

Title: High resolution epitope mapping of anti-Hu and anti-Yo autoimmunity by programmable phage display

Short title: Epitope mapping paraneoplastic autoimmunity

Brian O'Donovan¹, Caleigh Mandel-Brehm¹, Sara E. Vazquez¹, Jamin Liu^{1,2}, Audrey V. Parent³, Mark S. Anderson³, Travis Kassimatis¹, Anastasia Zekeridou⁴, Stephen L. Hauser⁵, Sean J Pittock⁴, Eric Chow¹, Michael R. Wilson^{5*}, Joseph L. DeRisi^{1,6*}

• Co-corresponding authors

1. Department of Biochemistry and Biophysics, University of California, San Francisco, CA
2. UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, Berkeley and San Francisco, CA
3. Diabetes Center, University of California, San Francisco, San Francisco, CA 94143
4. Department of Neurology, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester MN
5. Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, CA
6. Chan Zuckerberg Biohub, University of California, San Francisco, CA

© The Author(s) (2020). Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ABSTRACT

Paraneoplastic neurological disorders are immune-mediated diseases understood to manifest as part of a misdirected anti-tumor immune response. Paraneoplastic neurological disorder-associated autoantibodies can assist with diagnosis and enhance our understanding of tumor-associated immune processes. We designed a comprehensive library of 49 amino acid, overlapping peptides spanning the entire human proteome, including all splicing isoforms and computationally predicted coding regions. Using this library, we optimized a phage immunoprecipitation and sequencing protocol with multiple rounds of enrichment to create high-resolution epitope profiles in serum and CSF samples from patients suffering from two common paraneoplastic neurological disorders, the anti-Yo (n = 36 patients) and anti-Hu syndromes (n = 44 patients). All (100%) Anti-Yo patient samples yielded enrichment of peptides from the canonical anti-Yo (CDR2 and CDR2L) antigens, while 38% of anti-Hu patients enriched peptides deriving from the nELAVL (neuronal embryonic lethal abnormal vision like) family of proteins, the anti-Hu autoantigenic target. Among the anti-Hu patient samples that were positive for nELAVL, we noted a restricted region of immunoreactivity. To achieve single amino acid resolution, we designed a novel deep mutational scanning phage library encoding all possible single point mutants targeting the reactive nELAVL region. This analysis revealed a distinct preference for the degenerate motif, RLDxLL, shared by ELAVL2, 3 and 4. Lastly, phage immunoprecipitation sequencing identified several known autoantigens in these same patient samples, including peptides deriving from the cancer-associated antigens ZIC and SOX families of transcription factors. Overall, this optimized phage immunoprecipitation sequencing library and protocol yielded the highest resolution epitope mapping of the autoantigens targeted in anti-Yo and anti-Hu encephalitis patients to date. The

results presented here further demonstrate the utility and high-resolution capability of phage immunoprecipitation sequencing for both basic science and clinical applications and for better understanding the antigenic targets and triggers of paraneoplastic neurological disorders.

Keywords: phage display, autoimmunity, anti-Hu, anti-Yo, paraneoplastic

Abbreviations:

AA = amino acid, BSA = bovine serum albumin, CBA = cell based assay, CDR2 = cerebellar, degeneration related protein 2, CDR2L = cerebellar degeneration related protein 2 like, ELAVL = embryonic lethal abnormal vision like, FDR = false discovery rate, GP = generalized poisson, IEDB = Immune Epitope Database , IHC = immunohistochemistry, IP-MS = immunoprecipitation-mass spectrometry , mTECs = murine thymic medullary epithelial cells , NB = negative binomial, nELAVL = neuronal embryonic lethal abnormal vision like, NGS = next-generation sequencing, nt = nucleotide, PCA-1 = anti-Purkinje cell cytoplasmic antibody 1 , PCD = paraneoplastic cerebellar degeneration , pfu = plaque forming units, PhIP-Seq = phage display immunoprecipitation and sequencing, PND = paraneoplastic neurological disorder, rpK = reads per 100,000, SCLC = small cell lung cancer, TBS-T = tris buffered saline + tween

INTRODUCTION

Paraneoplastic neurological disorders (PNDs) affecting the central nervous system (CNS) are a collection of diseases characterized by an autoimmune response to proteins normally restricted to the CNS and triggered by ectopic expression of the antigen by a systemic neoplasm (Pignolet, Gebauer, and Liblau 2013). Existing PND pathogenesis models posit that an otherwise appropriate anti-tumor immune response to tumor antigens precipitates a breakdown of immune tolerance and subsequent infiltration of autoreactive lymphocytes and/or pathogenic autoantibodies into the CNS (Toothaker and Rubin 2009). The roles and culpability of cell-mediated and humoral immune responses in disease progression are disputed, as onconeurological

antibodies are often specific for intracellular proteins and are frequently detected in the sera of cancer patients in the absence of PND symptoms (Rosenfeld and Dalmau 2012).

Anti-Yo paraneoplastic cerebellar degeneration (PCD) is primarily associated with gynecological and breast malignancies (O'Brien et al. 1995)(Hasadsri et al. 2013). Anti-Yo antibodies, also known as anti-Purkinje cell cytoplasmic antibody 1 (PCA-1), target the two intracellular proteins CDR2 (cerebellar degeneration related protein 2) and/or CDR2L (CDR2-like). To date, the specific antigenic determinants on the CDR2 and CDR2L proteins have not been mapped, though recent studies implicated CDR2L as the primary antigen (Kråkenes et al. 2019).

Anti-Hu-related PND is most commonly associated with small cell lung cancer (SCLC) (Senties-Madrid and Vega-Boada 2001). Anti-Hu antibodies target members of the ELAVL (embryonic lethal abnormal vision like) family of intracellular RNA-binding proteins. There are 4 members of this family, ELAVL1, ELAVL2, ELAVL3, and ELAVL4 (also known as HuA/R, HuB, HuC, and HuD, respectively). ELAVL1 is the most divergent homolog and is ubiquitously expressed in all tissues. ELAVL2, 3, and 4 have restricted expression to the nervous system, and these three are understood to be the target of PND antibodies, and are often referred to as nELAVL proteins (neuronal ELAVL) (Pascale, Amadio, and Quattrone 2008).

The majority of currently available clinical autoimmune panels include antigen-specific ELISA-based methods, cell-based assays (CBAs), co-localization immunohistochemistry (IHC), and traditional western blots (Lancaster 2016). Unbiased and higher throughput methods exist in the research context and include immunoprecipitation-mass spectrometry (IP-MS), expression library screening, biochemical purification, peptide arrays, protein arrays, and phage display, each with its own advantages and limitations. Among these, variations on bacteriophage display technologies have been successfully utilized for decades in antigen discovery (Beghetto and

Gargano 2013), epitope mapping (Burrill et al. 1995), and antibody engineering (Frei and Lai 2016). Traditionally, tissue-specific cDNA libraries, degenerate or random synthetic oligonucleotides are cloned and expressed on the surface of phage capsids which are panned against immobilized antibody or other targets of interest. More recently, rationally designed libraries encompassing the entire human proteome have been implemented (Larman et al. 2011). With next generation sequencing (NGS) as a readout, researchers can quantify the enrichment of millions of individual phage clones simultaneously and identify sequences that bind to the target or antibody of interest. This technology, known as Phage Immunoprecipitation and Sequencing (PhIP-Seq), has successfully identified autoantigens and mapped epitopes in inclusion body myositis, rheumatoid arthritis, two other common PNDs, the anti-Ma and anti-Ri syndromes (Larman et al. 2013). More recently, our group used PhIP-Seq and other approaches to uncover a new PND defined by immunoreactivity to kelch-like protein 11 (Mandel-Brehm et al. 2019).

Here, we report the design and utilization of a comprehensive PhIP-seq library of endogenous human peptides, including all known and predicted human protein splice variants. Using this expanded “peptidome” library, we conducted nearly 800 IPs over 130 serum and CSF samples from patients suffering from anti-Hu and anti-Yo PNDs, yielding the highest resolution epitope mapping of the autoantigens targeted in these diseases to date.

MATERIALS AND METHODS

Patient Samples

Patient samples were collected by the Mayo Clinic Neuroimmunology Laboratory for routine paraneoplastic autoantibody testing. Seropositive samples were stored for quality assurance purposes. Clinical data were not available for most seropositive patients. Mayo Clinic’s IRB and

Biospecimens Committee approved provision of sending de-identified samples to UCSF. For Mayo Clinic patients for whom a limited data set (age, sex and tumor type) was provided to UCSF, consent for medical chart review was provided. Patient ID, sex, age and diagnoses data are available in Supplemental Table 1.

Computational Design of Library

We downloaded all human sequences in the NCBI protein database (Nov 2015); including all splicing isoforms and computationally predicted coding regions. Full length sequences were clustered on 99% sequence identity (CD-HIT v4.6) (Li and Godzik 2006) to remove duplicate and partial sequences before computationally dividing them into 49 amino acid peptides using a 24 AA sliding window approach starting at the N terminus. As such, each peptide shared a 25-residue overlap with the one preceding it on the primary protein sequence. Peptides encoding the final 49-residues at the c-terminus were substituted when the final sliding window resulted in shortened or truncated sequences (proteins not evenly divisible by 49). The resulting peptides were further clustered/collapsed on 95% identity such that peptides with 2 or fewer amino acid differences were combined by randomly choosing one representative sequence.

Amino acids sequences were converted to nucleotides using preferred *E. coli* codons (Supplemental Table 2) and relevant restriction sites (EcoRI, HindIII, BamHI, XhoI) were removed with synonymous mutations to facilitate cloning. To enable amplification from synthesized oligo pools, 21 nt universal priming sites were added to the 5' and 3' ends of each oligo. These sequences also serve to encode STREP and FLAG tag sequences (after addition of terminal codons by PCR) that allow for downstream assessment of proper, in frame cloning. The resulting library consists of 731,724 unique 189 mer oligonucleotide sequences encoding 49 AA

human peptides flanked by 8 AA STREP and FLAG sequences after amplification with appropriate primers.

Immunoprecipitation

All liquid handling steps were carried out on Biomek FX robotics platform (Beckman Coulter, Brea, CA). 96-well, full skirted, low profile PCR plates (BioRad Inc, Hercules ,CA) were incubated overnight with 180 μ L IP blocking buffer (3% BSA, PBS-T) to prevent nonspecific binding. Blocking buffer was replaced with 150 μ L of phage library (10^{11} pfu in storage buffer) and mixed with 2 μ L patients CSF (undiluted), patient sera (diluted 1:50 in blocking buffer), or 2 ng commercial antibody and incubated O/N at 4°C. Experimental batches were defined by sample type (CSF or serum only), with individual patient samples assigned randomly to wells. Replicates were run in separate batches.

