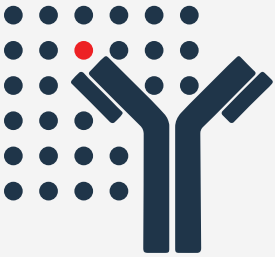




STUDY REPORT



VirScan® Report

Viral Proteome Phage ImmunoPrecipitation Sequencing (PhIP-Seq)

Pan-Viral Phage Display Immunoprecipitation-Sequencing (PhIP-Seq)

VirScan Study Report

Prepared For:

Vietnam

October 17, 2025

Contents

1	Overview of PhIP-Seq	3
2	Experimental Overview	3
3	Sample Screening	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	Seropositive Tests	22
13	Protein-Level Tests	26
14	Protein-Level Polyclonal Tests	30
15	Summary Heatmaps: Top Public	34
16	Summary Heatmaps: Fisher Exact Tests	35
17	Epitope Scanning Heatmaps	36
18	Individual Sample Barplots	37
19	References	38

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 validated pan-viral library representative of most known proteins from vertebrate, mosquito-borne, and tick-borne viral genomes according to the UniProt and RefSeq databases in 2017.^{3,4}

2 Experimental Overview

An *E. coli* codon-optimized library consisting of >480,000 oligonucleotides encoding 62-mer amino acid peptide inserts with 14-mer amino acid adjacent overlaps was prepared as-described³; this library covers most known proteins from vertebrate, mosquito-borne, and tick-borne viral genomes according to the UniProt and RefSeq databases in 2017. 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. The library was then packaged in vitro and expanded. The quality of library inserts was assessed by Illumina sequencing.

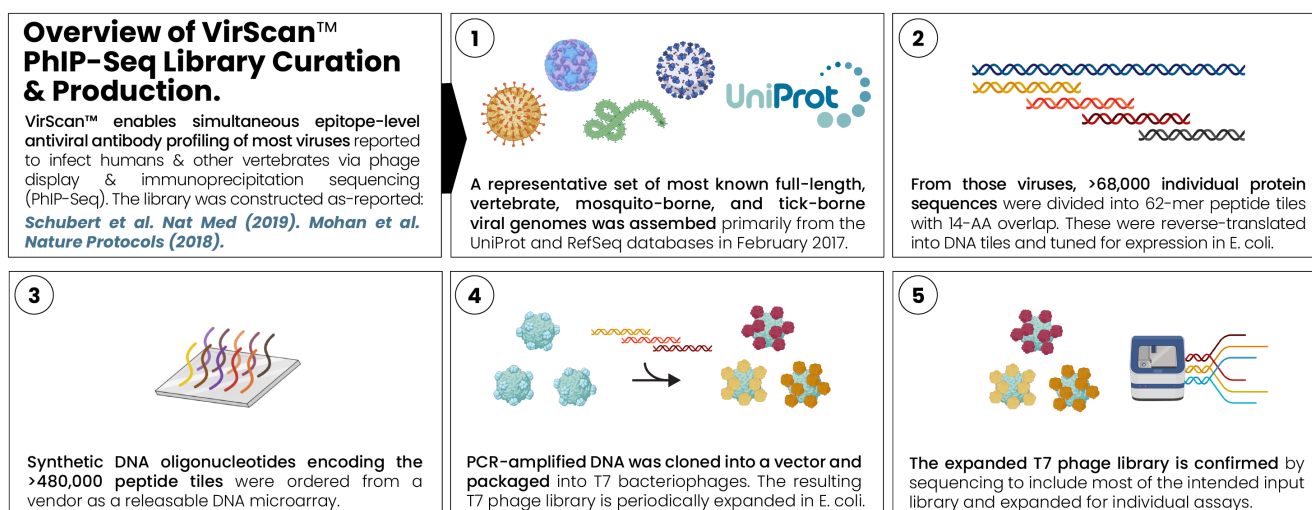


Figure by Tyler Hulett, PhD for CDI Labs; created with BioRender.

Overview of PhIP-Seq VirScan library construction. Pan-viral library creation was performed as-shown and as-reported: Schubert, R. et al. *Pan-viral serology implicates enteroviruses in acute flaccid myelitis*. *Nat. Med.* 25(11): 1748–1752. (2019). doi:10.1038/s41591-019-0613-1

3 Sample Screening

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. 4.8×10^{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.

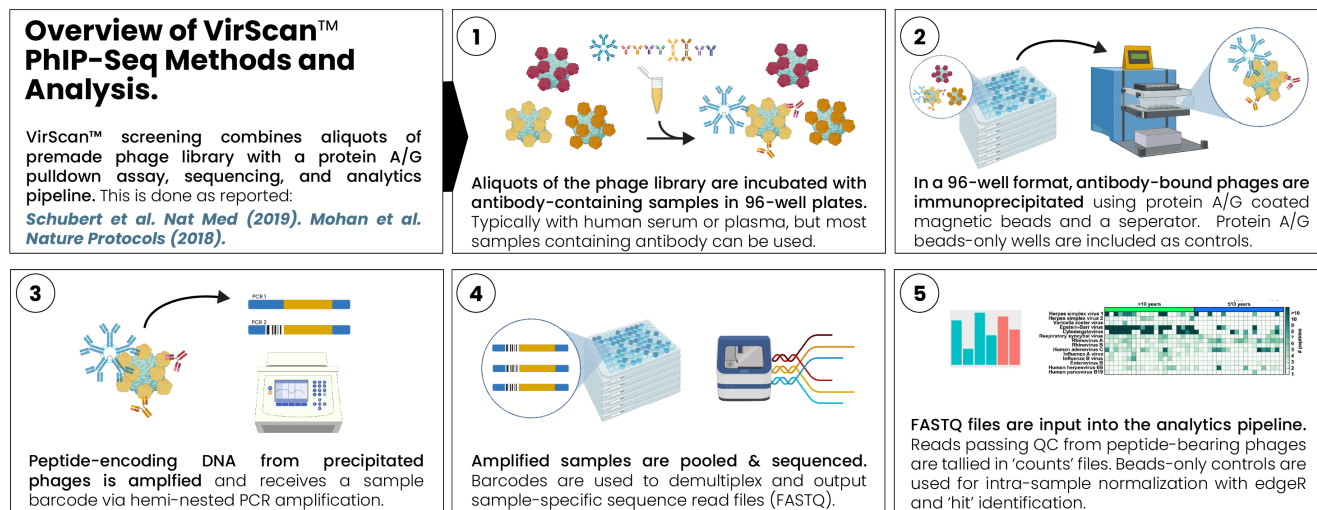


Figure by Tyler Hulett, PhD for CDI Labs; created with BioRender.

Overview of PhIP-Seq VirScan 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 viral 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	V2P01_A1
CDIV20000001 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_1	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	1	62	MKTIALSYLCLVFAQKLPNGDNSTATCLGHHAVPNGTIVTKIT	0
CDIV20000002 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_2	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	49	110	QIEVTNATELVQNSSTGEICDSPHQIDGKNCTLDALLGDPQCD	38
CDIV20000003 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_3	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	97	158	NKKWDLFVSRKSNFYDPVDPYASLRSLVASSGTLEFNESF	3
CDIV20000004 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_4	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	145	206	GCTVQNGHTSSACIRSKNSFFSRLNWLNTNFKYPALNVTMPNE	10
CDIV20000005 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_5	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	193	254	YIYGVHHPVTDKQDQIFLYAQASGRITVSTRKSQQTVPINSRPI	4
CDIV20000006 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_6	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	241	302	NIPRSISYIYWTIVKPGDILLINSTGNLIAPRYFKMQSGKSSIMRSD	0
CDIV20000007 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_7	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	337	398	ANPEVKQTRGIFGAIGFIENGWEGMVDGWYGFRRHNSSEGR	0
CDIV20000008 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_8	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	385	446	STQAIDQINGKLNRLKTNKFKHKEFSEVEGRIQDLKEYVE	23
CDIV20000009 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_9	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	433	494	KIDLWSYNAELVALENQHTIDLTDFEEMNKLFEKTKQLRENAEC	13
CDIV20000010 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_10	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	481	542	GCTKIYHKCDNACIGSIRNGTYDHNVYRDEALNRFQIKVELKS	0
CDIV20000011 Influenza_A_virus AFM71802.1 Orthomyxoviridae_Alphainfluenzavirus Seq_11	I6T1T5	hemagglutinin_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	505	566	NVYRDEALNRFQIKVELKSGYKDWILWYSFAISCFLLCVALG	13
CDIV20000012 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_1	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	1	62	MDVNPITLLFKVPAQNAISTTFPYTGDPYSGHGTGTGYTMDTVN	15
CDIV20000013 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_2	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	49	110	YSERGKWTNTTETGAPQLNIDPGLPEDNEPSGAYGTQDVCLEAH	10
CDIV20000014 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_3	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	97	158	ESHPIGFENSCLTETMAEQVTRVDKLTQGRQTYDWTNRRNP	12
CDIV20000015 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_4	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	145	206	NTIEFVSRNGLTANESGRIDFLKVMESMDKEEMEITTHFRQKI	27
CDIV20000016 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_5	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	193	254	DNMTKRMVYRTQRTGKIKKRVNKRGIYRLALTLNTMTKDAERGK	0
CDIV20000017 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_6	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	481	542	KSINIKTGTTFEFTFYRYGFVANFMSMELSPGVSGINESADMSIG	4
CDIV20000018 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_7	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	529	590	VIKNNMININDLPATAQMALQVFKDYRTRHGRDGTQIQTRI	5
CDIV20000019 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_8	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	577	638	KKLWDQTSRAGLLVSDGGPNLYNRIHLHPVCLKWLMDEN	2
CDIV20000020 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_9	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	625	686	CNLPNPFVSHKEIESVNNAVMPAHGPAKSMEDYDAVATTHSWI	0
CDIV20000021 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_10	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	673	734	SILNTSQRGILEDEQMYYKCCNLFEKFFPSSSYRPIGISSMVEAV	5
CDIV20000022 Influenza_A_virus AFM71810.1 Orthomyxoviridae_Alphainfluenzavirus Seq_11	I6S4X5	polymerase_P81_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	696	757	FEKFFPSSSYRPIGISSMVEAVMSRIRARLIDFESGRIKKEEFSE	1
CDIV20000023 Influenza_A_virus AFM71812.1 Orthomyxoviridae_Alphainfluenzavirus Seq_1	I6T1U5	polymerase_P82_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	49	110	WMAMKYPITADKRITEMVPERNEQQQLTLWSKMSDAGSDRI	0
CDIV20000024 Influenza_A_virus AFM71812.1 Orthomyxoviridae_Alphainfluenzavirus Seq_2	I6T1U5	polymerase_P82_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	145	206	VINPBGHADLSAKAQAQVMEVVFNEVGARILTSQSLTITKEKI	15
CDIV20000025 Influenza_A_virus AFM71812.1 Orthomyxoviridae_Alphainfluenzavirus Seq_3	I6T1U5	polymerase_P82_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	481	542	AAAAAASAAAAAASIDRFLRVDRDQRGNVLSPPEESVTPAQI	0
CDIV20000026 Influenza_A_virus AFM71812.1 Orthomyxoviridae_Alphainfluenzavirus Seq_4	I6T1U5	polymerase_P82_[Influenza_A_virus_(A/Victoria/210/2009(H3N2))]	529	590	ITYSSMMWWEINGPESVLNVTYQWIIRNWEAVKIQVWSQNPQM	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 (V for VirScan), version number (in this case ‘2’), unique phage number, virus name, reference library protein identifier, 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, UniProt, or other identifier of the parent protein used to generate this peptide.

Description: A longer description of the protein name.

