BRAIN COMMUNICATIONS

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

Brian O'Donovan,[1] Caleigh Mandel-Brehm,[1] Sara E. Vazquez,[1] Jamin Liu,[1,2] Audrey V. Parent,[3] Mark S. Anderson,[3] Travis Kassimatis,[1] Anastasia Zekeridou,[4,5] Stephen L. Hauser,[6,7] Sean J. Pittock,[4,5] Eric Chow,[1] Michael R. Wilson[6,7] and Joseph L. DeRisi[1,8]

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 cerebrospinal fluid (CSF) samples from patients suffering from two common paraneoplastic neurological disorders, the anti-Yo ($n = 36$ patients) and anti-Hu ($n = 44$ patients) syndromes. 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 high-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.

1 Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA
2 UC Berkeley-UCSF Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA 94158, USA
3 Department of Medicine, Diabetes Center, University of California, San Francisco, San Francisco, CA 94143, USA
4 Department of Neurology, Mayo Clinic, Rochester, MN 55902, USA
5 Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55902, USA
6 Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA
7 Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA
8 Chan Zuckerberg Biohub, University of California, San Francisco, San Francisco, CA 94158, USA

Correspondence to: Joseph L. DeRisi, Dr, Department of Biochemistry and Biophysics
University of California, San Francisco, San Francisco, CA 94158, USA
E-mail: joe@derisilab.ucsf.edu

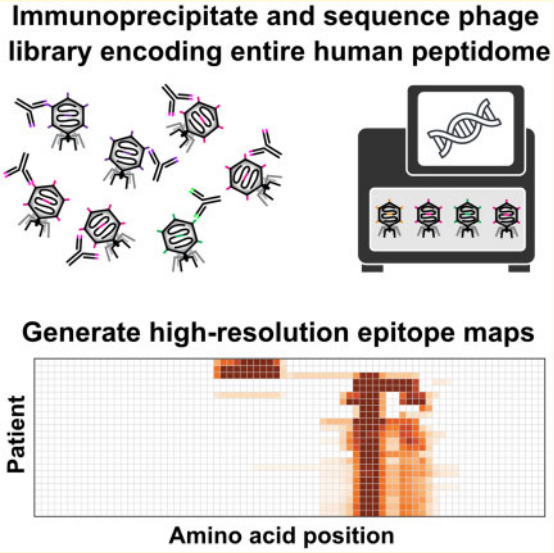
Received August 20, 2019. Revised March 08, 2020. Accepted March 12, 2020. Advance Access publication May 25, 2020
© The Author(s) (2020). Published by Oxford University Press on behalf of the Guarantors of Brain.
This is an Open Access article distributed under the terms of the Creative Commons Attribution 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

Correspondence may also be addressed to: Michael R. Wilson. E-mail: michael.wilson@ucsf.edu

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

Abbreviations: AA = amino acid; CDR2 = cerebellar, degeneration-related protein 2; CDR2L = cerebellar degeneration-related protein 2-like; ELAVL = embryonic lethal abnormal vision like; GPHN = gephyrin; nELAVL = neuronal embryonic lethal abnormal vision like; nt = nucleotide; PCA-1 = anti-Purkinje cell cytoplasmic antibody 1; PhIP-Seq = phage immunoprecipitation and sequencing; PND = paraneoplastic neurological disorder; rpK = reads per 100k

Graphical Abstract



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 *et al.*, 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 onconeuronal 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 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 (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 four 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 (neuronal ELAVL) proteins (Pascale *et al.*, 2008).

The majority of currently available clinical autoimmune panels include antigen-specific ELISA-based methods, cell-based assays, co-localization immunohistochemistry and traditional western blots (Lancaster, 2016). Unbiased and higher-throughput methods exist in the research context and include immunoprecipitation–mass spectrometry, expression library screening, biochemical purification, peptide arrays, protein arrays and phage display, each with its own advantages and limitations. Among these, variations in bacteriophage display technologies have been successfully utilized for decades in antigen discovery (Beghetto and Gargano, 2013), epitope mapping (Burritt

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 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 high-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 the 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 Supplementary Table 1.

Computational design of library

We downloaded all human sequences in the NCBI protein database (November 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 (AA) 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 two or fewer AA differences were combined by randomly choosing one representative sequence.

AA sequences were converted to nucleotides using preferred *Escherichia coli* codons (Supplementary 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-nucleotide (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 the 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, USA). The 96-well, full-skirted, low-profile PCR plates (BioRad Inc, Hercules, CA, USA) were incubated overnight with 180 µl immunoprecipitation (IP)-blocking buffer (3% bovine serum albumin, PBS-T) to prevent non-specific binding. Blocking buffer was replaced with 150 µl of phage library (10^{11} plaque-forming units in storage buffer), mixed with 2 µl patient CSF (undiluted), patient sera (diluted 1:50 in blocking buffer) or 2 ng commercial antibody and incubated overnight 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, USA) were mixed equally and 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). Ten microlitres of A/G bead slurry was added to each IP reaction (using wide bore pipette tips) and incubated for 1 h at 4°C. Bead–antibody complexes were washed three times by magnetic separation, removal of supernatant and resuspension in 150 µl TNP-40 washing buffer. Subsequent to final wash, beads were resuspended in 150 µl chilled LB and used to inoculate fresh *E. coli* cultures (400 µl at OD_{600} = 0.5) for *in vivo* amplification. Lysates were clarified by centrifugation at 3000 g (at 4°C) and 150 µl removed/stored for next-generation sequencing library preparation 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 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 IPs 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 the 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 <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 $\log_2$(rpK) in AG-only controls (Supplementary Fig. 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 (Supplementary Fig. 1D). To account for multiple tests, we calculated an adjusted *P*-value (P_{adj}) using the Benjamini–Hochberg procedure, enforcing a false discovery rate of 0.05. Peptides were reported as significant with a positive fold-change and a P_{adj} of $\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–Hochberg, false discovery rate $\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 µg/ml DNase I (Roche, Belmont, CA, USA) and 100 µg/ml Liberase™ (Sigma-Aldrich, St. Louis, MO, USA) 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% bovine serum albumin, 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 and