Protein A and G magnetic beads (Dynabeads – Invitrogen, Carlsbad, CA) were mixed equally, washed three times by magnetic separation and resuspension in equivalent volumes of TNP-40 wash buffer (150 mM NaCl, 50 mM Tris-HCL, 0.1% NP-40, pH 7.5). 10 μ L of A/G bead slurry was added to each IP reaction (using wide bore pipette tips) and incubated for 1 hour at 4°C. Bead-antibody complexes were washed three times by magnetic separation, removal of supernatant and resuspension in 150 μ L TNP-40 wash buffer. Subsequent to final wash, beads were re-suspended in 150 μ L chilled LB and used to inoculate fresh *E. coli* cultures (400 μ L at $OD_{600} = 0.5$) for *in vivo* amplification. Lysates were clarified by centrifugation at 3,000 g (at 4°C) and 150 μ L removed/stored for NGS library prep or additional rounds of IP.

Bioinformatic Analysis

Reads were quality filtered, paired-end reconciled (PEAR v0.9.8),(Zhang et al. 2014) and aligned to a reference database of the full library (bowtie2 v2.3.1) (Langmead and Salzberg 2012). Sequence alignment map (SAM) files were parsed using a suite of in-house analysis tools (Python/Pandas), and individual phage counts were normalized to reads per 100k (rpK) by dividing by the sum of counts and multiplying by 100,000. Peptide-level fold changes were calculated by dividing individual peptide rpK values with their expected abundance (mean rpK in mock IPs). For peptides never observed in any control IP, fold changes were calculated using the median rpK for all peptides derived from that gene in the mock IP.

We identified in our control IP's a set of the 100 most abundant peptides that also have a standard deviation less than their mean. This "internal control" set represented the most abundant and consistent phage carried along specifically by the protein-AG beads or other reagents, in absence of any antibody. We calculated a sample-specific scaling factor, defined as the ratio of median abundances (rpK) of these 100 peptides in our controls to their abundance in the given sample. Fold change values were multiplied by their sample-specific scaling factor.

To account for the bias introduced in the fold-change calculation stemming from under-represented peptides in the mock IP having inflated fold-changes despite only modest representation in an experimental IP, we normalized, within each experimental batch (defined as all samples on a given 96-well plate, never less than 50 samples), fold-change values using a correction factor defined as the solution to a linear regression fit to the fold-change values as a function of their mean $\log_2(\text{rpK})$ in AG-only controls (Supplementary Figure 1). Importantly, fold-change values were only corrected when the correction factor was positive (expected fold-change given the abundance in mock IPs was positive).

Statistical Analysis

We fit a normal distribution to the scaled and normalized fold-change values and assigned p-values by evaluating the survival function (1-CDF) at a given fold-change (Supplemental Figure 1D). To account for multiple tests, we calculated an adjusted p-value (p_{adj}) using the Benjamini-Hochberg procedure, enforcing a false discovery rate (FDR) of 0.05. Peptides were reported as significant with a positive fold-change and a $p_{\text{adj}} \leq 0.05$ in both technical replicates (independent IPs starting from single CSF or serum sample). Significance testing for associations between PND cohorts and healthy controls was performed using Fisher's Exact test with multiple testing correction (Benjamini-Hotchberg, $\text{FDR} \leq 0.05$).

ELAVL4 amplicon sequencing from bulk thymus and sorted human TECs

Sorted TECs were isolated from human thymus obtained from 18- to 22-gestational-week specimens under the guidelines of the University of California San Francisco Committee on Human Research. Tissue was washed, cut into small pieces using scissors, and gently mashed using the back of a syringe to extract thymocytes. To isolate TECs, remaining tissue pieces were digested at 37 °C using medium containing 100 $\mu\text{g/ml}$ DNase I (Roche, Belmont, CA) and 100 $\mu\text{g/ml}$ LiberaseTM (Sigma-Aldrich, St. Louis, MO) in RPMI. Fragments were triturated through a 5-ml pipette after 6 and 12 min to mechanically aid digestion. At 12 min, tubes were spun to pellet undigested fragments, and the supernatant was discarded. Fresh digestion medium was added to remaining fragments, and the digestion was repeated using a glass Pasteur pipette. Supernatant was again discarded. A third round of enzymatic digestion was performed using digestion medium supplemented with trypsin-EDTA for a final concentration of 0.05%. Remaining thymic fragments were digested for another 30 min, or until a single cell suspension was obtained. The

cells were moved to cold MACS buffer (0.5% BSA, 2 mM EDTA in PBS) to stop the enzymatic digestion. Following digestion, TECs were enriched by density centrifugation over a three-layer Percoll gradient with specific gravities of 1.115, 1.065 and 1.0. Stromal cells isolated from the Percoll-light fraction (between the 1.065 and 1.0 layers) were washed in MACS buffer. For surface staining, cells were blocked using a human Fc Receptor Binding Inhibitor Antibody (eBioscience) and incubated on ice for 30 min with the following antibodies:

PE anti-human Epcam (HEA-125) (Miltenyi 130-091-253) - dilution 1:50

Alexa Fluor 488 anti-human CD45 (HI30) (BioLegend 304017) - dilution 1:100

Epcam+ CD45- DAPI- cells were sorted using a FACS AriaII (BD Biosciences, San Jose, CA) directly into TRIzol LS (ThermoFisher Scientific, South San Francisco, CA). The aqueous phase was extracted using the manufacturer's instructions followed by RNA isolation using the RNeasy micro kit (QIAGEN, Redwood City, CA). RNA was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad, Emeryville, CA).

RNA samples from whole fetal and adult human thymus were purchased from Agilent (Santa Clara, CA) and Clontech (Mountain View, CA). RNA was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad, Emeryville, CA).

Amplicon libraries spanning the relevant exons were generated using the following primers:

ELAVL4_fwd: GAACCGATTACTGTGAAGTTTGCCAAC

ELAVL4_rev: GTAGACAAAGATGCACCACCCAGTTC

PCR products were purified using the QIAGEN Gel Extraction kit and library construction was conducted on the quantified DNA using the NEBNext® Ultra™ DNA Library Prep Kit (New England Biolabs, Ipswich, MA). The final library was sequenced using a MiSeq with single-end, 300-base reads.

Clinical PND Antibody Confirmation

Confirmation of anti-Yo and anti-Hu activity in patient samples was performed in the Mayo Clinic Neuroimmunology Laboratory with CLIA and New York State approved methodology. (Gadoth et al. 2017). Briefly, patients' sera and CSF were tested using cryosections of neural and non-neural murine tissues. Tissues were briefly fixed with 4% paraformaldehyde (1 minute), permeabilized with 0.5% CHAPS in PBS (1 minute) and 10% normal goat serum was applied for one hour (blocking of non-specific binding). Serum was preabsorbed with liver powder and applied at 1:240 dilutions; CSF was applied at 1:2 dilutions (doubling dilutions were used until endpoint). Secondary FITC-labeled antibodies used were from Southern Biotech (Birmingham, AL).

DATA AVAILABILITY

Sequencing data have been deposited in the NCBI SRA under BioProject **PRJNA507500**. Datasets Generated: Epitope mapping and antigen discovery by T7 phage display. Previously Published Datasets: Single-cell RNA-seq of murine medullary thymic epithelial cells (mTECs): Miragaia, R. J. et al., 2018, <https://www.ebi.ac.uk/ena/data/view/ERR1952294>, ERR1952294; Single cell RNAseq of electrophysiologically characterized neurons of the hippocampus: Földy, C. et al., 2016, <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP068021>, SRP068021; Single cell

isoform sequencing from whole mouse tissue: Gupta, I. et al., 2018, <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP128511>, SRP128511. Phage display library design files available at: <https://github.com/derisilab-ucsf/PhIP-PND-2018>

RESULTS

Design, Production, and Characterization of Human PhIP-Seq Library v2

The Human PhIP-Seq Library v2 contains all annotated human protein sequences from the NCBI protein database (November 2015), including all published and computationally predicted splice variants and coding regions (O’Leary et al. 2016) (Figure 1). Full-length sequences were clustered on 99% sequence identity (CD-HIT v4.6) (Li and Godzik 2006) to remove duplicate and partial entries, resulting in a set of 50,276 proteins. Each protein was computationally divided into 49 amino acid (AA) peptides using a 24 AA sliding window approach such that sequential peptides overlapped by 25-residues (see Methods). This resulted in a set of 1,256,684 peptide sequences encompassing the entire human proteome. To remove redundant sequences derived from identical regions of the various isoforms and homologs, we further clustered and collapsed the 49 AA peptides on 95% identity such that peptides with 2 or fewer AA differences were combined by choosing one representative sequence.

Amino acid sequences were converted to nucleotides using preferred *E. coli* codons (see Methods) (“Genetic Analysis of Pathogenic Bacteria: A Laboratory Manual.” n.d.). Restriction sites (EcoRI, HindIII, BamHI, XhoI) were removed with synonymous nucleotide changes to facilitate cloning. To enable amplification from synthesized oligo pools, 21 nucleotide (nt) universal priming sites were added to the 5’ and 3’ ends of each sequence. These priming sites also encoded STREP and FLAG tag sequences (after addition of a terminal codon by PCR). The

final library consisted of 731,724 unique 189 mer oligonucleotide sequences encoding 49 AA human peptides flanked by 5'-STREP and 3'-FLAG sequences. The library was synthesized by Agilent Technologies (Santa Clara, CA). The complete sequence of each oligo is available at <https://github.com/derisilab-ucsf/PhIP-PND-2018>.

After PCR amplification and size selection (Blue Pippin 3% Agarose cassette, Sage Science, Beverly, MA) but prior to cloning and packaging into phage, the synthesized oligos were evaluated by NGS with 70 million paired-end 125 nt reads (Figure 1B). 98.5% of all the synthesized sequences were full length, and 84% were error free. For those with errors, a majority (10% of all sequences) were deletions of 3 or fewer bases, while single base substitutions accounted for 4%, and the remaining 2% of oligos contained larger deletions (> 4 bases) or truncated/chimeric sequences. The commercially synthesized library yielded 722,436 unique peptide sequences with an estimated Chao1 diversity (Chao 1984) of 723,421, indicating that 99% of the library was present. The distribution and coverage of the library was uniform, with 99.9% of the sequences within ten-fold of the mean library proportion (Figure 1C).

Size-selected and restriction-digested oligos were cloned and packaged in 25 separate 10 μ L reactions according to the manufacturer's specifications (EMD, Burlington, MA). The efficiency of *in vitro* packaging was quantified by plaque assay. The diluted and quenched packaging reaction (3 ml total) contained 10^9 plaque forming units (pfu)/ml, roughly 1000x coverage of the entire library. The library of phage clones was sequenced (70 million paired-end 125 nt reads) to determine the fidelity of the packaged library. 77% of phage sequenced yielded error-free, full-sized inserts (Figure 1B). The majority of errors (21% of all sequences) were deletions or stop codons resulting in shorter expressed peptides. Within the 77% of correct sequences, there were 657,948 unique clones with an estimated Chao1 diversity of 658,476 –

indicating that 89% of this library was packaged and cloned without any errors. Allowing for synonymous mutations and single amino acid substitutions, 92% of the library was packaged successfully into phage.