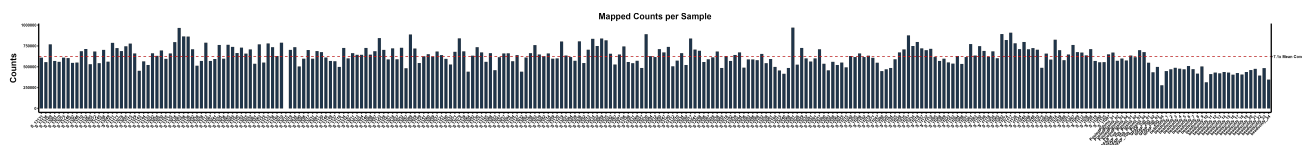
AA_Start: The amino acid position this peptide starts at related to the full parent protein sequence if-available.

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

Peptide: The literal peptide sequence carried by this phage.

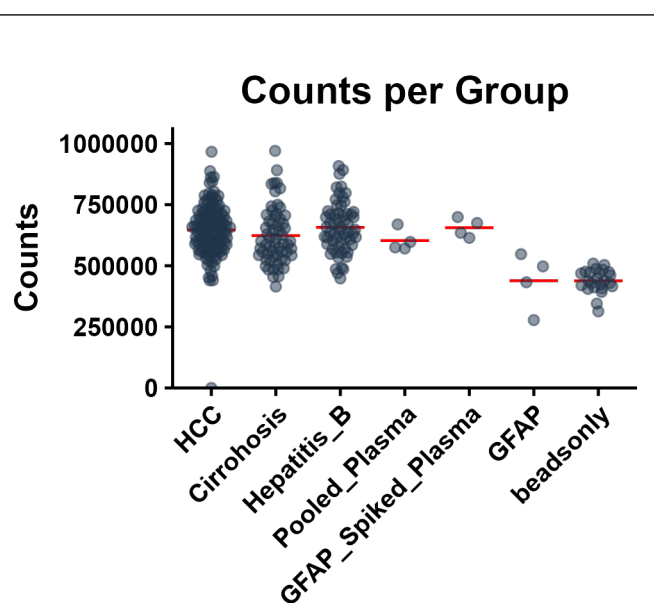
Sample Columns: All subsequent columns are samples in this assay. Any column strings containing ‘beadsonly’ are mock-immunoprecipitation controls used for hit calling. One column should exist for each sample provided to us in the *sampleManifest.csv* at assay onboarding. We have additionally added internal assay control columns including the ‘beadsonly’; both your 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.

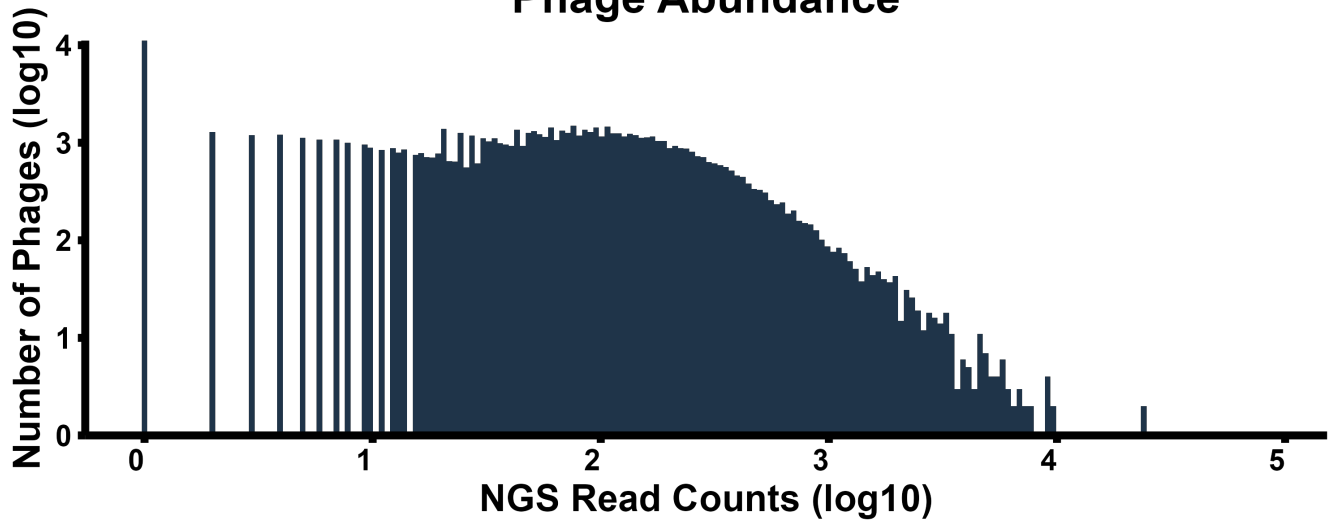
plots/counts_per_sample.png



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

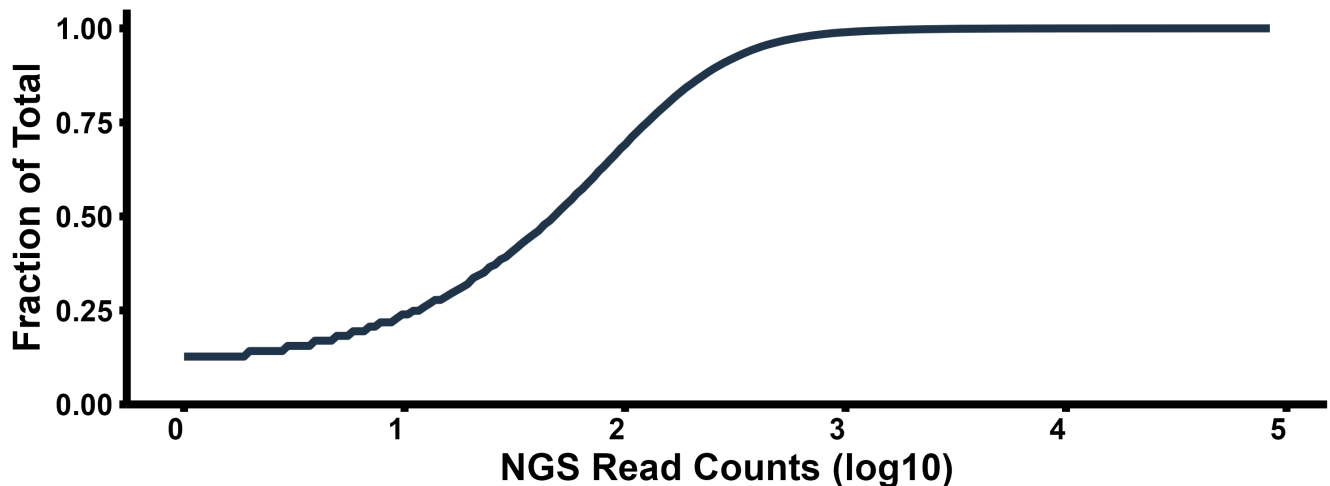
Histogram of Baseline Phage Abundance



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.