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

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

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

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

ELAVL4_fwd: GAACCGATTACTGTGAAGTTTGCCAAC and 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, USA). 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 min) and permeabilized with 0.5% CHAPS in PBS (1 min) and 10% normal goat serum was applied for 1 h (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, USA).

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: Miragaia *et al.*, 2018, https://www.ebi.ac.uk/ena/data/view/ERR1952294, ERR1952294; single-cell RNA-seq of electrophysiologically characterized neurons of the hippocampus: Földy *et al.*, 2016, https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP068021, SRP068021; single-cell isoform sequencing from whole mouse tissue: Gupta *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; Fig. 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-AA peptides using a 24-AA sliding window approach such that sequential peptides overlapped by 25 residues (see Materials and 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 two or fewer AA differences were combined by choosing one representative sequence.

AA sequences were converted to nucleotides using preferred *E. coli* codons (see Materials and methods; Maloy *et al.*, 1996). Restriction sites (EcoRI, HindIII, BamHI and XhoI) were removed with synonymous nucleotide changes 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 sequence. These priming sites also encoded STREP and FLAG tag sequences (after the 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. 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, USA) but prior to cloning and packaging into phage, the synthesized oligos were evaluated by next-generation sequencing with 70 million paired-end 125-nt reads (Fig. 1B). A total of 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 three 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 were uniform, with 99.9% of the sequences within 10-fold of the mean library proportion (Fig. 1C).

Size-selected and restriction-digested oligos were cloned and packaged in 25 separate 10-µl reactions according to the manufacturer's specifications (EMD, Burlington, MA, USA). 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/ml, roughly 1000× 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. A total of 77% of phage sequenced yielded error-free, full-sized inserts (Fig. 1B). The majority of errors (21% of all sequences) were deletions or stop codons resulting in shorter expressed peptides. Within the 77%

