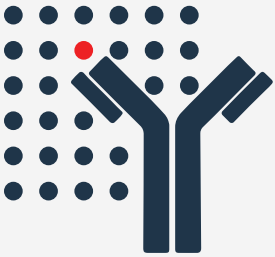




STUDY REPORT



HuScan® Report

Human Proteome Phage Display ImmunoPrecipitation Sequencing (PhIP-Seq)

Human Proteome Phage Display Immunoprecipitation-Sequencing (PhIP-Seq)

HuScan + Peptide Hormone Study Report

Prepared For:

Vietnam

October 19, 2025

Contents

1	Overview of PhIP-Seq	3
2	Library Creation	3
3	Experimental Overview	4
4	Informatics Methods	5
5	Core Pipeline Output Files	5
6	QC Figures from Counts File	7
7	Results Summary Figures from edgeR Enriched Hits . .	11
8	GFAP Positive Control Heatmaps	13
9	Processed Hits Pipeline Output Files	14
10	Fisher Exact Tests	15
11	Peptide-Level Tests	15
12	Protein-Level Tests	22
13	Protein-Level Polyclonal Tests	26
14	Summary Heatmaps: Top Public	30
15	Summary Heatmaps: Fisher Exact Tests	31
16	Epitope Scanning Heatmaps	32
17	Individual Sample Barplots	33
18	References	34

1 Overview of PhIP-Seq

Bacteriophage Display Immunoprecipitation-Sequencing (PhIP-Seq) is a powerful method of multiplexed seromics analysis that combines synthetic antigen libraries with an antibody-capturing pulldown and a high-throughput DNA sequencing readout. The first human autoantigen PhIP-Seq library was developed in 2011 with a viral library first reported in 2014.^{1,2}

There are many library and methods variations of PhIP-Seq. The data presented in this document were generated using a well-established human proteome library according to methods from reported protocols.^{3,4} It is enriched with an additional human peptide hormone and drugs database.⁵ All included peptide hormones have at least five redundant copies with alternative DNA sequences generating identical peptides; for hormones shorter than 33 amino acids, multiple display padding strategies are attempted.

2 Library Creation

An *E. coli* codon-optimized library consisting of >700,000 oligonucleotides encoding 49-mer amino acid peptide inserts with 25-mer amino acid adjacent overlaps was prepared as-described³; this library covers an entire RefSeq human proteome. The resulting oligonucleotides were synthesized on a releasable microarray, PCR amplified using appended adapter sequences, and cloned en masse into a modified T7-Select vector. This vector displays a library peptide as a fusion to the T7 phage capsid protein with flanking FLAG and STREP tags. The library was then packaged in vitro and expanded. The quality of library inserts was assessed by Illumina sequencing. A CDI-made peptide hormone library was added in addition to this via 33mer peptides and a similar insert strategy.

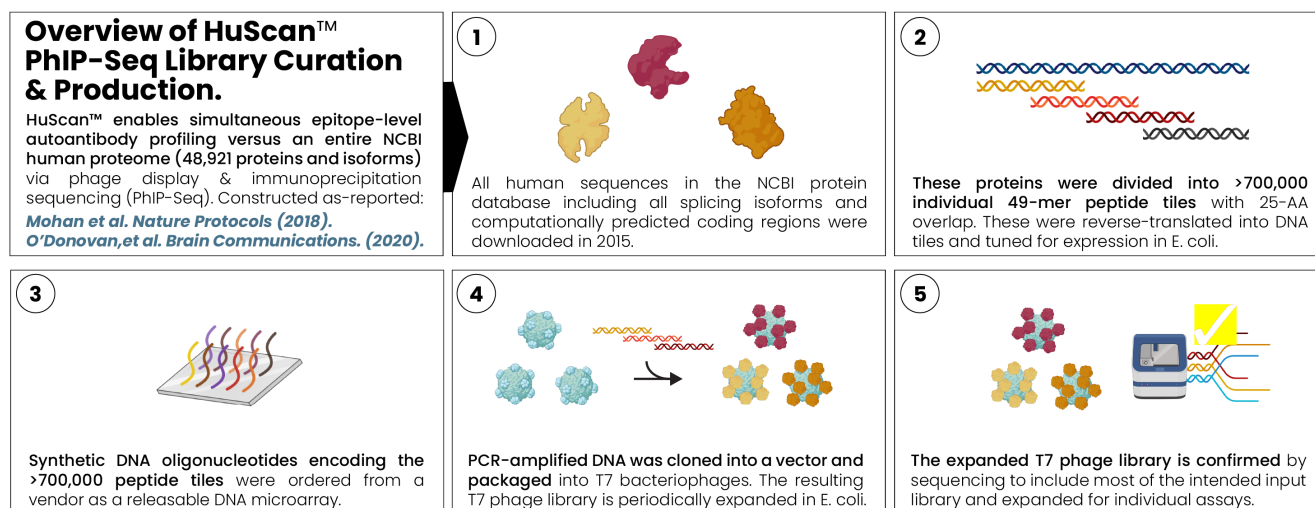


Figure by Tyler Hulett, PhD for CDI Labs; created with BioRender.

Overview of PhIP-Seq HuScan library construction. Human library creation was performed as-shown and as-reported: [O'Donovan, B. et al. High-resolution epitope mapping of anti-Hu and anti-Yo autoimmunity by programmable phage display. Brain Communications \(2020\)](#) <https://doi.org/10.1093/braincomms/fcaa059>.

3 Experimental Overview

Screens are performed in 0.5 mL PBS containing phage library at an average of 10^5 excess clones per unique phage in the input library (approx. 7×10^{10} total input plaque forming units) with 0.2 μ L serum or equivalent (approx. 2 μ g IgG) as-described.⁴ The mixture is rotated overnight to allow antibodies to bind any phage-displayed peptide targets. The next day, a Protein A/G coated magnetic bead slurry is added and rotated for an additional 4 hours to capture all IgG. A set of independent Protein A/G beads-only mock-immunoprecipitation buffer-alone samples are included as negative controls ('beadsonly'). All protein A/G beads are then washed and resuspended with a Herculanase II Fusion Polymerase master mix and amplified with PCR. Multiplex barcodes and sequencing adapters are then incorporated during a subsequent PCR reaction. Barcoded samples are pooled and sequenced using Illumina instruments to obtain paired-end reads of the amplified phage inserts.

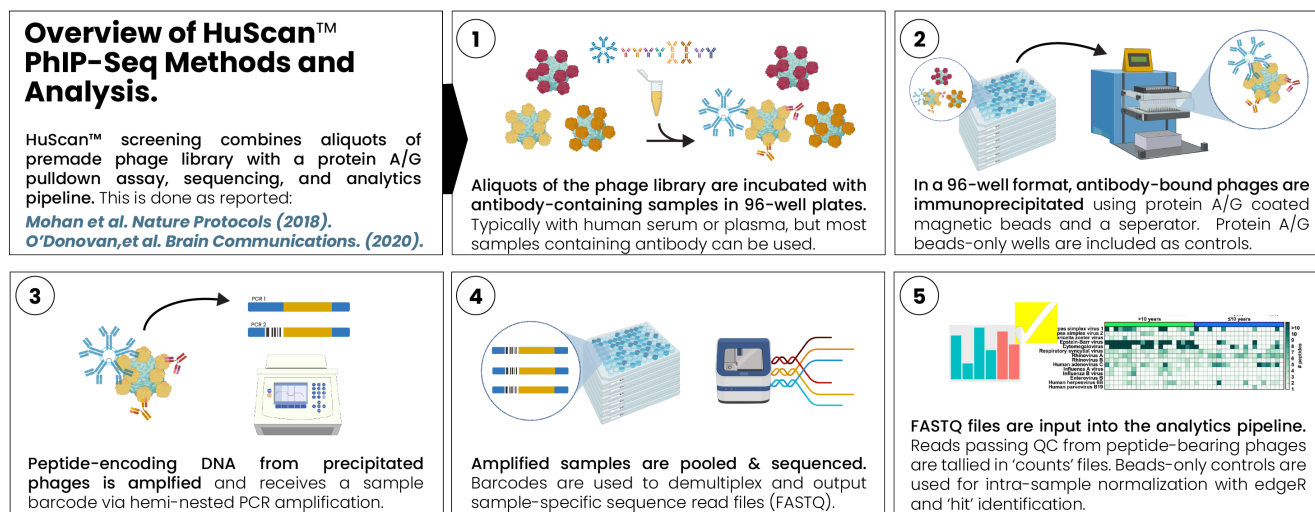


Figure by Tyler Hulett, PhD for CDI Labs; created with BioRender.

Overview of PhIP-Seq HuScan assay. Sample screening, PCR, and peptide read count data curation are performed as shown and as-described: *Mohan, D., et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. Nature Protocols. (2018)* <https://doi.org/10.1038/s41596-018-0025-6>.

4 Informatics Methods

Paired-end reads are demultiplexed, trimmed, and aligned to the human peptide library using bwa mem, requiring a minimum alignment score of 50 (“-T 50” argument). Reads aligning to each phage clone sequence are counted and recorded (*counts.csv*). Next, the edgeR exactTest function is used to estimate fold changes (*edgeR_log2foldchg.csv*) and p-values (*edgeR_log10pval.csv*) for each sample against the ‘beadsonly’ mock immunoprecipitations. For this calculation, dispersions for each phage clone are estimated from the mock immunoprecipitations only. Enriched ‘hit’ clones are defined as those with an edgeR-computed log2 fold change over the ‘beadsonly’ decoy controls of at least 2.0 (i.e. a >4x increase) and an unadjusted p-value of less than 10^{-4} (i.e. a $-\log_{10}(\text{p-value}) > 4$) to create the true/false (1/0) hit calls file (*edgeR_enriched.csv*) and the semi-continuous hit calls file of quantitative log2 fold change above the hit thresholds (*edgeR_enriched_log2foldchg.csv*). False discoveries are observed by calling hits for each ‘beadsonly’ sample versus the rest of the decoy pool excluding that sample; these data are recorded alongside study results. These core pipeline outputs were used to generate all downstream figures and tables presented in this report.

5 Core Pipeline Output Files

 *projectname_counts.csv*
 *projectname_edgeR_log2foldchg.csv*
 *projectname_edgeR_log10pval.csv*
 *projectname_edgeR_enriched.csv*
 *projectname_edgeR_enriched_log2foldchg.csv*

Core pipeline output files. These are produced from FASTQ files according to the pipeline described previously.

***projectname_counts.csv* - The data in its most raw format.** One column per sample; total counts of each phage reliably identified in that sample from its parent FASTQ file. Used for quality control plots and used to derive all downstream files and deliverables. For de novo normalization and hit calling - this is the file to start from.

***projectname_edgeR_log2foldchg.csv* -** The edgeR computed log2 fold change of each sample versus the ‘beadsonly’ background distribution. This file contains computed values used to derive hit calls.

***projectname_edgeR_log10pval.csv* -** EdgeR $-\log_{10}$ p-values for each counts reading versus ‘beadsonly’ background distribution. This file contains computed values used to derive hit calls.

***projectname_edgeR_enriched.csv* - Simple hits call file.** Record of entries that pass an edgeR-computed hit threshold for fold change and p-value. Passing entries have a log2 fold change over the ‘beadsonly’ controls of at least 2.0 (i.e. >4x increase) and a p-value of less than 10^{-4} (i.e. a $-\log_{10}(\text{pval}) > 4$). These are true/false (1/0) hit calls summarized from the *edgeR_log2foldchg.csv* and *edgeR_log10pval.csv* files.

***projectname_edgeR_enriched_log2foldchg.csv* - Main quantitative hits call file.** Values are edgeR called log2 fold changes versus ‘beadsonly’ background which pass the *edgeR_enriched.csv* hit call threshold; non-passing values are shown as zero creating a semi-continuous distribution. This file is used to compute many delivered plots and analyses. Unless you are performing your own normalizations and hit calling, this is the best file to start with for downstream analysis.

SeqID	Protein	Description	AA_Start	AA_Stop	Peptide	p4A02-beadonly
CDIH20000001 TSPAN6 NP_003261.1 Homo_sapiens Seq_1	NP_003261.1	tetraspanin-6 isoform a	1	49	MASPSRLQTKPVITCFKSVLLIYTFIFWITGVILLAVGIWGVSLNLY	0
CDIH20000002 TSPAN6 NP_003261.1 Homo_sapiens Seq_2	NP_003261.1	tetraspanin-6 isoform a	25	73	TFIFWITGVILLAVGIWGVSLNLYFSLNEKATNVPFVLIATGTVIIL	0
CDIH20000003 TSPAN6 NP_003261.1 Homo_sapiens Seq_3	NP_003261.1	tetraspanin-6 isoform a	49	97	YFSLNEKATNVPFVLIATGTVIILGTGFCFATCRASAWMLKYAMFL	1
CDIH20000004 TSPAN6 NP_003261.1 Homo_sapiens Seq_4	NP_003261.1	tetraspanin-6 isoform a	73	121	LLGTGFCFATCRASAWMLKYAMFLTLVFLVELVAAIVGFVFRHEIKNS	0
CDIH20000005 TSPAN6 NP_003261.1 Homo_sapiens Seq_5	NP_003261.1	tetraspanin-6 isoform a	97	145	LTLVFLVELVAAIVGFVFRHEIKNSFKNNYEKALKQYNSTGDYRSHAVD	0
CDIH20000006 TSPAN6 NP_003261.1 Homo_sapiens Seq_6	NP_003261.1	tetraspanin-6 isoform a	121	169	SFKNNYEKALKQYNSTGDYRSHAVDQIQLHCCGVTDRDWTDTNYS	0
CDIH20000007 TSPAN6 NP_003261.1 Homo_sapiens Seq_7	NP_003261.1	tetraspanin-6 isoform a	145	193	DKIQLHCCGVTDRDWTDTNYSSEKGFPSCKLEDCTPQRADAKVN	0
CDIH20000008 TSPAN6 NP_003261.1 Homo_sapiens Seq_8	NP_003261.1	tetraspanin-6 isoform a	169	217	SEKGFPSCKLEDCTPQRADAKVNNEGCFIKVMTIESEMGVAVAGISF	0
CDIH20000009 TSPAN6 NP_003261.1 Homo_sapiens Seq_9	NP_003261.1	tetraspanin-6 isoform a	193	241	NNEGCFIKVMTIESEMGVAVAGISFGVACFQLIGFLAYCLSRATNNQ	0
CDIH20000010 TSPAN6 NP_003261.1 Homo_sapiens Seq_10	NP_003261.1	tetraspanin-6 isoform a	197	245	CFIKVMTIESEMGVAVAGISFGVACFQLIGFLAYCLSRATNNQYEV	0
CDIH20000011 TSPAN6 NP_001265672.1 Homo_sapiens Seq_1	NP_001265672.1	tetraspanin-6 isoform d precursor	1	49	MLKYAMFLTLVFLVELVAAIVGFVFRHEIKNSFKNNYEKALKQYNSTG	0
CDIH20000012 TSPAN6 NP_001265672.1 Homo_sapiens Seq_2	NP_001265672.1	tetraspanin-6 isoform d precursor	25	73	VFRHEIKNSFKNNYEKALKQYNSTGDYRSHAVDQIQLHCCGVTDRD	0
CDIH20000013 TSPAN6 NP_001265672.1 Homo_sapiens Seq_3	NP_001265672.1	tetraspanin-6 isoform d precursor	49	97	GDYRSHAVDQIQLHCCGVTDRDWTDTNYSSEKGFPSCKLEDCTP	3
CDIH20000014 TSPAN6 NP_001265672.1 Homo_sapiens Seq_4	NP_001265672.1	tetraspanin-6 isoform d precursor	73	121	DWTDTNYSSEKGFPSCKLEDCTPQRADAKVNNELIGIFLAYCLSRAI	0
CDIH20000015 TSPAN6 NP_001265672.1 Homo_sapiens Seq_5	NP_001265672.1	tetraspanin-6 isoform d precursor	81	129	SEKGFPSCKLEDCTPQRADAKVNNELIGIFLAYCLSRATNNQYEV	0
CDIH20000016 TNMD NP_071427.2 Homo_sapiens Seq_1	NP_071427.2	tenomodulin	1	49	MAKNPPENCEDCHILNAEAFKSKIKSLKICLVFGILALTULVFWG	0
CDIH20000017 TNMD NP_071427.2 Homo_sapiens Seq_2	NP_071427.2	tenomodulin	25	73	ICKSLKICLVFGILALTULVFWGSKHFWPEVPKAYDMEHTFYSNGE	0
CDIH20000018 TNMD NP_071427.2 Homo_sapiens Seq_3	NP_071427.2	tenomodulin	49	97	GSKHFWPEVPKAYDMEHTFYSNGEKKIYMEIDPVTREIFRSGNGTD	0
CDIH20000019 TNMD NP_071427.2 Homo_sapiens Seq_4	NP_071427.2	tenomodulin	73	121	EKKKIYMEIDPVTREIFRSGNGTDLETHVDFKNGYTGIVFGLQKCF	0
CDIH20000020 TNMD NP_071427.2 Homo_sapiens Seq_5	NP_071427.2	tenomodulin	97	145	DETLEHDFKNGYTGIVFGLQKCFIKTQIKVPEFSEPEEIDEENEI	0
CDIH20000021 TNMD NP_071427.2 Homo_sapiens Seq_6	NP_071427.2	tenomodulin	121	169	FIKTQIKVPEFSEPEEIDEENEITTTFFEGSVIWWPAEKPIENRDL	0

Annotation and layout of the core pipeline output files. This is the start of a *counts.csv* file; all core pipeline deliverables start with this same structure. Metadata and layout remains identical, but the values for each sample are computed according to which file you are opening.

SeqID: Every row represents a unique identifier for a single phage in the assay with additional information separated by “|” delimiters. It starts with a CDI-ID that includes the library (H for HuScan), version number (in this case ‘2’), unique phage number, Gene_Name, NCBI Ref_Seq ID, species, and sequence position order. Not all sequence positions start at 1 due to removal for redundancy or low-complexity; proteins starting with higher sequence values often indicate the non-redundant portion of an isoform with a primary transcript covered elsewhere.

Protein: The RefSeq identifier of the parent protein used to generate this peptide.

Description: A longer description of the protein name.