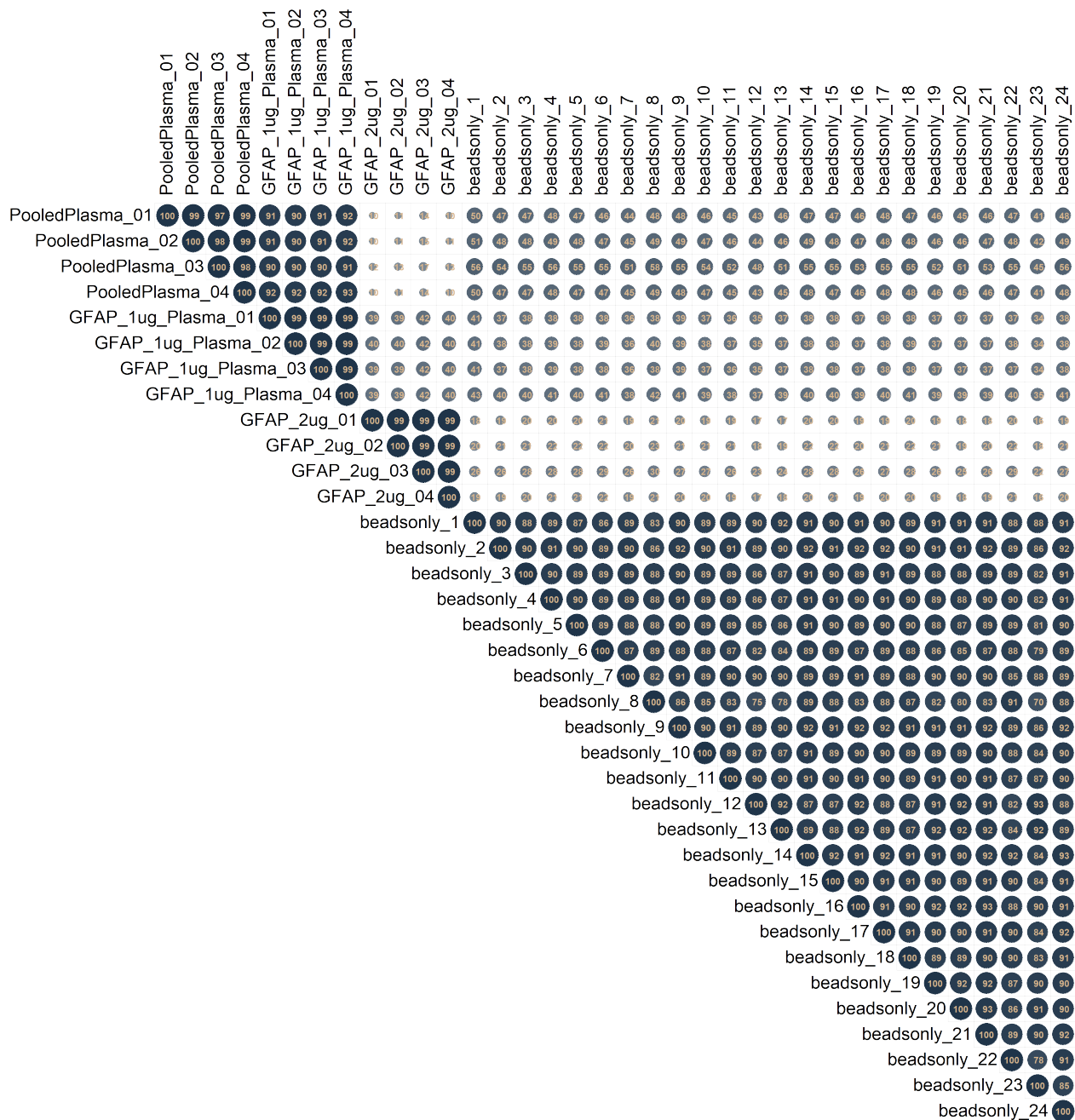
plots/beadonly_histogram_plot.png

Cumulative Sum of Baseline Phage Abundance



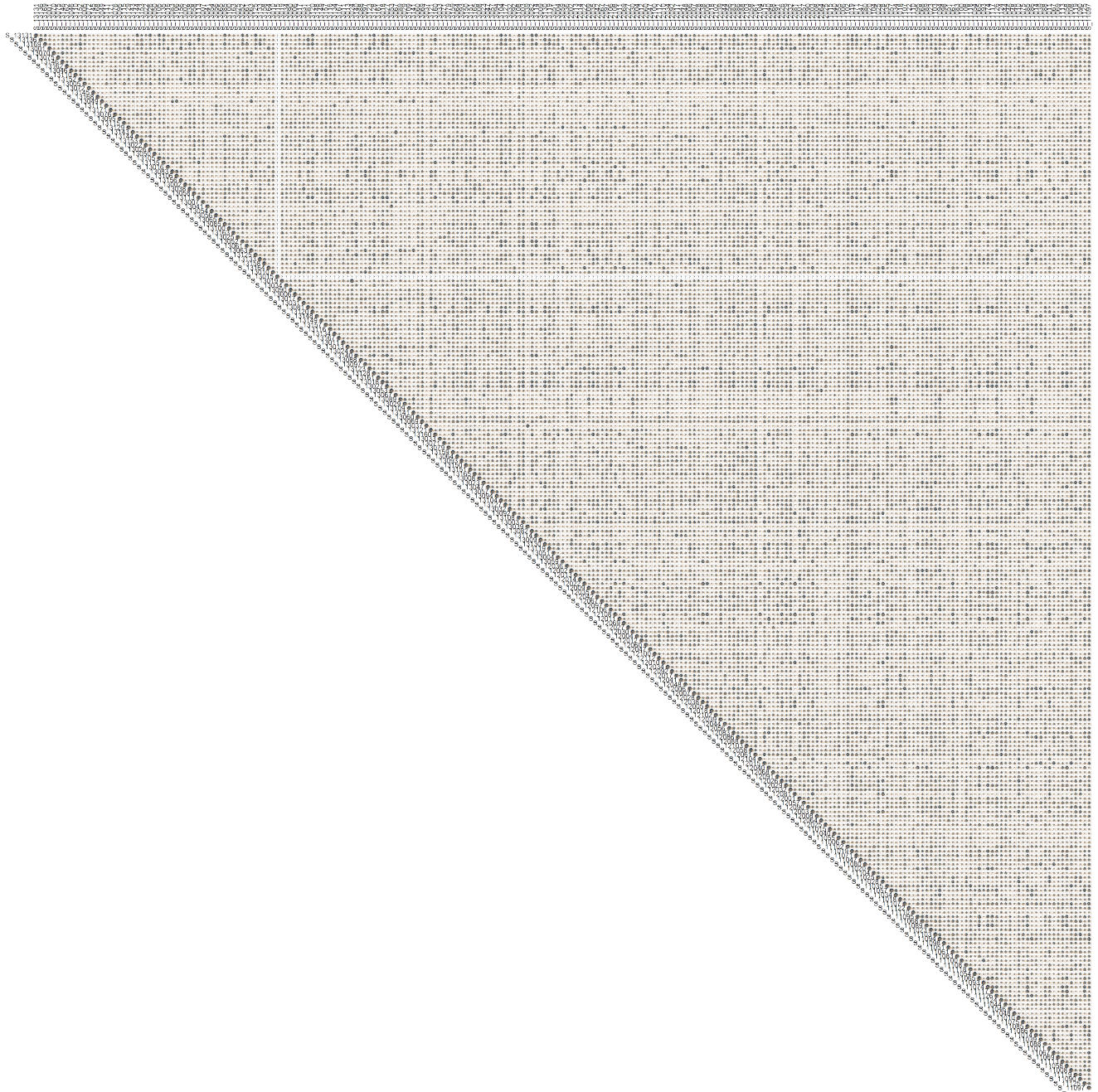
Data showing a cumulative sum of the ‘beadonly’ background phage distribution histogram. Demonstrates that most phages are present in the assay within one log step distance of the median frequency for this experiment. Typically, the median background frequency is around 100 copies, with most phages uniquely 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.

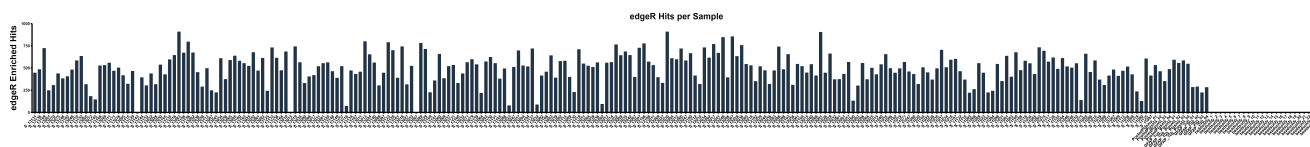
plots/correlation_plot_counts_controls.png



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.

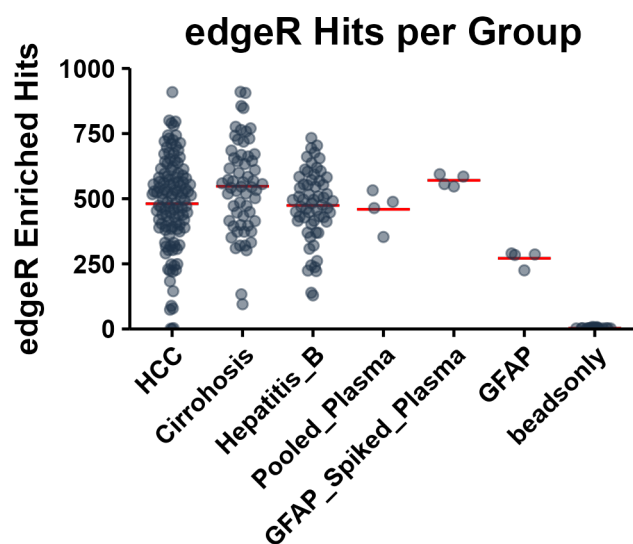
plots/correlation_plot_counts_cohort.png

7 Results Summary Figures from edgeR Enriched Hits



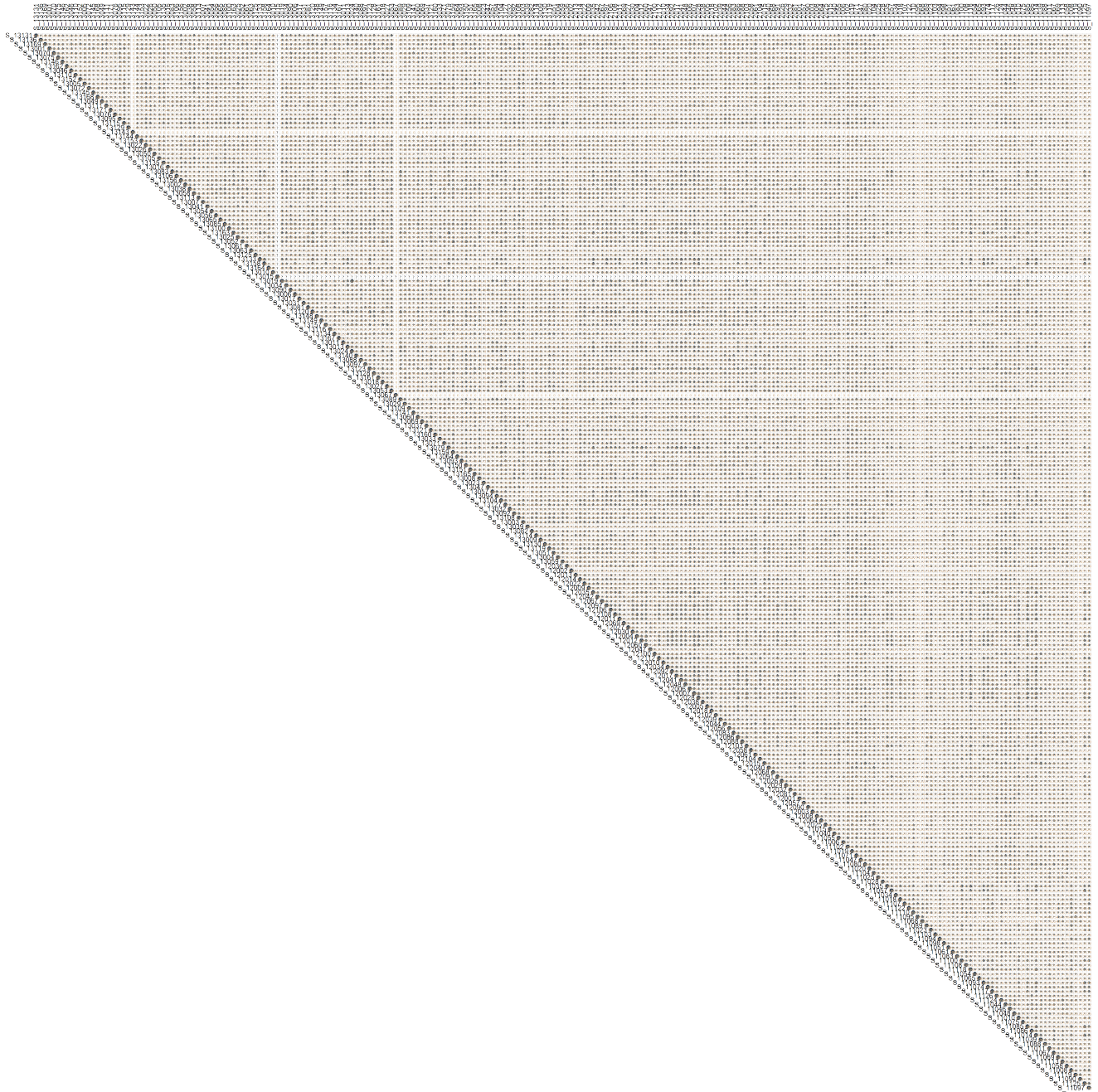
Number of peptide hit enrichments called in each sample. Each plotted bar represents the total number of unique peptide hits called for a single sample as-described in the informatics methods. Hits in the 'beadonly' decoy samples are additionally shown to demonstrate false discoveries in the assay background. 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.

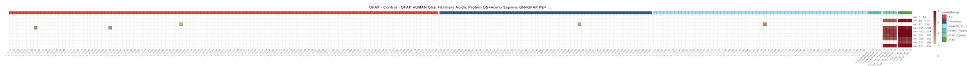
plots/hits_per_group.png



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 '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). A full resolution copy is available in the **scan_heatmaps** folder for this figure and for two other GFAP isoforms.

scan_heatmaps/heatmap_GFAP_Control_GFAP HUMAN Glial Fibrillary Acidic Protein OS=Homo Sapiens GN=GFAP PE=1 SV=1 (Human).png

scan_heatmaps/heatmap_GFAP_Isoform_3_Control_GFAP HUMAN Isoform 3 of Glial Fibrillary Acidic Protein OS=Homo Sapiens GN=GFAP.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.

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.

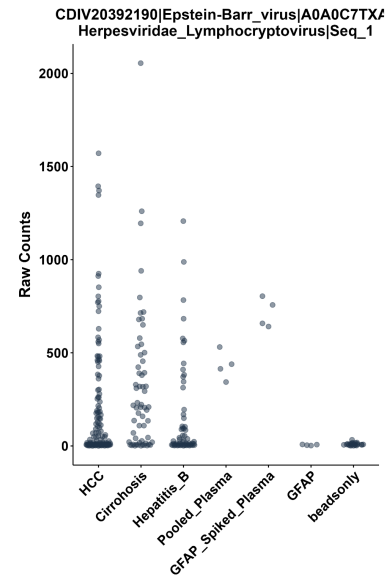
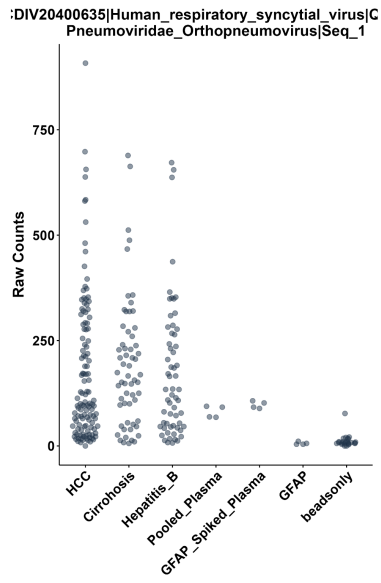
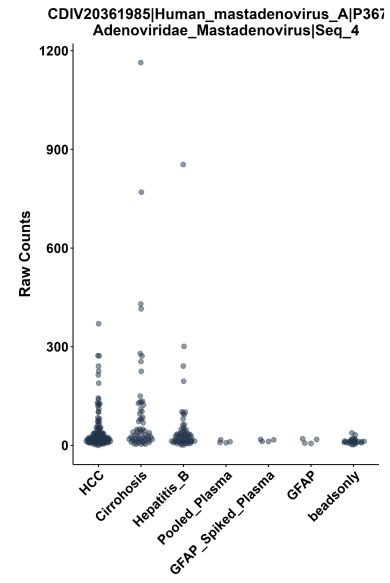
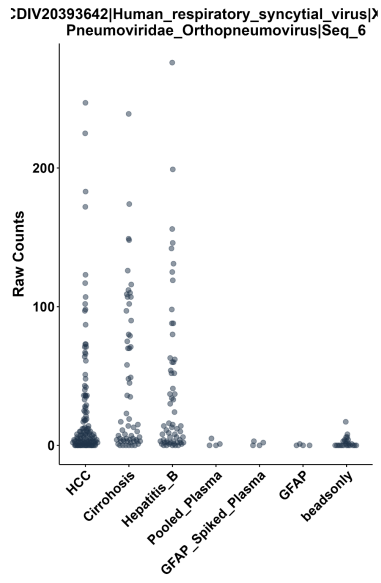
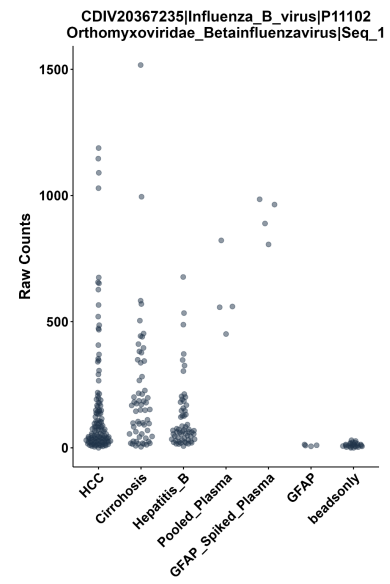
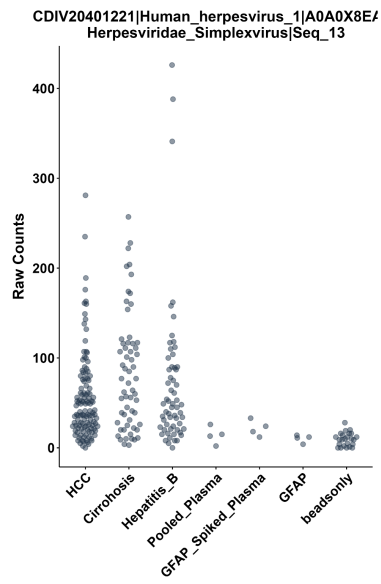
A whole-protein epitope-scanning heatmap is provided in the **scan_heatmaps** folder for each protein bearing one of the top 100 peptides enriched in each sample group.