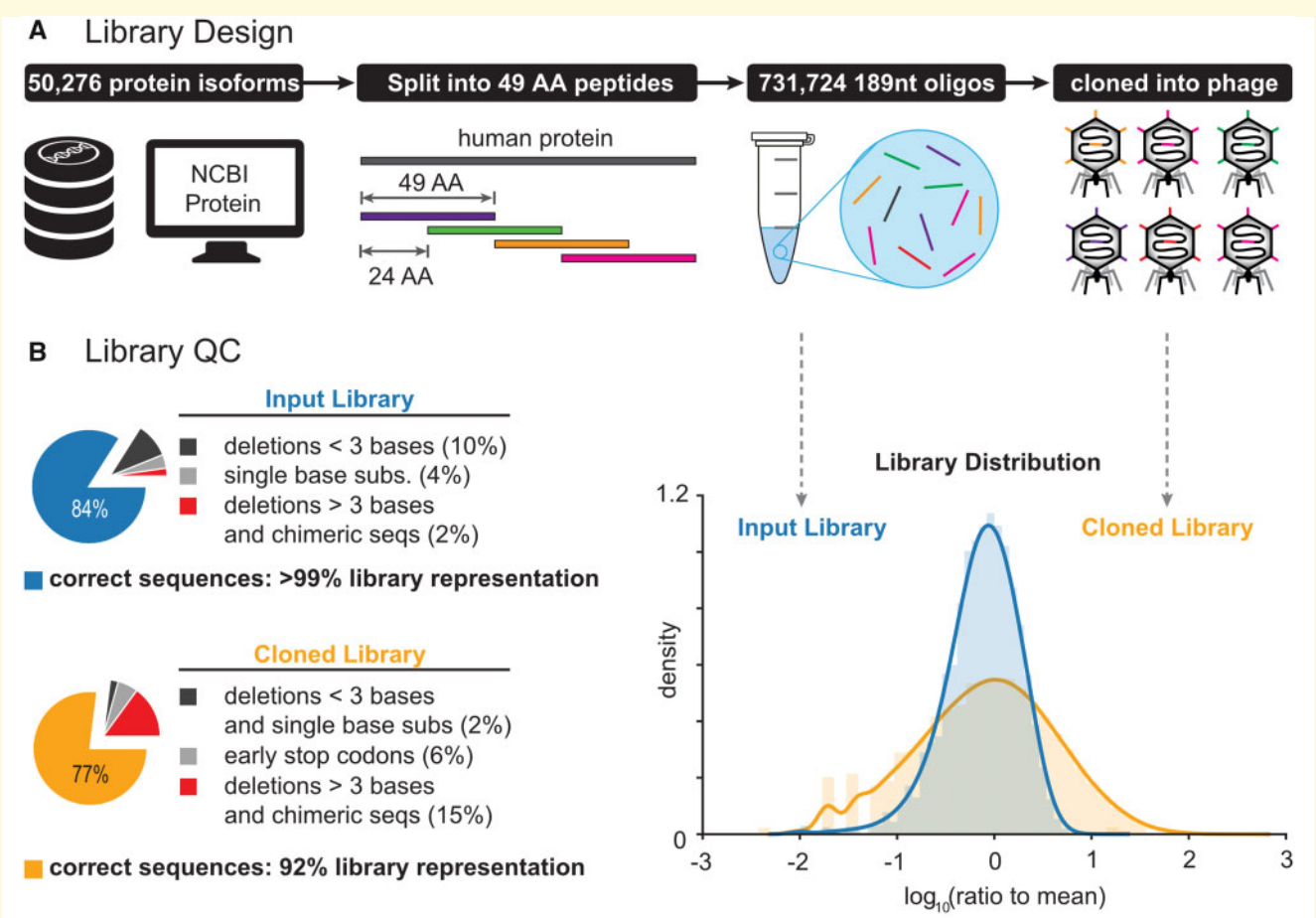


Figure 1 Design and characterization of PhIP-Seq library. (**A**) All human protein isoforms, variants and computationally predicted coding regions were downloaded from the NCBI Protein database. Full-length sequences were clustered on 99% sequence identity (CD-HIT v4.6) to remove duplicate and partial 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, USA). (**B**) Assessment of the library quality before (blue) and after (yellow) cloning and packaging.

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-AA 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 or negative binomial 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.

To identify phage significantly enriched after multiple consecutive rounds of selection with patient antibodies, a statistical framework was developed (Materials and 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 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 (Supplementary Fig. 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 Materials and methods; Du *et al.*, 2008). *P*-values were assigned to the fold-changes by evaluating the survival function of a normal distribution fit to the $\log_{10}$(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 and Hochberg, 1995).

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 (Agilent/DAKO, Carpinteria, CA, USA) and anti-gephyrin (GPHN) (Abcam, Burlingame, CA, USA) 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 three rounds of enrichment, a majority (50–70%) of the phage in the final population encoded peptides derived from the commercial antibody target (Fig. 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 (Fig. 2B). The anti-glial fibrillary acid protein IPs enriched 18 unique peptides representing >60% of all phage in the final library (in both replicates) and scaled fold-changes >300 000 × ($P_{adj} < 10^{-20}$) over the mock IPs. Similarly, the anti-GPHN commercial antibody enriched four unique C-terminal gephyrin peptides with 10 000-fold enrichment ($P_{adj} < 10^{-25}$) over control IPs.

Alignment of the significantly enriched peptides from the anti-glial fibrillary acid protein IP to the full-length protein revealed a distinct antigenic profile (Fig. 2C), suggesting that 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 tools Emini surface exposure calculation as having the highest B-cell antigenicity potential (Emini *et al.*, 1985; Vita *et al.*, 2015). The three significantly enriched GPHN peptides were all overlapping sequences with consensus at the final 32 residues of the surface-exposed C-terminus (Fig. 1D). The anti-GPHN IPs also consistently enriched a single peptide from chromogranin A at levels greater than or commensurate with that of the GPHN peptides ($P_{adj} < 10^{-30}$). Motif analysis with GLAM2 (Gapped Local Alignment of Motifs) (Frith *et al.*, 2008) and alignment of the chromogranin A and GPHN phage peptides revealed a discontinuous six-residue sequence (YxExxK) shared between chromogranin A and GPHN. Of the 256 peptides reported as significant in both replicates, 144 peptides (56%) contained the motif (Fig. 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

A total of 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 Materials and methods; Gadoth *et al.*, 2017). The PhIP-Seq protocol was performed on a Biomek FX automated liquid handler across three experimental batches (Materials and 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 ($P_{adj} \leq 0.05$) 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 ($P_{chi\text{-}square} = 0.23$). From 50 de-identified healthy donor sera, only a single CDR2L peptide from a single donor was significantly enriched (37 rpK, $P_{adj} = 0.0494$) reflecting the high specificity of anti-Hu and Yo antibodies in PND.

As shown in Fig. 3A for three representative patients, phage specific for CDR2/CDR2L dominated the population by the final round of selection in a majority of samples.

