, international cohorts with robust controls

As part of a multicenter, international collaboration, specimens of serum or plasma from patients with AIH (n = 115), NALFD (n = 178), or PBC (n = 26) were contributed by three well-characterized patient cohorts: Prospective Observational Study to Understand Liver Diseases (POSULD, San Francisco General Hospital, SF, CA, USA), FrAILT (UCSF Parnassus Hospital, SF, CA, UCSA), and the Eppendorf University cohort (Hamburg, Germany). De-identified healthy control samples were obtained from two sources: the New York Blood Center, or purchased through SeraCare (K2EDTA human plasma). Clinical characteristics are summarized in Table 1 (see methods for additional detail).

AIH disease status can be predicted using machine learning

A schematic overview of PhIP-Seq methodology is summarized in Figure 1a, as well a work-flow for the customized bioinformatic approach to analysis of this data, as previously reported[13] (Fig 1b). Applying these methods, we performed logistic regression with an 85%-15% train-test split ratio with 100 random iterations (Fig 1c). The resulting model was able to predict which patients

had AIH, or were healthy controls, on the basis of PhIP-Seq autoreactivity against all peptides (731,724 peptides; median AUC = 0.81). These results were significant, given that serologic diagnosis of AIH is quite heterogenous, as there is only one commonly identified autoantibody highly-specific specific to AIH, known as SLA, present in ~20% of AIH patients. Furthermore, patients may present with no positive autoantibodies at the time of clinical assessment[14]. When attempting to limit the logistic regression model to a smaller number features using only the top 1,000 weighted peptides, and refitting the model, predictive power was degraded (mean AUC=0.62), suggesting that there is substantial heterogeneity in the autoreactivity profiles of individual AIH patients, requiring the full breadth of peptides to successfully classify disease state.

Identification of individual peptide reactivities identifies novel, and known antibody targets in AIH

Given the known clinical heterogeneity in AIH, immunoreactivities to individual peptides were examined, with the goal of trying to identify targets that might inform our understanding of AIH pathogenesis. In order to do this, significantly enriched targets were selected by setting a Z-score threshold of >2.5, relative to the mean of healthy control samples, and further requiring that hits could *not* be significantly enriched in more than 2% of *all* controls (NAFLD, PBC, and healthy controls, 6/298 patients) and should be significantly enriched in at least 6% of the AIH patients (7/115 patients). This approach identified 120 highly-specific hits at the peptide level, representing 110 genes (see Supplemental Table 1 for a complete list). To further refine this list, liver specific targets were selected based on gene annotations using the ProteINSIDE online tool[15], which had available data on the tissue expression patterns of 102/110 genes. Annotations were used to divide genes into 3 categories: category 1, genes annotated as highly expressed or enriched liver tissue (8 peptides corresponding to 6 genes), category 2, genes annotated to be expressed ubiquitously or at low levels in liver (24 peptides corresponding to 22 genes) and category 3, genes annotated to not be expressed in liver (80 peptides corresponding to 74 genes).

Among the hits in category 1, genes enriched/highly expressed liver, SEPSECS, the gene encoding SLA/LP protein, was the most specific gene, and was not seen in any control patients (Fig 2b). The peptide hits for were localized to the C-terminus of the SLA/LP protein, starting at amino acid 393, which corresponds to the region reported to be essential for antibody reactivity to SLA in prior reports (amino acids 370-410)[6]. The remaining 5 hits in category 1 are all novel putative autoantibody targets. Among these, 3 were chosen for additional scrutiny because they derive from secreted proteins or cell surface receptors. Interestingly, deficiency in each of these 3 proteins, CFHR1, ADAMTS-7 and RXFP1, have been previously associated with inflammatory disease/autoimmunity or liver-related fibrosis. The complement Factor H Related 1 protein (CFHR1) has been associated with autoimmunity, specifically with the development of atypical hemolytic uremic syndrome[16]. The ADAMTS family member – a disintegrin-like and metalloproteinase with thrombospondin motif) – ADAMSTS-7, is a multi-domain secreted metalloproteinase that degrades cartilage oligomeric matrix protein (COMP), is part of a positive feedback loop with TNF-a[17], and protein deficiency has been reported to cause biliary fibrosis[18]. Of particular interest to AIH pathogenesis was relaxin family peptide receptor 1 (RXFP1), which is a cell surface receptor for the anti-fibrotic molecule, relaxin-2. Relaxin-2 signaling through RXFP1 on the surface of activated hepatic stellate cells has been shown to decrease their fibrogenic potential[19], and relaxin-2 has been trialed as an anti-fibrotic agent in phase 2 studies of patients with alcohol associated liver disease[20]. Autoantibody-mediated blockade of RXFP1 signaling, which is anti-fibrotic, has the potential to promote fibrogenesis. Among the 9 AIH patients found to be positive for antibodies against RXFP1 by PhIP-Seq, 8 patients (88%) had evidence of advanced fibrosis, F3 or greater. Therefore, RXFP1 was selected as a candidate for further validation and investigation.