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
CDIV20401221	Human_herpesvirus_1[AA0A0X8EA29]Herpesviridae_Simplexvirus[Seq_13	20	21	40	160	0.0002641
CDIV20367235	Influenza_B_virus[P11102]Orthomyxoviridae_Betainfluenzavirus[Seq_1	19	69	21	112	0.0003247
CDIV20393642	Human_respiratory_syncytial_virus[X4YAY5]Pneumoviridae_Orthopneu...	39	20	41	161	0.0004312
CDIV20361985	Human_mastadenovirus_A[P3671]Adenoviridae_Mastadenovirus[Seq_4	20	23	40	158	0.0007238
CDIV20400635	Human_respiratory_syncytial_virus[Q3MK18]Pneumoviridae_Orthopneu...	44	88	16	93	0.0009431
CDIV20392190	Epstein-Barr_virus[AA0A0C7TXA7]Herpesviridae_Lymphocryptovirus[Seq_1	41	78	19	103	0.0009602
CDIV20389841	Human_cytomegalovirus[AA0A0G21X8]Herpesviridae_Cytomegalovirus[Seq...	7	2	53	179	0.0010418
CDIV20346580	Human_immunodeficiency_virus_1[U42720]Retroviridae Lentivirus[Seq_8	6	1	54	180	0.0011011
CDIV20347881	Human_immunodeficiency_virus_1[AY173951]Retroviridae Lentivirus[Seq...	6	1	54	180	0.0011011
CDIV20042501	Human_mastadenovirus_B[AP_000553]Adenoviridae_Mastadenovirus[Seq_1	31	51	29	130	0.0015097
CDIV20360596	Severe_acute_respiratory_syndrome-related_coronavirus[Q0Q468]Cor...	10	7	50	174	0.0020801
CDIV20363550	Influenza_B_virus[Q36433]Orthomyxoviridae_Betainfluenzavirus[Seq_1	8	4	52	177	0.0020821
CDIV20119324	Influenza_B_virus[NP_056661]Orthomyxoviridae_Betainfluenzavirus[Seq...	36	67	24	114	0.0024778
CDIV20417051	Human_herpesvirus_1[Q09176]Herpesviridae_Simplexvirus[Seq_1	11	9	49	172	0.0025267
CDIV20399394	Human_respiratory_syncytial_virus[Q3MK39]Pneumoviridae_Orthopneu...	44	92	16	89	0.0025815
CDIV20415000	Enterovirus_D1E7FM39]Picornaviridae_Enterovirus[Seq_2	13	13	47	168	0.0033817
CDIV20353887	Human_immunodeficiency_virus_1[AF119819]Retroviridae Lentivirus[Seq...	6	2	54	179	0.0035382
CDIV20353727	Human_immunodeficiency_virus_1[AF286239]Retroviridae Lentivirus[Seq...	4	0	56	181	0.0035572
CDIV20355782	Human_immunodeficiency_virus_1[DQ085872]Retroviridae Lentivirus[Seq...	4	0	56	181	0.0035572
CDIV20261391	Wesselsbron_virus[YP_002922020]Flaviviridae_Flavivirus[Seq_5	5	1	55	180	0.0040636
CDIV20348756	Human_immunodeficiency_virus_1[EU697904]Retroviridae Lentivirus[Seq...	5	1	55	180	0.0040636
CDIV20457438	Human_respiratory_syncytial_virus[A7L3J6]Pneumoviridae_Orthopneu...	5	1	55	180	0.0040636
CDIV20410311	Epstein-Barr_virus[E2E7U4]Herpesviridae_Lymphocryptovirus[Seq_1	34	64	26	117	0.0041155
CDIV20362352	Enterovirus_B[Q66790]Picornaviridae_Enterovirus[Seq_13	43	90	17	91	0.0042017
CDIV20069829	Mupapillomavirus_1[NP_597739]Papillomaviridae_Mupapillomavirus[S...	8	5	52	176	0.0043378
CDIV20448624	Human_rhinovirus_sp.[Q4U254]Picornaviridae_Enterovirus[Seq_2	8	5	52	176	0.0043378
CDIV20475507	Human_respiratory_syncytial_virus[Q8V5H8]Pneumoviridae_Orthopneu...	20	28	40	153	0.0046678
CDIV20395186	Human_respiratory_syncytial_virus[Q38I82]Pneumoviridae_Orthopneu...	18	23	42	158	0.0047336
CDIV20469511	Human_respiratory_syncytial_virus[B0LS41]Pneumoviridae_Orthopneu...	16	19	44	162	0.0049934
CDIV20360597	Human_mastadenovirus_F[P19416]Adenoviridae_Mastadenovirus[Seq_1	12	12	48	169	0.0052372
CDIV20451187	Human_respiratory_syncytial_virus[AA0A0B5H44]Pneumoviridae_Ortho...	20	29	40	152	0.0055147
CDIV20376960	Influenza_A_virus[AAU4780]Orthomyxoviridae_Alphainfluenzavirus[Se...	9	7	51	174	0.0055302
CDIV20462971	Human_herpesvirus_1[AA0A0X8E9W2]Herpesviridae_Simplexvirus[Seq_2	24	38	36	143	0.0059490
CDIV20379369	Alphapapillomavirus_6[Q80958]Papillomaviridae_Alphapapillomaviru...	7	4	53	177	0.0061431
CDIV20442037	Human_respiratory_syncytial_virus[Q5XX87]Pneumoviridae_Orthopneu...	34	65	26	116	0.0062365
CDIV20398528	Human_respiratory_syncytial_virus[R9TD96]Pneumoviridae_Orthopneu...	28	49	32	132	0.0064827
CDIV20386540	Human_respiratory_syncytial_virus[AA0A059VDK9]Pneumoviridae_Ortho...	18	25	42	156	0.0064970
CDIV20410144	Human_respiratory_syncytial_virus[AA0A075BF85]Pneumoviridae_Ortho...	29	51	31	130	0.0067812
CDIV20390267	Human_respiratory_syncytial_virus[C1LP1]Pneumoviridae_Orthopneu...	21	32	39	149	0.0069459
CDIV20405470	Epstein-Barr_virus[AA0A03AZS7]Herpesviridae_Lymphocryptovirus[Seq_1	29	52	31	129	0.0071945
CDIV20417058	Human_herpesvirus_1[Q09176]Herpesviridae_Simplexvirus[Seq_8	19	27	41	154	0.0072839
CDIV20383428	Sindbis_virus[Q8Q0Z3]Togaviridae_Alphavirus[Seq_39	17	22	43	159	0.0075244
CDIV20399335	Human_respiratory_syncytial_virus[Q91941]Pneumoviridae_Orthopneu...	22	34	38	147	0.0075800
CDIV20420361	Human_respiratory_syncytial_virus[Q4KV07]Pneumoviridae_Orthopneu...	30	55	30	126	0.0078723
CDIV20360119	Influenza_A_virus[AAU6V2]Orthomyxoviridae_Alphainfluenzavirus[Se...	8	6	52	175	0.0081105
CDIV20414092	Human_mastadenovirus_B[Q2KRX1]Adenoviridae_Mastadenovirus[Seq_3	11	11	49	170	0.0081184
CDIV20118705	Myotis_gammaherpesvirus_8[YP_009229850]Herpesviridae_NA[Seq_11	6	3	54	178	0.0085311
CDIV20356084	Human_immunodeficiency_virus_1[AB286851]Retroviridae Lentivirus[Seq...	6	3	54	178	0.0085311
CDIV20399429	Human_immunodeficiency_virus_1[Q9Q6W4]Retroviridae Lentivirus[Seq_3	6	3	54	178	0.0085311
CDIV20349373	Human_immunodeficiency_virus_1[AY008715]Retroviridae Lentivirus[Seq...	15	19	45	162	0.0091542
CDIV20353679	Human_immunodeficiency_virus_1[AY588970]Retroviridae Lentivirus[Seq...	15	19	45	162	0.0091542
CDIV20399972	Human_respiratory_syncytial_virus[AA0A075BF97]Pneumoviridae_Ortho...	15	19	45	162	0.0091542
CDIV20464126	Cowpox_virus[Q72759]Poxviridae_Orthopoxvirus[Seq_31	15	19	45	162	0.0091542
CDIV20436223	Human_herpesvirus_1[AA0A0F7GT05]Herpesviridae_Simplexvirus[Seq_8	17	24	43	157	0.0099402
CDIV20408137	Human_respiratory_syncytial_virus[E3W9K0]Pneumoviridae_Orthopneu...	28	51	32	130	0.0109310
CDIV20417033	Human_respiratory_syncytial_virus[G8EIT1]Pneumoviridae_Orthopneu...	38	79	22	102	0.0109970
CDIV20426911	Human_respiratory_syncytial_virus[U3LLK0]Pneumoviridae_Orthopneu...	15	20	45	161	0.0111760
CDIV20375828	Human_respirovirus_3[P06159]Paramyxoviridae_Respirovirus[Seq_10	18	26	42	155	0.0112550
CDIV20403411	Human_respiratory_syncytial_virus[AA0A068B4W3]Pneumoviridae_Ortho...	18	26	42	155	0.0112550
CDIV20456592	Human_respiratory_syncytial_virus[E3WHY5]Pneumoviridae_Orthopneu...	29	53	31	128	0.0114090
CDIV20342899	Tanapox_virus[NP_073464]Poxviridae_Yatapoxvirus[Seq_2	5	2	55	179	0.0114700
CDIV20345555	HMO_Astrovirus_A[YP_003275951]Astroviridae_Mamastrovirus[Seq_19	5	2	55	179	0.0114700
CDIV20347803	Human_immunodeficiency_virus_1[FN392874]Retroviridae Lentivirus[Seq...	5	2	55	179	0.0114700
CDIV20348570	Human_immunodeficiency_virus_1[AY093603]Retroviridae Lentivirus[Seq...	5	2	55	179	0.0114700
CDIV20374531	Hepatitis_B_virus[Q69602]Hepadnaviridae_Orthohepadnavirus[Seq_5	5	2	55	179	0.0114700
CDIV20424284	Human_herpesvirus_1[AA0A0X8EA32]Herpesviridae_Simplexvirus[Seq_11	5	2	55	179	0.0114700
CDIV20428597	Epstein-Barr_virus[AA0A052YR54]Herpesviridae_Lymphocryptovirus[Seq_2	5	2	55	179	0.0114700
CDIV20346319	Human_immunodeficiency_virus_1[EU697908]Retroviridae Lentivirus[Seq...	13	16	47	165	0.0118240
CDIV20261182	Human_papillomavirus_KC5[YP_009134746]Papillomaviridae_Gammapapil...	7	5	53	176	0.0118280
CDIV20449139	Human_respiratory_syncytial_virus[AA0A0B5HKD1]Pneumoviridae_Ortho...	29	54	31	127	0.0119290
CDIV20379606	Rhinovirus_A[Q82122]Picornaviridae_Enterovirus[Seq_3	12	13	48	168	0.0122620
CDIV20449413	Human_herpesvirus_1[I3TC88]Herpesviridae_Simplexvirus[Seq_7	12	13	48	168	0.0122620
CDIV20391566	Human_respiratory_syncytial_virus[W5RNM1]Pneumoviridae_Orthopneu...	30	56	30	125	0.0123860
CDIV20450743	Human_respiratory_syncytial_virus[B7U384]Pneumoviridae_Orthopneu...	30	56	30	125	0.0123860
CDIV20456947	Human_respiratory_syncytial_virus[B2Z2Y7]Pneumoviridae_Orthopneu...	10	10	50	171	0.0126180
CDIV20418316	Human_respiratory_syncytial_virus[E9NW58]Pneumoviridae_Orthopneu...	21	34	39	147	0.0128200
CDIV20408746	Human_respiratory_syncytial_virus[Q9DLB2]Pneumoviridae_Orthopneu...	18	27	42	154	0.0128500
CDIV20400766	Human_respiratory_syncytial_virus[Q80R93]Pneumoviridae_Orthopneu...	25	44	35	137	0.0132610
CDIV20442002	Epstein-Barr_virus[R4QWP9]Herpesviridae_Lymphocryptovirus[Seq_9	53	132	7	49	0.0137530
CDIV20381238	Human_respiratory_syncytial_virus[Q50G60]Pneumoviridae_Orthopneu...	30	58	30	123	0.0138720
CDIV20261887	Rhinovirus_B[NP_041009]Picornaviridae_Enterovirus[Seq_5	8	7	52	174	0.0139270
CDIV20355524	Human_immunodeficiency_virus_1[E091932]Retroviridae Lentivirus[Seq...	14	18	46	163	0.0142190
CDIV20198883	Japanese_encephalitis_virus[NP_059434]Flaviviridae_Flavivirus[Se...	4	1	56	180	0.0144240
CDIV20346927	Human_immunodeficiency_virus_1[AY371159]Retroviridae Lentivirus[Seq...	4	1	56	180	0.0144240
CDIV20354519	Human_immunodeficiency_virus_1[HQ385479]Retroviridae Lentivirus[Seq...	4	1	56	180	0.0144240
CDIV20358297	Human_immunodeficiency_virus_1[AY588971]Retroviridae Lentivirus[Seq...	4	1	56	180	0.0144240
CDIV20358686	Human_immunodeficiency_virus_2[AB731742]Retroviridae Lentivirus[Seq...	4	1	56	180	0.0144240
CDIV20364867	Venezuelan_equine_encephalitis_virus[P36330]Togaviridae_Alphavir...	4	1	56	180	0.0144240
CDIV20375929	Rotavirus_B[Q86108]Reoviridae_Rotavirus[Seq_4	4	1	56	180	0.0144240
CDIV20391689	Guanarito_mammarenavirus[D1Z337]Arenaviridae_Mammarenavirus[Seq_7	4	1	56	180	0.0144240
CDIV20400135	Zika_virus[AA0A1D6XW9]Flaviviridae_Flavivirus[Seq_19	4	1	56	180	0.0144240
CDIV20411190	Human_mastadenovirus_A[T1UEZ7]Adenoviridae_Mastadenovirus[Seq_5	4	1	56	180	0.0144240
CDIV20439238	Human_respirovirus_3[T1UD36]Paramyxoviridae_Respirovirus[Seq_6	4	1	56	180	0.0144240
CDIV20060444	Human_mastadenovirus_B[YP_002213818]Adenoviridae_Mastadenovirus[Seq...	3	0	57	181	0.0148530
CDIV20070490	Bundibugyo_ebolavirus[YP_003815432]Filoviridae_Ebolavirus[Seq_11	3	0	57	181	0.0148530
CDIV20134553	Human_herpesvirus_6B[NP_050198]Herpesviridae_Roseolovirus[Seq_5	3	0	57	181	0.0148530
CDIV20145689	Modoc_virus[NP_619758]Flaviviridae_Flavivirus[Seq_40	3	0	57	181	0.0148530
CDIV20209042	Epstein-Barr_virus[YP_401659]Herpesviridae_Lymphocryptovirus[Seq_2	3	0	57	181	0.0148530
CDIV20346618	Human_immunodeficiency_virus_1[AF190127]Retroviridae Lentivirus[Seq...	3	0	57	181	0.0148530
CDIV20371288	Marburg_marburgvirus[Q1PD54]Filoviridae_Marburgvirus[Seq_37	3	0	57	181	0.0148530