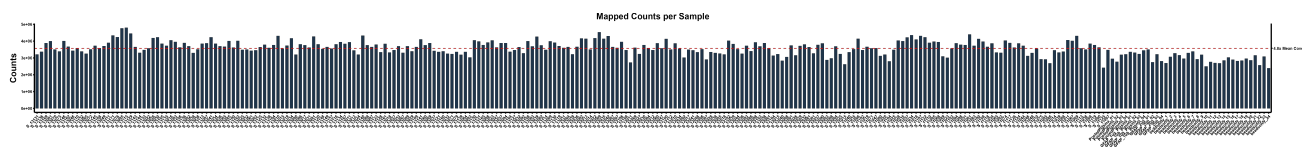
AA_Start: The amino acid position this peptide starts at related to the full parent RefSeq protein sequence.

AA_Stop: The amino acid position this peptide stops at related to the full parent RefSeq protein sequence.

Peptide: The literal peptide sequence carried by this phage.

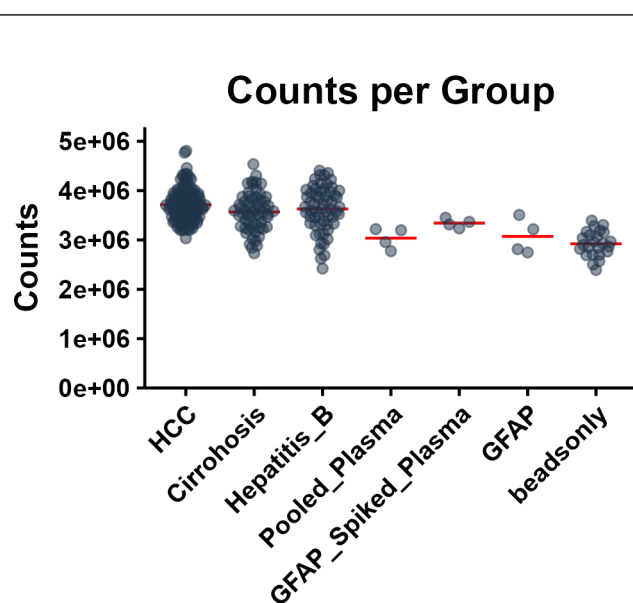
Sample Columns: All subsequent columns are samples in this assay. Any column strings containing ‘beadonly’ are mock-immunoprecipitation controls used for hit calling. One column should exist for each sample provided to us in the *sampleManifest.csv* at assay onboarding. We have additionally added internal assay control columns including the ‘beadonly’; both the sample labels and ours are presented together in the *studyManifest.csv* we use for downstream data processing. In the *counts.csv* file, sample column values represent literal tallies of the number of times each phage was reliably observed in this sample’s raw FASTQ data. Other delivery files contain values as-described.

6 QC Figures from Counts File



Number of mapped counts for each sample. Data showing the total mapped counts for each sample to the expected PhIP-Seq library; >95% of all samples – including ‘beadsonly’ and other provided assay controls – demonstrate read mapping to >1x coverage of the parent library and passes mapping quality control. You can recreate this plot from the *counts.csv* file by adding up the sum of each sample column.

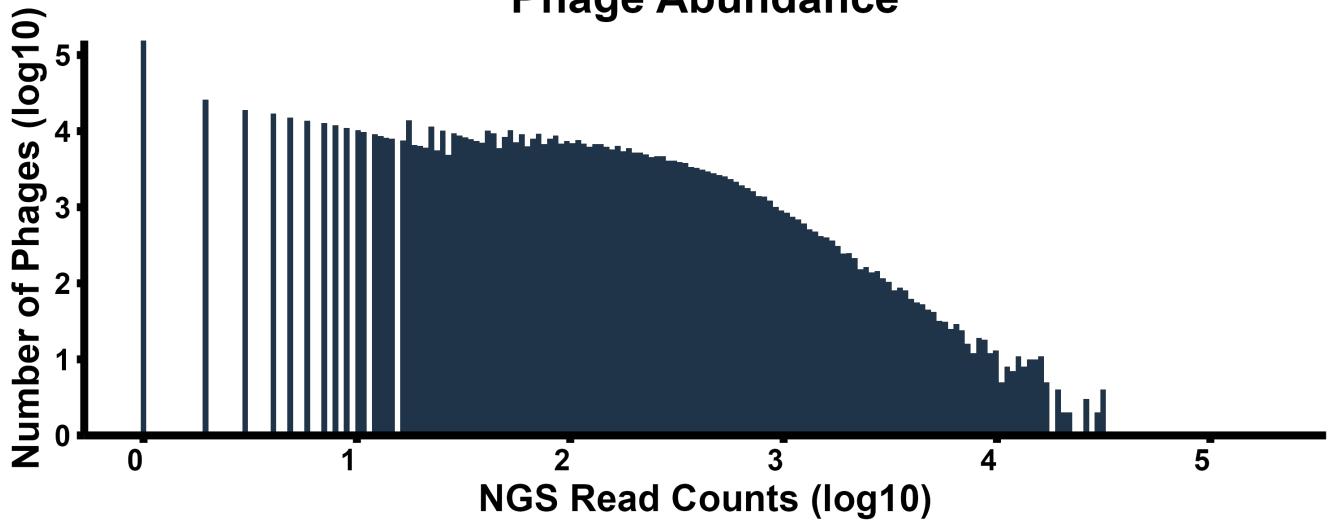
plots/counts_per_sample.png



Grouped number of mapped counts for each sample. Grouped results of the mapped counts according to provided study and control metadata.

plots/counts_per_group.png

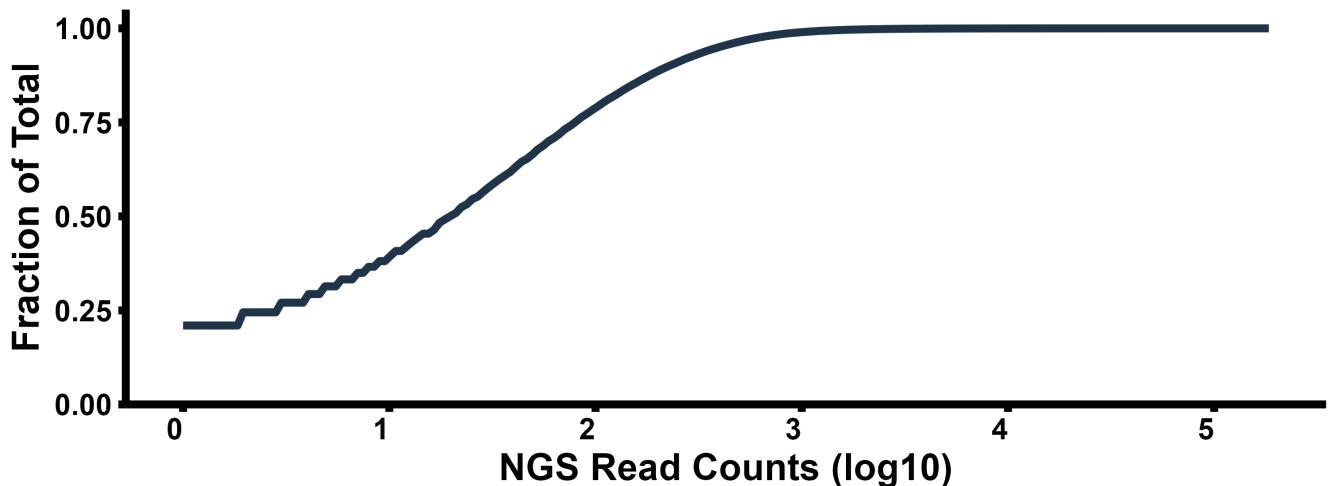
Histogram of Baseline Phage Abundance



Frequency histogram of the ‘beadonly’ background distribution. This histogram plots the number of unique phage clones present at different binned read count frequencies across the pool of ‘beadonly’ samples; demonstrates that most unique phages were successfully produced and are present in the mock-immunoprecipitation background for this experiment.

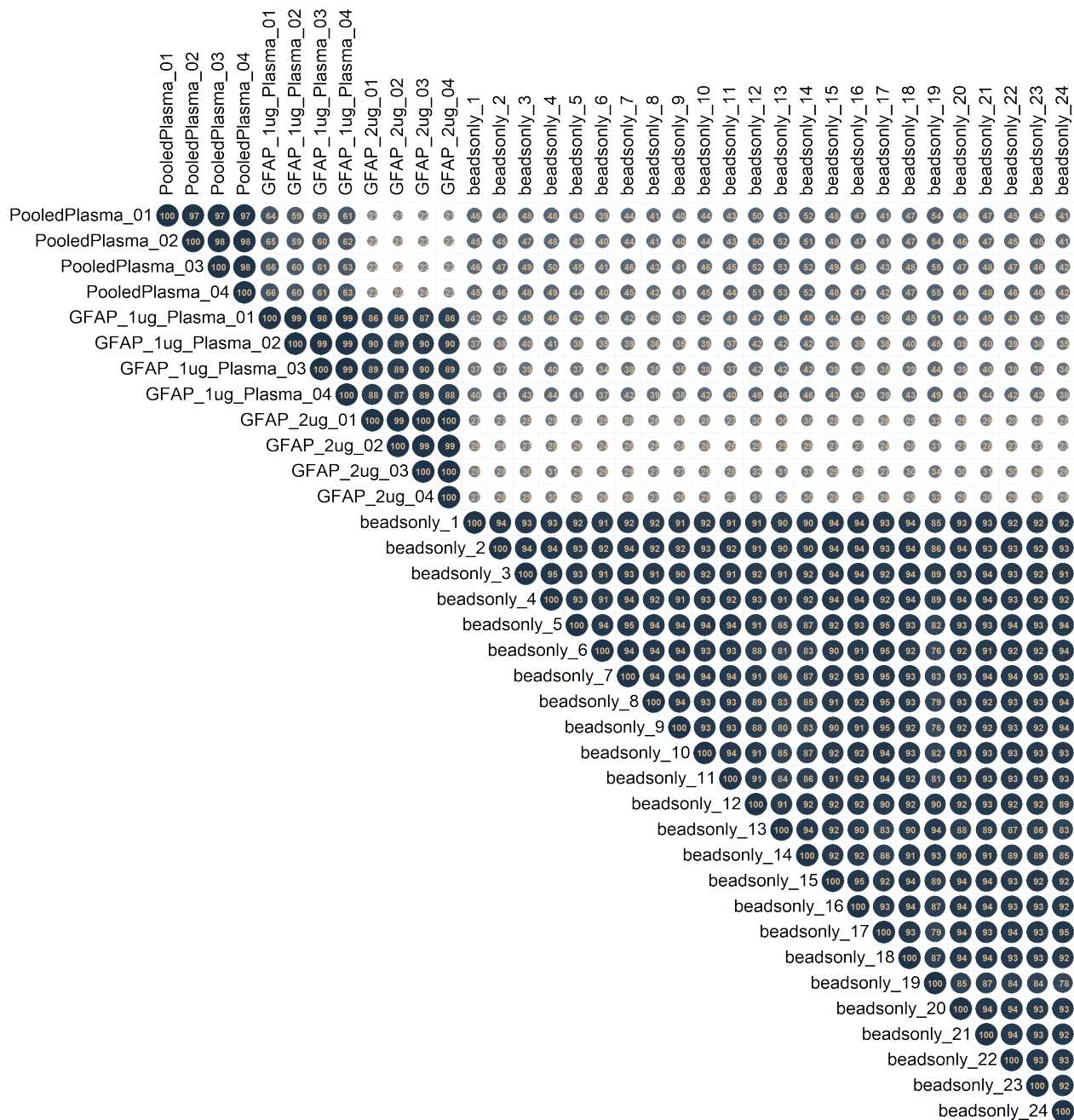
plots/beadonly_histogram_plot.png

Cumulative Sum of Baseline Phage Abundance



Data showing a cumulative sum of the ‘beadonly’ background phage distribution histogram. Demonstrates that most phages are present in the assay within one log step distance of the median frequency for this experiment. Typically, the median background frequency is around 100 copies, with most phages uniquely observed between ten and one thousand times in the mock-immunoprecipitation background (i.e. between 1 and 3 on a log10 scale).

plots/cumulative_sum_plot.png



Correlation matrix of baseline counts distribution for baseline phage abundance and controls. Demonstrates that the counts distribution for most background phages ('beadsonly') are highly correlated - this makes them a reliable decoy control from which to call real hits in the samples. Large dark circles indicate high correlation between samples. Additionally, pooled human plasma (Pooled_Plasma) and anti-GFAP polyclonal antibody (GFAP) readings both replicate well and are highly unique versus 'beadsonly' decoy samples. Anti-GFAP polyclonal was spiked into the pooled human plasma sample showing intermediate correlation between both (GFAP_Spiked_Plasma). This is a distribution from the actual experiment. Created by running the R package *corrplot* on all displayed columns in the *counts.csv* file.

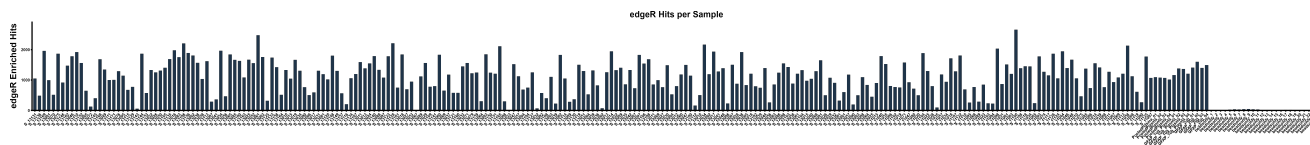
plots/correlation_plot_counts_controls.png



Correlation matrix of counts distribution for the study cohort. Demonstrates sample uniqueness and similarity. Large dark circles indicate high correlation between samples. If this study includes replicate samples or multi-timepoint samples a strong correlation should be observable within the raw counts data. Created by running the R package *corrplot* on all displayed columns in the *counts.csv* file.

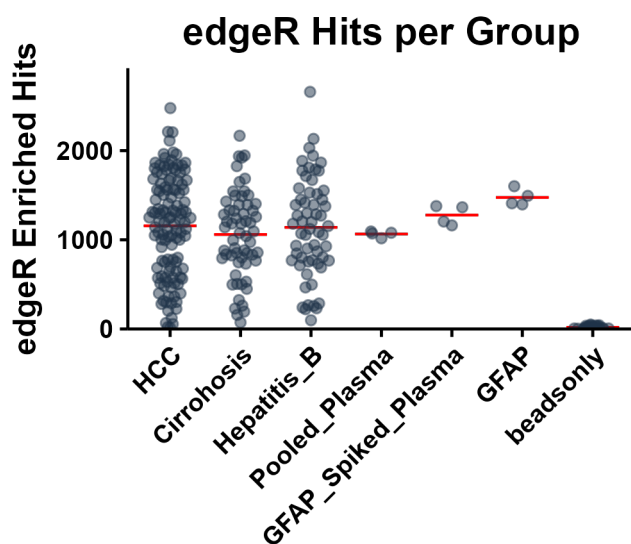
plots/correlation_plot_counts_cohort.png

7 Results Summary Figures from edgeR Enriched Hits



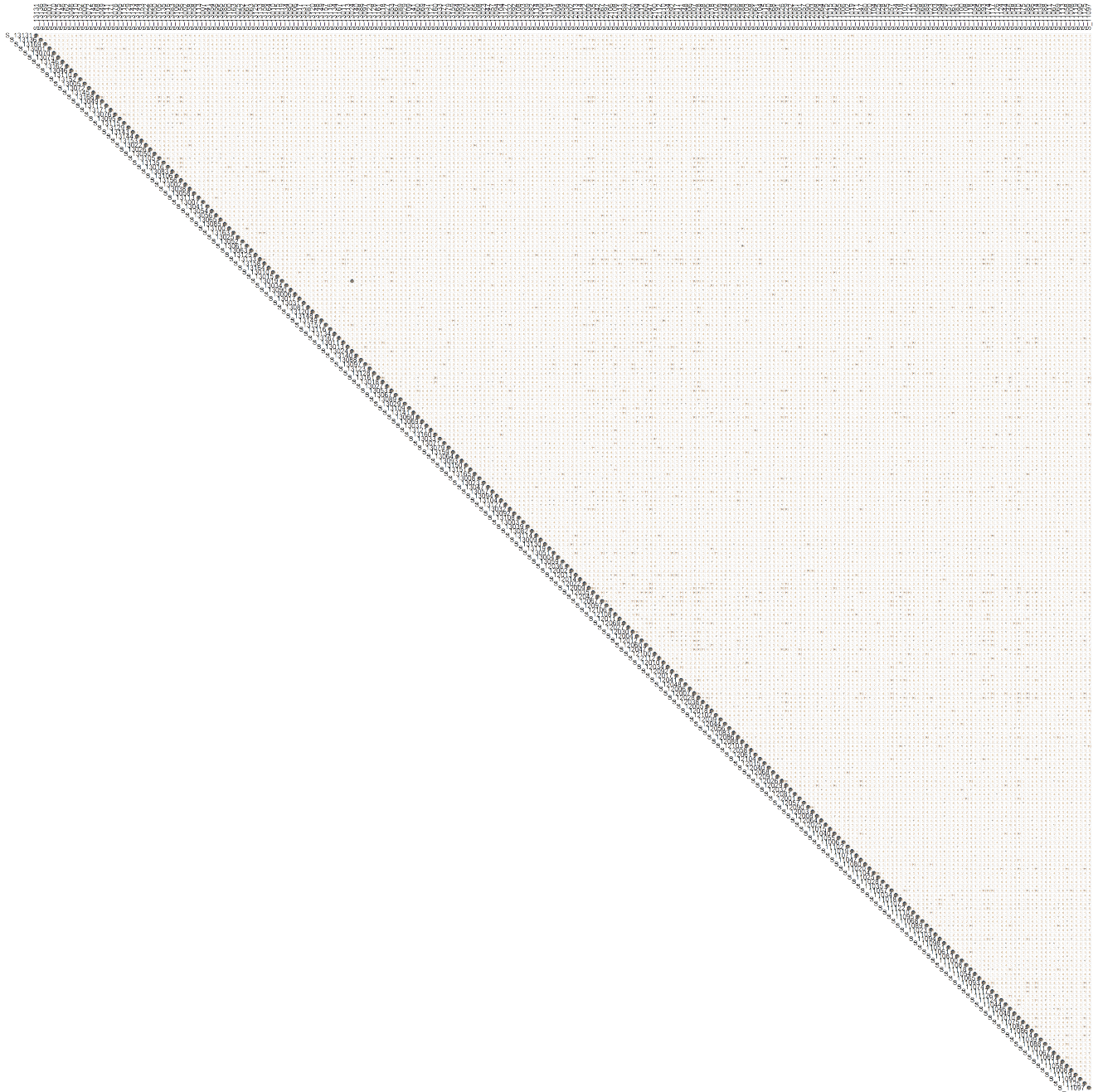
Number of peptide hit enrichments called in each sample. Each plotted bar represents the total number of unique peptide hits called for a single sample as-described in the informatics methods. Hits in the 'beadonly' decoy samples are additionally shown to demonstrate false discoveries in the assay background. You can recreate this plot by adding up the sum of each column in the *edgeR_enriched.csv* file; alternatively, it can be made from the *edgeR_enriched_log2foldchg.csv* file by counting the number of non-zero hit values in each column.

plots/hits_per_sample.png



Grouped number of unique peptide hits in each sample. Grouped results of the total hits called according to provided study and control metadata.