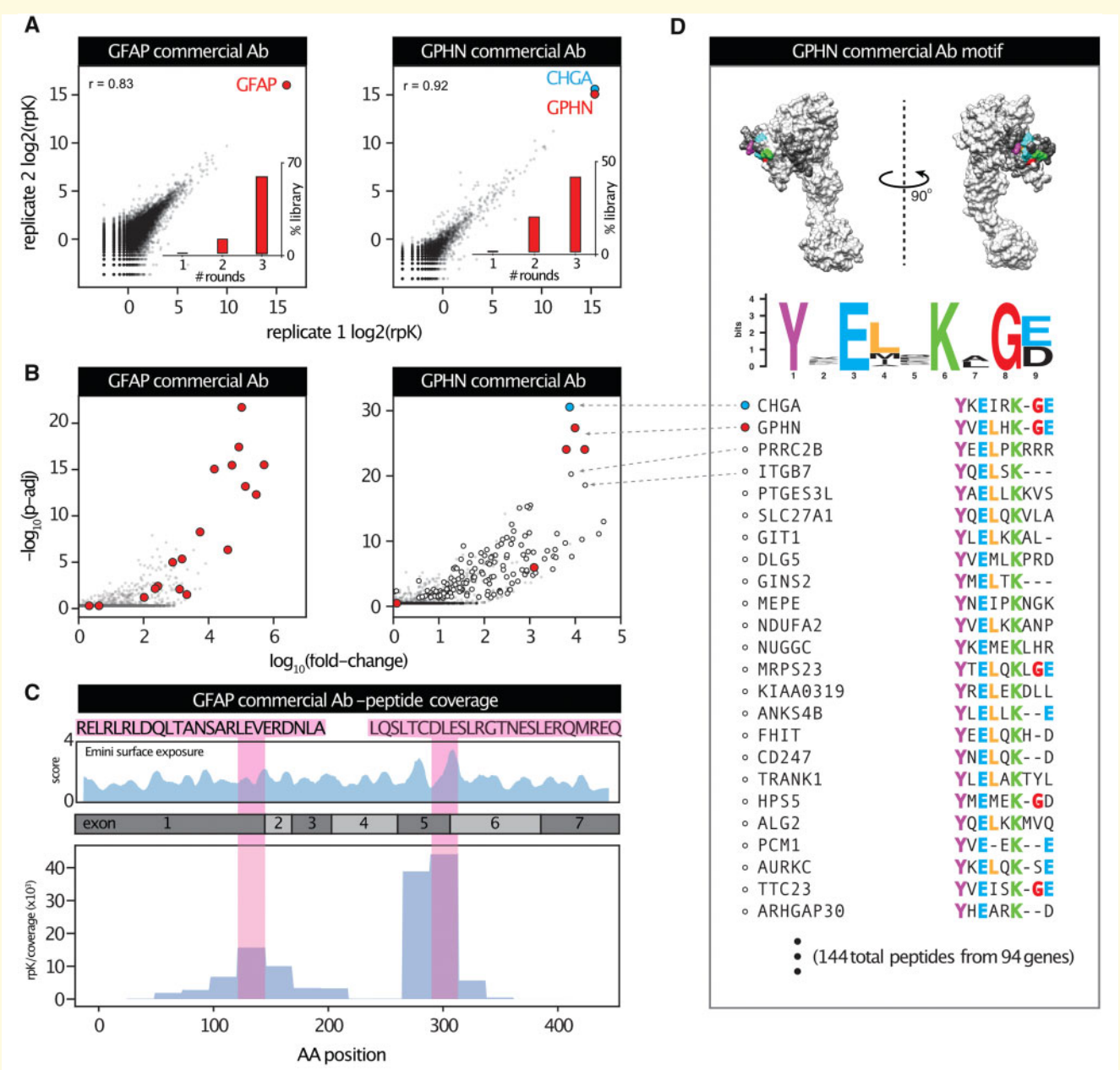


Figure 2 Validation of library by commercial antibody IPs. Phage library was subjected to three rounds of selection by polyclonal commercial antibodies to GFAP and GPHN. (**A**) Replicates show high correlation of gene-level counts for IPs performed on separate days. Final libraries are dominated by the commercial antibody target. The GPHN antibody also consistently enriched a single peptide from CHGA. (**B**) Scatterplots of peptide-level enrichments show that 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 reveals a short, discontiguous 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**. CHGA = chromogranin A; GFAP = glial fibrillary acid protein; IEDB = Immune Epitope Database.

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

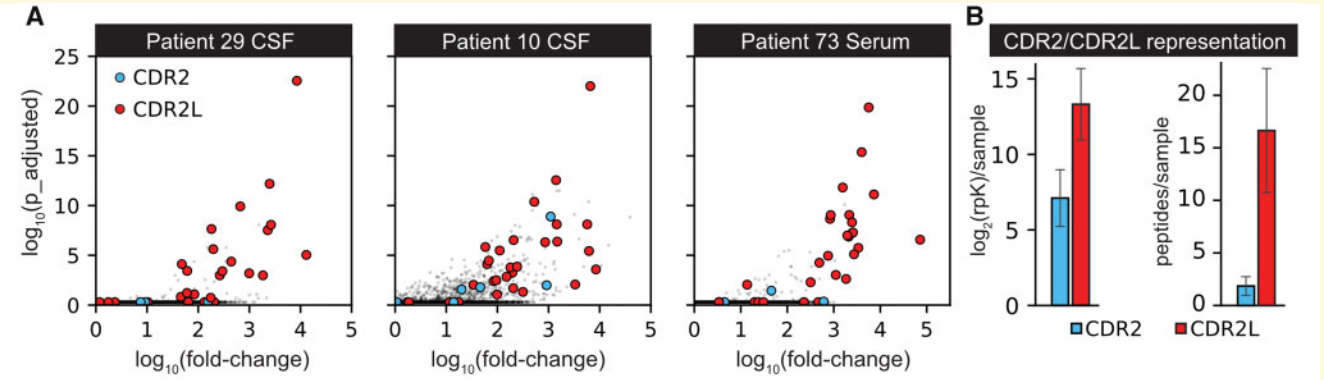


Figure 3 Specific enrichment of CDR2/CDR2L peptides by PhIP-Seq. (**A**) Representative data from three patients demonstrating the robust enrichment of CDR2/CDR2L peptides. CDR2/CDR2L peptides were the top five most abundant peptides in 44/51 samples. (**B**) While CDR2/CDR2L peptides represented 21%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