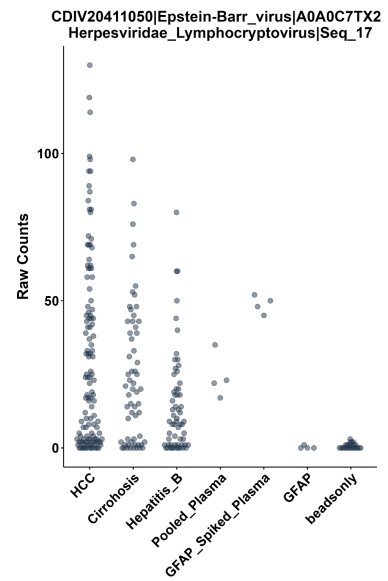
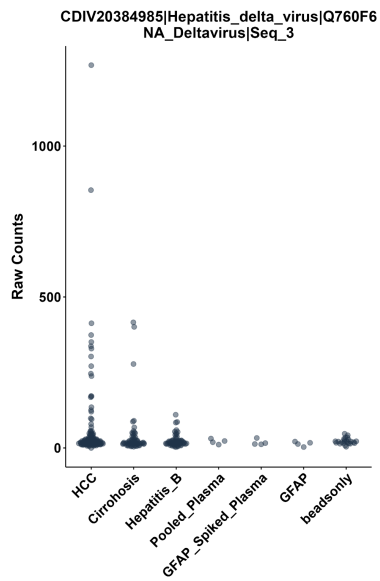
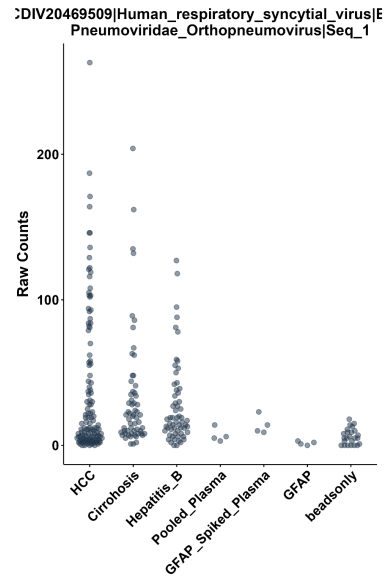
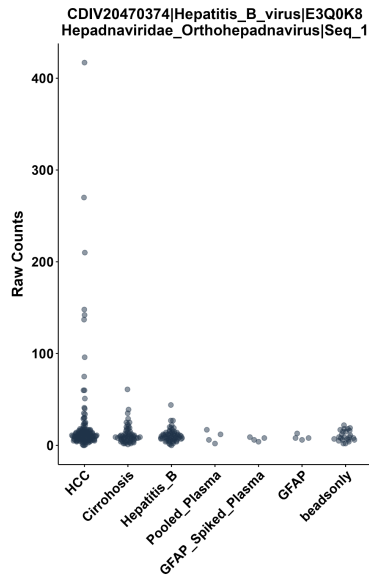
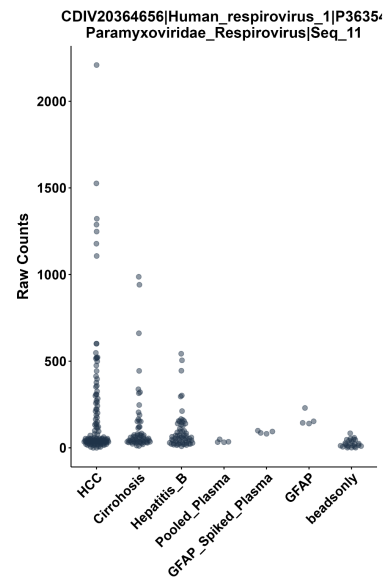
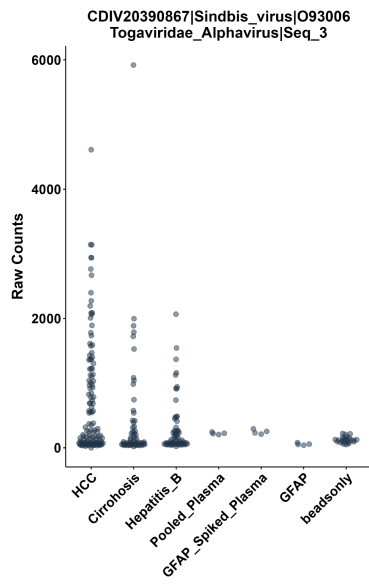
fisher_exact_tests/Vietnam_Cirrhosis_vs_cohort_fisher_exact_tests.csv