Examination of reactivities to putative autoantigens not found in liver identifies potential epitope homology to HHV6 U27 protein

Regarding the genes in category 3, those not annotated as being expressed in liver, the top 20 peptides are displayed in Fig 2c. Three peptides all derive from the same gene, DIP2A. These peptides all overlapped, sharing a 19 amino acid epitope common to each peptide. DIP2A is not known to be expressed in the liver, suggesting that the enrichment of these peptides was due to antibodies directed against some other similar target, perhaps of viral origin. To explore this hypothesis, a protein-level BLAST search against bacterial and viral taxa was performed using the enriched peptides as bait. Of the viruses that infect humans, the top match to the DIP2A 19 amino acid common epitope was human herpes virus 6B (HHV-6B), specifically the U27 gene (Fig 3d). U27 is a viral protein required for replication processivity, key to the HHV-6 viral lifecycle. Of note, the SLA protein has also been previously reported to have 41% homology to a peptide derived from the HHV-6 U14 protein[21] however, applying the same BLAST search methods to identify viral homology to the previously reported antigenic region of the SLA/LP epitope (amino acids 371-409) did not identify HHV-6 as a potential match. These differences may be accounted for by the degree of similarity between DIP2A and the HHV-6 region of homology, which was very high: 6 of the 8 amino acids were identical, a 7th with similar chemical properties between the two proteins, and only 1 mismatched amino acid (Fig 3d). While HHV6 is nearly ubiquitous, the virus itself establishes latency, and re-activation has been associated with hepatitis[22], and there are links between HHV-6 infection and the onset of AIH[23]. Using the MHC Class II search feature of the immune epitope database (IEDB)[24], the 8-amino acid region of U27 with homology to DIP2A is also predicted to bind with high affinity to HLA-DRB1 03*01, one of the two risk alleles associated with AIH in genome-wide association studies of patients across Europe and North America[7]. This is consistent with the notion that molecular mimicry to HHV-6, perhaps modulated by HLA type, could play a role in the development of autoimmune hepatitis.

Orthogonal validation of RXFP1 peptide reactivity by split luciferase binding assay confirms PhIP-Seq results, and correlation with patient metadata reveals clinical relevance

To validate the RXFP1 finding, we employed a split luciferase binding assay (SLBA), as recently reported[25]. Briefly, we used *in vitro* transcription and translation to generate the primary RXFP1 peptide identified by PhIP-Seq, with the addition of a HiBiT tag, which when complexed with the LgBiT protein, generates luminescence (Progema system). Immunoprecipitation of tagged peptides was performed with a subset of AIH patients (where larger volumes of serum or plasma were available), as well as with various controls. SLBA Positive control immunoprecipitation was performed using an anti-HiBiT antibody (Promega), and negative controls were performed using "mock IP" with buffer alone, absent patient serum. A sample was considered positive if the signal in the assay exceeded a cutoff value of the mean plus 3 standard deviations from the mean of all control signal (Fig 3a). This assay demonstrated that the same 9 patients positive by PhIP-Seq were positive by SLBA, and none of the liver disease controls or healthy controls were positive (Fig 3a). We included an additional control of another fibrotic autoimmune disease, systemic sclerosis (SSc), as relaxin-2 knockout mice develop systemic fibrosis similar to human SSc and relaxin-based therapy has been pursued in clinical trials of SSc as a therapeutic. We found evidence of only 1 positive SSc patient for anti-RXFP1 reactivity among the 30 patients assayed (Fig 3, green), further underscoring the specificity of anti-RXFP1 antibodies to AIH. Following this validation experiment, we returned to the PhiP-Seq data to further assess correlation among all patients reacting with RXFP1, and found that none of the 9 positive patients (Z-score of > 2.5) had quiescent disease (defined as normal AST, ALT and IgG levels); each positive patient had some evidence disease activity (defined by AST or ALT greater than 2x the upper limit of normal OR an elevated IgG level). Lastly, we searched for predicted MHC Class II binding using IEDB[24]. This search disclosed that more than half of the top, MHC Class II alleles predicted to bind the RXFP1 peptide of interest corresponded to HLA alleles DRB1 03*01 and 04*01, the two genes

associated with the development of AIH risk by GWAS among patients from Europe and North America[7] (Sup Table 2), suggesting a plausible mechanism for T cell binding.

Serum from AIH patients with anti-RXFP1 activity is able to inhibit relaxin-2 signaling through RXFP1 in an IgG dependent manner