Development of Statistical Approach to Analyze Peptide Enrichment

Several statistical approaches for analyzing PhIP-Seq data have been published, primarily based on Generalized Poisson (GP) or Negative Binomial (NB) count models fit to the distribution of unselected or input library phage populations (Xu et al. 2015)(Mohan et al. 2018). More recently, an algorithm leveraging z-scores of phage counts relative to AG bead only “mock” IPs was developed to account for non-specific phage binding to IP reagents and substrates (Yuan et al. 2018). These approaches were developed to analyze datasets generated from single rounds of immunoprecipitation and were optimized for sensitivity to low abundance phage.

In order to identify phage significantly enriched after multiple consecutive rounds of selection with patient antibodies, a statistical framework was developed (Methods) based on phage fold-change statistics relative to a collection of 40 independent mock IPs using protein-AG beads alone. Briefly, peptide counts from each control and sample IP were normalized to reads per 100,000 (rpK). Normalized peptide counts from each experimental IP were multiplied by a sample-specific scaling factor derived from the median rpK of the 100 most highly and consistently enriched AG-bead binding peptides (Supplemental Figure 1). Each peptide’s fold-change enrichment was then calculated by dividing its scaled rpK value by its mean representation in the AG controls. Given the nature of the fold-change calculation, small fluctuations in low abundance phage in control IPs can lead to inflated fold-change values and false positives. To address this, fold-change values were further transformed using a method inspired by smoothed quantile

normalization approaches in microarray data (see Methods) (Du, Kibbe, and Lin 2008). P-values were assigned to the fold-changes by evaluating the survival function of a normal distribution fit to the $\log_{10}(\text{fold-change})$ values in each experimental batch. Peptides reported as significant were those having an adjusted p-value (p_{adj}) ≤ 0.05 in at least two replicates after multiple test correction (Benjamini-Hochberg) (“Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR” n.d.).

Phage Library Performance Validation: Commercial Antibodies

Validation of our PhIP-Seq custom peptidome was performed using two commercial antibodies with known specificities: anti-glial fibrillary acid protein (GFAP) (Agilent/DAKO, Carpinteria, CA) and anti-gephyrin (GPHN) (Abcam, Burlingame, CA) polyclonal antibodies. Many commercial antibodies (to multiple antigens) were successfully screened and these two chosen as they represented both the most diverse and homogenous results with respect to peptide-level enrichment. Three rounds of enrichment were used to further select for the particular epitope, reduce requisite sequencing costs, and minimize the proportion of non-specific binders. After each round of enrichment, phage populations were assessed by sequencing to an average depth of 2 million paired-end 125 nt reads. After 3 rounds of enrichment, a majority (50-70%) of the phage in the final population encoded peptides derived from the commercial antibody target (Figure 2A). Replicate IPs were conducted on separate days and exhibited high reproducibility ($r > 0.8$) in phage counts between experiments.

When applied to these data, our statistical model correctly identified peptides from the target gene of interest as the most significantly enriched with minimal off-target or unrelated peptides/genes (Figure 2B). The anti-GFAP IPs enriched 18 unique peptides representing $>60\%$

of all phage in the final library (in both replicates) and scaled fold-changes greater than 300,000 x ($p_{\text{adj}} < 10^{-20}$) over the mock IPs. Similarly, the anti-GPHN commercial antibody enriched 4 unique C-terminal gephyrin peptides with 10,000-fold enrichment ($p_{\text{adj}} < 10^{-25}$) over control IPs.

Alignment of the significantly enriched peptides from the anti-GFAP IP to the full-length protein revealed a distinct antigenic profile (Figure 2C), suggesting the commercial antibodies had highest affinity for a ~27 AA region surrounding the exon 5-6 junction. Notably, this region was identified by the Immune Epitope Database (IEDB) tools Emini Surface Exposure calculation as having the highest B cell antigenicity potential (Emini et al. 1985). The three significantly enriched GPHN peptides were all overlapping sequences with consensus at the final 32 residues of the surface exposed C-terminus (Figure 1D). The anti-GPHN IPs also consistently enriched a single peptide from chromogranin A (CHGA) at levels greater than or commensurate with that of the GPHN peptides ($p_{\text{adj}} < 10^{-30}$). Motif analysis with GLAM2 (Gapped Local Alignment of Motifs) (Frith et al. 2008) and alignment of the CHGA and GPHN phage-peptides, revealed a discontinuous 6-residue sequence (YxE_xxK) shared between CHGA and GPHN. Of the 256 peptides reported as significant in both replicates, 144 peptides (56%), contained the motif (Figure 2D). The probability of recovering these motif-containing peptides by chance given their representation in the original library is infinitesimal ($<10^{-300}$, Fischer's exact test), indicating that a majority of the peptides in the final phage population were, in fact, true binders to the commercial antibody. These data also demonstrated that single AA resolution binding motifs could be recovered given sufficient representation in the original library.

PND Cohort Results

In total, 798 individual IPs (including all three rounds of enrichment) were performed on 130 CSF and serum samples from patients with anti-Yo or anti-Hu antibodies identified by the Mayo Clinic Neuroimmunology Laboratory with CLIA and New York State approved methodology, as previously described (see Methods) (Gadoth et al. 2017). The PhIP-Seq protocol was performed on a Biomek FX (Beckman Coulter, Brea, CA) automated liquid handler across three experimental batches (Methods). All samples were run blinded and in duplicate as technical replicates (fully independent IPs starting from single serum or CSF sample). A sample was reported as positive if peptides derived from the respective Yo (CDR2/CDR2L) or Hu (ELAVL2,3,4) antigens showed significant fold-changes ($p_{\text{adj}} \leq 0.05$) in both replicates.

Identification of CDR2/CDR2L antigens in anti-Yo sera and CSF

From 36 confirmed anti-Yo patients, a total of 53 samples were interrogated by PhIP-Seq, of which 51 (96%) were positive for anti-Yo peptides. When sample identities were unblinded, it was revealed that all 36 patients were represented by these positive samples. No significant differences were observed in sensitivity or binding profiles between CSF and serum, with 34/36 sera and 17/20 CSF testing positive ($p_{\text{chi-square}} = 0.23$). From 50 de-identified healthy donor sera, only a single CDR2L peptide from a single donor was significantly enriched (37 rpK, $p_{\text{adj}} = 0.0494$) reflecting the high specificity of anti-Hu and Yo antibodies in PND.

As shown in Figure 3A for three representative patients, phage specific for CDR2/CDR2L dominated the population by the final round of selection in a majority of samples. All patients showed enrichment for CDR2L while 30/36 (83%) were positive for both CDR2L and CDR2 peptides. No patients showed reactivity to CDR2 peptides alone. Across all patients, 29 unique

CDR2L peptides were identified and accounted for 21% of all phage sequenced. Conversely, 6 unique peptides from CDR2 were identified across all patient samples and accounted for 0.25% of all phage sequenced (Figure 3B), suggesting that the overall magnitude and complexity of peptide enrichment for CDR2L was greater than for CDR2 peptides.

The antigenic profile across the full length CDR2L protein, for all 36 patients, is shown in Figure 4A. Convergent epitopes were shared by subgroups of patients, particularly at residues 322-346, the most highly enriched epitope across all patients. In contrast to CDR2L, the antigenic profile of patient antibodies to CDR2 (Figure 4C) was primarily limited to 3 peptides spanning 100AA at the N-terminus. We further-examined the relationship between homology and enrichment (Figure 4B). Interestingly, the most highly enriched regions of CDR2L represented the most divergent regions between the two proteins, suggesting that the antibody responses to CDR2L and CDR2 are independent. This, combined with the observed differential enrichment between CDR2L and CDR2, support the notion that patient antibodies are, in fact, preferentially enriching for CDR2L over CDR2, consistent with a recent publication using cell-based assays (Kråkenes et al. 2019).

Identification of nELAVL antigens in anti-Hu sera and CSF

From 44 patients with laboratory confirmed anti-Hu PND, a total of 76 samples (32 paired serum/CSF, two CSF, and 10 serum) were subjected to PhIP-seq. A total of 19 samples from 13 patients resulted in significantly enriched peptides (in both replicates) derived from the nELAVL genes (ELAVL2, ELAVL3, and ELAVL4), the known targets of anti-Hu antibodies. Across these 19 samples, 20 unique nELAVL peptides were identified. While the majority (>80%) of enriched peptides were attributed to ELAVL4 (HuD). For the purpose of visualization, all nELAVL

peptides from each sample were aligned to the most common isoform of ELAVL4 (accession NP_001311142.1), shown in Figure 5A.

A majority (>90%) of the significantly enriched nELAVL peptides converged upon on a 17-residue sequence at AA positions 276-294 of ELAVL4, a sequence which is also common to variants of ELAVL2 and ELAVL3 (Figure 5B). The human peptidome PhIP-Seq library contained 18 peptides covering at least 10-residues of this short sequence, reflecting the large number of published isoforms in the NCBI Protein database for these highly spliced and studied proteins. Due to this redundancy, rpK and fold-change calculations were attenuated by as much as two orders of magnitude as individual peptide counts were distributed across multiple sequences containing the same epitope. To account for this, the analysis was re-run treating all 18 peptides containing this short sequence as a single entity, the anti-Hu “signature.” This resulted in the detection of an additional 8 samples from 4 patients yielding significant enrichment of the signature sequence. Notably, a single patient (patient 26) enriched only a single peptide in both serum and CSF derived from the N terminus of ELAVL2. The peptide is unique to ELAVL2 isoforms that had only been computationally predicted at the time of library design. These isoforms have since been confirmed experimentally and recognized as isoform c (Accession NP_001338384.1) (Berto et al. 2016). Importantly, none of 50 healthy, de-identified, patient sera samples yielded significant enrichment of nELAVL-derived peptides.

The convergent 17-residue signature sequence common to the majority of patient samples lies within a functionally significant hinge region between two RNA binding domains (RRM2 and RRM3). Expression of this region is under tight spatiotemporal control during fetal development and may contain nuclear localization and export sequences responsible for control of subcellular localization and neuronal differentiation (Wang et al. 2010). Though exon and isoform naming

conventions vary in the literature, this region is used to distinguish unique splice variants of ELAVL2, ELAVL3, and ELAVL4 (sv1, sv2, sv3) based on their inclusion or exclusion of relevant exons, specifically those commonly referred to as exons 6 and 7a. The identified motif terminates at the exon 6/7a junction (Figure 5B). The majority of patient samples preferentially enriched peptides that span the exon 6/7 junction, but do not include 7a, strongly suggesting that the binding determinants for the anti-Hu antibodies lie within this critical 17-residue anti-Hu signature region.