The antigenic profile across the full-length CDR2L protein, for all 36 patients, is shown in Fig. 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 (Fig. 4C) was primarily limited to three peptides spanning 100 AA at the N-terminus. We further examined the relationship between homology and enrichment (Fig. 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, supports 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, 2 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. 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), as shown in Fig. 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 (Fig. 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 rerun treating all 18 peptides containing this short sequence as a single entity, the anti-Hu 'signature'. This resulted in the detection of an additional eight samples from four 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

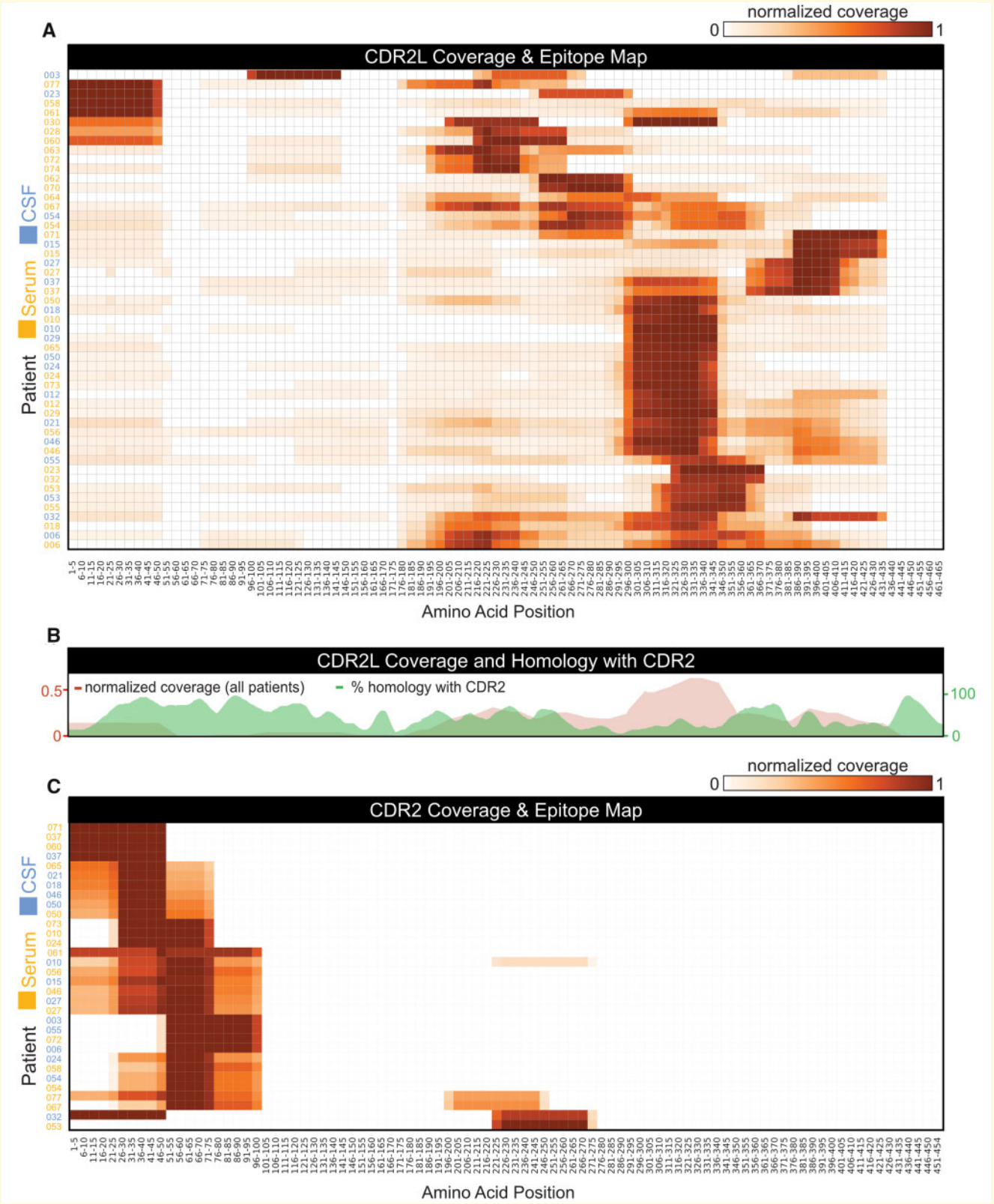


Figure 4 Anti-Yo cohort PhIP-Seq results. (**A**) Epitope map showing normalized coverage of each samples' significant CDR2L peptides aligned to the full-length gene (coverage is divided into five amino acid bins). Patient-binding signatures vary, though a majority converge on residues 322–346. (**B**) Patient antibodies targeting CDR2L bind to regions of least homology with CDR2, suggesting that 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. CDR2 = cerebellar degeneration-related protein 2; CDR2L = CDR2-like.