The functional implications of RXFP1 positivity were explored further to investigate the possibility of pathogenic autoantibodies. The proposed structure for RXFP1 is diagrammed in Fig 4a using UCSF ChimeraX[26], with annotation of various domains. The extracellular portion of RXFP1 (light gray, Fig 4a) is composed primarily of a large, leucine rich repeat (LRR) motif, present among a larger family of LRR containing G-protein coupled receptors (LGRs)[27]. This LRR repeat region is the site of relaxin-2 ligand binding. Further highlighted in red is the RXFP1 peptide target of patient reactivity, identified in PhIP-Seq and SLBA assays. At the core of this peptide is an LRRNT motif, an N-terminal capping region required to maintain stability of the leucine-rich repeat (LRR) region; the RXFP1 peptide is also at the end of a linker region connecting the LRRNT to an LDLa domain (diagrammed in red, Fig 4a inset panel). This linker has been demonstrated to be required for RXFP1 receptor activation[28]. Thus, based on the peptide's position within RXFP1, we hypothesized that antibodies targeting this peptide would able to functionally interfere with RXFP1 signaling. We tested this hypothesis using the cAMP-Glo signaling assay (Promega) in which ligand binding to its cognate G-protein coupled receptor (GPCR) leads to the production of cAMP, and cAMP production is coupled to luminescence, which can be measured as a sensitive read-out of ligand binding to the GPCR. We performed this assay in THP-1 cells, a human monocytic cell line, as cAMP production in response to relaxin-2 binding of RXFP1 has been well-characterized in this system[29]. To test the impact of sera from anti-RXFP1 positive or negative patients, we pre-incubated THP-1 cells for 2 hours with sera AIH patients positive for anti-RXFP1, or with sera with patients negative for anti-RXFP1. Using this approach, we were able to demonstrate that pre-incubation with patient serum positive for anti-RXFP1 led to a shift in the RXFP1 dose-response (representative patients, Fig 4b), and a significant increase in the half maximal concentration of relaxin-2 required to generate a cAMP response, known as EC50 (Fig 4c). AG-bead based depletion of IgG prior to pre-incubation with anti-RXFP1 positive serum abrogated this effect (Fig 4d), demonstrating IgG targeting RXFP1 is capable of inhibiting relaxin-2 signaling.

Discussion:

Employing PhIP-Seq to the study of human disease is powerful approach for discovery of novel autoantigens, and this methodology has been applied by our group and others[13,25,30–34] to identify several novel antibody targets with the potential to uncover disease pathogenesis. PhIP-Seq was used to study the role antibodies play in the pathogenesis of autoimmune hepatitis by studying patients from an international, multi-center collaboration. One unique aspect of this cohort is the inclusion of a large number of non-AIH liver disease controls including 178 patients with non-alcoholic fatty liver disease (NAFLD), a disease that often is coincident with AIH, and can be difficult to differentiate from AIH on the basis of serology given the high rates co-occurring autoantibodies, such as ANA and anti-smooth muscle antibodies, reported to be positive in 20-30% of NAFLD patients[4,35]. These controls facilitate identifying targets highly-specific to AIH, which may play a specific role in pathogenesis. Furthermore, the application of PhIP-Seq allows for unbiased, high-throughput, and sensitive detection of putative autoantibodies.

The use of machine learning methods was able to predict which patients had AIH (relative to healthy controls) however, this predictive power relied on building a model based on all PhIP-Seq peptides assayed. This data further supports a recent report demonstrating that detection of poly-reactive IgG was a more specific marker of AIH diagnosis than conventional autoantibody profiles[36], emphasizing the broad array of serologic reactivity characteristically seen in AIH.

To further leverage the comprehensive nature of PhIP-Seq performed on pateitns from a multi-center international cohort of AIH patents and controls, we turned our attention to the identification of both disease and tissue specific antibodies, with the goal of trying to identify new insights into AIH pathogenesis. SLA was identified as the most specific target using this approach, providing strong proof of concept for the reliability of this assay. We further identified several novel autoantigens, such IgG targeting RXFP1, which was capable of inhibiting relaxin-2 signaling via RXFP1. This inhibition has the potential to diminishing the anti-fibrotic properties of relaxin-2. Future directions include evaluating whether identification of anti-RXFP1 in patient serum may enable risk stratification of AIH patients for fibrosis progression, and has the potential to lead to the development of novel strategies for disease intervention.

We were also able to identify a possible role of non-liver specific of antibody reactivities, such as antibodies targeting a 19 amino acid epitope of DIP2A, which shares significant sequence homology with an 8 amino acid region of the HHV-6B protein U27. HHV6 has previously been implicated as a possible nidus of AIH-related molecular mimicry however, validation was unable to demonstrate that patients positive for the SLA antibody reacted to HHV-6 U14[21]. Further study is required to determine whether DIP2A positive patients react with the HHV-6 U27 protein based on the high degree of homology between the DIP2A epitope and HHV-6B.

This study has several limitations, including the small number of patients sharing any one particular set of reactivities. While this is in line with the proposed biology of AIH, it makes it difficult to determine generalizability of findings to larger populations. While the findings raise the specter of molecular mimicry, anti-viral antibody reactivities among patients was not measured. Finally, we understand the value of multiple different analysis approaches, and in that vein, remain committed to sharing the data generated using this approach for other groups with expertise in data analysis approaches to re-evaluate this data generated by PhIP-Seq to evaluate candidate hits. To facilite this, full PhIP-seq data will be linked to the ultimate publication and available for download at Dryad.

Methods

Patient cohorts

As part of a multicenter, international collaboration, specimens of serum or plasma from patients with AIH (n = 115), NALFD (n = 178), or PBC (n = 26) were contributed by three well-characterized patient cohorts: Prospective Observational Study to Understand Liver Diseases (POSULD, San Francisco General Hospital, SF, CA, USA), FrAILT (UCSF Parnassus Hospital, SF, CA, UCSA), and the Eppendorf University cohort (Hamburg, Germany). All patients provided informed consent in accord with institutional policies. Clinical data was obtained from existing databases and the medical record, including demographic information and disease activity, fibrosis stage, and medication regimen near the time of specimen collection. Patients with more than one diagnosis (eg overlap syndromes, AIH and NAFLD, etc.) were excluded from the current analysis. Coded specimens were analyzed in a deidentified fashion. Healthy

control samples were obtained as de-identified samples from two sources: the first was from the New York Blood Center, as part of retention tubes collected at the time of blood donation from volunteer donors who provided their informed consent for their samples to be used for research. The second source was patient plasma from donors obtained from FDA-licensed blood collection facilities, purchased through SeraCare (K2EDTA human plasma).