Vietnam HCC vs cohort

SeqID	Hits HCC	Hits Cohort	No Hits HCC	No Hits Cohort	p value
CDIV20390867 Sindbis_virus O93006 Togaviridae_Alphavirus Seq_3	53	22	68	98	0.0000253
CDIV20364656 Human_respirovirus_1 P36354 Paramyxoviridae_Respirovirus Seq_11	32	14	89	106	0.0049711
CDIV20470374 Hepatitis_B_virus E3Q0K8 Hepadnaviridae_Orthohepadnavirus Seq_1	8	0	113	120	0.0069331
CDIV20469509 Human_respiratory_syncytial_virus B0L541 Pneumoviridae_Orthopneu...	24	9	97	111	0.0079669
CDIV20384985 Hepatitis_delta_virus Q760F6 NA_Deltavirus Seq_3	16	4	105	116	0.0086798
CDIV20411050 Epstein-Barr_virus A0A0C7TX23 Herpesviridae_Lymphocryptovirus Se...	51	31	70	89	0.0096562
CDIV20362059 Human_respirovirus_1 P26590 Paramyxoviridae_Respirovirus Seq_11	10	1	111	119	0.0100030
CDIV20458675 Hepatitis_B_virus K7QGM2 Hepadnaviridae_Orthohepadnavirus Seq_1	12	2	109	118	0.0105660
CDIV20140001 Gammagapillomavirus_7 YP_004169296 Papillomaviridae_Gammagapillo...	23	9	98	111	0.0126640
CDIV20380972 Human_respiratory_syncytial_virus Q0QXN1 Pneumoviridae_Orthopneu...	7	0	114	120	0.0142920
CDIV20396304 Human_herpesvirus_1 F8RFY5 Herpesviridae_Simplexvirus Seq_1	71	51	50	69	0.0144050
CDIV20435269 Hepatitis_B_virus U3MTJ7 Hepadnaviridae_Orthohepadnavirus Seq_1	15	4	106	116	0.0148470
CDIV20413088 Human_respiratory_syncytial_virus A0A173DTM0 Pneumoviridae_Ortho...	11	2	110	118	0.0190950
CDIV20400339 Human_respiratory_syncytial_virus K4G8J7 Pneumoviridae_Orthopneu...	100	84	21	36	0.0233990
CDIV20387591 Human_cytomegalovirus A0A0A7CGS9 Herpesviridae_Cytomegalovirus S...	14	4	107	116	0.0249950
CDIV20472363 Human_respiratory_syncytial_virus A0A0K1H410 Pneumoviridae_Ortho...	14	4	107	116	0.0249950
CDIV20360560 Epstein-Barr_virus Q3KST1 Herpesviridae_Lymphocryptovirus Seq_1	79	61	42	59	0.0266690
CDIV20342809 Human_herpesvirus_3 NP_040161 Herpesviridae_Varicellovirus Seq_5	24	11	97	109	0.0272250
CDIV20063761 Human_mastadenovirus_E YP_006272954 Adenoviridae_Mastadenovirus ...	6	0	115	120	0.0293290
CDIV20103530 Epstein-Barr_virus YP_001129463 Herpesviridae_Lymphocryptovirus ...	6	0	115	120	0.0293290
CDIV20103535 Epstein-Barr_virus YP_001129463 Herpesviridae_Lymphocryptovirus ...	6	0	115	120	0.0293290
CDIV20385497 Epstein-Barr_virus A0A075FDV6 Herpesviridae_Lymphocryptovirus Seq_9	6	0	115	120	0.0293290
CDIV20388287 Human_respiratory_syncytial_virus B2Z2N1 Pneumoviridae_Orthopneu...	6	0	115	120	0.0293290
CDIV20390983 Hepatitis_B_virus A0A0G3VT08 Hepadnaviridae_Orthohepadnavirus Seq_1	6	0	115	120	0.0293290
CDIV20408485 Hepatitis_B_virus Q4R152 Hepadnaviridae_Orthohepadnavirus Seq_2	6	0	115	120	0.0293290
CDIV20103534 Epstein-Barr_virus YP_001129463 Herpesviridae_Lymphocryptovirus ...	12	3	109	117	0.0297690
CDIV20424719 Hepatitis_B_virus A0A142MK62 Hepadnaviridae_Orthohepadnavirus Seq_1	12	3	109	117	0.0297690
CDIV20007552 Torque_teno_mini_virus_8 YP_003587826 Anelloviridae_Betatorquevi...	21	9	100	111	0.0305500
CDIV20430090 Human_herpesvirus_3 Q0Z9V5 Herpesviridae_Varicellovirus Seq_1	21	9	100	111	0.0305500
CDIV20467057 Human_respiratory_syncytial_virus S5M3T2 Pneumoviridae_Orthopneu...	52	35	69	85	0.031960
CDIV20389859 Epstein-Barr_virus A0A0C7X583 Herpesviridae_Lymphocryptovirus Se...	10	2	111	118	0.0339850
CDIV20457569 Human_respiratory_syncytial_virus A0A0U4AV18 Pneumoviridae_Ortho...	10	2	111	118	0.0339850
CDIV20469752 Human_cytomegalovirus C8BZK25 Herpesviridae_Cytomegalovirus Seq_3	10	2	111	118	0.0339850
CDIV20458431 Human_respiratory_syncytial_virus X2C2D3 Pneumoviridae_Orthopneu...	99	84	22	36	0.0355660
CDIV20415983 Human_respiratory_syncytial_virus A0A127KVZ0 Pneumoviridae_Ortho...	8	1	113	119	0.0356130
CDIV20444489 Human_respiratory_syncytial_virus U6BZQ0 Pneumoviridae_Orthopneu...	74	57	47	63	0.0387840
CDIV20459136 Hepatitis_B_virus D2X485 Hepadnaviridae_Orthohepadnavirus Seq_1	13	4	108	116	0.0413640
CDIV20414500 Human_respiratory_syncytial_virus 11V4U9 Pneumoviridae_Orthopneu...	41	26	80	94	0.0438480
CDIV20455165 Human_respiratory_syncytial_virus Q6ECJ5 Pneumoviridae_Orthopneu...	92	77	29	43	0.0495120
CDIV20394063 Human_herpesvirus_1 G8H888 Herpesviridae_Simplexvirus Seq_60	11	3	110	117	0.0503700
CDIV20444786 Hepatitis_B_virus U3MSN1 Hepadnaviridae_Orthohepadnavirus Seq_1	11	3	110	117	0.0503700
CDIV20441460 Influenza_A_virus Q67352 Orthomyxoviridae_Alphainfluenzavirus Seq_7	25	13	96	107	0.0509210
CDIV20435016 Human_respiratory_syncytial_virus A0A075LYA3 Pneumoviridae_Ortho...	21	10	100	110	0.0530710
CDIV20412566 Human_respiratory_syncytial_virus V5N801 Pneumoviridae_Orthopneu...	71	55	50	65	0.0533670
CDIV20384135 Human_respiratory_syncytial_virus Q3MJZ5 Pneumoviridae_Orthopneu...	14	5	107	115	0.0534070
CDIV20408195 Hepatitis_delta_virus A0A0N6Y5S1 NA_Deltavirus Seq_3	14	5	107	115	0.0534070
CDIV20440027 Human_cytomegalovirus D2W9P1 Herpesviridae_Cytomegalovirus Seq_4	49	34	72	86	0.0576630
CDIV20406701 Human_respiratory_syncytial_virus A0A068B4C9 Pneumoviridae_Ortho...	22	11	99	109	0.0598090
CDIV20150653 Kysanur_forest_disease_virus NP_722551 Flaviviridae_Flavivirus ...	5	0	116	120	0.0599230
CDIV20167464 Human_respiratory_syncytial_virus NP_056858 Pneumoviridae_Orthop...	5	0	116	120	0.0599230
CDIV20235175 Human_herpesvirus_6B NP_050276 Herpesviridae_Roseolovirus Seq_1	5	0	116	120	0.0599230
CDIV20345881 Human_immunodeficiency_virus_1 J4302647 Retroviridae Lentivirus ...	5	0	116	120	0.0599230
CDIV20368424 Influenza_A_virus Q6XSW4 Orthomyxoviridae_Alphainfluenzavirus Seq_2	5	0	116	120	0.0599230
CDIV20389858 Epstein-Barr_virus A0A0C7X588 Herpesviridae_Lymphocryptovirus Se...	5	0	116	120	0.0599230
CDIV20395483 Human_polyomavirus_2 Q91NN6 Polyomaviridae_Betapolyomavirus Seq_8	5	0	116	120	0.0599230
CDIV20399753 Human_coronavirus_229E Q1HVN6 Coronaviridae_Alphacoronavirus Seq_5	5	0	116	120	0.0599230
CDIV20426041 Human_cytomegalovirus A0A140GX97 Herpesviridae_Cytomegalovirus S...	5	0	116	120	0.0599230
CDIV20427389 Hepatitis_B_virus I3VPK9 Hepadnaviridae_Orthohepadnavirus Seq_2	5	0	116	120	0.0599230
CDIV20448626 Human_rhinovirus_sp. Q4U254 Picornaviridae_Enterovirus Seq_4	5	0	116	120	0.0599230
CDIV20464615 Human_mastadenovirus_A D3J54 Adenoviridae_Mastadenovirus Seq_12	5	0	116	120	0.0599230
CDIV20384607 Human_cytomegalovirus Q6SWZ1 Herpesviridae_Cytomegalovirus Seq_4	40	26	81	94	0.0601460
CDIV20414769 Human_respiratory_syncytial_virus E9NW86 Pneumoviridae_Orthopneu...	52	37	69	83	0.0617030
CDIV20448558 Human_respiratory_syncytial_virus X2C1W3 Pneumoviridae_Orthopneu...	52	37	69	83	0.0617030
CDIV20163555 Yaba_monkey_tumor_virus NP_938294 Poxviridae_Yatapoxvirus Seq_1	7	1	114	119	0.0658030
CDIV20180908 Human_mastadenovirus_E YP_068034 Adenoviridae_Mastadenovirus Seq_1	7	1	114	119	0.0658030
CDIV20393118 Hepatitis_delta_virus E9KP24 NA_Deltavirus Seq_1	7	1	114	119	0.0658030
CDIV20393755 Human_cytomegalovirus V9PSV6 Herpesviridae_Cytomegalovirus Seq_3	7	1	114	119	0.0658030
CDIV20414418 Human_respiratory_syncytial_virus Q3MJZ8 Pneumoviridae_Orthopneu...	7	1	114	119	0.0658030
CDIV20419619 Hepatitis_B_virus B6D725 Hepadnaviridae_Orthohepadnavirus Seq_1	7	1	114	119	0.0658030
CDIV20466224 Human_respiratory_syncytial_virus Q9IUC5 Pneumoviridae_Orthopneu...	7	1	114	119	0.0658030
CDIV20468650 Hepatitis_B_virus A0A1B1ABP6 Hepadnaviridae_Orthohepadnavirus Seq_1	7	1	114	119	0.0658030
CDIV20470888 Human_polyomavirus_2 Q8JVB1 Polyomaviridae_Betapolyomavirus Seq_2	7	1	114	119	0.0658030
CDIV20473033 Human_herpesvirus_1 A0A1C37A4 Herpesviridae_Simplexvirus Seq_15	7	1	114	119	0.0658030
CDIV20099596 Tanapox_virus NP_073461 Poxviridae_Yatapoxvirus Seq_1	28	16	93	104	0.0659270
CDIV20425691 Human_rubulavirus_4 C4NTR1 Paramyxoviridae_Rubulavirus Seq_2	19	9	102	111	0.0690000
CDIV20364306 Human_mastadenovirus_C P03243-4 Adenoviridae_Mastadenovirus Seq_1	29	17	92	103	0.0707530
CDIV20457687 Epstein-Barr_virus A0A0U3THE3 Herpesviridae_Lymphocryptovirus Se...	46	32	75	88	0.0734240
CDIV20363340 Epstein-Barr_virus Q07286 Herpesviridae_Lymphocryptovirus Seq_1	30	18	91	102	0.0754690
CDIV20280002 Middle_East_respiratory_syndrome-related_coronavirus YP_00904721...	20	10	101	110	0.0778210
CDIV20427458 Human_respiratory_syncytial_virus A0A0A7NQZ4 Pneumoviridae_Ortho...	20	10	101	110	0.0778210
CDIV20450745 Human_respiratory_syncytial_virus L7QHM5 Pneumoviridae_Orthopneu...	20	10	101	110	0.0778210
CDIV20379855 Hepacivirus_C Q00269 Flaviviridae_Hepacivirus Seq_1	10	3	111	117	0.0836160
CDIV20448160 Human_cytomegalovirus Q6RXH9 Herpesviridae_Cytomegalovirus Seq_8	10	3	111	117	0.0836160
CDIV20465611 Hepatitis_delta_virus A0A0N6YRG4 NA_Deltavirus Seq_3	10	3	111	117	0.0836160
CDIV20457926 Human_herpesvirus_3 G9IX77 Herpesviridae_Varicellovirus Seq_3	26	15	95	105	0.0857370
CDIV20432641 Human_respiratory_syncytial_virus E9NW32 Pneumoviridae_Orthopneu...	81	67	40	53	0.0861720
CDIV20457436 Human_respiratory_syncytial_virus A7L3J6 Pneumoviridae_Orthopneu...	81	67	40	53	0.0861720
CDIV20400723 Human_respiratory_syncytial_virus W5RQ57 Pneumoviridae_Orthopneu...	93	80	28	40	0.0871280
CDIV20472220 Human_herpesvirus_1 G8H8J0 Herpesviridae_Simplexvirus Seq_1	78	64	43	56	0.0894470
CDIV20406331 Human_respiratory_syncytial_virus A0A075BF87 Pneumoviridae_Ortho...	27	16	94	104	0.0918430
CDIV20446409 Human_respiratory_syncytial_virus U3LVQ0 Pneumoviridae_Orthopneu...	74	60	47	60	0.0925360
CDIV20399152 Human_respiratory_syncytial_virus Q0QXF9 Pneumoviridae_Orthopneu...	44	31	77	89	0.0949750
CDIV20321692 Epstein-Barr_virus YP_401666 Herpesviridae_Lymphocryptovirus Seq_2	35	23	86	97	0.0971350
CDIV20400875 Human_cytomegalovirus Q91D82 Herpesviridae_Cytomegalovirus Seq_2	35	23	86	97	0.0971350
CDIV20367705 Human_rhinovirus_B Q82081 Picornaviridae_Enterovirus Seq_5	28	17	93	103	0.0977710
CDIV20404726 Epstein-Barr_virus K9UT27 Herpesviridae_Lymphocryptovirus Seq_20	14	6	107	114	0.1002100
CDIV20382106 Hepatitis_delta_virus A0A0N6YRF6 NA_Deltavirus Seq_3	18	9	103	111	0.1009900
CDIV20408127 Human_respiratory_syncytial_virus Q0QXK9 Pneumoviridae_Orthopneu...	18	9	103	111	0.1009900
CDIV20421115 Norwalk_virus Q9DU47 Caliciviridae_Norovirus Seq_21	18	9	103	111	0.1009900
CDIV20078839 Tanapox_virus NP_073481 Poxviridae_Yatapoxvirus Seq_3	8	2	113	118	0.1020800