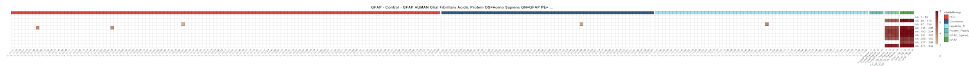
plots/hits_per_group.png



Correlation matrix of edgeR-called antibody hits for the study cohort. Demonstrates sample uniqueness and similarity. Large dark circles indicate high correlation between samples. If this study includes replicate samples or multi-timepoint samples a strong correlation should be observable within the processed edgeR-called hits. Created by running the R package *corrplot* on all displayed columns in the semi-continuous *edgeR_enriched_log2foldchg* hits file.

plots/correlation_plot_hits_cohort.png

8 GFAP Positive Control Heatmaps



Epitope scanning heatmap of pipeline-called anti-GFAP (glial fibrillary acidic protein) positive control hits alongside other study samples; lanes labeled GFAP have a polyclonal antibody spike-in against this protein and its isoforms which demonstrate positivity. This heatmap is created from the *edgeR_enriched_log2foldchg.csv* file and squares are colored according to the log2 fold change of the called hit versus the 'beadonly' distribution; non-enriched values are blank and log2 fold changes typically range from 2.0 (i.e. 4x increase above background) to >8.0 (i.e. >256x increase). A full resolution copy is available in the **heatmaps** folder for this figure and for two other GFAP isoforms.

scan_heatmaps/heatmap_NP_001229305.1_GFAP_glial fibrillary acidic protein isoform 3.png

scan_heatmaps/heatmap_NP_001124491.1_GFAP_glial fibrillary acidic protein isoform 2.png

scan_heatmaps/heatmap_NP_002046.1_GFAP_glial fibrillary acidic protein isoform 1.png

9 Processed Hits Pipeline Output Files

Processed hits pipeline output files. These files, as well as all subsequent figures and analysis, are derived from the semi-continuous *edgeR_enriched_log2foldchg.csv* hits file. Tables are provided in the **processed_outputs** folder.

projectname_top_public_peptides.csv - The most frequent publicly reactive peptides across the sample cohort; this file is output even for blinded studies. For each row in the *edgeR_enriched_log2foldchg.csv* file, the total hit frequency is calculated and sorted in descending order. Average log2 fold change of hits is used to break ties. These are used to create unbiased figures in the **summary_heatmaps** folder labeled as *top_public*. Additionally, a whole-protein epitope-scanning heatmap is provided in the **scan_heatmaps** folder for each protein in this list.

projectname_group_hit_counts.csv - Peptide-level group analysis. This is a true/false tally of the number of *edgeR_enriched_log2foldchg.csv* samples from each group which react to each peptide in the library. One *Hit* and *No_Hit* column pair per provided group described in the *studyManifest*; only one pair will be present if the study groups are blinded. If more than two groups are provided, these are used for Fisher exact tests.

projectname_group_protein_hit_counts.csv - Protein-level group analysis. This is a true/false tally of the number of *edgeR_enriched_log2foldchg.csv* samples from each group which react to at least one peptide from a given protein in the library. One *Protein_Hits* and *Protein_No_Hits* column pair per provided group described in the *studyManifest*; only one pair will be present if the study groups are blinded. If groups are provided, these are used for Fisher exact tests. Since this is a protein-level analysis, peptide-level metadata columns are not included.

projectname_group_polyclonal_hit_counts.csv - Protein-level group analysis. This is a true/false tally of the number of *edgeR_enriched_log2foldchg.csv* samples from each group which react to three or more peptides from a given protein in the library. One *Polyclonal_Hits* and *Polyclonal_No_Hits* column pair per provided group described in the *studyManifest*; only one pair will be present if the study groups are blinded. If groups are provided, these are used for Fisher exact tests. Since this is a protein-level analysis, peptide-level metadata columns are not included. **Note: For hormone peptides 'polyclonal' refers to 3 or more hits to multiple variants of a single transcript, a high confidence result but not necessarily 'polyclonal.'**

10 Fisher Exact Tests

The following pages contain case versus cohort statistical outputs according to the provided **studyGroup** column entries; results are output at both the peptide level and as simple protein-level and 'polyclonal' tests. One output file is provided for each unique *studyGroup*. For all samples in a group, their status as an edgeR called *Hit* or *No_Hit* is identified and tallied. If an explicit 'Control' group was provided, all groups are additionally compared to 'Control' only without the rest of the study cohort. One output file per test is provided for each unique *studyGroup*. For all samples in a group, their status as a *Hit* or *No_Hit* is identified and tallied.

11 Peptide-Level Tests

Peptide-level Fisher exact tests are performed on a 2x2 matrix of **Hits** and **No_Hits** columns for each provided group using the *fisher.test* function in R. Resulting p-value calculations are exported and saved in the **fisher_exact_tests** folder to create *vs_cohort_fisher_exact_tests.csv* files. To calculate directionality of the p-values, the log odds ratio of the groups is calculated and output in the *log_odds_ratio* column. Positive log odds ratios represent hits over-represented in the *studyGroup* labeled for this set of samples; negative log odds ratios represent hits overrepresented in remainder of the cohort. Note that if hits exist in one group and not the other, an infinite (Inf) or negative infinite (-Inf) ratio results. These values appear appropriately in a text editor or R, but -Inf often displays incorrectly in Excel.

On the following pages you will find a table of the top p-value outlier hits enriched in each provided group versus the rest of the cohort. A sample-level unclustered, clustered, and row clustered visualization heatmap is provided for the top 100 rows of each table in the **summary_heatmaps** folder.

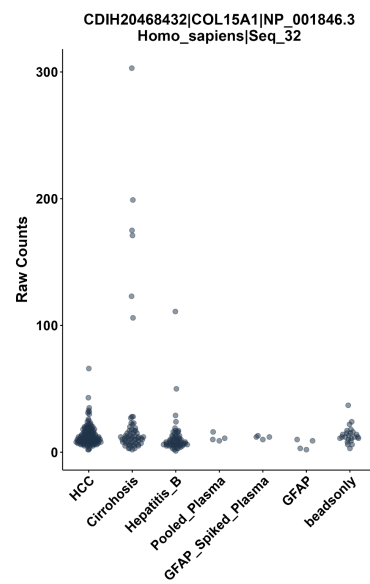
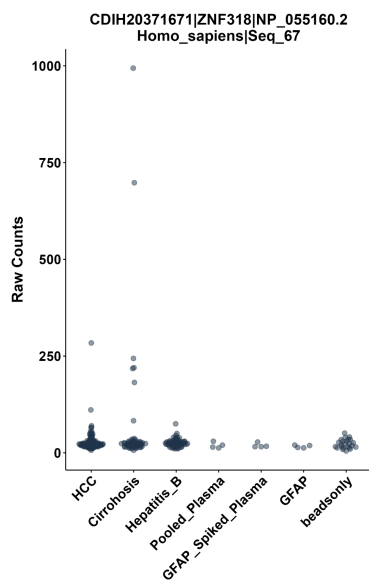
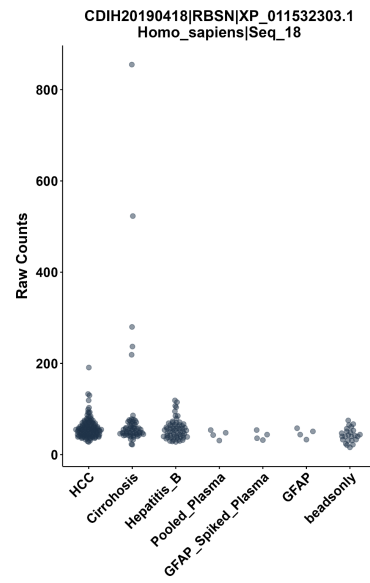
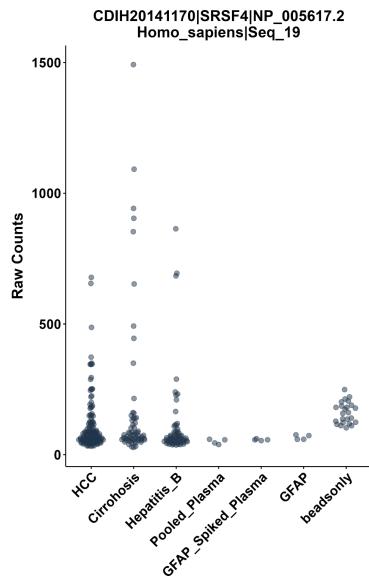
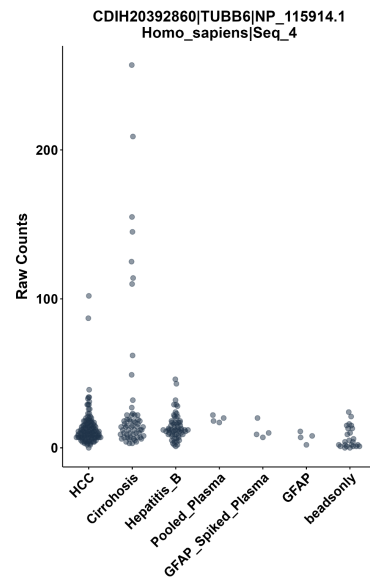
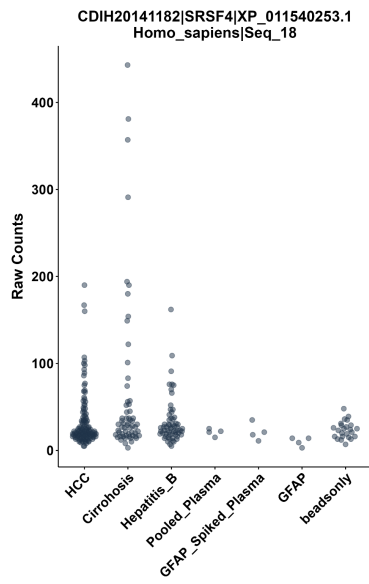
A whole-protein epitope-scanning heatmap is provided in the **scan_heatmaps** folder for each protein bearing one of the top 100 peptides enriched in each sample group.

For each of the top 100 peptide enrichments, a plot of the raw counts data for that peptide is provided in the **peptide_plots** folder. These provide a visualization of raw sequence counts data from the *counts.csv* file for each of the top Fisher exact test called differences. The top x6 p-value peptides are additionally included as a page within this report.

Vietnam Cirrhosis vs cohort

SeqID	Hits Cirrhosis	Hits Cohort	No Hits Cirrhosis	No Hits Cohort	p value
CDIH20141182[SRSF4]XP_011540253.1 Homo_sapiens Seq_18	10	4	50	177	0.0002078
CDIH20392860[TUBB6]NP_115914.1 Homo_sapiens Seq_4	7	1	53	180	0.0002887
CDIH20141170[SRSF4]NP_005617.2 Homo_sapiens Seq_19	5	0	55	181	0.0008405
CDIH20190418[RBSN]XP_011532803.1 Homo_sapiens Seq_18	5	0	55	181	0.0008405
CDIH20371671[ZNF318]NP_055160.2 Homo_sapiens Seq_67	6	1	54	180	0.0011011
CDIH20468432[COL15A1]NP_001846.3 Homo_sapiens Seq_32	6	1	54	180	0.0011011
CDIH2052223[KIF1A]XP_005247079.1 Homo_sapiens Seq_61	6	1	54	180	0.0011011
CDIH20620739[SRPK2]XP_011514833.1 Homo_sapiens Seq_3	6	1	54	180	0.0011011
CDIHDB100324[BNP5-31]hor003004 Homo_sapiens Seq_1d_right	6	1	54	180	0.0011011
CDIH20303920[ZC3H18]XP_011521165.1 Homo_sapiens Seq_19	20	26	40	155	0.0021594
CDIH20063613[NOP14]NP_001278907.1 Homo_sapiens Seq_1	13	12	47	169	0.0023802
CDIH20231567[GIT2]NP_476510.1 Homo_sapiens Seq_21	7	3	53	178	0.0027850
CDIH20331488[RELL2]NP_776189.3 Homo_sapiens Seq_6	7	3	53	178	0.0027850
CDIH20650938[NECTIN2]XP_011525494.1 Homo_sapiens Seq_12	7	3	53	178	0.0027850
CDIH20194378[ZRANB2]NP_976225.1 Homo_sapiens Seq_12	9	6	51	175	0.0030207
CDIH20081673[MTFHD1]NP_005947.3 Homo_sapiens Seq_25	6	2	54	179	0.0035382
CDIH20259684[MLIP]NP_001268675.1 Homo_sapiens Seq_9	6	2	54	179	0.0035382
CDIH20342648[SPINT1]NP_857593.1 Homo_sapiens Seq_7	6	2	54	179	0.0035382
CDIH20412076[HHLPL1]NP_001120730.1 Homo_sapiens Seq_28	6	2	54	179	0.0035382
CDIH20590620[CLASP1]XP_011509149.1 Homo_sapiens Seq_34	6	2	54	179	0.0035382
CDIH20672711[SNX25]XP_011530604.1 Homo_sapiens Seq_31	6	2	54	179	0.0035382
CDIH20004801[MAP3K9]NP_001271159.1 Homo_sapiens Seq_34	4	0	56	181	0.0035572
CDIH20043113[ZFYVE26]NP_056161.2 Homo_sapiens Seq_11	4	0	56	181	0.0035572
CDIH20222656[SEMA6D]XP_011520379.1 Homo_sapiens Seq_39	4	0	56	181	0.0035572
CDIH20263456[FRMPD3]XP_011529358.1 Homo_sapiens Seq_67	4	0	56	181	0.0035572
CDIH20337463[LRFN5]NP_689660.2 Homo_sapiens Seq_17	4	0	56	181	0.0035572
CDIH20497681[PHRF1]NP_001273510.1 Homo_sapiens Seq_43	4	0	56	181	0.0035572
CDIH20581468[EP300]XP_006724228.1 Homo_sapiens Seq_20	4	0	56	181	0.0035572
CDIH20628669[RALGPS1]XP_011517529.1 Homo_sapiens Seq_9	4	0	56	181	0.0035572
CDIH20637957[LOC101929304]XP_011521101.1 Homo_sapiens Seq_8	4	0	56	181	0.0035572
CDIH20677114[COL7A1]XP_011531638.1 Homo_sapiens Seq_40	4	0	56	181	0.0035572
CDIH20682166[SLC4A7]XP_011532566.1 Homo_sapiens Seq_3	4	0	56	181	0.0035572
CDIH20697638[ZRANB1]XP_011538189.1 Homo_sapiens Seq_9	4	0	56	181	0.0035572
CDIH20722123[FAM83H]XP_011544834.1 Homo_sapiens Seq_47	4	0	56	181	0.0035572
CDIHDB103183[Relaxin_B_chain]hor002225 Homo_sapiens Seq_1c_center	4	0	56	181	0.0035572
CDIHDB104688[Leumorphin]hor0030281 Homo_sapiens Seq_1c_repeat	4	0	56	181	0.0035572
CDIHDB105552[Anakinra]hd00071 Drug Seq_1b_right	4	0	56	181	0.0035572
CDIH20083528[RIMS4]NP_001192246.1 Homo_sapiens Seq_5	5	1	55	180	0.0040636
CDIH20115266[SLAIN2]XP_005248178.1 Homo_sapiens Seq_18	5	1	55	180	0.0040636
CDIH20130870[PRLR]NP_001191244.1 Homo_sapiens Seq_14	5	1	55	180	0.0040636
CDIH20173126[MCF2L]XP_011535784.1 Homo_sapiens Seq_43	5	1	55	180	0.0040636
CDIH20566849[SCN5A]XP_006713345.1 Homo_sapiens Seq_25	5	1	55	180	0.0040636
CDIH20382800[CHD2]NP_001262.3 Homo_sapiens Seq_46	8	5	52	176	0.0043378
CDIH20030001[WNK1]NP_055638.2 Homo_sapiens Seq_68	9	7	51	174	0.0055302
CDIH20360766[THBS3]XP_011508228.1 Homo_sapiens Seq_27	9	7	51	174	0.0055302
CDIH20086947[GUCY2F]NP_001513.2 Homo_sapiens Seq_29	11	10	49	171	0.0060485
CDIH20011261[MBTPS2]NP_056968.1 Homo_sapiens Seq_5	7	4	53	177	0.0061431
CDIH20300042[RNF111]XP_006720638.1 Homo_sapiens Seq_35	7	4	53	177	0.0061431
CDIH20687569[UTRN]XP_011534403.1 Homo_sapiens Seq_86	7	4	53	177	0.0061431
CDIH20550459[LUZP4]XP_005268400.1 Homo_sapiens Seq_4	8	6	52	175	0.0081105
CDIH20654813[MYO1F]XP_011526328.1 Homo_sapiens Seq_42	8	6	52	175	0.0081105
CDIH20139361[SRSF7]NP_001026854.1 Homo_sapiens Seq_8	6	3	54	178	0.0085311
CDIH20204365[TUT4]NP_001009881.1 Homo_sapiens Seq_60	6	3	54	178	0.0085311
CDIH20231579[GIT2]XP_006719770.1 Homo_sapiens Seq_20	6	3	54	178	0.0085311
CDIH20253790[STK11P]XP_011508828.1 Homo_sapiens Seq_43	6	3	54	178	0.0085311
CDIH20353711[SRRM2]NP_057417.3 Homo_sapiens Seq_23	6	3	54	178	0.0085311
CDIH20448979[WNK3]XP_011529104.1 Homo_sapiens Seq_65	6	3	54	178	0.0085311
CDIH20465371[L1CAM]NP_001265045.1 Homo_sapiens Seq_6	6	3	54	178	0.0085311
CDIH20525156[MAP3K1]XP_005248576.2 Homo_sapiens Seq_47	6	3	54	178	0.0085311
CDIH20644936[RECQL5]XP_011523786.1 Homo_sapiens Seq_21	6	3	54	178	0.0085311
CDIH20688504[PPP1R13B]XP_011534896.1 Homo_sapiens Seq_47	6	3	54	178	0.0085311
CDIH20116318[NSD2]XP_005248058.1 Homo_sapiens Seq_52	10	9	50	172	0.0097190
CDIH20550537[LARP1]XP_005268461.1 Homo_sapiens Seq_36	10	9	50	172	0.0097190
CDIH20136039[PDE1A]XP_011509627.1 Homo_sapiens Seq_19	5	2	55	179	0.0114700
CDIH20146162[CENPF]XP_011507384.1 Homo_sapiens Seq_126	5	2	55	179	0.0114700
CDIH20231421[GAS2L3]NP_777602.1 Homo_sapiens Seq_16	5	2	55	179	0.0114700
CDIH20255118[MYH15]XP_011510861.1 Homo_sapiens Seq_61	5	2	55	179	0.0114700
CDIH20316602[SYNC]NP_001155180.1 Homo_sapiens Seq_6	5	2	55	179	0.0114700
CDIH20343197[NOLC1]XP_005270330.1 Homo_sapiens Seq_10	5	2	55	179	0.0114700
CDIH20368037[POLH]NP_006493.1 Homo_sapiens Seq_15	5	2	55	179	0.0114700
CDIH20411403[TNRC18]NP_001073964.2 Homo_sapiens Seq_109	5	2	55	179	0.0114700
CDIH20450220[CRACDL]XP_005263980.1 Homo_sapiens Seq_8	5	2	55	179	0.0114700
CDIH20462607[ECI2]NP_996667.2 Homo_sapiens Seq_3	5	2	55	179	0.0114700
CDIH20467756[GIGYF2]NP_056390.2 Homo_sapiens Seq_37	5	2	55	179	0.0114700
CDIH20622551[PTK2]XP_011515488.1 Homo_sapiens Seq_36	5	2	55	179	0.0114700
CDIH20644960[RECQL5]XP_011523787.1 Homo_sapiens Seq_21	5	2	55	179	0.0114700
CDIH20678623[DNAH1]XP_011531879.1 Homo_sapiens Seq_138	5	2	55	179	0.0114700
CDIH20687152[NHSL1]XP_011534273.1 Homo_sapiens Seq_20	5	2	55	179	0.0114700
CDIH20710075[MAP3K1]XP_011541709.1 Homo_sapiens Seq_48	5	2	55	179	0.0114700
CDIH20149453[PPL]XP_006720965.1 Homo_sapiens Seq_56	7	5	53	176	0.0118280
CDIH20161391[CYREN]NP_001230681.1 Homo_sapiens Seq_2	7	5	53	176	0.0118280
CDIH20252527[LNPX]NP_001291938.1 Homo_sapiens Seq_10	7	5	53	176	0.0118280
CDIH20285823[MIA3]NP_940953.2 Homo_sapiens Seq_25	7	5	53	176	0.0118280
CDIH20296055[SCAF4]NP_001138916.1 Homo_sapiens Seq_15	7	5	53	176	0.0118280
CDIHDB100511[Cocaine-and_amphetamine-regulated_transcript_protein]hor000698 Homo_sapiens Seq_3a_right	7	5	53	176	0.0118280
CDIHDB100512[Cocaine-and_amphetamine-regulated_transcript_protein]hor000698 Homo_sapiens Seq_3b_right	7	5	53	176	0.0118280
CDIHDB100515[Cocaine-and_amphetamine-regulated_transcript_protein]hor000698 Homo_sapiens Seq_3e_right	7	5	53	176	0.0118280
CDIHDB100733[Gastrin]hor001002 Homo_sapiens Seq_1c_right	7	5	53	176	0.0118280
CDIHDB100735[Gastrin]hor001002 Homo_sapiens Seq_1e_right	7	5	53	176	0.0118280
CDIH20463711[ALPK2]NP_443179.3 Homo_sapiens Seq_48	10	10	50	171	0.0126180
CDIH20599209[PRPF40A]XP_011509748.1 Homo_sapiens Seq_13	18	27	42	154	0.0128500
CDIH20226393[SEN7]XP_011511342.1 Homo_sapiens Seq_7	8	7	52	174	0.0139270
CDIH20597822[AFF3]XP_011509474.1 Homo_sapiens Seq_16	8	7	52	174	0.0139270
CDIH20516278[GIGYF1]NP_072096.2 Homo_sapiens Seq_41	12	14	48	167	0.0143830
CDIH20001404[PLXND1]XP_011510892.1 Homo_sapiens Seq_27	4	1	56	180	0.0144240
CDIH20011587[KDM5D]XP_011529770.1 Homo_sapiens Seq_54	4	1	56	180	0.0144240
CDIH20013672[PHLDB1]XP_011541005.1 Homo_sapiens Seq_34	4	1	56	180	0.0144240
CDIH20028200[RC3H2]NP_001094058.1 Homo_sapiens Seq_4	4	1	56	180	0.0144240
CDIH20090837[NFAT5]NP_006590.1 Homo_sapiens Seq_44	4	1	56	180	0.0144240
CDIH20116576[KLHL5]NP_057074.3 Homo_sapiens Seq_30	4	1	56	180	0.0144240