Phage immunopercipitaion sequencing (PhIP-Seq)

PhIP-seq was performed as previously reported[13,30,35] and PhIP-seq protocols described in detail are available at protocols.io, located at the links below.

- Multichannel-based scaled: https://www.protocols.io/view/scaled-moderate-throughput-multichannel-phip-proto-8epv5zp6dv1b/v1
- Library Preparation: https://www.protocols.io/view/phage-display-library-prep-method-rm7vz3945gx1/v1

Briefly, blood from individuals with Type 1 AIH and controls were analyzed using PhIP-Seq. Phage were cloned to express >700,000 overlapping peptides spanning the human proteome. The following text adapted from Zamecnik *et al*[33]*:* 96-well, 2mL deep well polypropylene plates were incubated with a blocking buffer (3% BSA in TBST) overnight at 4°C to prevent nonspecific binding. Blocking buffer was then replaced with 500 µL of freshly grown phage library and 1 µL of human sera 1:1 storage buffer (PBS with 0.04% NaN3, 40% Glycerol, 40mM HEPES). To facilitate antibody-phage binding, the deep well plates with library and sample were incubated overnight at 4°C on a rocker platform. 10 µL of each of Pierce Protein A and G Beads (ThermoFisher Scientific, 10002D & 10004D) slurry were aliquoted per reaction and washed 3 times in TNP-40 (140mM NaCl, 10mM Tris-HCL, 0.1% NP40). After the final wash, beads were resuspended in TNP-40 in half the slurry volume (20uL) and added to the phage-patient antibody mixture and incubated on the rocker at 4°C for 1 hour. Beads were then washed in RIPA buffer, and then the immunoprecipitated solution was resuspended in 150 µL of LB-Carb and then added to 0.5mL of log-phase BL5403 E. coli for amplification (OD600 = 0.4-0.6) until lysis was complete (approximately 2h) on an 800 rpm shaker. After amplification, sterile 5M NaCl was added to lysed E. coli to a final concentration of 0.5M NaCl to ensure complete lysis. The lysed solution was spun at 3220 rcf for 20 minutes and the top 500 µL was filtered to remove remaining cell debris. Filtered solution was transferred to a new pre-blocked deep-well plate where patient sera was added and subjected to another round of immunoprecipitation and amplification, and 3 total rounds of immunoprecipitation were completed. The final lysate was spun at 3220 rcf for 30 minutes, with supernatant then filtered and stored at 4°C for subsequent NGS library prep. Phage DNA from each sample was barcoded and amplified (Phusion PCR) and then underwent Next-Generation Sequencing on an Illumina NovaSeq Instrument (Illumina, San Diego, CA).

Split Luciferase Binding Assay (SLBA)

This assay was performed as recently reported[25]; and a detailed SLBA protocol is available on protocols.io at dx.doi.org/10.17504/protocols.io.4r3l27b9pg1y/v1.

Briefly, the target peptide of relaxin family peptide was identified by PhIP-Seq, with the following peptide sequence: VGSVPVQCLCQGLELDCDETNLRAVPSVSSNVTAMSLQWNLIRKLPPDC. The following text was adapted from Rackaityte *et al*.[25]: The nucleic acid sequence of this

construct was inserted into a split luciferase construct containing a terminal HiBiT tag and synthesized (IDT) as DNA oligomers. Constructs were amplified by PCR using 5'-AAGCAGAGCTCGTTTAGTGAACCGTCAGA-3' and 5'-GGCCGGCCGTTTAAACGCTGATCTT-3' primer pair. Unpurified PCR product was used as input to rabbit reticulocyte transcription translation system (Promega) and Nano-Glo HiBit Lytic Detection System (Promega Cat No. N3040) was used to measure relative luciferase units (RLU) of translated peptides in a luminometer. Peptides were normalized to 5e6 RLU input, incubated overnight with patient sera, and immunoprecipitated with protein A and protein G sepharose beads (Millipore Sigma). After thoroughly washing beads with SLBA buffer (0.15M NaCl, 0.02M Tris-HCl pH7.4, 1% w/v sodium azide, 1% w/v bovine serum albumin, and 0.15% v/v Tween-20), luminescence remaining on beads was measured using Nano-Glo HiBit Lytic Detection System (Promega Cat No. N3040) in a luminometer. Anti-HiBiT antibody (Promega) was used as a positive control for each peptide. An patient was considered positive by SLBA if the RLU exceeded the mean of all controls + 3 standard deviations.