To further characterize the precise determinants of antibody binding, a new deep mutational scanning phage library was designed, encoding all possible single point mutants across the 17-residue signature motif. Samples from patients 01 and 38 were profiled by deep mutational scanning PhIP-seq, as they represented the samples with the most and least complex peptide representation at this region, respectively. At each position, the fold-change enrichment of each amino acid substitution against an AG-only IP and relative to the abundance of the reference sequence is shown in Figure 5C, providing a landscape of the mutational tolerance at each residue. Examination of mutations that result in STOP codons (denoted by * in figure) are particularly informative, revealing the minimal sequence length required for antibody binding. Patient 01 antibodies required a minimal epitope containing residues up to and including positions 245-255 (QAQRFRLDNLL) with a shorter subsequence (QRFRLDNLL) being least tolerant of mutation. Patient 038's antibodies appear to target a closely overlapping motif, RLDNLLN-AYG, with residues downstream not influencing peptide enrichment.

In light of this refined anti-Hu signature sequence, patient 01's original PhIP-Seq data was re-analyzed using GLAM2 to identify enriched motifs. Of the 36 peptides identified as significant in both replicates, 34 (94%) share a short, 7-residue motif, identical to the RLDxxLL identified by the mutational scanning approach (Figure 6). These include 16 peptides that were not derived from

nELAVL genes, demonstrating that the dominant epitope in patient 01 is present in many unrelated protein sequences and explains the majority of peptides derived from non-nELAVL enriched genes. The biological significance of promiscuous binding to unrelated proteins harboring this motif is unknown.

Consensus motif contains subsequences with predicted and experimentally demonstrated T-Cell Antigenicity

Examination of the refined anti-Hu signature motif and the full ELAVL4 sequence using the IEDB suite of sequence analysis tools (Supplemental Figure 2) suggests that the motif has low predicted probability of being a linear B cell epitope (Bepipred 2.0). The motif does, however, contain several sequences predicted to have high affinity for major histocompatibility complex type 1 (MHC-I) alleles (A1, A2, and A3 supertypes) and high prediction scores for MHC processing and presentation, with two predicted proteasomal cleavage sites flanking the region at amino acid positions 246 and 264. Interestingly, peptides derived from this region have previously been identified as anti-Hu T cell antigens (Rousseau et al. 2005). Indeed, peptides containing this exact motif have been shown to bind with high affinity to HLA-A1 (RLDNLLNMAY) and HLA-A2/B18 (NLLNMAYGV) and also to activate autologous CD8⁺ T cells obtained from Hu-positive patients (Roberts et al. 2009, 2).

For paraneoplastic autoimmune diseases, the question of how immune tolerance is broken in the absence of new mutations remains to be addressed adequately. Selective exclusion of exons during expression in the thymus has been previously shown in mouse models as a possible route to break tolerance (Klein et al. 2000). In the case of the anti-Hu ELAVL family, splice variants have been previously shown to be controlled by a concentration-dependent autoregulatory mechanism wherein ELAVL proteins bind to AU-rich elements on nascent transcripts, protecting

them from splicing and exclusion (Zaharieva et al. 2015). Therefore, it may be possible that protein levels in the thymus during the development of central tolerance might fail to reach the requisite levels, resulting in exclusion of certain exons which in turn could affect processing and presentation. To further investigate this speculation, publicly available RNAseq datasets derived from bulk and sorted murine thymus and thymic medullary epithelial cells (mTECs) were examined. While most datasets lacked sufficient nELAVL coverage necessary for analysis of splice variants, a single mTEC from a single cell RNAseq dataset contained >50X coverage across ELAVL4, revealing the exclusion of exon 7a (Supplemental Figure 2) (Miragaia et al. 2018). Comparatively, sorted murine hippocampal and cerebellar neurons show essentially even coverage across all exons (Földy et al. 2016)(Gupta et al. 2018).

To further investigate exon inclusion and exclusion of nELAVL splice variants during the development of central tolerance, amplicon libraries spanning exons 5, 6, 7a and 7 were generated from cDNA libraries derived from sorted human fetal and bulk adult thymic epithelial cells. An average of 1 million 125-nt paired-end reads (in triplicate) were obtained from each amplicon library. Exon 7a was represented in <0.5% of all reads in both fetal and adult TECs, supporting the notion that certain nELAV exons, such as exon 7a, are largely excluded from thymic expression. Interestingly, exon 7a is adjacent to exon 6 containing the critical anti-Hu signature motif. The impact of exclusion of exon 7a with respect to processing and thymic presentation of the motif present in exon 6 remains unknown, but taken together, these data suggest a potential mechanistic connection to central thymic tolerance whereby incomplete T cell tolerance to selective exons within nELAVL helps promote an autoantibody response against this region of the protein.

PhIP-Seq Identifies Additional Known and Potentially Novel Autoantigens in PND Samples

Antigens, other than those classically thought to be anti-Hu or anti-Yo were also investigated. Anti-CRMP5 (DPYSL5/CV-2) antibodies are a well-established clinical biomarker for SCLC and malignant thymoma and can be found together with anti-Hu antibodies in patients with a PND (Honnorat et al. 2009). Two of the anti-Hu patients (patients 51 and 52) had previously tested positive for anti-CRMP5 antibodies with a clinical assay (indirect immunofluorescence screening on mouse tissue substrate) (Yu et al. 2001). This was corroborated by PhIP-Seq (experimenters were blinded to previous diagnoses), with both patients' CSF samples yielding significant enrichment of CRMP5 peptides. Patient 51 showed enrichment of 3 overlapping sequences at the C-terminus of the protein (AA residues 481-564) while patient 52's CSF enriched for two peptides spanning residues 73-217. Though not contiguous on the primary sequence, these regions of the protein are in contact on the surface of the published three-dimensional structure (Figure 7).

Zic and Sox Family Enrichment

Among proteins that were significantly enriched, many have been previously identified as serological markers of cancer. Antibodies to ZIC (Zinc Fingers of the Cerebellum) and SOX (SRY-related HMG-box) families of transcription factors have a well-established clinical association with SCLC and PND (Titulaer et al. 2009)(Kazarian and Laird-Offringa 2011). Consistent with this fact, enrichment of peptides derived from the ZIC family (ZIC1, 2, 3, 4 and 5) was significant in 13/44 anti-Hu patients and 0/50 healthy controls ($p_{\text{Fisher's exact}} = 0.000017$) while 19/44 patients and 2/50 healthy controls tested positive for SOX family peptides ($p_{\text{Fisher's exact}} = 0.000004$). Though a majority of the peptides were derived from SOX1 and SOX2, the most common sequence,

present in 13/19 patients with SOX reactivity, was a 34-residue sequence common to SOX1, 2, 3, 14, and 21. ZIC and SOX were the two most highly correlated genes in the anti-Hu patient data, with 7/13 patients with ZIC reactivity also enriching for a concurrent SOX peptide.

DISCUSSION

Here, we designed and implemented an expanded PhIP-Seq library, encompassing the entire human proteome and including all NCBI predicted splice variants and isoforms. We used this new library to investigate patient CSF and serum samples from two common PNDs, the anti-Yo and anti-Hu syndromes. While the autoantigens in these PNDs have been previously identified, high-resolution epitope mapping afforded by PhIP-Seq or similar methods has, to our knowledge, not yet been achieved. In addition, we reasoned that the high degree of alternative splicing in these neuronal antigens would help us leverage the unique design of our library.

PhIP-Seq correctly identified the canonical anti-Yo antigens, CDR2 and CDR2L in 36/36 patients. Given this highly concordant result and the fact that the peptides bound in these experiments were derived from phage infection of *E. coli*, and thus lack canonical human post translational modifications (PTMs), these results suggests that anti-Yo PND antibodies present in the patient sera assayed here readily IP linear epitopes without a requirement for specific PTMs, or significant secondary or tertiary structure. In general, PhIP-Seq can be insensitive for detecting epitopes containing PTMs or conformational epitopes including some cell surface proteins like NMDA receptor subunits in patients with NMDA receptor encephalitis and aquaporin-4 in patients with neuromyelitis optica (data not shown).

CDR2L peptides were much more enriched than CDR2 peptides. Indeed, across all samples, phage expressing CDR2L peptides represented 21% of all phage sequenced, and these

peptides were represented in the top 5 most significantly enriched sequences in > 80% of samples tested. While these data do not provide a quantitative measurement of antibody affinity (especially to the native form *in vivo*), they suggest that CD2RL is the immunodominant antigen in patients suffering from anti-Yo PND. This also supports previous studies arguing that the antibody responses to each protein are independent (i.e., not cross-reactive) and that CDR2L is the primary antigen (Eichler et al. 2013). Further experiments, including cloning CDR2L and CDR2 reactive patient antibodies for deep characterization, will likely be required for further investigation.

Within the anti-Hu cohort, PhIP-Seq yielded significant enrichment of nELAVL peptides in 17/34 patients. Those patients that were positive for nELAVL peptides yielded a remarkably convergent antigenic signature, with antibodies binding a short 17-residue sequence (QAQRFRLDNLLNMAYGVK) shared by ELAVL2, 3 and 4 but absent in ELAVL1. This short sequence lies at the junction of exons 6 and 7a in ELAVL4, and high-resolution deep mutational scanning followed by motif analysis led to a further refined signature motif sequence (RLDNLLNMAY) as most deterministic for antibody binding. We note that this short sequence has been previously characterized as a T-cell epitope in anti-Hu patients (Rousseau et al. 2005)(Roberts et al. 2009). Additionally, this rare exon, limited to isoforms expressed during fetal development and in small populations of terminally differentiated neurons, demonstrates the utility of our expanded library encompassing all splice variants in the NCBI protein database. Regarding the anti-Hu patients that did not yield peptides from nELAVL genes, it may be the case that the autoantibodies in these patients require specific secondary and tertiary conformations, somatic mutations, or PTMs not emulated by our peptide library.

Interestingly, our analysis of publicly available RNAseq datasets and our own amplicon sequencing of human thymic epithelial cells reveals that exon 7a, immediately adjacent to our

identified motif, is absent in >99% of ELAVL4 transcripts in the thymus but abundant in specific neuronal subtypes and brain regions in mice. While investigation of exon exclusion in the thymus and the mechanisms underpinning autoimmunity are beyond the scope of this initial survey, we speculate that the absence of these short exons during the development of central tolerance could potentially influence the rise of autoreactive T-cells when the full-length protein is encountered in the CNS or expressed by peripheral tumors. Further, exclusion of exons bearing proteasomal cleavage sites (as with exon 7a) could bias MHC processing and presentation. Lineages of B cells with B cell receptors specific for ELAVL4 isoforms containing these excluded exons could presumably be maintained and expanded through interactions with CD4⁺ T cells specific for this short sequence when displayed on the surface of a B cell acting in its capacity as an antigen presenting cell (Klein et al. 2000)(Blachère et al. 2014).