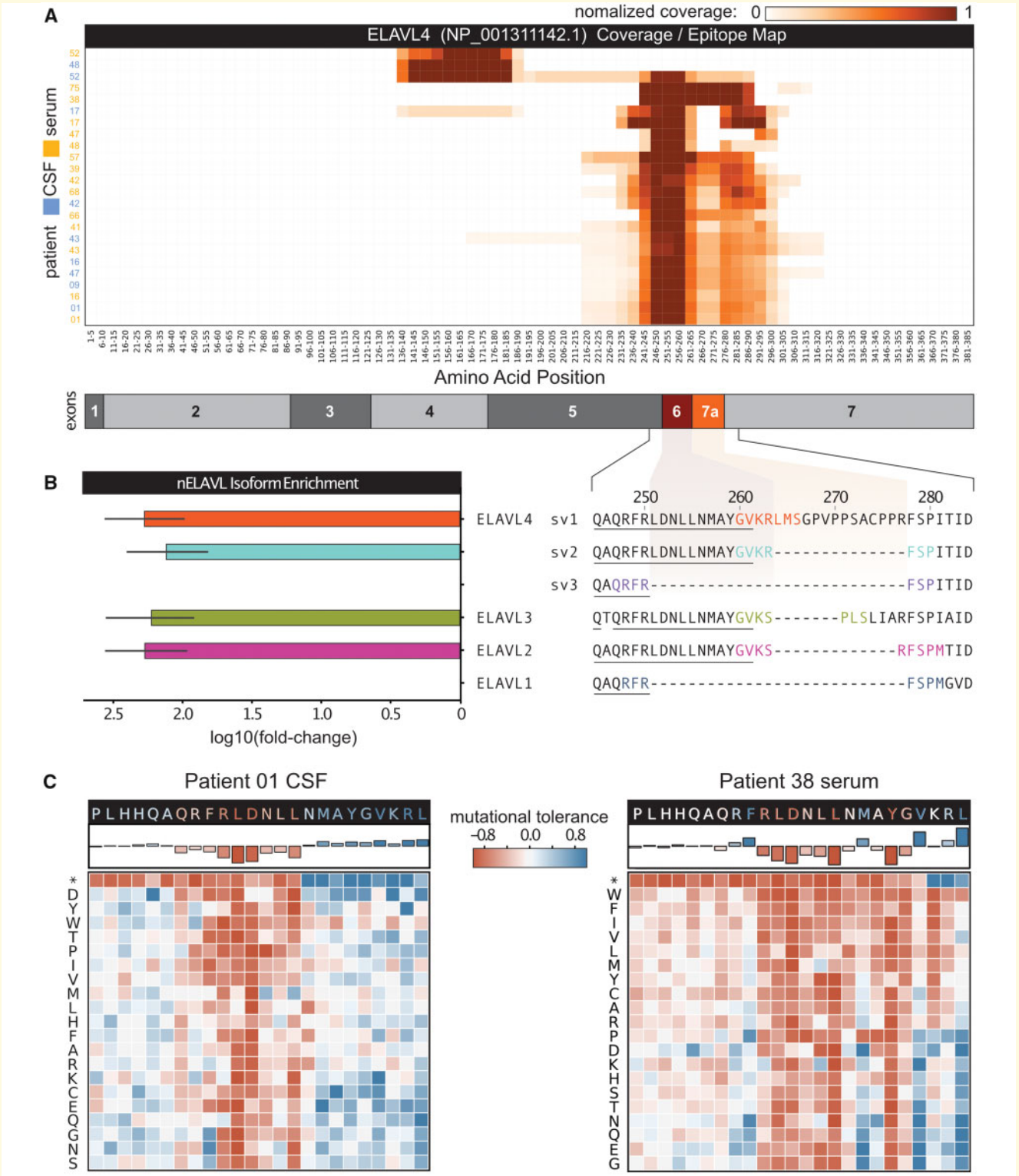


Figure 5 Anti-Hu PhIP-Seq results. (**A**) Antibody-binding signatures for samples that enriched nELAVL peptides. A majority of patients show a markedly convergent binding preference for a short 17-residue sequence at the exon 6/7a junction of ELAVL4. (**B**) Mean fold-change values for peptides spanning 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 *) revealing that 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.

junction (Fig. 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 AA substitution against an AG-only IP and relative to the abundance of the reference sequence is shown in Fig. 5C, providing a landscape of the mutational tolerance at each residue. Examination of mutations that results in STOP codons (denoted by * in figure) is 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 were re-analyzed using GLAM2 to identify enriched motifs. Of the 36 peptides identified as significant in both replicates, 34 (94%) share a short, seven-residue motif, identical to the RLDxxLL identified by the mutational scanning approach (Fig. 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.

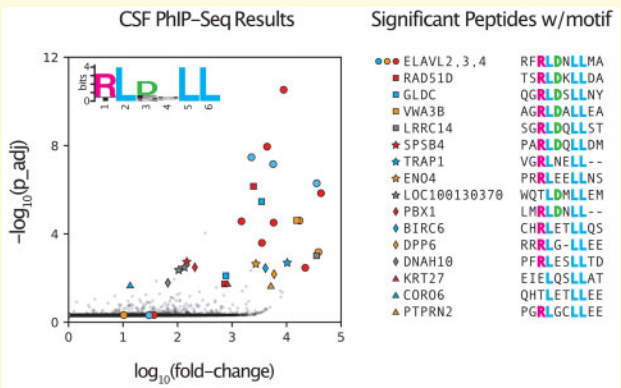


Figure 6 Immunodominant nELAVL motif in Patient 01 CSF sample is that identified by mutational scanning. The short sequence, dominated by RLDxLL, is present in 34 of the 36 peptides identified in both replicates. Apart from nELAVL proteins, the patient's antibodies enrich for peptides from 16 additional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

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 Immune Epitope Database suite of sequence analysis tools (Supplementary Fig. 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 AA 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 the exclusion of certain exons, which in turn could affect processing and presentation. To further investigate this speculation, publicly available RNA-seq 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 the analysis of splice variants, a single murine thymic medullary epithelial cell from a single-cell RNA-seq dataset contained >50X coverage across ELAVL4, revealing the exclusion of exon 7a (Supplementary Fig. 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) was 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 small-cell lung cancer and malignant thymoma and can be found together with anti-Hu antibodies in patients with a PND (Honnorat *et al.*, 2008). 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 three 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 (Fig. 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 small-cell lung cancer 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.