Relaxin-2 signaling/cAMP-Glo assay/IgG depletion

Recombinant human relaxin H2 was purchased from R&D systems (catalog # 6586-RN-025/CF) and resuspended in 1x sterile PBS with 1% BSA at a concentration of 100 ul/ml. In order to block cyclic nucleotide phosphodiesterases during the cAMP-Glo assay, serial dilutions of relaxin were made up in induction buffer composed of 1x PBS with 500 µM 3-isobutyl-1-methylxanthine (IBMX, Sigma Aldrich), and 500 µM Ro 20-1724 (Cayman Chemical). Relaxin concentrations ranged from 0.0488 – 50 ng/ul of ligand, and the 12th dilution was "untreated" control, of just induction buffer. Dilutions were made in sterile 96 well plates in order to apply to THP1 cells to study signaling. THP1 cells were obtained via ATCC, and seeded at a density of $1*10^6$ cells/well of a 96-well plate. Prior to the addition of relaxin-2, THP1 cells were pre-incubated with patient serum from RXFP1 positive patients or RXPF1 negative patients at a dilution of [1:100] in RPMI with 10% FBS/1% PSG for 2 hours in a 37 °C incubator. Following this pre-incubation, relaxin was added at each of the 11 pre-diluted concentrations to pre-incubated THP-1 cells, and 96-well plates were returned to the 37 °C incubator for two hours. All reactions were performed in triplicate. Following this incubation, cells were assayed for cAMP production using the cAMP-Glo assay (Promega catalog # V1502) was performed per the manufacturer's instructions, with the following modifications. Lysis of cells was performed using 20 ul of cAMP-Glo lysis buffer for 30 minutes in standard tissue culture plates, and then transferred to opaque white 96-well plates (Nunc) for the remainder of the assay in order to facilitate plate reading of relative light units (RLU) in a luminometer (Promega). The remainder was as per manufacturer's instructions. Results were then normalized to the fraction of untreated RLU, and EC50 values were calculated using GraphPad Prism software. For depletion experiemnts, prior to cAMP-Glo assay, as described above, sera was pre-incubated for 2 hours at room temperature with Pierce Protein A and G Beads (Pierce) at a ratio of 20 ul of A/20 ul G to 1 ul of serum, with gentle rocking. Samples were ultimately pre-incubated with THP1 cells at a dilution of [1:250] prior to addition of relaxin-2.

Bioinformatic and statistical analysis

Raw sequencing reads were aligned to the input peptide library using Bowtie2, and aligned at the protein level using RAPsearch, as previously described[30]. Aligned reads were controlled for varying read depth by aligning to 100,000 reads per K-mer (RPK). Fold-change for each peptide was generated from the mean RPK of the controls, and a z-score was calculated from the background distribution. An autoantigen was considered a hit if enrichment was >2.5 standard

deviations from the mean of healthy controls and present in at least 7 AIH patients (6% sensitivity) and not present in more than 2% of control patients (98% specificity). Logistic regression was performed using the Scikit-learn package in Python[36], using the liblinear solver and L1 regularization. The model was evaluated with and 85-15 train-test-split ratio, and performed with 100 iterations, plotting the data corresponding to the max, min, and median AUC.

Annotation of protein sequences was performed using the proteINSIDE analysis platform (https://www.proteinside.org) for tissue level annotation[15].

Molecular graphics and analyses performed with UCSF ChimeraX (https://www.cgl.ucsf.edu/chimerax/), developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

HLA Class II binding predictions were performed using the Immune Epitope Database (IEDB, www.iedb.org)[37].

Funding and Acknowledgements

NIH/NIDDK T32 DK060414, NIH/NIDDK R21, Chan Zuckerberg Biohub/Initiative, UCSF Liver Center AASLD Autoimmune Liver Diseases Pilot Research Award, UCSF Liver Center/Pilot Feasibility Award, UCSF Department of Medicine Cohort Development Grant, UCSF Precision Medicine in Rheumatology Grant, D. Montgomery Bissell, Emily D. Crawford

Figures
Table 1. Patient Characteristics

Table 1a. Characteristics of the study population assayed by PhIP-Seq					
Diagnosis	**n**	**Mean Age**	**Female**	**Caucasian**	**Latinex**
AIH	115	58	70%	96%	14%
NAFLD	178	60	60%	82%	67%
PBC	26	58%	94%	84%	21%
Healthy Control	94	--	--	--	--

Table 1b. Characteristics of AIH in the patients assayed by PhIP-Seq							
	F3/4 fibrosis	**On steroids**	**ANA +**	**ASMA +**	**Mean ALT**	**Mean AST**	**Mean IgG**
AIH Patients	60%	18%	88%	83%	79	--	1520

Figure 1. Phage immunoprecipitation sequencing enables AIH disease prediction by logistic regression.
(**a**) Phage display library design, and an overview of methods to apply PhIP-Seq to evaluate the autoimmune hepatitis and control cohort.
(**b**) Summary of the customized bioinformatic analysis pipeline to applied to analyze next-generation sequencing output of PhIP-Seq data.
(**c**) Receiver operating characteristic (ROC) curve analysis for prediction of AIH vs healthy control disease status, area under the curve (AUC).