Ultimately, as comprehensive, proteome-wide serological screens are more widely adopted, we expect the diagnostic, prognostic, and research potential of such assays to be fully realized. Indeed, the third most common category of autoantibody-mediated causes of encephalitis belonged to patients with unidentified autoantigens that have defied identification by more conventional methods (Dubey et al. 2018). Furthermore, high-resolution datasets characterizing the co-occurrence of specific antigens and antigenic determinants to specific diseases, symptoms, and clinical outcomes may ultimately serve to identify early biomarkers, stratify patient outcomes and inform treatment decisions. Finally, higher resolution epitope mapping and analysis of antigenic signatures across multiple patients should better inform the currently incomplete models describing onco-immunologic processes and the loss of self-tolerance underpinning the rise of PNDs and autoimmune disease in general.

FUNDING INFORMATION

Chan Zuckerberg Biohub: Brian O'Donovan, Caleigh Mandel-Brehm, Sara E Vazquez, Jamin Liu, Joseph L. DeRisi; Sandler Foundation: Brian O'Donovan, Caleigh Mandel-Brehm, Sara E Vazquez, Jamin Liu, Michael R. Wilson, Joseph L. DeRisi; Bioengineering Graduate Program, University of California, San Francisco (UCSF): Jamin Liu, T32GM008155-31; Medical Science Training Program (MSTP), UCSF: Sara E Vazquez, T32GM007618; Department of Health and Human Services (HHS) | National Institutes of Health (NIH) | National Institute of Neurological Disorders and Stroke (NINDS): Michael R. Wilson, K08NS096117; William K. Bowes, Jr. Foundation: Brian O'Donovan, Caleigh Mandel-Brehm, Sara E Vazquez, Jamin Liu, Michael R. Wilson, Joseph L. DeRisi; Bioinformatics Graduate Program, UCSF: Brian O'Donovan, T32GM067547; HHS | NIH | NINDS: Sean J Pittock, NS065829-01 The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

COMPETING INTERESTS

No competing interests declared.

FIGURE LEGENDS

FIGURE 1: Design and characterization of PhIP-Seq Library. A.) All human protein isoforms, variants, and computationally predicted coding regions were downloaded from the NCBI Protein database. Full length sequences were clustered on 99% sequence identity (CD-HIT v4.6) to remove duplicate and partial sequences. Each protein was computationally divided into 49 amino acid peptides using a 24 AA sliding window. Redundant or duplicate sequences were removed by further clustering on 95% sequence identity. The resulting 731,724 sequences were synthesized and cloned into the T7 select vector (Millipore, Burlington, MA). B.) Assessment of the library quality before (blue) and after (yellow) cloning and packaging.

FIGURE 2: Validation of library by commercial antibody IP's. Phage library was subjected to three rounds of selection by polyclonal commercial antibodies to GFAP and GPHN. A.) Replicates show high correlation of gene-level counts for IPs performed on separate days. Final libraries are dominated by the commercial antibody target. The GPHN antibody also consistently enriched a single peptide from chromogranin A (CHGA). B.) Scatterplots of peptide-level enrichments show 18 unique GFAP peptides and 4 GPHN peptides were identified as antibody binders (red markers). Peptides sharing a motif with both GPHN and CHGA (white, see D). C.) Alignment of GFAP peptides to the full-length protein reveals two distinct regions of antibody affinity corresponding to regions of high solvent exposure and B cell antigenicity as predicted by IEDB Tools Emini surface exposure and Bepipred algorithms. D.) Motif analysis of the significant peptides in the GPHN IP reveal a short, discontinuous motif shared by 144 significantly enriched peptides in the final population (from 94 unique genes/proteins), including the most abundant CHGA peptide. The motif is highlighted on the crystal structure of GPHN (top) and peptides sharing this motif are marked by a white circle in B.

Figure 3: A.) Representative data from 3 patients demonstrating the robust enrichment of CDR2/CDR2L peptides. CDR2/CDR2L peptides were the top 5 most abundant peptides in 44/51 samples. B.) While CDR2/CDR2L peptides represented 21% of all phage, CDR2L enrichment is more pronounced and complex (more unique peptides per sample) than CDR2.

FIGURE 4: Anti-Yo cohort PhIP-Seq Results. A.) Epitope map showing normalized coverage of each samples' significant CDR2L peptides aligned to the full-length gene (coverage is divided into 5 amino acid bins). Patient binding signatures vary, though a majority converge on residues 322-346. B.) Patient antibodies targeting CDR2L bind to regions of least homology with CDR2, suggesting the antibody responses to each protein are independent. C.) Epitope map showing normalized peptide coverage across the full length CDR2 sequence. Patient antibodies enrich for peptides primarily restricted to the first 100 amino acids at the N terminus. Abbreviations CDR2: cerebellar degeneration related protein 2, CDR2L: CDR2-like.

FIGURE 5: Anti-Hu PhIP-Seq Results. A.) Antibody binding signatures for samples that enriched nELAVL peptides. A majority of patients show a markedly convergent binding preference for a short 17-residue sequence at the exon 6/7a junction of ELAVL4. B.) Mean fold-change values for peptides spanning the unique regions defining ELAV4 splice variants and those

from ELAVL2 and ELAVL3, indicating lack of enrichment beyond the exon 6/7a junction. C.) Mutational scan using smaller phage library with all possible point mutations at each location spanning the motif highlighted in (B). Mutational tolerance is calculated for each position, defined as the fold-change over unselected (AG-bead only IP) relative to the reference sequence. Most telling are the mutations encoding stop codons (denoted *) that reveal patient 01's antibodies require a minimal sequence up to residues 255 (RLDDNLL) with the subsequence QRFRLDNLL least amenable to mutation. Patient 38's antibodies bind to residues up to and including position 260 (RLDNLLNMAYGV), with highest affinity for RLDNLLN-AYG.

FIGURE 6: Immunodominant nELAVL motif in Patient 01 CSF sample is that identified by mutational scanning. The short sequence, dominated by RLDxLL is present in 34 of the 36 peptides identified in both replicates. Apart from nELAVL proteins, the patient's antibodies enrich for peptides from 16 additional genes containing the motif. The biological consequences of such promiscuous autoantibody binding are unknown.

FIGURE 7: Anti-CRMP5 Antibody Epitope Mapping with PhIP-Seq. Two anti-Hu PND patient (patients 51 and 52) CSF samples also enriched CRMP5 peptides. CRMP5 is a well-established clinical biomarker for SCLC and malignant thymoma and anti-CRMP5 antibodies are often found concurrently with anti-Hu antibodies in patients with a PND. Though the peptides identified in each patient were discontinuous on the primary amino acid sequence, they are in close proximity in the 3D protein structure of CRMP5 at the region surrounding the C-terminus and the interface of the CRMP5 homodimer.

REFERENCES

- Beghetto, Elisa, and Nicola Gargano. 2013. "Antigen Discovery Using Whole-Genome Phage Display Libraries." *Methods in Molecular Biology (Clifton, N.J.)* 1061: 79–95. https://doi.org/10.1007/978-1-62703-589-7_4.
- Berto, Stefano, Noriyoshi Usui, Genevieve Konopka, and Brent L. Fogel. 2016. "ELAVL2-Regulated Transcriptional and Splicing Networks in Human Neurons Link Neurodevelopment and Autism." *Human Molecular Genetics* 25 (12): 2451–64. <https://doi.org/10.1093/hmg/ddw110>.
- Blachère, Nathalie E., Dana E. Orange, Bianca D. Santomasso, Jessica Doerner, Patricia K. Foo, Margaret Herre, John Fak, et al. 2014. "T Cells Targeting a Neuronal Paraneoplastic Antigen Mediate Tumor Rejection and Trigger CNS Autoimmunity with Humoral Activation." *European Journal of Immunology* 44 (11): 3240–51. <https://doi.org/10.1002/eji.201444624>.
- Burritt, J. B., M. T. Quinn, M. A. Jutila, C. W. Bond, and A. J. Jesaitis. 1995. "Topological Mapping of Neutrophil Cytochrome b Epitopes with Phage-Display Libraries." *The Journal of Biological Chemistry* 270 (28): 16974–80.
- Chao, Anne. 1984. "Nonparametric Estimation of the Number of Classes in a Population." *Scandinavian Journal of Statistics* 11 (4): 265–70.
- "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR." n.d. Accessed November 14, 2018. https://www.jstor.org/stable/2346101?seq=1#metadata_info_tab_contents.
- Du, Pan, Warren A. Kibbe, and Simon M. Lin. 2008. "Lumi: A Pipeline for Processing Illumina Microarray." *Bioinformatics (Oxford, England)* 24 (13): 1547–48. <https://doi.org/10.1093/bioinformatics/btn224>.
- Dubey, Divyanshu, Sean J. Pittock, Cecilia R. Kelly, Andrew McKeon, Alfonso Sebastian Lopez-Chiriboga, Vanda A. Lennon, Avi Gadoth, et al. 2018. "Autoimmune Encephalitis Epidemiology and a Comparison to Infectious Encephalitis." *Annals of Neurology* 83 (1): 166–77. <https://doi.org/10.1002/ana.25131>.
- Eichler, Tilo W., Cecilie Totland, Mette Haugen, Tor H. Qvale, Kibret Mazengia, Anette Storstein, Bjørn I. Haukanes, and Christian A. Vedeler. 2013. "CDR2L Antibodies: A New Player in Paraneoplastic Cerebellar Degeneration." Edited by Prasun Datta. *PLoS ONE* 8 (6): e66002. <https://doi.org/10.1371/journal.pone.0066002>.
- Emini, E A, J V Hughes, D S Perlow, and J Boger. 1985. "Induction of Hepatitis A Virus-Neutralizing Antibody by a Virus-Specific Synthetic Peptide." *Journal of Virology* 55 (3): 836–39.
- Földy, Csaba, Spyros Darmanis, Jason Aoto, Robert C. Malenka, Stephen R. Quake, and Thomas C. Südhof. 2016. "Single-Cell RNAseq Reveals Cell Adhesion Molecule Profiles in Electrophysiologically Defined Neurons." *Proceedings of the National Academy of Sciences of the United States of America* 113 (35): E5222–5231. <https://doi.org/10.1073/pnas.1610155113>.
- Frei, J. C., and J. R. Lai. 2016. "Protein and Antibody Engineering by Phage Display." *Methods in Enzymology* 580: 45–87. <https://doi.org/10.1016/bs.mie.2016.05.005>.