fisher_exact_tests/Vietnam_HCC_vs_cohort_fisher_exact_tests.csv

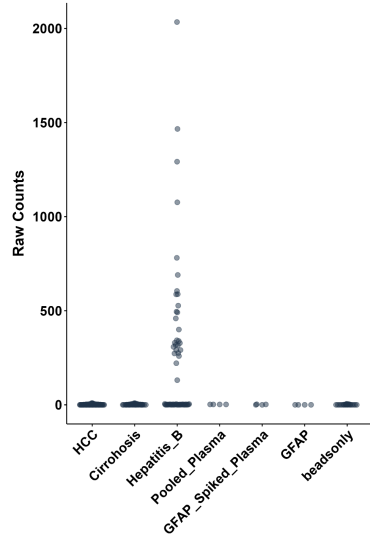


Vietnam Hepatitis B vs cohort

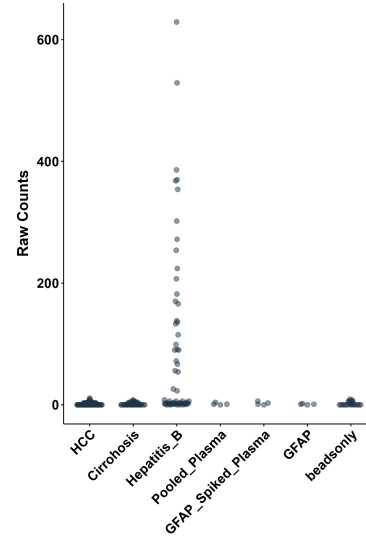
SeqID	Hits Hepatitis B	Hits Cohort	No Hits Hepatitis B	No Hits Cohort	p value
CDIV20057555 Eastern_equine_encephalitis_virus YP_006491233 Togaviridae_Alpha...	28	0	32	181	0.0000e+00
CDIV20177780 Variola_virus NP_042071 Poxviridae_Orthopoxvirus Seq_2	22	0	38	181	0.0000e+00
CDIV20375744 Influenza_A_virus P03448 Orthomyxoviridae_Alphainfluenzavirus Se...	25	0	35	181	0.0000e+00
CDIV20431012 Hepatitis_B_virus D6QW13 Hepadnaviridae_Orthohepadnavirus Seq_3	28	0	32	181	0.0000e+00
CDIV20432553 Human_cytomegalovirus A0A0G2TP82 Herpesviridae_Cytomegalovirus S...	28	0	32	181	0.0000e+00
CDIV20453668 West_Nile_virus A0A076E7A3 Flaviviridae_Flavivirus Seq_7	26	0	34	181	0.0000e+00
CDIV20336727 Human_associated_cyclovirus_8 YP_009047065 Circoviridae_Cyclovir...	21	0	39	181	1.0000e-14
CDIV20247795 Enterovirus_D NP_040760 Picornaviridae_Enterovirus Seq_32	19	0	41	181	2.8000e-13
CDIV20261645 Human_mastadenovirus_B YP_002213767 Adenoviridae_Mastadenovirus ...	19	0	41	181	2.8000e-13
CDIV20350994 Human_immunodeficiency_virus_1 FN392876 Retroviridae Lentivirus ...	19	0	41	181	2.8000e-13
CDIV20448125 West_Nile_virus A0A067Z569 Flaviviridae_Flavivirus Seq_36	19	0	41	181	2.8000e-13
CDIV20357808 Human_immunodeficiency_virus_1 EU697906 Retroviridae Lentivirus ...	18	0	42	181	1.5100e-12
CDIV20371479 Rhinovirus_A P23008 Picornaviridae_Enterovirus Seq_18	17	0	43	181	7.8600e-12
CDIV20412255 Human_herpesvirus_1 A0A110AV91 Herpesviridae_Simplexvirus Seq_6	17	0	43	181	7.8600e-12
CDIV20369692 Human_cytomegalovirus P16847 Herpesviridae_Cytomegalovirus Seq_2	16	0	44	181	4.0210e-11
CDIV20179538 Human_coronavirus_229E NP_073556 Coronaviridae_Alphacoronavirus ...	15	0	45	181	2.0195e-10
CDIV20404897 Human_herpesvirus_2 A0A0K0KN18 Herpesviridae_Simplexvirus Seq_23	15	0	45	181	2.0195e-10
CDIV20470693 Hepatitis_delta_virus Q70E10 NA_Deltavirus Seq_2	15	0	45	181	2.0195e-10
CDIV20354013 Human_immunodeficiency_virus_2 JN863898 Retroviridae Lentivirus ...	14	0	46	181	9.9656e-10
CDIV20439895 Human_herpesvirus_8 G3KH28 Herpesviridae_Rhadinovirus Seq_3	14	0	46	181	9.9656e-10
CDIV20448701 Hepatitis_B_virus I0DG61 Hepadnaviridae_Orthohepadnavirus Seq_1	14	0	46	181	9.9656e-10
CDIV20454131 Human_herpesvirus_2 A0A0Y0QNP6 Herpesviridae_Simplexvirus Seq_4	14	0	46	181	9.9656e-10
CDIV20477146 Hepatitis_B_virus U5JQH2 Hepadnaviridae_Orthohepadnavirus Seq_2	14	0	46	181	9.9656e-10
CDIV20466908 Hepatitis_B_virus A0A1B1ABA1 Hepadnaviridae_Orthohepadnavirus Seq_3	15	1	45	180	2.6280e-09
CDIV20020094 Human_mastadenovirus_D YP_003038612 Adenoviridae_Mastadenovirus ...	13	0	47	181	4.8344e-09
CDIV20350053 Human_immunodeficiency_virus_1 AF457051 Retroviridae Lentivirus ...	12	0	48	181	2.3064e-08
CDIV20369988 Betapapillomavirus_1 P50770 Papillomaviridae_Betapapillomavirus ...	12	0	48	181	2.3064e-08
CDIV20064350 Epstein-Barr_virus YP_401693 Herpesviridae_Lymphocryptovirus Seq_3	11	0	49	181	1.0826e-07
CDIV20379909 Alphapapillomavirus_11 Q82007 Papillomaviridae_Alphapapillomavir...	11	0	49	181	1.0826e-07
CDIV20400760 Influenza_A_virus A4GXG3 Orthomyxoviridae_Alphainfluenzavirus Seq_3	11	0	49	181	1.0826e-07
CDIV20139276 Human_herpesvirus_8 YP_001129378 Herpesviridae_Rhadinovirus Seq_2	10	0	50	181	5.0016e-07
CDIV20222327 Mayaro_virus NP_579968 Togaviridae_Alphavirus Seq_34	10	0	50	181	5.0016e-07
CDIV20338449 Mamastrovirus_1 YP_009275018 Astroviridae_Mamastrovirus Seq_6	10	0	50	181	5.0016e-07
CDIV20345540 HMO_Astrovirus_A YP_003275951 Astroviridae_Mamastrovirus Seq_4	10	0	50	181	5.0016e-07
CDIV20351537 Human_immunodeficiency_virus_1 EU581825 Retroviridae Lentivirus ...	10	0	50	181	5.0016e-07
CDIV20391118 Hepatitis_B_virus C1K1Z4 Hepadnaviridae_Orthohepadnavirus Seq_15	10	0	50	181	5.0016e-07
CDIV20424069 Aichi_virus Q8B6B2 Picornaviridae_Kobuvirus Seq_2	10	0	50	181	5.0016e-07
CDIV20440822 Human_cytomegalovirus Q6RXD9 Herpesviridae_Cytomegalovirus Seq_2	10	0	50	181	5.0016e-07
CDIV20368320 Human_betaherpesvirus_6A P24443 Herpesviridae_Roseolovirus Seq_6	9	0	51	181	2.2752e-06
CDIV20397029 Zika_virus A0A166XWL4 Flaviviridae_Flavivirus Seq_43	9	0	51	181	2.2752e-06
CDIV20309552 Cosavirus_E YP_002956072 Picornaviridae_Cosavirus Seq_30	10	1	50	180	4.4192e-06
CDIV20038778 Zaire_ebolavirus NP_066244 Filoviridae_Ebolavirus Seq_6	8	0	52	181	1.0195e-05
CDIV20353851 Human_immunodeficiency_virus_1 AY900572 Retroviridae Lentivirus ...	8	0	52	181	1.0195e-05
CDIV20358815 Human_immunodeficiency_virus_1 GQ372987 Retroviridae Lentivirus ...	8	0	52	181	1.0195e-05
CDIV20378946 Alphapapillomavirus_8 P36727 Papillomaviridae_Alphapapillomaviru...	8	0	52	181	1.0195e-05
CDIV20380645 Human_respiratory_syncytial_virus G8EIZ0 Pneumoviridae_Orthopneu...	8	0	52	181	1.0195e-05
CDIV20380935 Human_cytomegalovirus A0A0G2U480 Herpesviridae_Cytomegalovirus S...	8	0	52	181	1.0195e-05
CDIV20398854 Human_herpesvirus_1 A0A0X8EA33 Herpesviridae_Simplexvirus Seq_6	8	0	52	181	1.0195e-05
CDIV20168469 Hepacivirus_C YP_001469630 Flaviviridae_Hepacivirus Seq_49	7	0	53	181	4.5011e-05
CDIV20348814 Human_immunodeficiency_virus_2 U05357 Retroviridae Lentivirus Se...	7	0	53	181	4.5011e-05
CDIV20350433 Human_immunodeficiency_virus_1 U88826 Retroviridae Lentivirus Se...	7	0	53	181	4.5011e-05
CDIV20414307 Cowpox_virus G0XW74 Poxviridae_Orthopoxvirus Seq_7	7	0	53	181	4.5011e-05
CDIV20108903 Human_cytomegalovirus YP_081561 Herpesviridae_Cytomegalovirus Seq_4	23	25	37	156	0.00012507
CDIV20063748 Human_mastadenovirus_E YP_006272954 Adenoviridae_Mastadenovirus ...	6	0	54	181	0.00019588
CDIV20122350 Rosavirus_A YP_009028557 Picornaviridae_Rosavirus Seq_7	6	0	54	181	0.00019588
CDIV20199062 Rio_Bravo_virus NP_620044 Flaviviridae_Flavivirus Seq_52	6	0	54	181	0.00019588
CDIV20349581 Human_immunodeficiency_virus_1 AY423387 Retroviridae Lentivirus ...	6	0	54	181	0.00019588
CDIV20349942 Human_immunodeficiency_virus_1 EU735535 Retroviridae Lentivirus ...	6	0	54	181	0.00019588
CDIV20358438 Human_immunodeficiency_virus_1 EU581828 Retroviridae Lentivirus ...	6	0	54	181	0.00019588
CDIV20359208 Human_immunodeficiency_virus_1 AY173951 Retroviridae Lentivirus ...	6	0	54	181	0.00019588
CDIV20360949 Influenza_A_virus Q64968 Orthomyxoviridae_Alphainfluenzavirus Seq_6	6	0	54	181	0.00019588
CDIV20361786 Rotavirus_A Q9E780 Reoviridae_Rotavirus Seq_10	6	0	54	181	0.00019588
CDIV20370089 Influenza_A_virus P03476 Orthomyxoviridae_Alphainfluenzavirus Seq_9	6	0	54	181	0.00019588
CDIV20377936 Epstein-Barr_virus P03235 Herpesviridae_Lymphocryptovirus Seq_4	6	0	54	181	0.00019588
CDIV20403070 Human_cytomegalovirus A0A126NBR7 Herpesviridae_Cytomegalovirus S...	6	0	54	181	0.00019588
CDIV20407908 Cowpox_virus U5TB11 Poxviridae_Orthopoxvirus Seq_1	6	0	54	181	0.00019588
CDIV20412938 Hepatitis_B_virus F2X840 Hepadnaviridae_Orthohepadnavirus Seq_4	6	0	54	181	0.00019588
CDIV20429728 Cowpox_virus G0XW68 Poxviridae_Orthopoxvirus Seq_7	6	0	54	181	0.00019588
CDIV20385084 West_Nile_virus A0A067ZR50 Flaviviridae_Flavivirus Seq_39	13	10	47	171	0.00064185
CDIV20389627 Hepatitis_B_virus A0A075BZB0 Hepadnaviridae_Orthohepadnavirus Seq_2	9	4	51	177	0.0006723
CDIV20064106 Cosavirus_B YP_002956075 Picornaviridae_Cosavirus Seq_15	5	0	55	181	0.00084051
CDIV20207525 Marburg_marburgvirus YP_009055228 Filoviridae_Marburgvirus Seq_19	5	0	55	181	0.00084051
CDIV20312144 Human_herpesvirus_6B NP_050255 Herpesviridae_Roseolovirus Seq_14	5	0	55	181	0.00084051
CDIV20347026 Human_immunodeficiency_virus_2 U27200 Retroviridae Lentivirus Seq_1	5	0	55	181	0.00084051
CDIV20352986 Human_immunodeficiency_virus_1 L20571 Retroviridae Lentivirus Seq_7	5	0	55	181	0.00084051
CDIV20361455 Colorado_tick_fever_virus Q9ENK8 Reoviridae_Coltivirus Seq_3	5	0	55	181	0.00084051
CDIV20367811 Alphapapillomavirus_6 Q02515 Papillomaviridae_Alphapapillomaviru...	5	0	55	181	0.00084051
CDIV20381427 Human_herpesvirus_2 A0A0Y0QNI6 Herpesviridae_Simplexvirus Seq_54	5	0	55	181	0.00084051
CDIV20381880 Orf_virus F1AXE0 Poxviridae_Parapoxvirus Seq_1	5	0	55	181	0.00084051
CDIV20382850 Human_cytomegalovirus A0A0G2TC35 Herpesviridae_Cytomegalovirus S...	5	0	55	181	0.00084051
CDIV20385087 West_Nile_virus A0A067ZR50 Flaviviridae_Flavivirus Seq_42	5	0	55	181	0.00084051
CDIV20386201 Human_mastadenovirus_C E7DY96 Adenoviridae_Mastadenovirus Seq_1	5	0	55	181	0.00084051
CDIV20406720 Hepatitis_B_virus F5C0M6 Hepadnaviridae_Orthohepadnavirus Seq_11	5	0	55	181	0.00084051
CDIV20406944 Enterovirus_C Q7T7P4 Picornaviridae_Enterovirus Seq_18	5	0	55	181	0.00084051
CDIV20432788 Human_coronavirus_HKU1 Q0ZJG7 Coronaviridae_Betacoronavirus Seq_20	5	0	55	181	0.00084051
CDIV20445016 Hepatitis_B_virus X2FVR0 Hepadnaviridae_Orthohepadnavirus Seq_1	5	0	55	181	0.00084051
CDIV20354925 Human_immunodeficiency_virus_1 AF289550 Retroviridae Lentivirus ...	6	1	54	180	0.0011011
CDIV20445275 Human_coronavirus_HKU1 Q0ZJ1 Coronaviridae_Betacoronavirus Seq_24	6	1	54	180	0.0011011
CDIV20459900 West_Nile_virus B5U2T6 Flaviviridae_Flavivirus Seq_12	6	1	54	180	0.0011011
CDIV20388038 Human_rubulavirus_4 A0A097F13 Paramyxoviridae_Rubulavirus Seq_4	14	13	46	168	0.0015004
CDIV20351455 Human_immunodeficiency_virus_2 L07625 Retroviridae Lentivirus Se...	26	40	34	141	0.0024031
CDIV20004131 Epstein-Barr_virus YP_001129449 Herpesviridae_Lymphocryptovirus ...	6	2	54	179	0.0035382
CDIV20125919 Human_betaherpesvirus_6A NP_042924 Herpesviridae_Roseolovirus Se...	6	2	54	179	0.0035382
CDIV20200293 Human_herpesvirus_6B NP_050212 Herpesviridae_Roseolovirus Seq_13	6	2	54	179	0.0035382
CDIV20144171 Gyrovirus_GyV3 YP_005454312 Anelloviridae_Gyrovirus Seq_3	4	0	56	181	0.0035572
CDIV20345442 Epstein-Barr_virus YP_401715 Herpesviridae_Lymphocryptovirus Seq_16	4	0	56	181	0.0035572
CDIV20356287 Human_immunodeficiency_virus_1 AF492624 Retroviridae Lentivirus ...	4	0	56	181	0.0035572
CDIV20368950 Severe_acute_respiratory_syndrome-related_coronavirus Q315 S Cor...	4	0	56	181	0.0035572
CDIV20376975 Alphapapillomavirus_4 P22162 Papillomaviridae_Alphapapillomaviru...	4	0	56	181	0.0035572
CDIV20379790 Alphapapillomavirus_7 P24830 Papillomaviridae_Alphapapillomaviru...	4	0	56	181	0.0035572