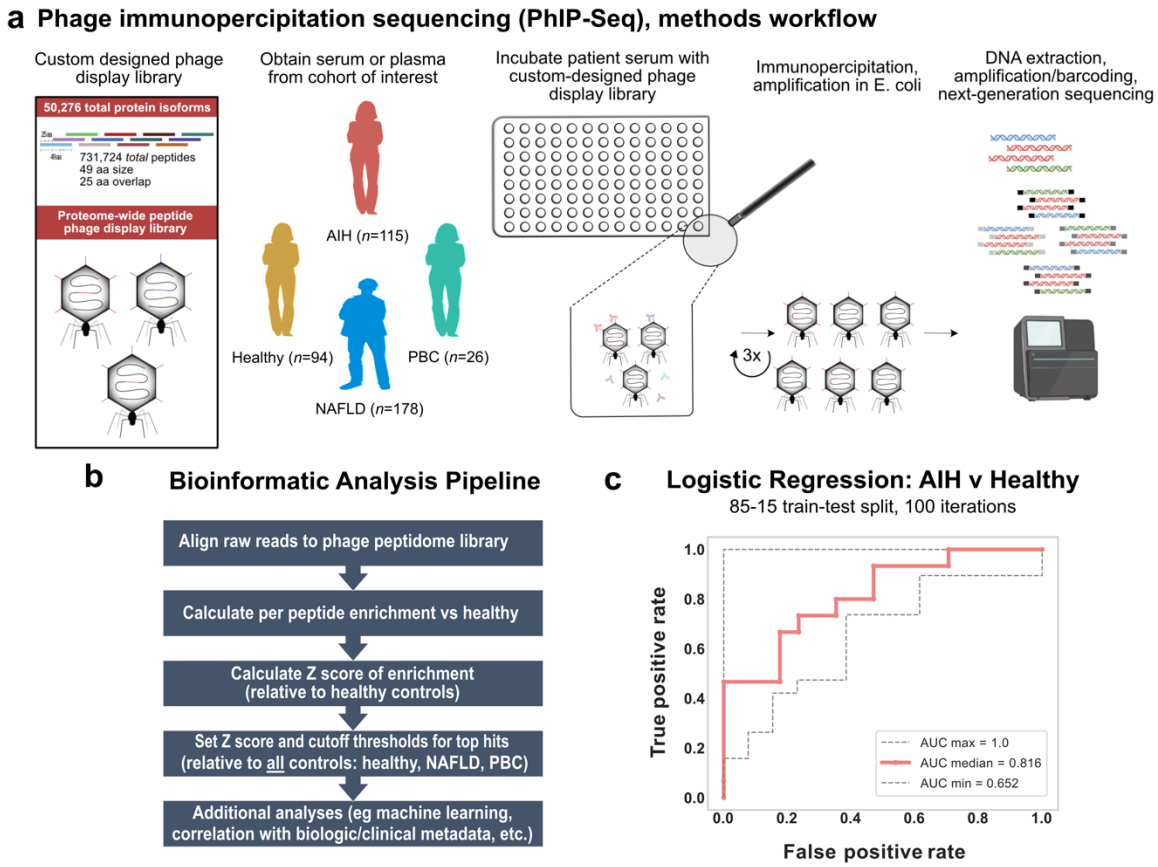


Figure 2. PhIP-Seq identifies novel (and known) antigenic targets present in AIH, which ar largely absent in control patients.

(**a**) Heatmap of liver-enriched or highly expressed genes, based on ProteinINSIDE annotations, which were identified as putative hits in AIH patients, shaded by Z-score of enrichment (legend, panel e).

(**b**) Heatmap of liver-enriched or highly expressed genes, based on ProteinINSIDE annotations, which were identified as putative hits in control patients; top legend indicates whether controls had NALFD (blue), PBC (green), or were healthy controls (yellow), shaded by Z-score of enrichment (legend, panel e).

(**c**) Heatmap of the top 20 peptides identified as hits in PhIP-Seq, not noted to be expressed in liver based on ,based on ProteinINSIDE annotations, shaded by Z-score of enrichment (legend, panel e).

(**d**) Clustal Omega alignement of the identified DIP2A epitope with the U27 protien of HHV6.