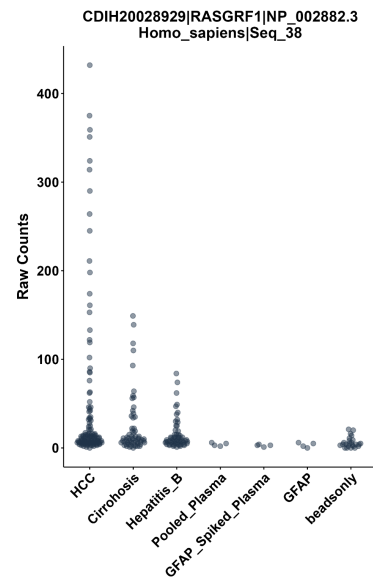
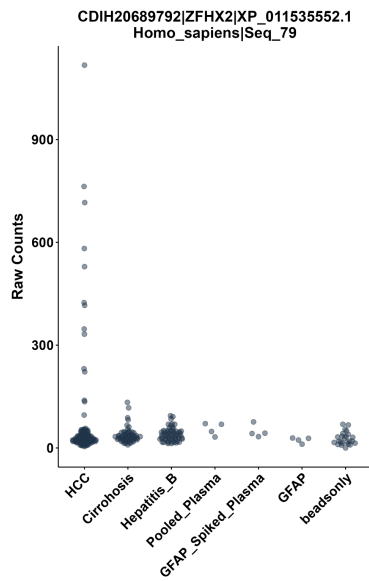
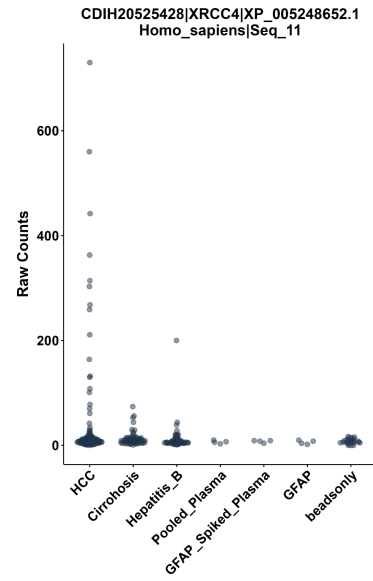
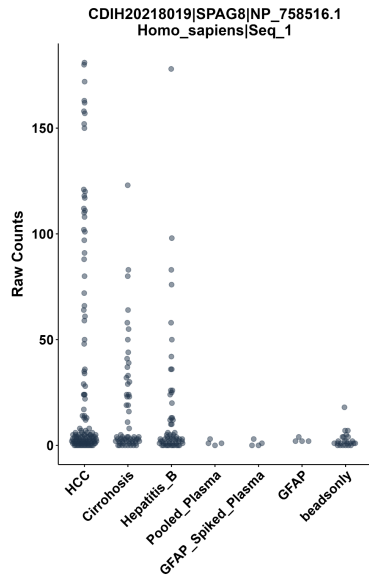
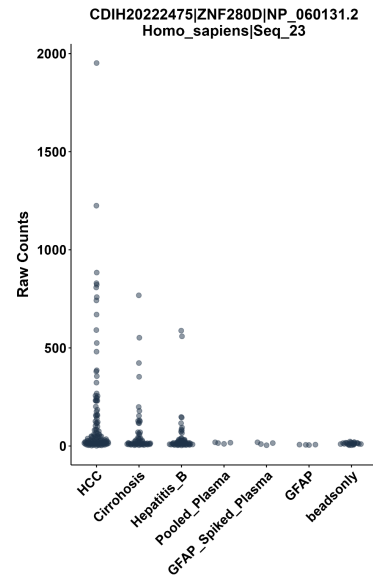
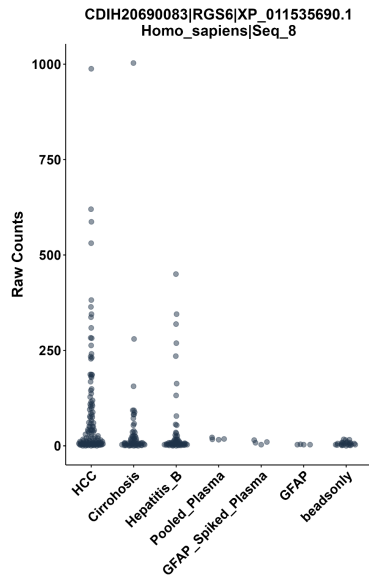
fisher_exact_tests/Vietnam_Cirrhosis_vs_cohort_fisher_exact_tests.csv



Vietnam HCC vs cohort

SeqID	Hits HCC	Hits Cohort	No Hits HCC	No Hits Cohort	p value
CDIH20690083 RGS6 XP_011535690.1 Homo_sapiens Seq_8	41	16	80	104	0.0002356
CDIH20222475 ZNF280D NP_060131.2 Homo_sapiens Seq_23	42	18	79	102	0.0005494
CDIH20218019 SPAG8 NP_758516.1 Homo_sapiens Seq_1	26	7	95	113	0.0005688
CDIH20525428 XRCC4 XP_005248652.1 Homo_sapiens Seq_11	14	1	107	119	0.0006908
CDIH20689792 ZFHX4 XP_011535552.1 Homo_sapiens Seq_79	11	0	110	120	0.0007700
CDIH20028929 RASGRF1 NP_002882.3 Homo_sapiens Seq_38	21	5	100	115	0.0014091
CDIH20481367 ATF6B NP_004372.3 Homo_sapiens Seq_23	25	8	96	112	0.0023153
CDIH20015547 GCLM NP_002052.1 Homo_sapiens Seq_4	9	0	112	120	0.0033480
CDIH20653968 MAST1 XP_011526107.1 Homo_sapiens Seq_47	15	3	106	117	0.0055670
CDIH20084064 DID01 XP_011526811.1 Homo_sapiens Seq_89	13	2	108	118	0.0057648
CDIH20436072 CYP26C1 NP_899230.2 Homo_sapiens Seq_20	20	6	101	114	0.0060305
CDIH20027983 ITI4 NP_002209.2 Homo_sapiens Seq_30	8	0	113	120	0.0069331
CDIH20044277 ACADVL NP_001257376.1 Homo_sapiens Seq_16	8	0	113	120	0.0069331
CDIH20243395 SCYL1 NP_065731.3 Homo_sapiens Seq_32	8	0	113	120	0.0069331
CDIH2038542 AMER2 NP_954589.1 Homo_sapiens Seq_16	8	0	113	120	0.0069331
CDIH20027982 ITI4 NP_002209.2 Homo_sapiens Seq_29	10	1	111	119	0.0100030
CDIH20220451 BUD13 NP_001153208.1 Homo_sapiens Seq_15	10	1	111	119	0.0100030
CDIH20590661 FLG2 XP_011507833.1 Homo_sapiens Seq_37	10	1	111	119	0.0100030
CDIH20336233 GAPVD1 NP_001269608.1 Homo_sapiens Seq_24	12	2	109	118	0.0105660
CDIH20531175 TLN2 XP_005254765.1 Homo_sapiens Seq_76	17	5	104	115	0.0123440
CDIH20413755 RGS6 NP_004287.3 Homo_sapiens Seq_9	23	9	98	111	0.0126640
CDIH20697242 ANK3 XP_011538007.1 Homo_sapiens Seq_84	23	9	98	111	0.0126640
CDIH20712118 KDM6A XP_011542261.1 Homo_sapiens Seq_26	23	9	98	111	0.0126640
CDIH20022788 ARHGAP6 NP_038286.2 Homo_sapiens Seq_14	7	0	114	120	0.0142920
CDIH20024734 DKK3 NP_056965.3 Homo_sapiens Seq_13	7	0	114	120	0.0142920
CDIH20201958 LAMTOR5 NP_006393.2 Homo_sapiens Seq_7	7	0	114	120	0.0142920
CDIH20310693 MED27 NP_004260.2 Homo_sapiens Seq_6	7	0	114	120	0.0142920
CDIH20316111 USP24 XP_005270747.1 Homo_sapiens Seq_82	7	0	114	120	0.0142920
CDIH20332060 FNDC1 NP_115921.2 Homo_sapiens Seq_32	7	0	114	120	0.0142920
CDIH20346366 ZNF592 NP_055445.2 Homo_sapiens Seq_46	7	0	114	120	0.0142920
CDIH20408638 MUC16 NP_078966.2 Homo_sapiens Seq_4	7	0	114	120	0.0142920
CDIH20599064 RIF1 XP_011509696.1 Homo_sapiens Seq_64	7	0	114	120	0.0142920
CDIH20663386 SEC14L4 XP_011528465.1 Homo_sapiens Seq_15	7	0	114	120	0.0142920
CDIH20695005 PFIA2 XP_011537204.1 Homo_sapiens Seq_32	7	0	114	120	0.0142920
CDIH20726141 DLGAP2 XP_011545671.1 Homo_sapiens Seq_30	7	0	114	120	0.0142920
CDIH20653764 DHPS XP_011526072.1 Homo_sapiens Seq_14	15	4	106	116	0.0148470
CDIH20138071 KDM3A NP_060903.2 Homo_sapiens Seq_50	18	6	103	114	0.0164560
CDIH20118971 AMBRA1 NP_001254712.1 Homo_sapiens Seq_27	13	3	108	117	0.0172910
CDIH20248203 ADAMTSL4 XP_011507953.1 Homo_sapiens Seq_1	13	3	108	117	0.0172910
CDIH20723167 ASB2 XP_011544947.1 Homo_sapiens Seq_9	13	3	108	117	0.0172910
CDIH20051833 PPP2R5C NP_848701.1 Homo_sapiens Seq_20	9	1	112	119	0.0189960
CDIH20123631 RFX4 NP_998759.1 Homo_sapiens Seq_22	9	1	112	119	0.0189960
CDIH20126961 ARFGEF3 NP_065073.3 Homo_sapiens Seq_51	9	1	112	119	0.0189960
CDIH20229202 LEF1 XP_005263103.1 Homo_sapiens Seq_14	9	1	112	119	0.0189960
CDIH20236380 TICRR NP_689472.3 Homo_sapiens Seq_68	9	1	112	119	0.0189960
CDIH20427026 KPNA7 XP_011514517.1 Homo_sapiens Seq_4	9	1	112	119	0.0189960
CDIH20437064 SSX7 NP_775494.1 Homo_sapiens Seq_6	9	1	112	119	0.0189960
CDIH20514306 MST1L NP_001258662.1 Homo_sapiens Seq_16	9	1	112	119	0.0189960
CDIH20565582 ANKRD36B XP_006712724.1 Homo_sapiens Seq_52	9	1	112	119	0.0189960
CDIH20575190 PFIA2 XP_006719732.1 Homo_sapiens Seq_30	9	1	112	119	0.0189960
CDIH20597720 ANKRD36 XP_011509432.1 Homo_sapiens Seq_75	9	1	112	119	0.0189960
CDIH20034073 SPEN NP_055816.2 Homo_sapiens Seq_143	11	2	110	118	0.0190950
CDIH20408788 MUC16 NP_078966.2 Homo_sapiens Seq_154	11	2	110	118	0.0190950
CDIH20496097 NAP1L4 NP_005960.1 Homo_sapiens Seq_1	11	2	110	118	0.0190950
CDIH20543684 CCDC74A XP_005263892.1 Homo_sapiens Seq_7	11	2	110	118	0.0190950
CDIH20687118 NHS1L XP_011534267.1 Homo_sapiens Seq_56	11	2	110	118	0.0190950
CDIH20191422 NPHP4 XP_011539518.1 Homo_sapiens Seq_46	31	16	90	104	0.0222360
CDIH20043725 AFF4 XP_006714650.1 Homo_sapiens Seq_15	14	4	107	116	0.0249950
CDIH20228279 LARP1B XP_011530363.1 Homo_sapiens Seq_12	14	4	107	116	0.0249950
CDIH20303082 RIBC1 XP_005262047.1 Homo_sapiens Seq_8	14	4	107	116	0.0249950
CDIH20539606 SCUBE1 XP_005261809.1 Homo_sapiens Seq_24	14	4	107	116	0.0249950
CDIH20656318 LMNB2 XP_011526681.1 Homo_sapiens Seq_8	17	6	104	114	0.0265450
CDIH20037056 PFKP NP_002618.1 Homo_sapiens Seq_15	6	0	115	120	0.0293290
CDIH20064854 DOCK3 NP_004938.1 Homo_sapiens Seq_80	6	0	115	120	0.0293290
CDIH20082521 EFS NP_115835.1 Homo_sapiens Seq_12	6	0	115	120	0.0293290
CDIH20128525 LAMA4 NP_001098676.2 Homo_sapiens Seq_21	6	0	115	120	0.0293290
CDIH20134750 INO80D XP_011509671.1 Homo_sapiens Seq_9	6	0	115	120	0.0293290
CDIH20141361 S100PBP XP_011540265.1 Homo_sapiens Seq_14	6	0	115	120	0.0293290
CDIH20177675 TUBGCP6 NP_065194.2 Homo_sapiens Seq_59	6	0	115	120	0.0293290
CDIH20180236 VPS13C NP_060154.3 Homo_sapiens Seq_16	6	0	115	120	0.0293290
CDIH20196912 MTUS2 XP_011533324.1 Homo_sapiens Seq_6	6	0	115	120	0.0293290
CDIH20201957 LAMTOR5 NP_006393.2 Homo_sapiens Seq_6	6	0	115	120	0.0293290
CDIH20241396 CSNK1D NP_001884.2 Homo_sapiens Seq_16	6	0	115	120	0.0293290
CDIH20248235 ADAMTSL4 NP_079284.2 Homo_sapiens Seq_8	6	0	115	120	0.0293290
CDIH20290287 TTN NP_597681.4 Homo_sapiens Seq_914	6	0	115	120	0.0293290
CDIH20303839 PPP1R9A NP_001159632.1 Homo_sapiens Seq_26	6	0	115	120	0.0293290
CDIH20373790 CAMTA1 XP_011539385.1 Homo_sapiens Seq_34	6	0	115	120	0.0293290
CDIH20385852 MGA XP_005254300.1 Homo_sapiens Seq_71	6	0	115	120	0.0293290
CDIH20390974 TSGA10P NP_689975.2 Homo_sapiens Seq_4	6	0	115	120	0.0293290
CDIH20401518 C10orf53 NP_872360.2 Homo_sapiens Seq_4	6	0	115	120	0.0293290
CDIH20435365 RSBN1L NP_940869.2 Homo_sapiens Seq_32	6	0	115	120	0.0293290
CDIH20447277 ARMCX4 NP_001243084.2 Homo_sapiens Seq_30	6	0	115	120	0.0293290
CDIH20470514 PKHD1L1 NP_803875.2 Homo_sapiens Seq_175	6	0	115	120	0.0293290
CDIH20481831 AGER NP_00193858.1 Homo_sapiens Seq_7	6	0	115	120	0.0293290
CDIH20487675 ACTN3 NP_001095.2 Homo_sapiens Seq_8	6	0	115	120	0.0293290
CDIH20515110 C5orf60 NP_001292317.1 Homo_sapiens Seq_7	6	0	115	120	0.0293290
CDIH20519070 DUSP27 XP_005245660.1 Homo_sapiens Seq_16	6	0	115	120	0.0293290
CDIH20520438 DYNC1I2 XP_005246420.1 Homo_sapiens Seq_19	6	0	115	120	0.0293290
CDIH20552254 USP33 XP_005270706.1 Homo_sapiens Seq_17	6	0	115	120	0.0293290
CDIH20574844 NACA XP_006719477.1 Homo_sapiens Seq_69	6	0	115	120	0.0293290
CDIH20590090 POGZ XP_011507632.1 Homo_sapiens Seq_41	6	0	115	120	0.0293290
CDIH20590996 GON4L XP_011507960.1 Homo_sapiens Seq_57	6	0	115	120	0.0293290
CDIH20604427 PKP4 XP_011510318.1 Homo_sapiens Seq_9	6	0	115	120	0.0293290
CDIH20641625 ZNF624 XP_011522273.1 Homo_sapiens Seq_7	6	0	115	120	0.0293290
CDIH20650249 RASIP1 XP_011525355.1 Homo_sapiens Seq_34	6	0	115	120	0.0293290
CDIH20706629 CROCC XP_011540771.1 Homo_sapiens Seq_22	6	0	115	120	0.0293290
CDIH20728254 INPP5D XP_011547153.1 Homo_sapiens Seq_40	6	0	115	120	0.0293290
CDIHD8100119 Angiotensin_1-4 hor004209 Homo_sapiens Seq_1d_right	6	0	115	120	0.0293290
CDIHD8104033 Pituitary_adenylate_cyclase-activating_polypeptide_27 hor001969 Homo_sapiens Seq_1c_left	6	0	115	120	0.0293290
CDIHD8104975 Pituitary_adenylate_cyclase-activating_polypeptide_27 hor001969 Homo_sapiens Seq_1e_repeat	6	0	115	120	0.0293290