- Frith, Martin C., Neil F. W. Saunders, Bostjan Kobe, and Timothy L. Bailey. 2008. "Discovering Sequence Motifs with Arbitrary Insertions and Deletions." Edited by Gary Stormo. *PLoS Computational Biology* 4 (5): e1000071. <https://doi.org/10.1371/journal.pcbi.1000071>.
- Gadoth, Avi, Thomas J. Kryzer, Jim Fryer, Andrew McKeon, Vanda A. Lennon, and Sean J. Pittock. 2017. "Microtubule-Associated Protein 1B: Novel Paraneoplastic Biomarker." *Annals of Neurology* 81 (2): 266–77. <https://doi.org/10.1002/ana.24872>.
- "Genetic Analysis of Pathogenic Bacteria: A Laboratory Manual." n.d. Accessed November 14, 2018. <https://www.cabdirect.org/cabdirect/abstract/19972201503>.
- Gupta, Ishaan, Paul G. Collier, Bettina Haase, Ahmed Mahfouz, Anoushka Joglekar, Taylor Floyd, Frank Koopmans, et al. 2018. "Single-Cell Isoform RNA Sequencing Characterizes Isoforms in Thousands of Cerebellar Cells." *Nature Biotechnology*, October. <https://doi.org/10.1038/nbt.4259>.
- Hasadsri, Linda, James Lee, Bonnie H. Wang, Lalitha Yekkirala, and Mingtao Wang. 2013. "Anti-Yo Associated Paraneoplastic Cerebellar Degeneration in a Man with Large Cell Cancer of the Lung." *Case Reports in Neurological Medicine* 2013: 1–5. <https://doi.org/10.1155/2013/725936>.
- Honorat, J., S. Cartalat-Carel, D. Ricard, J. Ph Camdessanche, A. F. Carpentier, V. Rogemond, F. Chapuis, et al. 2009. "Onco-Neural Antibodies and Tumour Type Determine Survival and Neurological Symptoms in Paraneoplastic Neurological Syndromes with Hu or CV2/CRMP5 Antibodies." *Journal of Neurology, Neurosurgery, and Psychiatry* 80 (4): 412–16. <https://doi.org/10.1136/jnnp.2007.138016>.
- Kazarian, Meleeneh, and Ite A Laird-Offringa. 2011. "Small-Cell Lung Cancer-Associated Autoantibodies: Potential Applications to Cancer Diagnosis, Early Detection, and Therapy." *Molecular Cancer* 10 (1): 33. <https://doi.org/10.1186/1476-4598-10-33>.
- Klein, Ludger, Matthias Klugmann, Klaus-Armin Nave, V. K. Tuohy, and Bruno Kyewski. 2000. "Shaping of the Autoreactive T-Cell Repertoire by a Splice Variant of Self Protein Expressed in Thymic Epithelial Cells." *Nature Medicine* 6 (1): 56–61. <https://doi.org/10.1038/71540>.
- Kråkenes, Torbjørn, Ida Herdlevaer, Margrethe Raspotnig, Mette Haugen, Manja Schubert, and Christian A. Vedeler. 2019. "CDR2L Is the Major Yo Antibody Target in Paraneoplastic Cerebellar Degeneration." *Annals of Neurology* 86 (2): 316–21. <https://doi.org/10.1002/ana.25511>.
- Lancaster, Eric. 2016. "The Diagnosis and Treatment of Autoimmune Encephalitis." *Journal of Clinical Neurology (Seoul, Korea)* 12 (1): 1–13. <https://doi.org/10.3988/jcn.2016.12.1.1>.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Larman, H. Benjamin, Uri Laserson, Luis Querol, Katrijn Verhaeghen, Nicole L. Solimini, George Jing Xu, Paul L. Klarenbeek, et al. 2013. "PhIP-Seq Characterization of Autoantibodies from Patients with Multiple Sclerosis, Type 1 Diabetes and Rheumatoid Arthritis." *Journal of Autoimmunity* 43 (June): 1–9. <https://doi.org/10.1016/j.jaut.2013.01.013>.
- Larman, H. Benjamin, Zhenming Zhao, Uri Laserson, Mamie Z. Li, Alberto Ciccica, M. Angelica Martinez Gakidis, George M. Church, et al. 2011. "Autoantigen Discovery with a Synthetic Human Peptidome." *Nature Biotechnology* 29 (6): 535–41. <https://doi.org/10.1038/nbt.1856>.

- Li, Weizhong, and Adam Godzik. 2006. "Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences." *Bioinformatics (Oxford, England)* 22 (13): 1658–59. <https://doi.org/10.1093/bioinformatics/btl158>.
- Mandel-Brehm, Caleigh, Divyanshu Dubey, Thomas J. Kryzer, Brian D. O'Donovan, Baouyen Tran, Sara E. Vazquez, Hannah A. Sample, et al. 2019. "Kelch-like Protein 11 Antibodies in Seminoma-Associated Paraneoplastic Encephalitis." *New England Journal of Medicine* 381 (1): 47–54. <https://doi.org/10.1056/NEJMoa1816721>.
- Miragaia, Ricardo J., Xiuwei Zhang, Tomás Gomes, Valentine Svensson, Tomislav Illicic, Johan Henriksson, Gozde Kar, and Tapio Lönnberg. 2018. "Single-Cell RNA-Sequencing Resolves Self-Antigen Expression during MTEC Development." *Scientific Reports* 8 (1): 685. <https://doi.org/10.1038/s41598-017-19100-4>.
- Mohan, Divya, Daniel L. Wansley, Brandon M. Sie, Muhammad S. Noon, Alan N. Baer, Uri Laserson, and H. Benjamin Larman. 2018. "PhIP-Seq Characterization of Serum Antibodies Using Oligonucleotide-Encoded Peptidomes." *Nature Protocols* 13 (9): 1958. <https://doi.org/10.1038/s41596-018-0025-6>.
- O'Brien, T. J., B. Pasaliaris, A. D'Apice, and E. Byrne. 1995. "Anti-Yo Positive Paraneoplastic Cerebellar Degeneration: A Report of Three Cases and Review of the Literature." *Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia* 2 (4): 316–20.
- O'Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufu, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. "Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation." *Nucleic Acids Research* 44 (D1): D733–745. <https://doi.org/10.1093/nar/gkv1189>.
- Pascale, A., M. Amadio, and A. Quattrone. 2008. "Defining a Neuron: Neuronal ELAV Proteins." *Cellular and Molecular Life Sciences: CMLS* 65 (1): 128–40. <https://doi.org/10.1007/s00018-007-7017-y>.
- Pignolet, Béatrice SL, Christina MT Gebauer, and Roland S Liblau. 2013. "Immunopathogenesis of Paraneoplastic Neurological Syndromes Associated with Anti-Hu Antibodies: A Beneficial Antitumor Immune Response Going Awry." *Oncolmmunology* 2 (12): e27384. <https://doi.org/10.4161/onci.27384>.
- Roberts, Wendy K., Ilana J. Deluca, Ashby Thomas, John Fak, Travis Williams, Noreen Buckley, Athanasios G. Dousmanis, Jerome B. Posner, and Robert B. Darnell. 2009. "Patients with Lung Cancer and Paraneoplastic Hu Syndrome Harbor HuD-Specific Type 2 CD8+ T Cells." *Journal of Clinical Investigation*, June. <https://doi.org/10.1172/JCI36131>.
- Rosenfeld, Myrna R., and Josep Dalmau. 2012. "Paraneoplastic Neurologic Disorders: A Brief Overview." *Memo* 5 (3): 197–200. <https://doi.org/10.1007/s12254-012-0034-z>.
- Rousseau, A., B. Benyahia, J. Dalmau, F. Connan, J. -G. Guillet, J. -Y. Delattre, and J. Choppin. 2005. "T Cell Response to Hu-D Peptides in Patients with Anti-Hu Syndrome." *Journal of Neuro-Oncology* 71 (3): 231–36. <https://doi.org/10.1007/s11060-004-1723-1>.
- Senties-Madrid, H., and F. Vega-Boada. 2001. "Paraneoplastic Syndromes Associated with Anti-Hu Antibodies." *The Israel Medical Association Journal: IMAJ* 3 (2): 94–103.
- Titulaer, Maarten J., Rinse Klooster, Marko Potman, Lidia Sabater, Francesc Graus, Ingrid M. Hegeman, Peter E. Thijssen, et al. 2009. "SOX Antibodies in Small-Cell Lung Cancer and Lambert-Eaton Myasthenic Syndrome: Frequency and Relation with Survival." *Journal of*

- Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27 (26): 4260–67. <https://doi.org/10.1200/JCO.2008.20.6169>.
- Toothaker, Thomas B., and Michael Rubin. 2009. "Paraneoplastic Neurological Syndromes: A Review." *The Neurologist* 15 (1): 21–33. <https://doi.org/10.1097/NRL.0b013e3181870aa2>.
- Vita, Randi, James A. Overton, Jason A. Greenbaum, Julia Ponomarenko, Jason D. Clark, Jason R. Cantrell, Daniel K. Wheeler, et al. 2015. "The Immune Epitope Database (IEDB) 3.0." *Nucleic Acids Research* 43 (Database issue): D405–12. <https://doi.org/10.1093/nar/gku938>.
- Wang, Huiwen, Jill Molfenter, Hui Zhu, and Hua Lou. 2010. "Promotion of Exon 6 Inclusion in HuD Pre-mRNA by Hu Protein Family Members." *Nucleic Acids Research* 38 (11): 3760–70. <https://doi.org/10.1093/nar/gkq028>.
- Xu, George J., Tomasz Kula, Qikai Xu, Mamie Z. Li, Suzanne D. Vernon, Thumbi Ndung'u, Kiat Ruxrungtham, et al. 2015. "Comprehensive Serological Profiling of Human Populations Using a Synthetic Human Virome." *Science (New York, N.Y.)* 348 (6239): aaa0698. <https://doi.org/10.1126/science.aaa0698>.
- Yu, Z., T. J. Kryzer, G. E. Griesmann, K. Kim, E. E. Benarroch, and V. A. Lennon. 2001. "CRMP-5 Neuronal Autoantibody: Marker of Lung Cancer and Thymoma-Related Autoimmunity." *Annals of Neurology* 49 (2): 146–54.
- Yuan, Tiezheng, Divya Mohan, Uri Laserson, Ingo Ruczinski, Alan Baer, and H. Benjamin Larman. 2018. "Improved Analysis of Phage ImmunoPrecipitation Sequencing (PhIP-Seq) Data Using a Z-Score Algorithm," April. <https://doi.org/10.1101/285916>.
- Zaharieva, Emanuela, Irmgard U. Haussmann, Ulrike Bräuer, and Matthias Soller. 2015. "Concentration and Localization of Co-Expressed ELAV/Hu Proteins Control Specificity of mRNA Processing." *Molecular and Cellular Biology*, June, MCB.00473-15. <https://doi.org/10.1128/MCB.00473-15>.
- Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. "PEAR: A Fast and Accurate Illumina Paired-End ReAd MergeR." *Bioinformatics* 30 (5): 614–20. <https://doi.org/10.1093/bioinformatics/btt593>.