(**e**) Legend of Z-score of enrichment

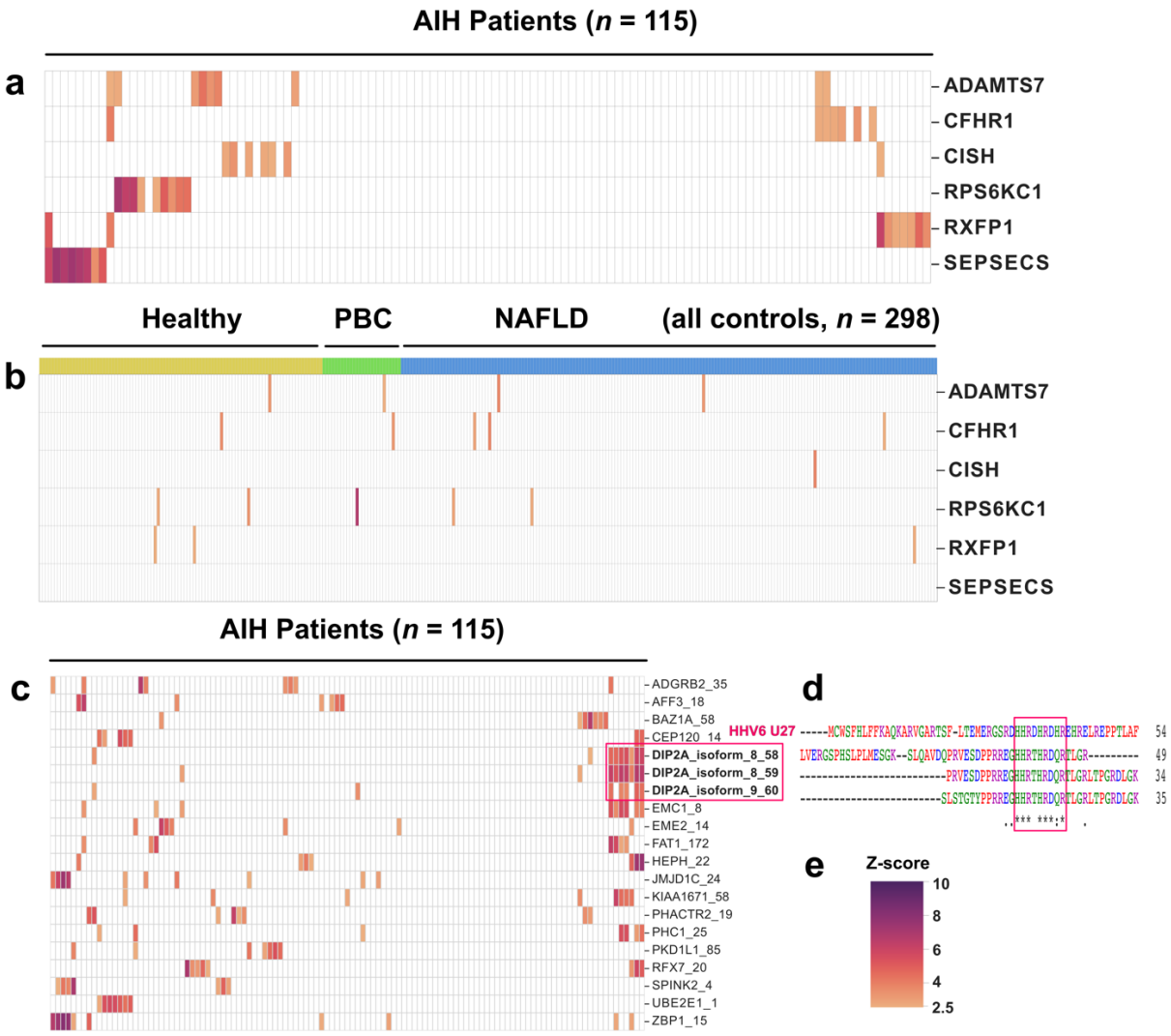

AIH Patients (n = 115)
a
ADAMTS7
CFHR1
CISH
RPS6KC1
RXFP1
SEPSECS
Healthy
PBC
NAFLD
(all controls, n = 298)
b
ADAMTS7
CFHR1
CISH
RPS6KC1
RXFP1
SEPSECS
AIH Patients (n = 115)
c
ADGRB2_35
AFF3_18
BAZ1A_58
CEP120_14
DIP2A_isoform_8_58
DIP2A_isoform_8_59
DIP2A_isoform_9_60
EMC1_8
EME2_14
FAT1_172
HEPH_22
JMJD1C_24
KIAA1671_58
PHACTR2_19
PHC1_25
PKD1L1_85
RFX7_20
SPINK2_4
UBE2E1_1
ZBP1_15
d
HHV6 U27
e
Z-score
10
8
6
4
2.5

Figure 3.
(**a**) SLBA validation of anti-RXFP1 peptide reactivity in various patients groups (x-axis), as measured by enrichment of relative light units (RLU, y-axis); the cutoff for positivity was set at the mean + 3 standard deviations of all controls (blue dashed line).
(**b**) PhIP-Seq data plotting z-score of anti-RXFP1 peptide reactivity among AIH patients (y-axis); AIH patients were separated into groups of active vs inactive AIH (x-axis); patients considered as positivity reactive against anti-RXFP1 had a Z-score of enrichment >2.5 (denoted by blue dashed line, highlighted in pink).