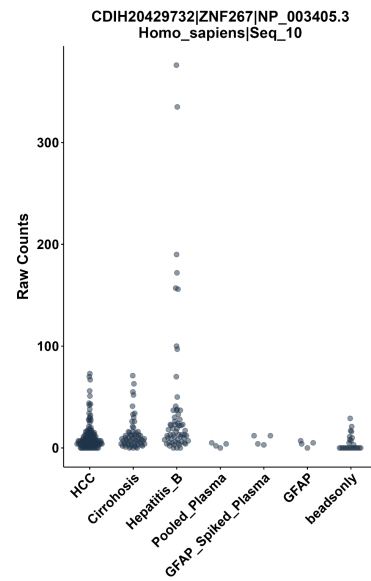
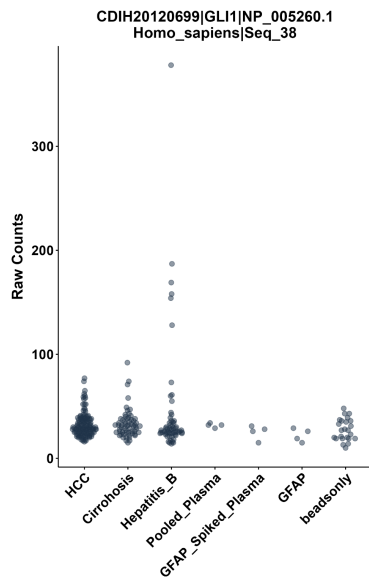
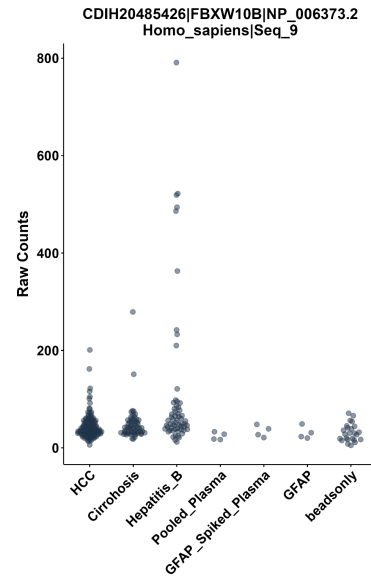
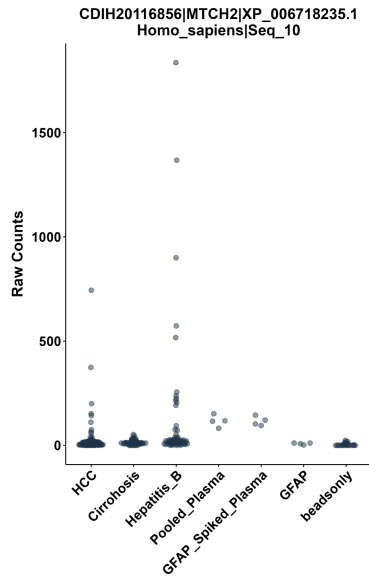
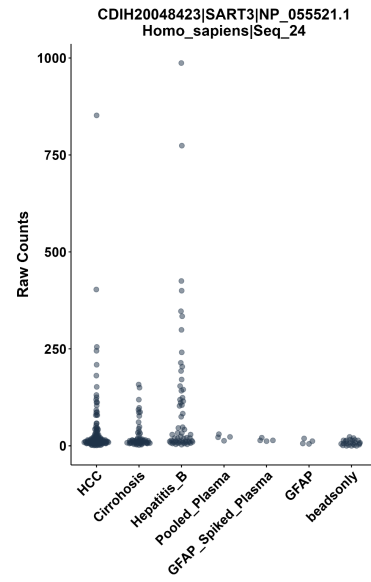
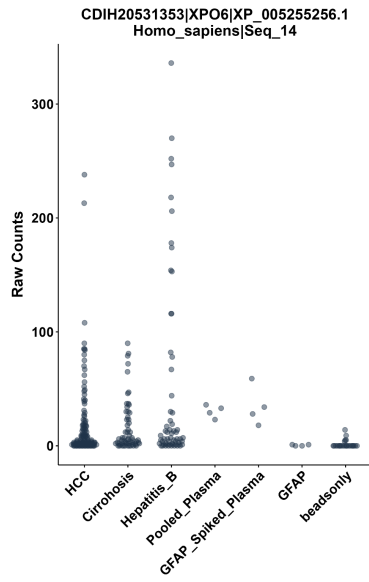
fisher_exact_tests/Vietnam_HCC_vs_cohort_fisher_exact_tests.csv



Vietnam Hepatitis B vs cohort

SeqID	Hits Hepatitis B	Hits Cohort	No Hits Hepatitis B	No Hits Cohort	p value
CDIH20531353 XPO6 XP_005255256.1 Homo_sapiens Seq_14	13	2	47	179	3.3166e-07
CDIH20048423 SART3 NP_055521.1 Homo_sapiens Seq_24	21	18	39	163	2.2997e-05
CDIH20116856 MTCH2 XP_006718235.1 Homo_sapiens Seq_10	11	4	49	177	6.1768e-05
CDIH20485426 FBXW10B NP_006373.2 Homo_sapiens Seq_9	9	2	51	179	8.0494e-05
CDIH20120699 GLI1 NP_005260.1 Homo_sapiens Seq_38	6	0	54	181	0.00019588
CDIH20429732 ZNF267 NP_003405.3 Homo_sapiens Seq_10	6	0	54	181	0.00019588
CDIH20280633 MEPE NP_001171623.1 Homo_sapiens Seq_14	10	4	50	177	0.00020785
CDIH20021911 DSG2 NP_001934.2 Homo_sapiens Seq_38	9	3	51	178	0.00025811
CDIH20337470 LRFN5 NP_689660.2 Homo_sapiens Seq_24	9	4	51	177	0.0006723
CDIH20120728 GLI1 XP_011536492.1 Homo_sapiens Seq_36	5	0	55	181	0.00084051
CDIH20242042 MBD1 NP_001191065.1 Homo_sapiens Seq_12	5	0	55	181	0.00084051
CDIH20174963 CTAG2 NP_066274.2 Homo_sapiens Seq_6	8	3	52	178	0.00086611
CDIH20535130 LRFN1 XP_005259159.1 Homo_sapiens Seq_5	8	3	52	178	0.00086611
CDIH20118561 KIAA1549L XP_005252905.1 Homo_sapiens Seq_56	7	2	53	179	0.0010418
CDIH20165807 PCK1 NP_002582.3 Homo_sapiens Seq_20	7	2	53	179	0.0010418
CDIH20178126 RNF112 XP_006721634.1 Homo_sapiens Seq_15	7	2	53	179	0.0010418
CDIH20303189 NRG2 NP_053585.1 Homo_sapiens Seq_2	7	2	53	179	0.0010418
CDIH20397677 ZBED2 NP_078784.2 Homo_sapiens Seq_7	10	6	50	175	0.0010688
CDIH20005322 GTF2IRD1 NP_005676.3 Homo_sapiens Seq_38	6	1	54	180	0.0011011
CDIH20005371 GTF2IRD1 XP_006716245.1 Homo_sapiens Seq_38	6	1	54	180	0.0011011
CDIH20266832 HMCN2 NP_001278744.1 Homo_sapiens Seq_68	6	1	54	180	0.0011011
CDIH20441910 BEND4 NP_997289.2 Homo_sapiens Seq_22	6	1	54	180	0.0011011
CDIH20526082 DST XP_005249367.1 Homo_sapiens Seq_127	6	1	54	180	0.0011011
CDIH20581006 LTN1 XP_006724050.1 Homo_sapiens Seq_3	6	1	54	180	0.0011011
CDIH20123073 ATN1 NP_001931.2 Homo_sapiens Seq_21	19	23	41	158	0.0014886
CDIH20618448 CFTR XP_011514053.1 Homo_sapiens Seq_31	9	5	51	176	0.0015083
CDIH20250818 PYCR2 NP_001258610.1 Homo_sapiens Seq_7	10	7	50	174	0.0020801
CDIH20625056 PTPRD XP_011516295.1 Homo_sapiens Seq_61	14	14	46	167	0.0021821
CDIH20580989 ADARB1 XP_006724017.1 Homo_sapiens Seq_23	11	9	49	172	0.0025267
CDIH20122641 NOP2 NP_001245237.1 Homo_sapiens Seq_14	7	3	53	178	0.002785
CDIH20290140 TTN NP_597681.4 Homo_sapiens Seq_767	7	3	53	178	0.002785
CDIH20322334 TEX55 NP_689752.2 Homo_sapiens Seq_11	7	3	53	178	0.002785
CDIH20371376 KSR2 XP_011536526.1 Homo_sapiens Seq_16	7	3	53	178	0.002785
CDIH20440650 RALGAP2 XP_006723661.1 Homo_sapiens Seq_34	7	3	53	178	0.002785
CDIH20719727 LOC100996415 XP_011544458.1 Homo_sapiens Seq_8	7	3	53	178	0.002785
CDIH20592723 PPFIA4 XP_011508371.1 Homo_sapiens Seq_26	12	10	48	171	0.0029986
CDIH20160524 ZMIZ2 XP_005249923.1 Homo_sapiens Seq_2	9	6	51	175	0.0030207
CDIH20684326 CDK8 XP_011533165.1 Homo_sapiens Seq_8	9	6	51	175	0.0030207
CDIH20176565 WFIKK1 NP_444514.1 Homo_sapiens Seq_14	17	20	43	161	0.0031073
CDIH20063689 TXNDC16 NP_065835.2 Homo_sapiens Seq_12	6	2	54	179	0.0035382
CDIH20235044 TLE3 XP_011520278.1 Homo_sapiens Seq_5	6	2	54	179	0.0035382
CDIH20350044 DPEP2 XP_011521570.1 Homo_sapiens Seq_18	6	2	54	179	0.0035382
CDIH20423917 PRR16 NP_001287712.1 Homo_sapiens Seq_3	6	2	54	179	0.0035382
CDIH20496494 LILRA2 NP_006857.2 Homo_sapiens Seq_18	6	2	54	179	0.0035382
CDIH20535747 CDC114 XP_005259471.1 Homo_sapiens Seq_8	6	2	54	179	0.0035382
CDIH20540857 STAG2 XP_005262414.1 Homo_sapiens Seq_44	6	2	54	179	0.0035382
CDIH20551621 TCF7L2 XP_005270128.1 Homo_sapiens Seq_25	6	2	54	179	0.0035382
CDIH20662268 PHF21B XP_011528158.1 Homo_sapiens Seq_1	6	2	54	179	0.0035382
CDIH20664932 LOC101927859 XP_011528928.1 Homo_sapiens Seq_53	6	2	54	179	0.0035382
CDIH20665047 LOC100996335 XP_011528942.1 Homo_sapiens Seq_12	6	2	54	179	0.0035382
CDIH20054951 COL5A3 NP_056534.2 Homo_sapiens Seq_31	4	0	56	181	0.0035572
CDIH20101035 PLEKHA4 XP_011525459.1 Homo_sapiens Seq_21	4	0	56	181	0.0035572
CDIH20136449 NDUFS7 NP_077718.3 Homo_sapiens Seq_2	4	0	56	181	0.0035572
CDIH20170503 DDX31 NP_073616.6 Homo_sapiens Seq_11	4	0	56	181	0.0035572
CDIH20218525 PIM1 NP_001230115.1 Homo_sapiens Seq_2	4	0	56	181	0.0035572
CDIH20254655 CAND2 XP_011531805.1 Homo_sapiens Seq_24	4	0	56	181	0.0035572
CDIH20257543 ANKRD31 XP_011541603.1 Homo_sapiens Seq_54	4	0	56	181	0.0035572
CDIH20289996 TTN NP_597681.4 Homo_sapiens Seq_623	4	0	56	181	0.0035572
CDIH20389911 CKAP5 NP_055571.2 Homo_sapiens Seq_77	4	0	56	181	0.0035572
CDIH20422245 MAML2 XP_011541325.1 Homo_sapiens Seq_33	4	0	56	181	0.0035572
CDIH20441049 CC2D2B NP_001123918.2 Homo_sapiens Seq_3	4	0	56	181	0.0035572
CDIH20452538 KANK2 NP_056308.3 Homo_sapiens Seq_32	4	0	56	181	0.0035572
CDIH20461737 NCOA6 XP_011527023.1 Homo_sapiens Seq_32	4	0	56	181	0.0035572
CDIH20482375 MUC13 NP_543146.2 Homo_sapiens Seq_41	4	0	56	181	0.0035572
CDIH20498871 KIR2DL5B NP_001018091.1 Homo_sapiens Seq_10	4	0	56	181	0.0035572
CDIH20501641 MUC6 NP_005952.2 Homo_sapiens Seq_64	4	0	56	181	0.0035572
CDIH20659046 GDF5 XP_011527377.1 Homo_sapiens Seq_16	4	0	56	181	0.0035572
CDIH20659402 SIRPB1 XP_011527432.1 Homo_sapiens Seq_5	4	0	56	181	0.0035572
CDIH20692603 PRPF40B XP_011536437.1 Homo_sapiens Seq_15	4	0	56	181	0.0035572
CDIH20014042 ADGRA2 NP_116166.9 Homo_sapiens Seq_13	5	1	55	180	0.0040636
CDIH2011780 PBLD NP_071412.2 Homo_sapiens Seq_9	5	1	55	180	0.0040636
CDIH20112887 CAMTA2 NP_001164638.1 Homo_sapiens Seq_14	5	1	55	180	0.0040636
CDIH20156385 CHRNA2 NP_000733.2 Homo_sapiens Seq_16	5	1	55	180	0.0040636
CDIH20200776 ZFC3H1 NP_659419.3 Homo_sapiens Seq_4	5	1	55	180	0.0040636
CDIH20205410 DZIP1 NP_945319.1 Homo_sapiens Seq_27	5	1	55	180	0.0040636
CDIH20540254 AMOT XP_005262144.1 Homo_sapiens Seq_5	5	1	55	180	0.0040636
CDIH20586153 LOC728715 XP_011506650.1 Homo_sapiens Seq_2	5	1	55	180	0.0040636
CDIH20589096 LOC100996335 XP_011507252.1 Homo_sapiens Seq_8	5	1	55	180	0.0040636
CDIH20620834 ZYX XP_011514869.1 Homo_sapiens Seq_12	5	1	55	180	0.0040636
CDIH20627872 WINK2 XP_011517235.1 Homo_sapiens Seq_91	5	1	55	180	0.0040636
CDIH20644352 TMUB2 XP_011523501.1 Homo_sapiens Seq_6	5	1	55	180	0.0040636
CDIH20664871 LOC101927859 XP_011528926.1 Homo_sapiens Seq_40	5	1	55	180	0.0040636
CDIH20665043 LOC100996335 XP_011528942.1 Homo_sapiens Seq_8	5	1	55	180	0.0040636
CDIH20178761 CCDC136 NP_073579.4 Homo_sapiens Seq_4	8	5	52	176	0.0043378
CDIH20154638 TTLL2 NP_114155.4 Homo_sapiens Seq_5	14	15	46	166	0.0047564
CDIH20675122 RGPD1 XP_011531149.1 Homo_sapiens Seq_24	9	7	51	174	0.0055302
CDIH20049697 RAP1GAP NP_001139129.1 Homo_sapiens Seq_19	11	10	49	171	0.0060485
CDIH20052840 OSBPL6 NP_001188411.1 Homo_sapiens Seq_31	7	4	53	177	0.0061431
CDIH20060947 ATRX XP_005262211.2 Homo_sapiens Seq_49	7	4	53	177	0.0061431
CDIH20606458 OSBPL1 XP_011510686.1 Homo_sapiens Seq_5	7	4	53	177	0.0061431
CDIH20657487 ASXL1 XP_011526951.1 Homo_sapiens Seq_34	7	4	53	177	0.0061431
CDIH20692278 CUX2 XP_011536363.1 Homo_sapiens Seq_28	7	4	53	177	0.0061431
CDIH20719787 LOC105369191 XP_011544466.1 Homo_sapiens Seq_6	7	4	53	177	0.0061431
CDIH20721502 LOC105374297 XP_011544763.1 Homo_sapiens Seq_2	7	4	53	177	0.0061431
CDIH20718812 TNRC6A XP_011544095.1 Homo_sapiens Seq_52	17	22	43	159	0.0075244
CDIH20246512 POGK XP_011508117.1 Homo_sapiens Seq_19	13	14	47	167	0.0076425
CDIH20009892 FMO1 NP_001269621.1 Homo_sapiens Seq_22	8	6	52	175	0.0081105
CDIH20091034 DHODH NP_001352.2 Homo_sapiens Seq_1	8	6	52	175	0.0081105
CDIH20643190 LGALS9 XP_011523098.1 Homo_sapiens Seq_9	8	6	52	175	0.0081105
CDIH20010724 ANLN NP_001271230.1 Homo_sapiens Seq_31	6	3	54	178	0.0085311

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_fisher_exact_tests.csv



12 Protein-Level Tests

Protein-level Fisher exact tests are performed on a 2x2 matrix of *Protein_Hits* and *Protein_No_Hits* columns for each provided group versus all other samples in the study cohort using the *fisher.test* function in R. These tests only require a single-reactive peptide within a full protein; outputs are otherwise analogous to peptide-level tests.