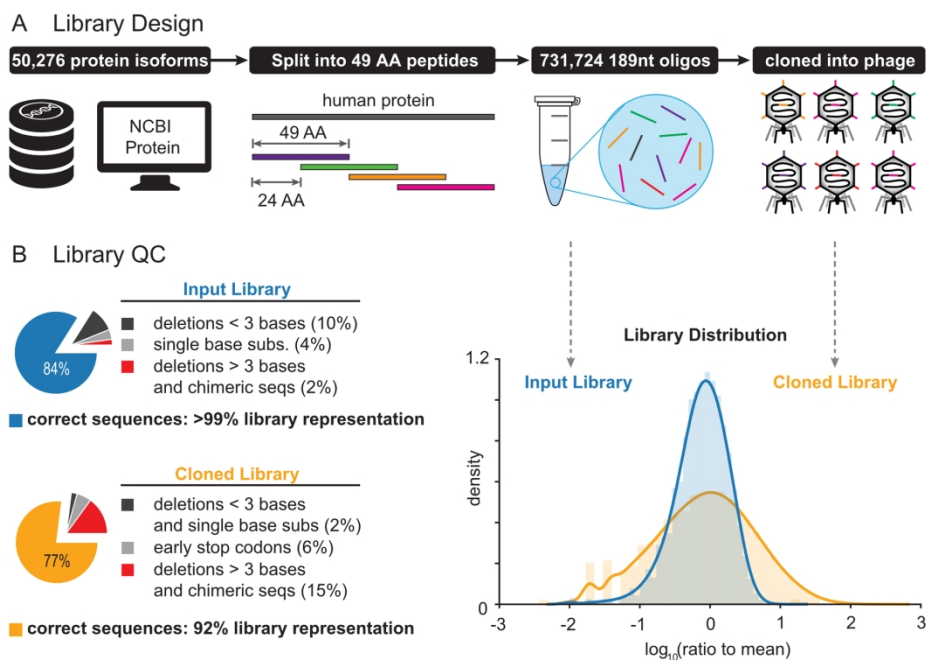


FIGURE 1: Design and characterization of PhIP-Seq Library. A.) All human protein isoforms, variants, and computationally predicted coding regions were downloaded from the NCBI Protein database. Full length sequences were clustered on 99% sequence identity (CD-HIT v4.6) to remove duplicate and partial sequences. Each protein was computationally divided into 49 amino acid peptides using a 24 AA sliding window. Redundant or duplicate sequences were removed by further clustering on 95% sequence identity. The resulting 731,724 sequences were synthesized and cloned into the T7 select vector (Millipore, Burlington, MA). B.) Assessment of the library quality before (blue) and after (yellow) cloning and packaging.

180x127mm (300 x 300 DPI)

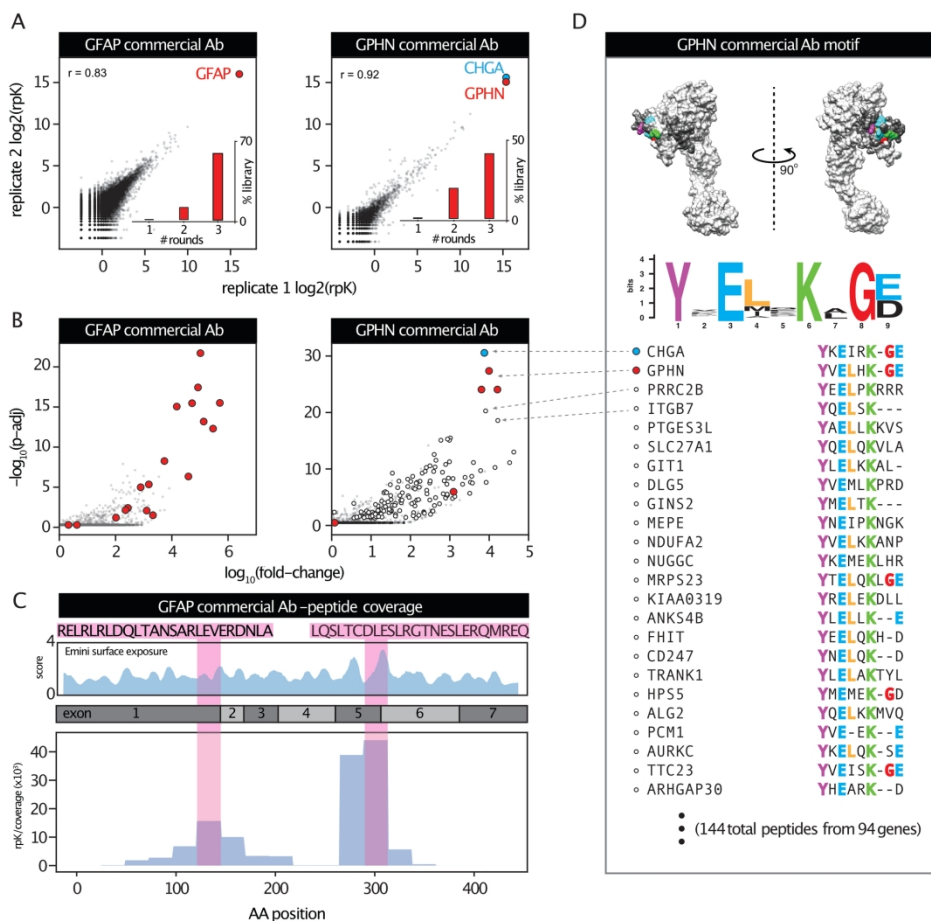


FIGURE 2: Validation of library by commercial antibody IP's. Phage library was subjected to three rounds of selection by polyclonal commercial antibodies to GFAP and GPHN. A.) Replicates show high correlation of gene-level counts for IPs performed on separate days. Final libraries are dominated by the commercial antibody target. The GPHN antibody also consistently enriched a single peptide from chromogranin A (CHGA). B.) Scatterplots of peptide-level enrichments show 18 unique GFAP peptides and 4 GPHN peptides were identified as antibody binders (red markers). Peptides sharing a motif with both GPHN and CHGA (white, see D). C.) Alignment of GFAP peptides to the full-length protein reveals two distinct regions of antibody affinity corresponding to regions of high solvent exposure and B cell antigenicity as predicted by IEDB Tools Emini surface exposure and Bepiped algorithms. D.) Motif analysis of the significant peptides in the GPHN IP reveal a short, discontinuous motif shared by 144 significantly enriched peptides in the final population (from 94 unique genes/proteins), including the most abundant CHGA peptide. The motif is highlighted on the crystal structure of GPHN (top) and peptides sharing this motif are marked by a white circle in B.

199x190mm (300 x 300 DPI)

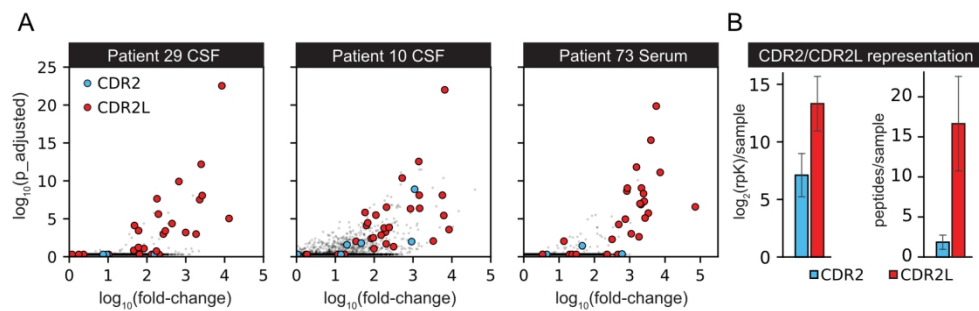


Figure 3: A.) Representative data from 3 patients demonstrating the robust enrichment of CDR2/CDR2L peptides. CDR2/CDR2L peptides were the top 5 most abundant peptides in 44/51 samples. B.) While CDR2/CDR2L peptides represented 21% of all phage, CDR2L enrichment is more pronounced and complex (more unique peptides per sample) than CDR2.

169x53mm (300 x 300 DPI)

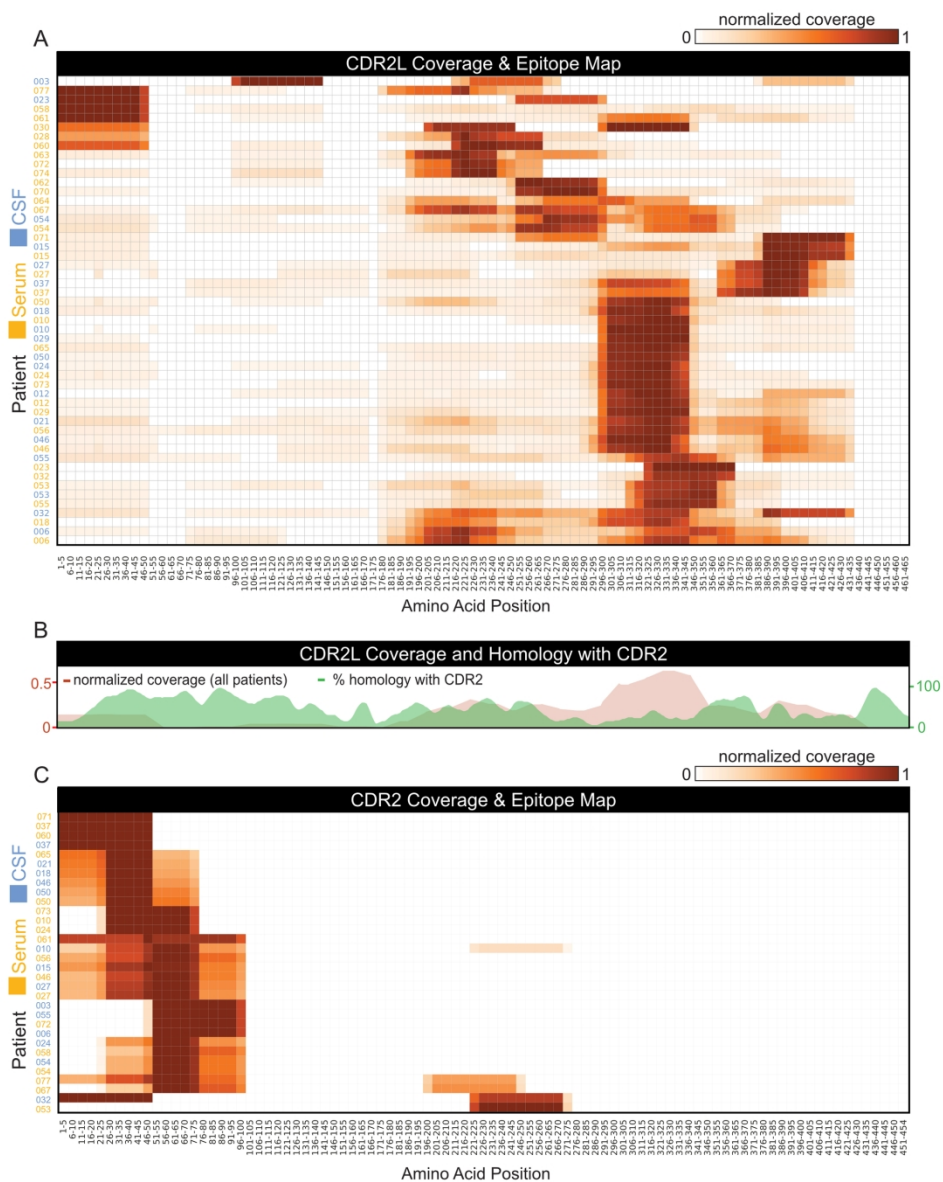


FIGURE 4: Anti-Yo cohort PhIP-Seq Results. A.) Epitope map showing normalized coverage of each samples' significant CDR2L peptides aligned to the full-length gene (coverage is divided into 5 amino acid bins). Patient binding signatures vary, though a majority converge on residues 322-346. B.) Patient antibodies targeting CDR2L bind to regions of least homology with CDR2, suggesting the antibody responses to each protein are independent. C.) Epitope map showing normalized peptide coverage across the full length CDR2 sequence. Patient antibodies enrich for peptides primarily restricted to the first 100 amino acids at the N terminus. Abbreviations CDR2: cerebellar degeneration related protein 2, CDR2L: CDR2-like.