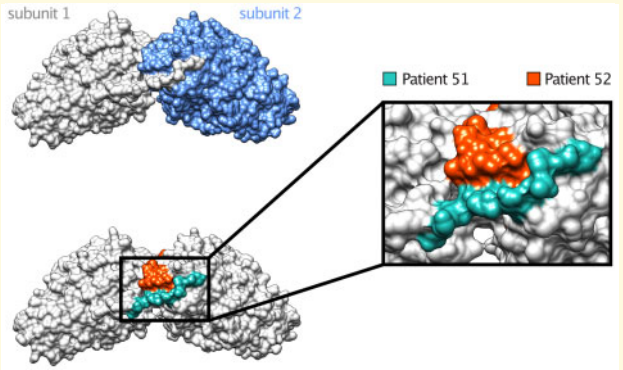


Figure 7 Anti-CRMP5 antibody epitope mapping with PhIP-Seq. Two anti-Hu PND patient (Patients 51 and 52) CSF samples also enriched CRMP5 peptides. CRMP5 is a well-established clinical biomarker for SCLC and malignant thymoma and anti-CRMP5 antibodies are often found concurrently with anti-Hu antibodies in patients with a PND. Though the peptides identified in each patient were discontiguous 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. SCLC = small-cell lung cancer.

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, these results suggest that anti-Yo PND antibodies present in the patient sera assayed here readily IP linear epitopes without a requirement for specific post-translational modifications, or significant secondary or tertiary structure. In general, PhIP-Seq can be insensitive for detecting epitopes containing post-translational modifications 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 five 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 who 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). In addition, 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 post-translational modifications not emulated by our peptide library.

Interestingly, our analysis of publicly available RNA-seq datasets and our own amplicon sequencing of human thymic epithelial cells reveal 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. Furthermore, 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, high-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.

Supplementary material

Supplementary material is available at *Brain Communications* online.

Funding

Chan Zuckerberg Biohub: Brian O'Donovan, Caleigh Mandel-Brehm, Sara E. Vazquez, Jamin Liu and Joseph L. DeRisi; Sandler Foundation: Brian O'Donovan, Caleigh Mandel-Brehm, Sara E. Vazquez, Jamin Liu, Michael R. Wilson and 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 and Joseph L. DeRisi; Bioinformatics Graduate Program, UCSF: Brian O'Donovan, T32GM067547; HHS|NIH|NINDS: Sean J. Pittock, NS065829-01. UCSF Marcus Program Transformative Integrated Research Award: Michael R. Wilson and Joseph L. DeRisi. 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.

References

Beghetto E, Gargano N. Antigen discovery using whole-genome phage display libraries. Methods Mol Biol 2013; 1061: 79–95.

Berto S, Usui N, Konopka G, Fogel BL. ELAVL2-regulated transcriptional and splicing networks in human neurons link neurodevelopment and autism. Hum Mol Genet 2016; 25: 2451–64.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 1995; 57: 289–300.

Blachère NE, Orange DE, Santomasso BD, Doerner J, Foo PK, Herre M, et al. T cells targeting a neuronal paraneoplastic antigen mediate tumor rejection and trigger CNS autoimmunity with humoral activation. Eur J Immunol 2014; 44: 3240–51.

Burritt JB, Quinn MT, Jutila MA, Bond CW, Jesaitis AJ. Topological mapping of neutrophil cytochrome b epitopes with phage-display libraries. J Biol Chem 1995; 270: 16974–80.

Chao A. Nonparametric estimation of the number of classes in a population. Scand J Stat 1984; 11: 265–70.

Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing illumina microarray. Bioinformatics 2008; 24: 1547–8.

Dubey D, Pittock SJ, Kelly CR, McKeon A, Lopez-Chiriboga AS, Lennon VA, et al. Autoimmune encephalitis epidemiology and a comparison to infectious encephalitis. Ann Neurol 2018; 83: 166–77.

Eichler TW, Totland C, Haugen M, Qvale TH, Mazengia K, Storstein A, et al. CDR2L antibodies: a new player in paraneoplastic cerebellar degeneration. PLoS One 2013; 8: e66002.

Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 1985; 55: 836–9.

Földy C, Darmanis S, Aoto J, Malenka RC, Quake SR, Südhof TC. Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc Natl Acad Sci USA 2016; 113: E5222–5231.

Frei JC, Lai JR. Protein and antibody engineering by phage display. Methods Enzymol 2016; 580: 45–87.

Frith MC, Saunders NFW, Kobe B, Bailey TL. Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput Biol 2008; 4: e1000071.

Gadoth A, Kryzer TJ, Fryer J, McKeon A, Lennon VA, Pittock SJ. Microtubule-associated protein 1B: novel paraneoplastic biomarker. Ann Neurol 2017; 81: 266–77.

Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol 2018; 36: 1197–202.

Hasadsri L, Lee J, Wang BH, Yekkirala L, Wang M. Anti-Yo associated paraneoplastic cerebellar degeneration in a man with large cell cancer of the lung. Case Rep Neurol Med 2013; 2013: 1–5.