A sample-level unclustered, clustered, and row clustered visualization heatmap is provided for the top 100 rows of each table in the **summary_heatmaps** folder. These figures display the top peptide reaction per-protein.

A whole-protein epitope-scanning heatmap is provided in the **scan heatmaps** folder for each protein bearing one of the top 100 peptides enriched in each sample group.

Vietnam Cirrhosis vs cohort protein

Gene Name	Protein	Protein Hits Cirrhosis	Protein Hits Cohort	Protein No Hits Cirrhosis	Protein No Hits Cohort	p value
KAZN	NP_001018000.1	8	1	52	180	0.0000736
Anakinra	hd00071	11	5	49	176	0.0001587
SRPK2	XP_011514833.1	6	1	54	180	0.0011011
KIF1A	XP_005247079.1	11	8	49	173	0.0014184
CHD2	NP_001262.3	13	11	47	170	0.0018129
SPINT1	NP_857593.1	7	3	53	178	0.0027850
HHIP1	NP_001120730.1	7	3	53	178	0.0027850
NECTIN2	XP_011525494.1	7	3	53	178	0.0027850
SNX25	XP_011530604.1	7	3	53	178	0.0027850
BNP5-31	hor003004	7	3	53	178	0.0027850
NBEAL1	XP_006712761.1	21	29	39	152	0.0029621
MED12	XP_005262374.1	32	56	28	125	0.0030582
RALGPS1	XP_011517529.1	4	0	56	181	0.0035572
SLC4A7	XP_011532566.1	4	0	56	181	0.0035572
Relaxin_B_chain	hor002225	4	0	56	181	0.0035572
RIMS4	NP_001192246.1	5	1	55	180	0.0040636
KLHL5	NP_057074.3	5	1	55	180	0.0040636
PRLR	NP_001191244.1	5	1	55	180	0.0040636
LGR6	NP_001017403.1	5	1	55	180	0.0040636
WTAP	NP_004897.2	5	1	55	180	0.0040636
MTMR12	XP_011512360.1	5	1	55	180	0.0040636
EXTL3	NP_001431.1	8	5	52	176	0.0043378
ZRANB2	NP_976225.1	9	7	51	174	0.0055302
UPP1	NP_001274355.1	7	4	53	177	0.0061431
C1orf174	NP_997239.2	7	4	53	177	0.0061431
AGBL2	XP_005253195.1	7	4	53	177	0.0061431
BLTP3A	NP_060224.3	28	49	32	132	0.0064827
BUD13	XP_011541337.1	21	32	39	149	0.0069459
MCF2L	XP_011535784.1	13	14	47	167	0.0076425
REPS1	NP_001273540.1	8	6	52	175	0.0081105
EEF1D	XP_006725127.1	8	6	52	175	0.0081105
CRACDL	XP_005263980.1	11	11	49	170	0.0081184
TPM3	XP_011508254.1	11	11	49	170	0.0081184
EPB41L1	XP_011526969.1	11	11	49	170	0.0081184
SRSF7	NP_001026854.1	6	3	54	178	0.0085311
GIMAP8	XP_005250008.1	6	3	54	178	0.0085311
PTK2	XP_011515488.1	6	3	54	178	0.0085311
NSUN6	XP_011517681.1	6	3	54	178	0.0085311
PEX16	XP_011518776.1	6	3	54	178	0.0085311
NEUROG1	NP_006152.2	13	15	47	166	0.0092193
GUCY2F	NP_001513.2	11	12	49	169	0.0114040
MAP3K1	XP_011541709.1	11	12	49	169	0.0114040
TRIM41	NP_291027.3	5	2	55	179	0.0114700
TCTN2	NP_079085.2	5	2	55	179	0.0114700
OLIG2	XP_005260965.1	5	2	55	179	0.0114700
RAMP3	XP_006715694.1	5	2	55	179	0.0114700
LOC101929304	XP_011521101.1	5	2	55	179	0.0114700
ZFPM1	XP_011521215.1	5	2	55	179	0.0114700
RECQL5	XP_011523787.1	5	2	55	179	0.0114700
FGFR2	XP_011537812.1	5	2	55	179	0.0114700
Progonadoliberin-1	hor002039	5	2	55	179	0.0114700
DENND4B	XP_005245735.1	13	16	47	165	0.0118240
MAP7D1	XP_005271084.1	13	16	47	165	0.0118240
POMGNT1	NP_001277058.1	7	5	53	176	0.0118280
CYREN	NP_001230681.1	7	5	53	176	0.0118280
LNPK	NP_001291938.1	7	5	53	176	0.0118280
STK11IP	XP_011508828.1	7	5	53	176	0.0118280
IQCJ-SCHIP1	NP_055390.1	7	5	53	176	0.0118280
GAB1	XP_005262938.1	7	5	53	176	0.0118280
AXIN1	XP_011520987.1	7	5	53	176	0.0118280
ZNF536	XP_011525861.1	7	5	53	176	0.0118280
NSD2	XP_005248058.1	16	22	44	159	0.0129840
SMAP1	NP_001037770.1	8	7	52	174	0.0139270
USP36	XP_011523367.1	12	14	48	167	0.0143830
PDCD6IP	NP_001155901.1	12	14	48	167	0.0143830
PLXND1	XP_011510892.1	4	1	56	180	0.0144240
MAN2B2	NP_056089.1	4	1	56	180	0.0144240
TJAP1	NP_001139488.1	4	1	56	180	0.0144240
RANBP10	XP_005256127.1	4	1	56	180	0.0144240
DNAH3	NP_060009.1	4	1	56	180	0.0144240
NTM	NP_057606.1	4	1	56	180	0.0144240
PADI6	NP_997304.3	4	1	56	180	0.0144240
USP17L7	NP_001243798.1	4	1	56	180	0.0144240
KY	XP_005247475.1	4	1	56	180	0.0144240
GSTCD	XP_005263279.1	4	1	56	180	0.0144240
DLG1	XP_005269358.1	4	1	56	180	0.0144240
LOC101930115	XP_011506874.1	4	1	56	180	0.0144240
KCNE1	XP_011527859.1	4	1	56	180	0.0144240
STEEP1	NP_001164041.1	3	0	57	181	0.0148530
ATP1A2	NP_000693.1	3	0	57	181	0.0148530
F7	NP_001254483.1	3	0	57	181	0.0148530
REC8	XP_011535723.1	3	0	57	181	0.0148530
MGME1	XP_005260924.1	3	0	57	181	0.0148530
ZNF20	NP_066966.2	3	0	57	181	0.0148530
PTPN22	XP_011539523.1	3	0	57	181	0.0148530
PTPN22	NP_036543.4	3	0	57	181	0.0148530
KRT80	NP_872313.2	3	0	57	181	0.0148530
HLA-DQB1	NP_002114.3	3	0	57	181	0.0148530
ZNF429	NP_001001415.2	3	0	57	181	0.0148530
ZNF69	XP_006722936.1	3	0	57	181	0.0148530
HLA-C	NP_002108.4	3	0	57	181	0.0148530
TOP3A	NP_004609.1	3	0	57	181	0.0148530
CCDC148	XP_005246376.1	3	0	57	181	0.0148530
IL1RAP	XP_005247490.1	3	0	57	181	0.0148530
UBAC2	XP_011519387.1	3	0	57	181	0.0148530
OCA2	XP_011519941.1	3	0	57	181	0.0148530
EMC8	XP_011521113.1	3	0	57	181	0.0148530
RABL2B	XP_011528970.1	3	0	57	181	0.0148530
LOC105373244	XP_011529383.1	3	0	57	181	0.0148530
GP2	XP_011544099.1	3	0	57	181	0.0148530

fisher_exact_tests/Vietnam_Cirrhosis_vs_cohort_protein_fisher_exact_tests.csv

Vietnam HCC vs cohort protein

Gene Name	Protein	Protein Hits HCC	Protein Hits Cohort	Protein No Hits HCC	Protein No Hits Cohort	p value
GON4L	NP_001269785.1	20	2	101	118	0.0000601
RG56	XP_011535690.1	42	16	79	104	0.0001389
XRCC4	XP_005248652.1	16	1	105	119	0.0001718
OBSCN	XP_005273344.1	72	44	49	76	0.0004871
SPAG8	NP_758516.1	26	7	95	113	0.0005688
ASB2	XP_011544947.1	19	4	102	116	0.0015902
LOC400499	XP_006721060.2	28	10	93	110	0.0023319
PLXNB3	NP_001156729.1	20	5	101	115	0.0024756
ITIH4	NP_002209.2	12	1	109	119	0.0026814
RG56	NP_004287.3	26	9	95	111	0.0030175
ERICH1	NP_997215.1	16	3	105	117	0.0030923
ACADVL	NP_001257376.1	14	2	107	118	0.0031049
GCLM	NP_002052.1	9	0	112	120	0.0033480
NPNT	XP_011530122.1	9	0	112	120	0.0033480
AGER	NP_001193858.1	9	0	112	120	0.0033480
DEDD2	XP_011524871.1	9	0	112	120	0.0033480
RASGRF1	NP_002882.3	24	8	97	112	0.0038566
DCC	NP_005206.2	19	5	102	115	0.0042907
ATF6B	NP_004372.3	28	11	93	109	0.0046834
MOK	NP_055041.1	29	12	92	108	0.0056323
TRIP12	XP_005247015.1	23	8	98	112	0.0063338
OBSCN	XP_005273359.1	45	25	76	95	0.0068356
DKK3	NP_056965.3	8	0	113	120	0.0069331
RSBN1L	NP_940869.2	8	0	113	120	0.0069331
CLEC4F	XP_005264229.1	8	0	113	120	0.0069331
SNAPC4	XP_006717307.1	8	0	113	120	0.0069331
TRIM46	XP_011508305.1	8	0	113	120	0.0069331
WIPF1	XP_011510082.1	8	0	113	120	0.0069331
INPP5D	XP_011547153.1	8	0	113	120	0.0069331
BUD13	NP_001153208.1	16	4	105	116	0.0086798
MADD	XP_011518730.1	16	4	105	116	0.0086798
ADAMTSL4	NP_079284.2	14	3	107	117	0.0098842
MST1L	NP_001258662.1	14	3	107	117	0.0098842
GNL2	XP_011539602.1	14	3	107	117	0.0098842
ZNF410	NP_001229857.1	10	1	111	119	0.0100030
URB1	NP_055640.2	10	1	111	119	0.0100030
KHSRP	XP_005259725.1	10	1	111	119	0.0100030
TICAM1	NP_891549.1	12	2	109	118	0.0105660
LPIN3	XP_011527298.1	12	2	109	118	0.0105660
ZNF74	NP_003417.2	12	2	109	118	0.0105660
CCDC74A	XP_005263892.1	12	2	109	118	0.0105660
PRPF40B	XP_011536439.1	12	2	109	118	0.0105660
ZNF280D	NP_060131.2	45	26	76	94	0.0107680
GTF3C1	XP_011544117.1	17	5	104	115	0.0123440
KIAA1217	NP_001091970.1	35	18	86	102	0.0124230
AFF4	XP_006714650.1	20	7	101	113	0.0130600
CYP26C1	NP_899230.2	20	7	101	113	0.0130600
PPP1R9A	NP_001159632.1	31	15	90	105	0.0134020
SPDL1	NP_060255.3	7	0	114	120	0.0142920
PRPF6	NP_036601.2	7	0	114	120	0.0142920
STX2	NP_001971.2	7	0	114	120	0.0142920
LAMTOR5	NP_006393.2	7	0	114	120	0.0142920
FCRL5	XP_011508332.1	7	0	114	120	0.0142920
ANKRD30A	XP_011518059.1	7	0	114	120	0.0142920
PLCB3	NP_001171812.1	7	0	114	120	0.0142920
CACNB4	NP_001139270.1	7	0	114	120	0.0142920
ANKRD20A1	XP_011543983.1	7	0	114	120	0.0142920
MTUS2	XP_011533324.1	15	4	106	116	0.0148470
UBAP1	NP_001164672.1	15	4	106	116	0.0148470
ADAMTS20	XP_011537056.1	15	4	106	116	0.0148470
GNL2	NP_037417.1	24	10	97	110	0.0152350
LIMCH1	XP_011511950.1	21	8	100	112	0.0163410
PDCD11	XP_011537840.1	18	6	103	114	0.0164560
LMNB2	XP_011526681.1	18	6	103	114	0.0164560
TRIP10	XP_006723003.1	13	3	108	117	0.0172910
RAP1GAP2	NP_001093868.1	13	3	108	117	0.0172910
BRD1	XP_011528366.1	13	3	108	117	0.0172910
PPP2R5C	NP_848701.1	9	1	112	119	0.0189960
CTCF	NP_001255973.1	9	1	112	119	0.0189960
LEF1	XP_005263103.1	9	1	112	119	0.0189960
ZNF653	NP_620138.2	9	1	112	119	0.0189960
SSX7	NP_775494.1	9	1	112	119	0.0189960
EVC2	XP_011511694.1	9	1	112	119	0.0189960
IKZF2	XP_011509118.1	11	2	110	118	0.0190950
MED26	NP_004822.2	11	2	110	118	0.0190950
MAP3K3	NP_002392.2	11	2	110	118	0.0190950
NAP1L4	NP_005960.1	11	2	110	118	0.0190950
ZNF160	XP_011525765.1	11	2	110	118	0.0190950
ZFXH2	XP_011535552.1	11	2	110	118	0.0190950
TSC2	NP_001107854.1	22	9	99	111	0.0198260
FRMD4A	XP_011517843.1	16	5	105	115	0.0204590
CHTF18	XP_005255528.1	26	12	95	108	0.0207300
USP9X	XP_005272732.1	19	7	102	113	0.0209170
LAMB2	XP_005265184.1	36	20	85	100	0.0217170
TCEA3	XP_011540357.1	23	10	98	110	0.0234680
SETBP1	NP_056374.2	14	4	107	116	0.0249950
DCAKD	NP_001275583.1	14	4	107	116	0.0249950
KRT73	NP_778238.1	20	8	101	112	0.0256410
NUP188	NP_056169.1	17	6	104	114	0.0265450
CD6	NP_006716.3	6	0	115	120	0.0293290
NRIP2	NP_113662.1	6	0	115	120	0.0293290
DOCK3	NP_004938.1	6	0	115	120	0.0293290
OSBP	NP_002547.1	6	0	115	120	0.0293290
S100PBP	XP_011540265.1	6	0	115	120	0.0293290
RNF19B	XP_006710419.1	6	0	115	120	0.0293290
CLCC1	NP_001041675.1	6	0	115	120	0.0293290
STRIP2	NP_065755.1	6	0	115	120	0.0293290
MTUS1	NP_001001931.1	6	0	115	120	0.0293290
PAK4	XP_011524618.1	6	0	115	120	0.0293290
KIF3A	NP_001287720.1	6	0	115	120	0.0293290