180x225mm (300 x 300 DPI)

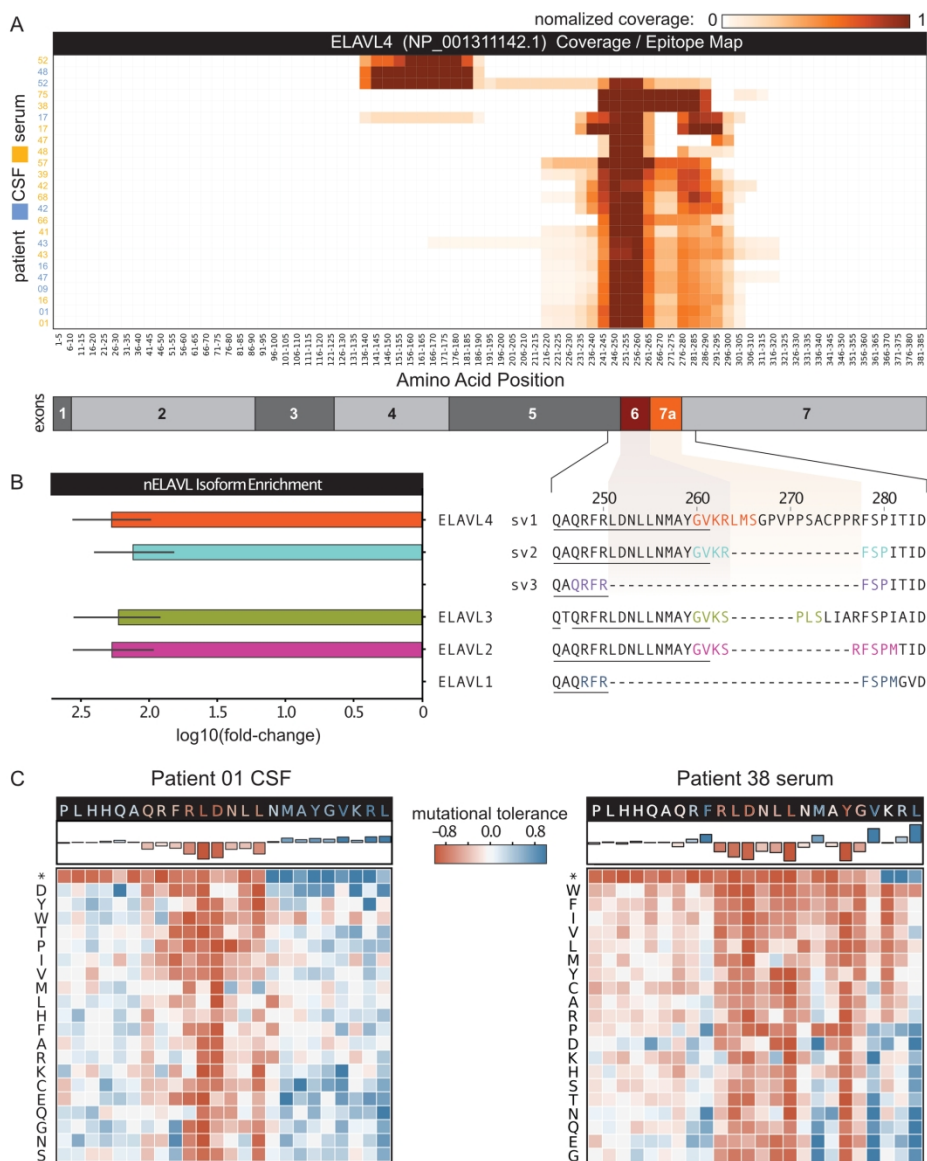


FIGURE 5: Anti-Hu PhIP-Seq Results. A.) Antibody binding signatures for samples that enriched nELAVL peptides. A majority of patients show a markedly convergent binding preference for a short 17-residue sequence at the exon 6/7a junction of ELAVL4. B.) Mean fold-change values for peptides spanning the unique regions defining ELAV4 splice variants and those from ELAVL2 and ELAVL3, indicating lack of enrichment beyond the exon 6/7a junction. C.) Mutational scan using smaller phage library with all possible point mutations at each location spanning the motif highlighted in (B). Mutational tolerance is calculated for each position, defined as the fold-change over unselected (AG-bead only IP) relative to the reference sequence. Most telling are the mutations encoding stop codons (denoted *) that reveal patient 01's antibodies require a minimal sequence up to residues 255 (RLDDNLL) with the subsequence QRFRDNLN least amenable to mutation. Patient 38's antibodies bind to residues up to and including position 260 (RLDNLNLMAYGV), with highest affinity for RLDNLLN-AYG.

193x244mm (300 x 300 DPI)

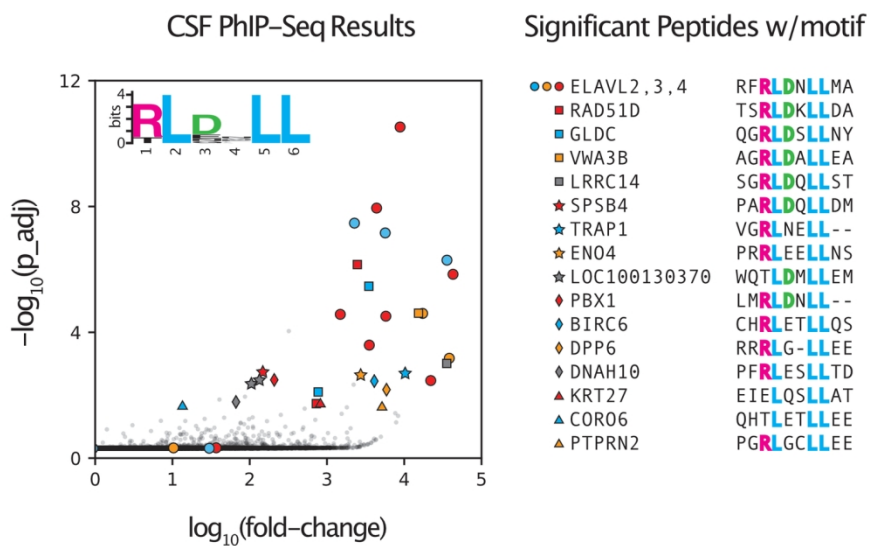


FIGURE 6: Immunodominant nELAVL motif in Patient 01 CSF sample is that identified by mutational scanning. The short sequence, dominated by RLDxLL is present in 34 of the 36 peptides identified in both replicates. Apart from nELAVL proteins, the patient's antibodies enrich for peptides from 16 additional genes containing the motif. The biological consequences of such promiscuous autoantibody binding are unknown.

147x95mm (300 x 300 DPI)

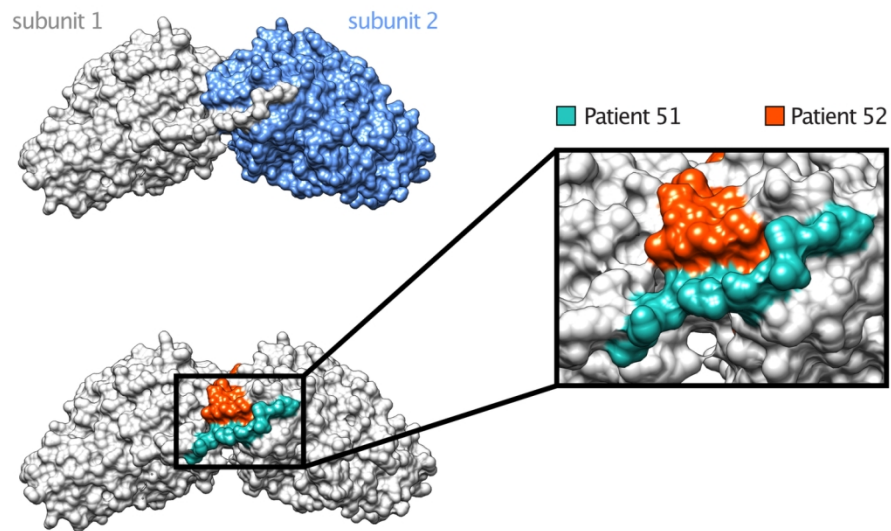
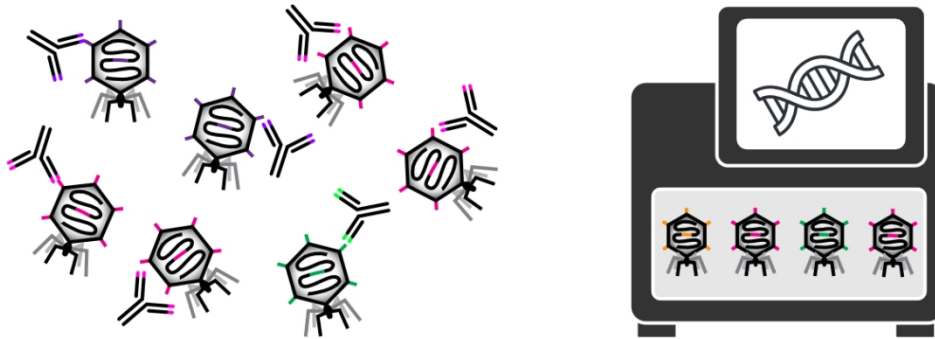


FIGURE 7: Anti-CRMP5 Antibody Epitope Mapping with PhIP-Seq. Two anti-Hu PND patient (patients 51 and 52) CSF samples also enriched CRMP5 peptides. CRMP5 is a well-established clinical biomarker for SCLC and malignant thymoma and anti-CRMP5 antibodies are often found concurrently with anti-Hu antibodies in patients with a PND. Though the peptides identified in each patient were discontinuous on the primary amino acid sequence, they are in close proximity in the 3D protein structure of CRMP5 at the region surrounding the C-terminus and the interface of the CRMP5 homodimer.

139x91mm (300 x 300 DPI)

Phage display and immunoprecipitation sequencing (PhIP-Seq) was used to screen the peptide binding affinities of 80 patients suffering from anti-Hu or anti-Yo paraneoplastic disease. Patient binding signatures implicate CDR2L as the primary antigen in anti-Yo paraneoplastic disease and suggest rare exons/isoforms play a role in the anti-Hu oncoimmune response.

Immunoprecipitate and sequence phage library encoding entire human peptidome



Generate high-resolution epitope maps