Honnorat J, Cartalat-Carel S, Ricard D, Camdessanche JP, Carpentier AF, Rogemond V, et al. Onco-neural antibodies and tumour type determine survival and neurological symptoms in paraneoplastic neurological syndromes with Hu or CV2/CRMP5 antibodies. J Neurol Neurosurg Psychiatry 2008; 80: 412–6.

Kazarian M, Laird-Offringa IA. Small-cell lung cancer-associated autoantibodies: potential applications to cancer diagnosis, early detection, and therapy. Mol Cancer 2011; 10: 33.

Klein L, Klugmann M, Nave K-A, Tuohy VK, Kyewski B. Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. Nat Med 2000; 6: 56–61.

Kråkenes T, Herdlevær I, Raspotnig M, Haugen M, Schubert M, Vedeler CA. CDR2L is the major Yo antibody target in paraneoplastic cerebellar degeneration. Ann Neurol 2019; 86: 316–21.

Lancaster E. The diagnosis and treatment of autoimmune encephalitis. J Clin Neurol 2016; 12: 1–13.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012; 9: 357–9.

Larman HB, Laserson U, Querol L, Verhaeghen K, Solimini NL, Xu GJ, et al. PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. J Autoimmun 2013; 43: 1–9.

Larman HB, Benjamin Z, Zhao U, Laserson MZ, Li A, Ciccia MA, Gakidis Martinez, et al. Autoantigen discovery with a synthetic human peptidome. Nat Biotechnol 2011; 29: 535–41.

Li W, Godzik A. Cd-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006; 22: 1658–9.

Maloy SR, Stewart VJ, Taylor RK. Genetic analysis of pathogenic bacteria: a laboratory manual. Cold Spring Harbor Laboratory Press; 1996.

Mandel-Brehm C, Dubey D, Kryzer TJ, O'Donovan BD, Tran B, Vazquez SE, et al. Kelch-like protein 11 antibodies in seminoma-associated paraneoplastic encephalitis. N Engl J Med 2019; 381: 47–54.

Miragaia RJ, Zhang X, Gomes T, Svensson V, Ilicic T, Henriksson J, et al. Single-cell RNA-sequencing resolves self-antigen expression during MTEC development. Sci Rep 2018; 8: 685.

Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. Nat Protoc 2018; 13: 1958–78.

O'Brien TJ, Pasaliaris B, D'Apice A, Byrne E. Anti-Yo positive paraneoplastic cerebellar degeneration: a report of three cases and review of the literature. J Clin Neurosci 1995; 2: 316–20.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 2016; 44: D733–745.

Pascale A, Amadio M, Quattrone A. Defining a neuron: neuronal ELAV proteins. Cell Mol Life Sci 2008; 65: 128–40.

Pignolet BS, Gebauer CM, Liblau RS. Immunopathogenesis of paraneoplastic neurological syndromes associated with anti-Hu antibodies: a beneficial antitumor immune response going awry. OncoImmunology 2013; 2: e27384.

Roberts WK, Ilana J, Deluca A, Thomas J, Fak T, Williams N, Buckley, et al. Patients with lung cancer and paraneoplastic Hu syndrome harbor HuD-specific type 2 CD8+ T cells. J Clin Investig 2009; 119: 2042–51.

Rosenfeld MR, Dalmau J. Paraneoplastic neurologic disorders: a brief overview. Memo 2012; 5: 197–200.

Rousseau A, Benyahia B, Dalmau J, Connan F, Guillet J-G, Delattre J-Y, et al. T cell response to Hu-D peptides in patients with anti-Hu syndrome. J Neurooncol 2005; 71: 231–6.

Senties-Madrid H, Vega-Boada F. Paraneoplastic syndromes associated with anti-Hu antibodies. Isr Med Assoc J 2001; 3: 94–103.

Titulaer MJ, Klooster R, Potman M, Sabater L, Graus F, Hegeman IM, et al. SOX antibodies in small-cell lung cancer and Lambert-

Eaton myasthenic syndrome: frequency and relation with survival. JCO 2009; 27: 4260–7.

Toothaker TB, Rubin M. Paraneoplastic neurological syndromes: a review. Neurol 2009; 15: 21–33.

Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The Immune Epitope Database (IEDB) 3.0. Nucleic Acids Res 2015; 43: D405–12.

Wang H, Molfenter J, Zhu H, Lou H. Promotion of exon 6 inclusion in HuD Pre-MRNA by Hu protein family members. Nucleic Acids Res 2010; 38: 3760–70.

Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, et al. Comprehensive serological profiling of human populations using a synthetic human virome. Science 2015; 348: aaa0698–aaa0698.

Yu Z, Kryzer TJ, Griesmann GE, Kim K, Benarroch EE, Lennon VA. CRMP-5 neuronal autoantibody: marker of lung cancer and thymoma-related autoimmunity. Ann Neurol 2001; 49: 146–54.

Yuan T, Mohan D, Laserson U, Ruczinski I, Baer A, Benjamin Larman H. Improved analysis of Phage ImmunoPrecipitation Sequencing (PhIP-Seq) data using a Z-score algorithm. 2018. https://doi.org/10.1101/285916.

Zaharieva E, Haussmann IU, Bräuer U, Soller M. Concentration and localization of co-expressed ELAV/Hu proteins control specificity of MRNA processing. Mol Cell Biol 2015; 35: 3104–15.

Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate illumina paired-end ReAd MergeR. Bioinformatics 2014; 30: 614–20.