fisher_exact_tests/Vietnam_HCC_vs_cohort_protein_fisher_exact_tests.csv

Vietnam Hepatitis B vs cohort protein

Gene Name	Protein	Protein Hits Hepatitis B	Protein Hits Cohort	Protein No Hits Hepatitis B	Protein No Hits Cohort	p value
SART3	NP_055521.1	24	22	36	159	7.4917e-06
XPO6	XP_005255256.1	14	6	46	175	1.0864e-05
SH2D3C	NP_005480.2	12	5	48	176	4.8409e-05
LOC100996335	XP_011528942.1	14	8	46	173	5.8404e-05
MTCH2	XP_006718235.1	11	4	49	177	6.1768e-05
ZNF267	NP_003405.3	6	0	54	181	0.00019588
FBXW10B	NP_006373.2	9	3	51	178	0.00025811
AXIN2	XP_011523624.1	8	2	52	179	0.00029475
MEPE	NP_001171623.1	11	6	49	175	0.00036098
LOC101927859	XP_011528926.1	28	41	32	140	0.00083997
PEX6	XP_005249237.1	5	0	55	181	0.00084051
RNF112	XP_006721634.1	7	2	53	179	0.0010418
HAPLN3	NP_001294881.1	7	2	53	179	0.0010418
PHF21B	XP_011528158.1	7	2	53	179	0.0010418
GTF2IRD1	XP_006716245.1	6	1	54	180	0.0011011
EFCAB5	XP_011523061.1	6	1	54	180	0.0011011
ZNF264	XP_011525823.1	6	1	54	180	0.0011011
STAG2	XP_005262414.1	11	8	49	173	0.0014184
ANKRD6	NP_001229742.1	16	17	44	164	0.001834
LOC100996415	XP_011544458.1	16	17	44	164	0.001834
ZMIZ2	XP_005249923.1	10	7	50	174	0.0020801
PYCR2	NP_001258610.1	10	7	50	174	0.0020801
ZBED2	NP_078784.2	10	7	50	174	0.0020801
CTAG2	NP_066274.2	8	4	52	177	0.0020821
ENAH	XP_006711860.1	8	4	52	177	0.0020821
DMD	XP_006724535.1	8	4	52	177	0.0020821
LOC100996335	XP_011507252.1	8	4	52	177	0.0020821
ADARB1	XP_006724017.1	14	14	46	167	0.0021821
PTPRD	XP_011516295.1	14	14	46	167	0.0021821
NRG2	NP_053585.1	11	9	49	172	0.0025267
CCDC114	XP_005259471.1	11	9	49	172	0.0025267
KALRN	NP_008995.2	7	3	53	178	0.002785
KBTBD11	XP_011544870.1	7	3	53	178	0.002785
PCK1	NP_002582.3	9	6	51	175	0.0030207
CLDN8	NP_955360.1	9	6	51	175	0.0030207
EPHA1	NP_005223.4	13	13	47	168	0.0033817
OBSL1	XP_011509156.1	13	13	47	168	0.0033817
TXNDC16	NP_065835.2	6	2	54	179	0.0035382
CHRNA2	NP_000733.2	6	2	54	179	0.0035382
DPEP2	XP_011521570.1	6	2	54	179	0.0035382
PRR16	NP_001287712.1	6	2	54	179	0.0035382
MAGEE2	NP_619648.1	6	2	54	179	0.0035382
ZNF628	NP_149104.3	6	2	54	179	0.0035382
LILRA2	NP_006857.2	6	2	54	179	0.0035382
SIRPB1	XP_011527432.1	6	2	54	179	0.0035382
EP300	NP_001420.2	4	0	56	181	0.0035572
WASL	NP_003932.3	4	0	56	181	0.0035572
PIM1	NP_001230115.1	4	0	56	181	0.0035572
RCAN1	NP_001272320.2	4	0	56	181	0.0035572
GAB3	XP_011529405.1	4	0	56	181	0.0035572
TXNRD2	NP_006431.2	4	0	56	181	0.0035572
CC2D2B	NP_001123918.2	4	0	56	181	0.0035572
SMCO1	NP_001071125.1	4	0	56	181	0.0035572
ARPIN-AP352	NP_001185987.1	4	0	56	181	0.0035572
DIRAS3	XP_006711090.2	4	0	56	181	0.0035572
CDK8	XP_011533165.1	10	8	50	173	0.0037493
RIMBP2	XP_011536405.1	5	1	55	180	0.0040636
PBLD	NP_071412.2	5	1	55	180	0.0040636
PTCD3	NP_060422.4	5	1	55	180	0.0040636
SEMA6C	NP_112175.2	5	1	55	180	0.0040636
WDPCP	XP_005264405.1	5	1	55	180	0.0040636
ALPP	NP_001623.3	5	1	55	180	0.0040636
HEXD	NP_775891.2	5	1	55	180	0.0040636
FSCB	NP_115511.3	5	1	55	180	0.0040636
MRNIP	NP_001017987.1	5	1	55	180	0.0040636
IRF7	XP_006725200.1	5	1	55	180	0.0040636
PPP2R2C	XP_011511797.1	5	1	55	180	0.0040636
ZYX	XP_011514869.1	5	1	55	180	0.0040636
LOC729461	XP_011544460.1	19	26	41	155	0.0042126
MUC4	NP_612154.2	37	72	23	109	0.004284
DLX6	NP_005213.3	8	5	52	176	0.0043378
NPR1	XP_006711405.1	8	5	52	176	0.0043378
PTPDC1	XP_011516544.1	8	5	52	176	0.0043378
TLL2	NP_114155.4	16	19	44	162	0.0049934
ZNF638	XP_011531070.1	12	12	48	169	0.0052372
TXNDC2	NP_001091999.1	9	7	51	174	0.0055302
NEB	XP_011509529.1	9	7	51	174	0.0055302
PPFIA4	XP_011508371.1	14	16	46	165	0.0058996
LTN1	XP_006724050.1	7	4	53	177	0.0061431
LOC105375305	XP_011513988.1	7	4	53	177	0.0061431
WFIKN1	NP_444514.1	17	22	43	159	0.0075244
POLD1	XP_011525340.1	8	6	52	175	0.0081105
LOC102724988	XP_011544756.1	8	6	52	175	0.0081105
RAP1GAP	NP_001139129.1	11	11	49	170	0.0081184
ASCC2	XP_011528744.1	11	11	49	170	0.0081184
AGRN	XP_011539732.1	19	28	41	153	0.0084585
ADAMTS8	NP_008968.4	6	3	54	178	0.0085311
TLE3	XP_011520278.1	6	3	54	178	0.0085311
TCF7L2	XP_005270128.1	6	3	54	178	0.0085311
SNTG2	XP_011508666.1	6	3	54	178	0.0085311
POGK	XP_011508117.1	13	15	47	166	0.0092193
CD163	NP_004235.4	13	15	47	166	0.0092193
RGPD1	XP_011531149.1	10	9	50	172	0.009719
WWC3	NP_056506.2	11	12	49	169	0.011404
CUX2	XP_011536363.1	11	12	49	169	0.011404
ARSF	XP_011543824.1	5	2	55	179	0.01147
NDUF57	NP_077718.3	5	2	55	179	0.01147
TPP2	XP_005254127.1	5	2	55	179	0.01147
ITIH2	NP_002207.2	5	2	55	179	0.01147
CDC25C	NP_001274512.1	5	2	55	179	0.01147

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_protein_fisher_exact_tests.csv

13 Protein-Level Polyclonal Tests

To count as a ‘polyclonal hit’ - a sample must have ≥ 3 unique edgeR called hits within a single protein. Fisher exact tests are then performed on a 2x2 matrix of *Polyclonal_Hits* and *Polyclonal_No_Hits* columns for each provided group versus all other samples in the study cohort using the *fisher.test* function in R; outputs are otherwise analogous to peptide-level and protein-level tests. **Note: For hormone peptides ‘polyclonal’ refers to 3 or more hits to multiple variants of a single transcript, a high confidence result but not necessarily ‘polyclonal.’**

A sample-level unclustered, clustered, and row clustered visualization heatmap is provided for the top 100 rows of each table in the **summary_heatmaps** folder. These figures display the top peptide reaction per-protein.

A whole-protein epitope-scanning heatmap is provided in the **scan heatmaps** folder for each protein bearing one of the top 100 peptides enriched in each sample group.

Vietnam Cirrhosis vs cohort polyclonal

Gene Name	Protein	Polyclonal Hits Cirrhosis	Polyclonal Hits Cohort	Polyclonal No Hits Cirrhosis	Polyclonal No Hits Cohort	p value
GIPC3	XP_005259549.1	6	3	54	178	0.0085311
Anakirra	hd00071	4	1	56	180	0.0144240
PIEZO2	XP_011524027.1	3	0	57	181	0.0148530
MLLT3	NP_004520.2	3	0	57	181	0.0148530
CDKL5	XP_011543872.1	3	0	57	181	0.0148530
Cocaine-and_amphetamine-regulated_transcript_protein	hor000698	7	6	53	175	0.0205680
MAML2	NP_115803.1	9	10	51	171	0.0263150
FOX P2	NP_683698.2	8	8	52	173	0.0309370
BNP5-31	hor003004	4	2	56	179	0.0351440
Promotilin	hor002414	4	2	56	179	0.0351440
PRX	XP_011525473.1	2	0	58	181	0.0612030
MYO18B	XP_011528760.1	2	0	58	181	0.0612030
NHSL1	NP_001137532.1	2	0	58	181	0.0612030
CABYR	NP_722452.1	2	0	58	181	0.0612030
LOC100996274	XP_011524591.1	2	0	58	181	0.0612030
SHROOM3	XP_011530460.1	2	0	58	181	0.0612030
Big_gastrin	hor001001	2	0	58	181	0.0612030
Insulin_growth_factor-like_family_member_1	hor006758	2	0	58	181	0.0612030
Potential_peptide	hor003951	2	0	58	181	0.0612030
NCKAP5	XP_006712553.1	4	3	56	178	0.0667090
HUWE1	XP_011529048.1	4	3	56	178	0.0667090
BNP5-29	hor003003	4	3	56	178	0.0667090
Corticotropin	hor003942	4	3	56	178	0.0667090
Neuropeptide_5	hor003050	4	3	56	178	0.0667090
ANKHD1-EIF4EBP3	NP_065741.3	3	2	57	179	0.1002100
TEX13C	XP_005262409.1	3	2	57	179	0.1002100
PHC1	XP_011518904.1	3	2	57	179	0.1002100
BNP2-31	hor002994	3	2	57	179	0.1002100
BNP5-32	hor003005	3	2	57	179	0.1002100
Neuromedin-U-25	hor003028	3	2	57	179	0.1002100
Corticotropin	hd00063	3	2	57	179	0.1002100
SVIL	XP_011517932.1	4	4	56	177	0.1087300
TLN1	NP_006280.3	5	5	55	176	0.1261900
Acyl-CoA-binding_domain-containing_protein_7	hor006763	5	5	55	176	0.1261900
Apelin-31	hor006494	16	32	44	149	0.1389400
TRIOBP	NP_001034230.1	5	6	55	175	0.1475000
PAXIP1	XP_005249596.1	5	6	55	175	0.1475000
PHLDB1	XP_011541005.1	2	1	58	180	0.1539000
RIMS1	NP_001161882.1	2	1	58	180	0.1539000
ARHGEF17	NP_055601.2	2	1	58	180	0.1539000
YLP M1	NP_062535.2	2	1	58	180	0.1539000
NHSL1	NP_065197.1	2	1	58	180	0.1539000
FRMPD3	XP_011529358.1	2	1	58	180	0.1539000
CCDC88B	NP_115627.6	2	1	58	180	0.1539000
PRDM10	NP_064613.2	2	1	58	180	0.1539000
LARP1	XP_005268461.1	2	1	58	180	0.1539000
MED15	XP_006724326.1	2	1	58	180	0.1539000
LOC105371409	XP_011507010.1	2	1	58	180	0.1539000
ZBTB21	XP_011527889.1	2	1	58	180	0.1539000
NHSL1	XP_011534267.1	2	1	58	180	0.1539000
Cholecystokinin-25	hor006109	2	1	58	180	0.1539000
Cholecystokinin-33	hor000997	2	1	58	180	0.1539000
EA-92	hor000765	2	1	58	180	0.1539000
Glicentin	hor006378	2	1	58	180	0.1539000
Neuroendocrine_regulatory_peptide-1	hor000039	2	1	58	180	0.1539000
Neuropeptide_2	hor006258	2	1	58	180	0.1539000
Trefoil_factor_1	hor005578	2	1	58	180	0.1539000
RSRP1	NP_064713.3	3	3	57	178	0.1652800
Lipotropin_beta	hor003944	3	3	57	178	0.1652800
Corticotropin	hd000004	3	3	57	178	0.1652800
Neuromedin-S	hor003027	10	18	50	163	0.1678100
VGF-derived_peptide_TLQP-62	hor000047	5	7	55	174	0.1792700
FLG2	NP_001014364.1	15	30	45	151	0.1802500
Apelin-28	hor006495	13	26	47	155	0.2241800
Neuropeptide_gamma	hor005795	30	73	30	108	0.2285300
TNRC18	NP_001073964.2	4	5	56	176	0.2320500
Neuropeptide_NPVF	hor001622	6	10	54	171	0.2381200
RBM5	XP_006712980.1	1	0	59	181	0.2489600
ARX	NP_620689.1	1	0	59	181	0.2489600
GTF2IRD1	NP_001186136.1	1	0	59	181	0.2489600
CRAMP1	NP_065876.3	1	0	59	181	0.2489600
MRE11	NP_005582.1	1	0	59	181	0.2489600
CNTLN	XP_006716856.1	1	0	59	181	0.2489600
LTBP1	XP_011531162.1	1	0	59	181	0.2489600
USE1	NP_060937.1	1	0	59	181	0.2489600
LAMA3	NP_001121189.2	1	0	59	181	0.2489600
SYNE2	NP_055995.4	1	0	59	181	0.2489600
NFYC	XP_005270952.3	1	0	59	181	0.2489600
N4BP2	XP_011512020.1	1	0	59	181	0.2489600
REXO1	XP_011526446.1	1	0	59	181	0.2489600
ZNF324	XP_005258770.1	1	0	59	181	0.2489600
PLEKHG2	XP_006723397.1	1	0	59	181	0.2489600
SETD1A	NP_055527.1	1	0	59	181	0.2489600
TRIP11	NP_004230.2	1	0	59	181	0.2489600
MYBL2	NP_002457.1	1	0	59	181	0.2489600
NFAT5	NP_006590.1	1	0	59	181	0.2489600
GABPB1	NP_002032.2	1	0	59	181	0.2489600
CSPP1	XP_011515900.1	1	0	59	181	0.2489600
MED26	NP_004822.2	1	0	59	181	0.2489600
TJP3	NP_001254490.1	1	0	59	181	0.2489600
DNAH11	NP_001264044.1	1	0	59	181	0.2489600
IQCE	NP_689771.3	1	0	59	181	0.2489600
KMT5B	NP_060105.3	1	0	59	181	0.2489600
SOX6	NP_001139291.1	1	0	59	181	0.2489600
GLI1	NP_005260.1	1	0	59	181	0.2489600
TBP	NP_003185.1	1	0	59	181	0.2489600
HMGXB3	NP_055798.2	1	0	59	181	0.2489600
MARK1	XP_011507863.1	1	0	59	181	0.2489600
TTL5	NP_055887.3	1	0	59	181	0.2489600
LMO7	NP_056667.2	1	0	59	181	0.2489600

fisher_exact_tests/Vietnam_Cirrhosis_vs_cohort_polyclonal_fisher_exact_tests.csv

Vietnam HCC vs cohort polyclonal

Gene Name	Protein	Polyclonal Hits HCC	Polyclonal Hits Cohort	Polyclonal No Hits HCC	Polyclonal No Hits Cohort	p value
ZFXH3	NP_008816.3	9	1	112	119	0.018996
ARMCX4	NP_001243084.2	26	12	95	108	0.020730
Lepirudin	hd00045	10	2	111	118	0.033985
TTN	NP_597681.4	24	12	97	108	0.045807
TTN	XP_011510031.1	31	18	90	102	0.054189
Angiotensin_1-7	hor004211	5	0	116	120	0.059923
Met-enkephalin-Arg-Gly-Leu	hor003284	7	1	114	119	0.065803
OBSCN	XP_005273344.1	20	10	101	110	0.077821
TTN	XP_011510032.1	23	13	98	107	0.103040
Low_molecular_weight_growth-promoting_factor	hor005763	6	1	115	119	0.119550
MYCBP2	NP_055872.4	4	0	117	120	0.121890
KMT2A	NP_001184033.1	4	0	117	120	0.121890
MGA	NP_001074010.2	4	0	117	120	0.121890
Corticotropin-like_intermediary_peptide	hor003943	4	0	117	120	0.121890
Substance_P	hor004394	4	0	117	120	0.121890
FAM9A	XP_011543765.1	12	5	109	115	0.129220
Romiplostim	hd00049	10	4	111	116	0.166730
STARD9	XP_011520136.1	7	2	114	118	0.171470
FLG2	XP_011507833.1	28	19	93	101	0.193270
STARD9	XP_011520133.1	5	1	116	119	0.212890
CAMK2N1	NP_061054.2	5	1	116	119	0.212890
TTBK1	XP_011513247.1	5	1	116	119	0.212890
Gonadoliberin-1	hor002038	5	1	116	119	0.212890
Leu-enkephalin	hor003280	5	1	116	119	0.212890
Neuropeptide_FF	hor001619	5	1	116	119	0.212890
Progonadoliberin-2	hor002042	5	1	116	119	0.212890
Spexin	hor002515	5	1	116	119	0.212890
VGF-derived_peptide_TLQP-21	hor000046	8	3	113	117	0.215730
gamma-melanocyte-stimulating_hormone	hor003210	8	3	113	117	0.215730
OBSCN	XP_006711891.1	12	6	109	114	0.219810
MUC16	NP_078966.2	25	17	96	103	0.234540
TCOF1	NP_001128716.1	3	0	118	120	0.246890
SPEG	XP_011508781.1	3	0	118	120	0.246890
AFF4	XP_005272020.1	3	0	118	120	0.246890
MAST2	XP_011539361.1	3	0	118	120	0.246890
TNRC6B	NP_001155973.1	3	0	118	120	0.246890
KIAA1549L	XP_005252905.1	3	0	118	120	0.246890
ZNF280D	NP_060131.2	3	0	118	120	0.246890
AHCTF1	XP_006711821.1	3	0	118	120	0.246890
CACNB2	NP_001161417.1	3	0	118	120	0.246890
TSGA10IP	NP_689975.2	3	0	118	120	0.246890
ZNF609	NP_055857.1	3	0	118	120	0.246890
ROBO2	XP_011532283.1	3	0	118	120	0.246890
TCF20	XP_006724376.1	3	0	118	120	0.246890
PLEKHG4B	XP_011512271.1	3	0	118	120	0.246890
ADARB1	XP_011527727.1	3	0	118	120	0.246890
ANKRD50	XP_011530448.1	3	0	118	120	0.246890
MICAL3	XP_011544442.1	3	0	118	120	0.246890
Early_placenta_insulin-like_peptide_A_chain	hor006260	3	0	118	120	0.246890
Gastrin-releasing_peptide	hor000275	3	0	118	120	0.246890
Insulin_A_chain	hor002222	3	0	118	120	0.246890
Insulin_A	hd00079a	3	0	118	120	0.246890
SRCAP	NP_006653.2	6	2	115	118	0.280810
NCOR2	NP_001070729.2	6	2	115	118	0.280810
BNP3-32	hor002997	10	5	111	115	0.286160
LOC101927260	XP_011510561.1	16	10	105	110	0.299290
NPP	hor003949	18	12	103	108	0.329510
SVIL	XP_011517933.1	7	3	114	117	0.333370
Melanotropin_gamma	hor003948	7	3	114	117	0.333370
AFF3	XP_011509472.1	4	1	117	119	0.369780
TOX3	XP_005255950.1	4	1	117	119	0.369780
OBSCN	XP_005273351.1	4	1	117	119	0.369780
OBSCN	XP_011542597.1	4	1	117	119	0.369780
Substance_P	hor004260	4	1	117	119	0.369780
Angiotensin_1-5	hor004210	4	1	117	119	0.369780
Angiotensin-1	hor004213	4	1	117	119	0.369780
Bradykinin	hor000295	4	1	117	119	0.369780
Katacalcin	hor000682	4	1	117	119	0.369780
Melanocyte-stimulating_hormone_alpha	hor003946	4	1	117	119	0.369780
Pituitary_adenylate_cyclase-activating_polypeptide_38	hor001970	4	1	117	119	0.369780
Lanreotide	hd000016	4	1	117	119	0.369780
Cetorelix	hd000017	4	1	117	119	0.369780
Pituitary_adenylate_cyclase-activating_polypeptide_27	hor001969	8	4	113	116	0.375380
Lysyl-bradykinin	hor000296	9	5	112	115	0.409850
NPQ_53-70	hor002512	46	39	75	81	0.419190
MUC19	NP_775871.2	10	6	111	114	0.438730
RIMS1	NP_055804.2	5	2	116	118	0.446230
PALLD	XP_011530070.1	5	2	116	118	0.446230
BPTF	XP_005257213.1	5	2	116	118	0.446230
KMT2C	XP_011514754.1	5	2	116	118	0.446230
MPRIIP	XP_011522063.1	5	2	116	118	0.446230
Deltorphin_I	hor003208	5	2	116	118	0.446230
Angiotensin-3	hor004215	5	2	116	118	0.446230
Angiotensin-4	hor004216	5	2	116	118	0.446230
Neurokinin_A	hor004392	5	2	116	118	0.446230
Neuropeptide_SF	hor001620	5	2	116	118	0.446230
N-Tyr-MIF-1	hor002527	5	2	116	118	0.446230
Tail_peptide	hor005797	5	2	116	118	0.446230
WE-14	hor000775	5	2	116	118	0.446230
Cerebellin	hor000134	5	2	116	118	0.446230
Ganirelix	hd000018	5	2	116	118	0.446230
Palifermin	hd000073	5	2	116	118	0.446230
Nociceptin	hor003287	11	7	110	113	0.463340
Apelin-36	hor006496	20	15	101	105	0.465150
ZFXH4	NP_078997.4	2	0	119	120	0.497930
DICER1	NP_803187.1	2	0	119	120	0.497930
DIDO1	XP_011526806.1	2	0	119	120	0.497930
NALCN	XP_011519371.1	2	0	119	120	0.497930
ZNF821	NP_001188482.1	2	0	119	120	0.497930
AMBRA1	NP_001287660.1	2	0	119	120	0.497930