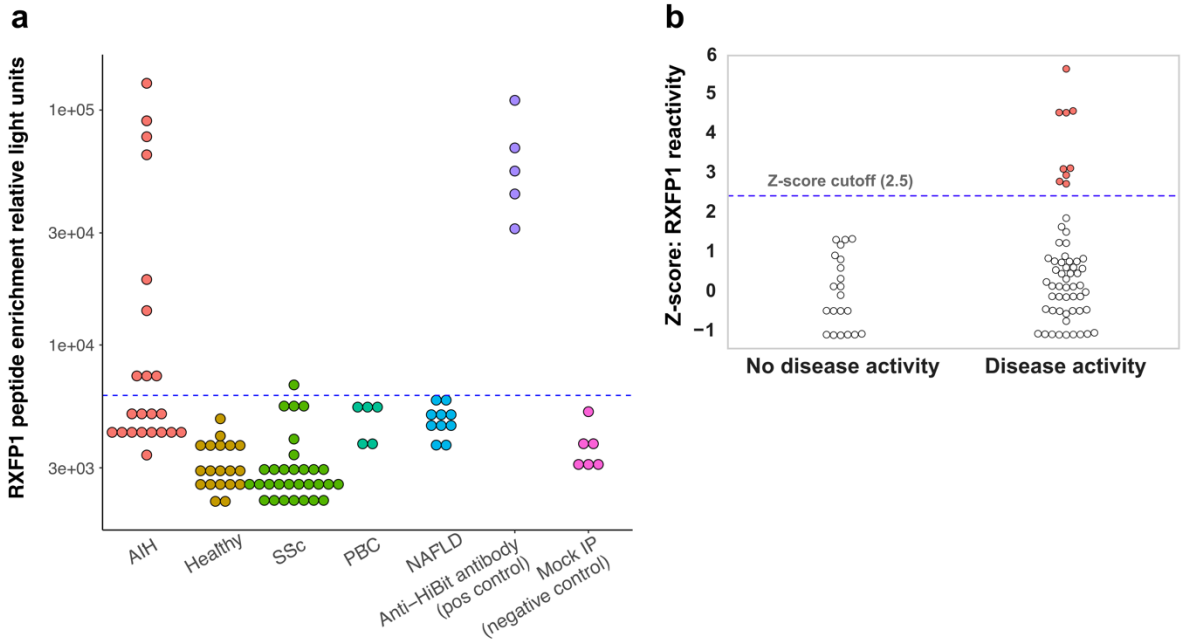
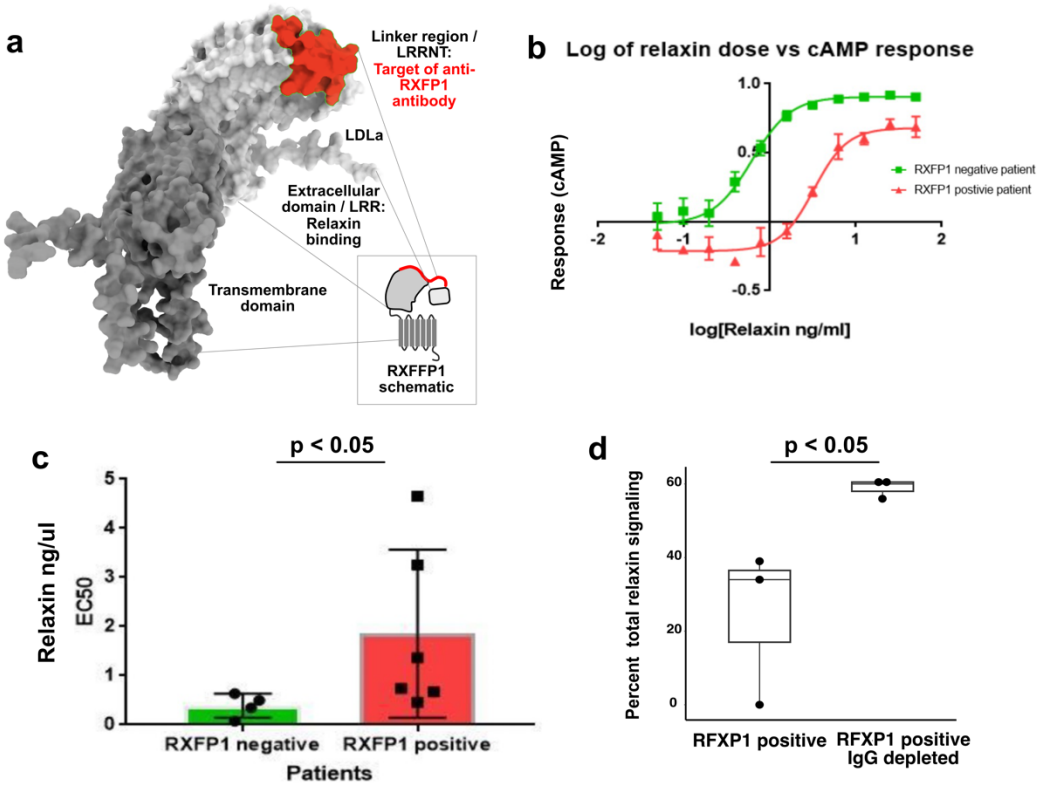


Figure 4.
(**a**) Putative structure of RXFP1, as depicted using ChimeraX; the region corresponding to the RXFP1 peptide identified by PhIP-Seq is highlighted in red, along with annotation of functional domains (for schematic representation, see panel inset).
(**b**) Assay of relaxin-2-induced induction of cAMP by RXFP1, in TH1 cells pre-incubated with [1:100] dilution of patient serum negative (green) or positive (red) for RXFP1 antibodies; relaxin concentration (x-axis), cAMP response reported as a percent of untreated control signal, y-axis.
(**c**) Measurement of relaxin-2 EC50 in ng/ul (y-axis) for patient serum negative (green) or positive (red) for RXFP1 antibodies.
(**d**) Depletion of IgG using protein A-G beads (x-axis, right) or mock-depleted serum (x-axis, left) was performed prior to incubating THP1 cells with patient serum at [1:250]; resultant impact on relaxin-2 signal expressed as a percent of untreated signal (y-axis).


a
Linker region / LRRNT:
Target of anti-RXFP1 antibody
LDLa
Extracellular domain / LRR: Relaxin binding
Transmembrane domain
RXFFP1 schematic
b
Log of relaxin dose vs cAMP response
Response (cAMP)
log[Relaxin ng/ml]
RXFP1 negative patient
RXFP1 postivie patient
c
p < 0.05
Relaxin ng/ul EC50
RXFP1 negative
RXFP1 positive
Patients
d
p < 0.05
Percent total relaxin signaling
RFXP1 positive
RFXP1 positive IgG depleted