fisher_exact_tests/Vietnam_HCC_vs_cohort_polyclonal_fisher_exact_tests.csv

Vietnam Hepatitis B vs cohort polyclonal

Gene Name	Protein	Polyclonal Hits Hepatitis B	Polyclonal Hits Cohort	Polyclonal No Hits Hepatitis B	Polyclonal No Hits Cohort	p value
CACNA1D	NP_000711.1	3	0	57	181	0.014853
USF3	NP_001009899.3	3	0	57	181	0.014853
MED12L	XP_011510688.1	3	0	57	181	0.014853
ANK3	NP_066267.2	5	3	55	178	0.024690
LOC101927859	XP_011528926.1	10	13	50	168	0.041423
LOC105369189	XP_011544559.1	5	4	55	177	0.044888
PLEC	XP_005251033.1	3	1	57	180	0.048739
LOC100996335	XP_011507252.1	3	1	57	180	0.048739
CACNA1D	XP_011532397.1	3	1	57	180	0.048739
MAST4	XP_011541684.1	3	1	57	180	0.048739
CREBBP	XP_005255182.1	2	0	58	181	0.061203
CLASRP	XP_011524698.1	2	0	58	181	0.061203
SLC4A3	XP_011509967.1	2	0	58	181	0.061203
LAMA5	XP_006723859.1	2	0	58	181	0.061203
CENPE	XP_011529846.1	2	0	58	181	0.061203
CACNA1D	NP_001122311.1	2	0	58	181	0.061203
ZC3H18	XP_011521165.1	2	0	58	181	0.061203
PLIN4	XP_011526535.1	2	0	58	181	0.061203
DST	XP_005249367.1	2	0	58	181	0.061203
RBM33	XP_005249591.1	2	0	58	181	0.061203
ANK2	XP_011530193.1	2	0	58	181	0.061203
IOSEC1	XP_011532606.1	2	0	58	181	0.061203
SH3BP5L	XP_011542586.1	2	0	58	181	0.061203
KAT6B	XP_011547078.1	2	0	58	181	0.061203
FCGBP	XP_011547112.1	2	0	58	181	0.061203
Early_placenta_insulin-like_peptide_B_chain	hor006261	2	0	58	181	0.061203
Insulin-like_growth_factor_I	hor002218	2	0	58	181	0.061203
ZC3H13	XP_005266365.1	9	12	51	169	0.062661
HMCN2	NP_001278744.1	8	10	52	171	0.083860
AMOT	XP_005262144.1	3	2	57	179	0.100210
LOC100996335	XP_011528942.1	3	2	57	179	0.100210
LOC100996415	XP_011544458.1	3	2	57	179	0.100210
LOC105375612	XP_011544801.1	3	2	57	179	0.100210
Atrial_natriuretic_factor	hor002986	3	2	57	179	0.100210
HMW-guanylin	hor006249	3	2	57	179	0.100210
SS-18	hor000772	4	4	56	177	0.108730
OBSCN	XP_005273354.1	15	28	45	153	0.118830
MUC17	NP_001035194.1	5	5	55	176	0.126190
MDC1	XP_005274956.1	7	10	53	171	0.142480
TRIOBP	NP_001034230.1	5	6	55	175	0.147500
NINL	XP_011527488.1	2	1	58	180	0.153900
CLASRP	NP_001265368.1	2	1	58	180	0.153900
IQCE	NP_001274429.1	2	1	58	180	0.153900
PHC1	NP_004417.2	2	1	58	180	0.153900
ZRANB2	NP_005446.2	2	1	58	180	0.153900
CACNA1D	XP_011532396.1	2	1	58	180	0.153900
CHD9	XP_005256232.2	2	1	58	180	0.153900
PPRC1	XP_005269713.1	2	1	58	180	0.153900
ENAH	XP_005273244.1	2	1	58	180	0.153900
PLIN4	XP_006722929.1	2	1	58	180	0.153900
KIAA1210	XP_006724837.1	2	1	58	180	0.153900
FAM120B	XP_006725078.1	2	1	58	180	0.153900
SCAPER	XP_011519950.1	2	1	58	180	0.153900
KDM3A	XP_011531286.1	2	1	58	180	0.153900
LOC105378193	XP_011535833.1	2	1	58	180	0.153900
HERC2	XP_011547261.1	2	1	58	180	0.153900
Cholecystokinin-58_desnonopeptide	hor006110	2	1	58	180	0.153900
Urotilatin	hor002989	2	1	58	180	0.153900
CCDC13	XP_005264955.1	3	3	57	178	0.165280
TLN2	XP_005254765.1	3	3	57	178	0.165280
KMT2C	XP_011514753.1	5	7	55	174	0.179270
LOC105376351	XP_011518069.1	15	30	45	151	0.180250
ZC3H13	XP_005266358.1	8	13	52	168	0.184270
FLG	NP_002007.1	18	39	42	142	0.219560
NRXN3	XP_011535665.1	1	0	59	181	0.248960
BCLAF1	NP_055554.1	1	0	59	181	0.248960
THRAP3	NP_005110.2	1	0	59	181	0.248960
WNK1	NP_001171914.1	1	0	59	181	0.248960
SFSWAP	NP_001248340.1	1	0	59	181	0.248960
SFSWAP	XP_011536955.1	1	0	59	181	0.248960
ANKS1A	XP_011512733.1	1	0	59	181	0.248960
BLTP3A	NP_060224.3	1	0	59	181	0.248960
TCOF1	NP_001128717.1	1	0	59	181	0.248960
SPAG5	NP_006452.3	1	0	59	181	0.248960
MECOM	XP_005247270.1	1	0	59	181	0.248960
RBM27	XP_005268523.1	1	0	59	181	0.248960
CARD10	NP_055365.2	1	0	59	181	0.248960
SRSF5	XP_011535379.1	1	0	59	181	0.248960
FMR1	NP_002015.1	1	0	59	181	0.248960
SALL1	NP_002959.2	1	0	59	181	0.248960
MCM4	NP_005905.2	1	0	59	181	0.248960
CACTIN	NP_001074012.1	1	0	59	181	0.248960
AKNA	NP_110394.3	1	0	59	181	0.248960
DNM1	NP_001275668.1	1	0	59	181	0.248960
EIF3A	NP_003741.1	1	0	59	181	0.248960
NSMCE4A	NP_060085.2	1	0	59	181	0.248960
ANKRD26	NP_055730.2	1	0	59	181	0.248960
ATN1	NP_001931.2	1	0	59	181	0.248960
PTK7	NP_002812.2	1	0	59	181	0.248960
EIF4G1	NP_001181875.1	1	0	59	181	0.248960
PIKFYVE	XP_011509080.1	1	0	59	181	0.248960
TANC1	XP_006712877.1	1	0	59	181	0.248960
APC2	XP_005259532.1	1	0	59	181	0.248960
MARK1	NP_001273053.1	1	0	59	181	0.248960
MAP7D1	NP_060537.3	1	0	59	181	0.248960
ARID1A	NP_006006.3	1	0	59	181	0.248960
ACOT2	NP_006812.3	1	0	59	181	0.248960
COIL	NP_004636.1	1	0	59	181	0.248960
ZSCAN20	NP_060281.2	1	0	59	181	0.248960
LRP1	NP_002323.2	1	0	59	181	0.248960

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_polyclonal_fisher_exact_tests.csv

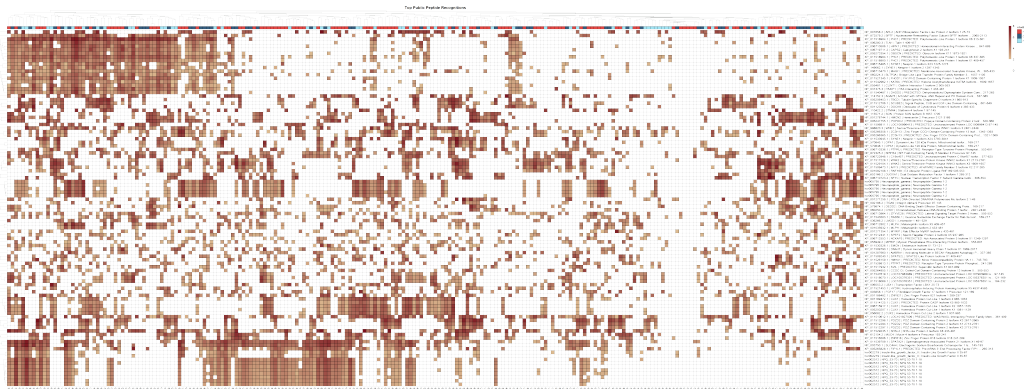
14 Summary Heatmaps: Top Public

Unbiased summary heatmaps of the top public reactive peptides from the top_public_peptides.csv output rows. Full resolution figures are provided in the **summary_heatmaps** folder. These figures are often good at independently identifying and clustering sample pairs and replicates via observation of the study's highest frequency semi-continuous hits. Note: most public epitopes are unrelated with conditions of interest and this plot often will not cluster disease groups.

summary_heatmaps/top_public_unclustered_heatmap.png - Created in R with the *pheatmap* package and no clustering. Samples and groups are arranged according to the *studyManifest*.

summary_heatmaps/top_public_clustered_heatmap.png - Created in R with the *pheatmap* package with default row and column clustering settings (Euclidean distance as the similarity measure and clustering samples based on the 'complete' method).

summary_heatmaps/top_public_row_clustered_heatmap.png - Created in R with the *pheatmap* package with default row clustering settings (Euclidean distance as the similarity measure). Samples and groups are unclustered and remain arranged according to the *studyManifest*.



summary_heatmaps/top_public_clustered_heatmap.png

15 Summary Heatmaps: Fisher Exact Tests

Summary heatmaps of each of the top 100 peptide-level, protein-level, and protein-level polyclonal Fisher exact test outliers shown on the preceding pages. Full resolution figures are provided in the **summary_heatmaps** folder. These figures are all directional for top p-values associated with each test and attempt to visualize co-occurring autoantibody signature subgroups. All visualized hits are from the **edgeR_enriched_log2foldchg.csv** semi-continuous hits data. These values are directly plotted log2 fold change for peptide-level hits. For protein-level and protein-level polyclonal hits these represent the maximum log2 fold change peptide hit occurring within that protein.

summary_heatmaps/filename_unclustered_heatmap.png - One created for each directional peptide, protein, and polyclonal-level test in the **fisher_exact_tests** folder. Created in R with the *pheatmap* package and no clustering. Samples and groups are arranged according to the *studyManifest*.

summary_heatmaps/filename_clustered_heatmap.png - One created for each directional peptide, protein, and polyclonal-level test in the **fisher_exact_tests** folder. Created in R with the *pheatmap* package with default row and column clustering settings (Euclidean distance as the similarity measure and clustering samples based on the 'complete' method).

summary_heatmaps/filename_row_clustered_heatmap.png - One created for each directional peptide, protein, and polyclonal-level test in the **fisher_exact_tests** folder. Created in R with the *pheatmap* package with default row clustering settings (Euclidean distance as the similarity measure). Samples and groups are unclustered and remain arranged according to the *studyManifest*.

16 Epitope Scanning Heatmaps

Protein-level epitope-scanning heatmaps are created and shared in the **scan_heatmaps** folder. These are created from the *edgeR_enriched_log2foldchg.csv* file. All hits pass a p-value and fold change threshold versus the 'beadsonly' distribution described in the **Informatics Methods** section. Squares are colored according to their status as a hit and the log2 fold change of the called hit versus the 'beadsonly' distribution; non-enriched values are blank and log2 fold changes typically range from 2.0 (i.e. 4x increase above background) to >8.0 (i.e. >256x increase). The amino acid start and stop position of each peptide within the parent protein is labeled if-available.

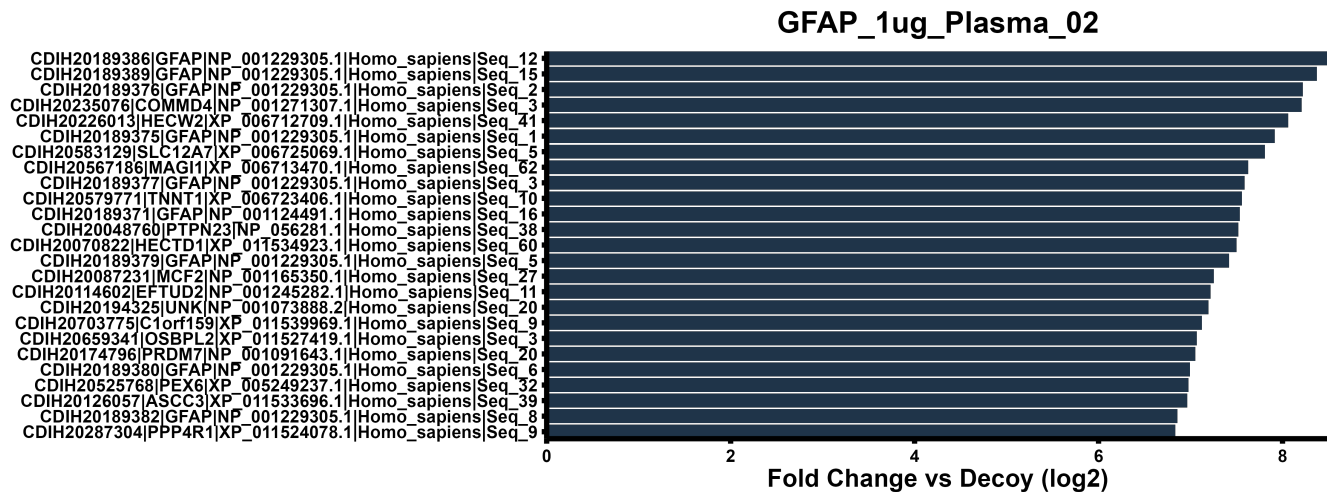
Heatmaps have been created for all proteins containing a peptide represented in the top_public_peptides file. Note: as some proteins will typically include two or more peptides in the top variants list there are often fewer than 100 heatmap files made.

Heatmaps have been created for all proteins represented in the each of the top 100 peptide-level, protein-level, and protein-level polyclonal Fisher exact test outliers shown on the preceding pages for each provided *studyGroup*. Note: as some proteins will typically include two or more peptides in the top variants list there are often fewer than 100 heatmap files made per file.

17 Individual Sample Barplots

Peptide-level top outlier fold changes for each sample are provided in the **barplots** folder.

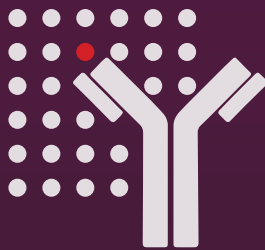
These barplots are created from the *edgeR_enriched_log2foldchg.csv* file and represent the top x25 enriched fold change signals versus 'beadsonly' decoy controls for each individual sample. If fewer than x25 significant hits are available, fewer peptides are plotted. An example of one of these images is shown below.



GFAP_1ug_Plasma_02_enriched_foldchange_barplot.png

18 References

1. Larman, H. B. et al. Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* 29, 535–541 (2011)
2. Xu, G. J. et al. Comprehensive serological profiling of human populations using a synthetic human virome. *Science* 348, aaa0698 (2015).
3. O'Donovan, B. et al. High-resolution epitope mapping of anti-Hu and anti-Yo autoimmunity by programmable phage display. *Brain Commun.* 2, fcaa059 (2020).
4. Mohan, D. et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* 13, 1958–1978 (2018).
5. Zhu, N. et al. HORDB a comprehensive database of peptide hormones. *Nat. Scientific Data.* 9, 187 (2022).



CDI Labs - Mayagüez

12 W Mendez Vigo Avenue
2nd Floor
Mayagüez, PR 00680
USA

CDI Labs - Baltimore

1 N. Haven Street
Suite B001
Baltimore, MD 21224
USA

Sales

Email: info@cdilabs.com

Support

Email: support@cdilabs.com

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

© **CDI Labs**. All rights reserved.

CDI Labs and HuScan® are trademarks and registered trademarks respectively, of CDI Labs.