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_fisher_exact_tests.csv

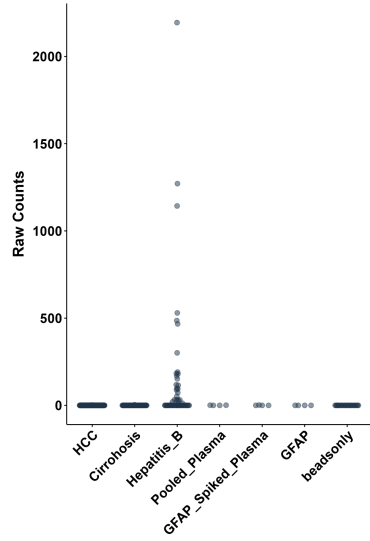
IV20057555|Eastern_equine_encephalitis_virus|YP_
Togaviridae_Alphavirus|Seq_16



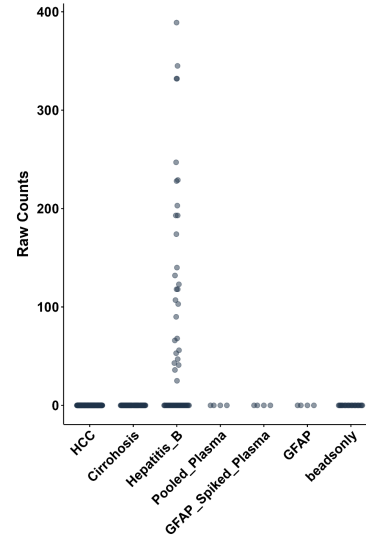
CDIV20177780|Variola_virus|NP_042071
Poxviridae_Orthopoxvirus|Seq_2



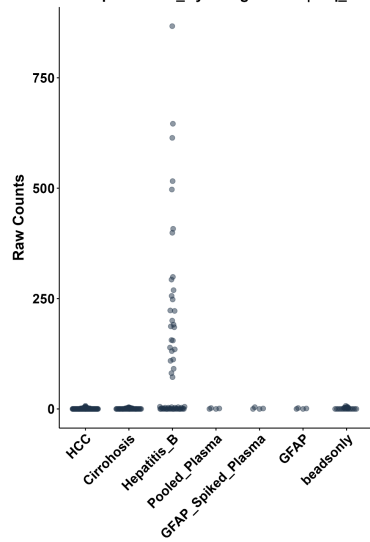
CDIV20375744|Influenza_A_virus|P03448
Orthomyxoviridae_Alphainfluenzavirus|Seq_1



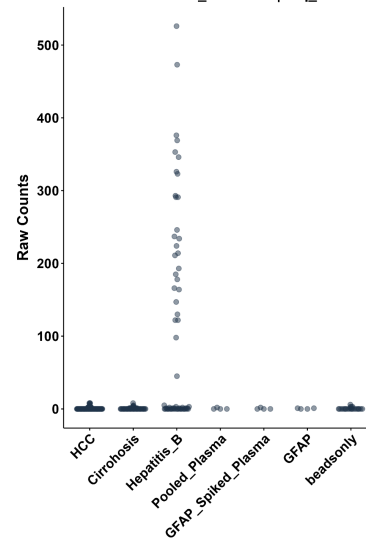
CDIV20431012|Hepatitis_B_virus|D6QW13
Hepadnaviridae_Orthohepadnavirus|Seq_3



CDIV20432553|Human_cytomegalovirus|A0A0G21
Herpesviridae_Cytomegalovirus|Seq_3



CDIV20453668|West_Nile_virus|A0A076E7A3
Flaviviridae_Flavivirus|Seq_7

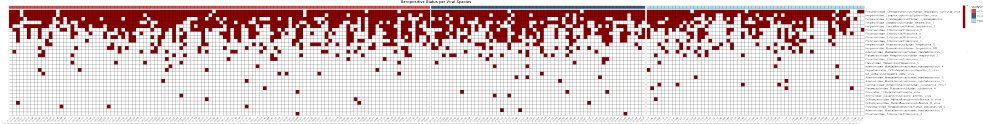


12 Seropositive Tests

Due to the large number of peptides included in the VirScan assay, patients typically generate real, reproducible hits to peptides from viruses which they have never been exposed to. Thus, a VirScan assay should never be considered a replacement for clinical screening. Nonetheless, we have found value calling patients ‘seropositive’ for a virus when there are 2 or more hits against 2 or more proteins per virus. Outputs are analogous to peptide-level and protein-level tests. These are recorded at the level of individual proteins *seropositive_proteins.csv* and individual viral species *seropositive_viruses.csv* in the **Processed_Outputs** folder.

A sample-level unclustered, clustered, and row clustered visualization heatmap is provided for all viruses seropositive for two or more patients in the **summary_heatmaps** folder. These figures display a binary true / false output of seropositivity for each patient.

Seropositive-level Fisher exact tests are performed on a 2x2 matrix of *Seropositive_Hits* and *Seropositive_No_Hits* columns for each provided group versus all other samples in the study cohort using the *fisher.test* function in R. Outputs are analogous to peptide-level and protein-level tests. A record of these outputs is also provided in the *group_seropositive_hit_counts.csv* file in the **Processed_Outputs** folder. If an explicit ‘Control’ group is provided, control tests are also performed.



Vietnam Cirrohosis vs cohort seropositive

Family Genus Species	Seropositive Hits Cirrohosis	Seropositive Hits Cohort	Seropositive No Hits Cirrohosis	Seropositive No Hits Cohort	p value
Picornaviridae_Enterovirus[Rhinovirus_B	29	54	31	127	0.011929
Picornaviridae_Enterovirus[Rhinovirus_A	30	64	30	117	0.048338
Poxviridae_Orthopoxvirus[Cowpox_virus	3	1	57	180	0.048739
Adenoviridae_Mastadenovirus[Human_mastadenovirus_F	2	0	58	181	0.061203
Picornaviridae_Enterovirus[Enterovirus_C	19	36	41	145	0.075429
Paramyxoviridae_Respirovirus[Human_respirovirus_3	10	16	50	165	0.097560
Adenoviridae_Mastadenovirus[Human_mastadenovirus_B	3	2	57	179	0.100210
Arteriviridae_Equartevirus[Equine_arteritis_virus	2	1	58	180	0.153900
Papillomaviridae_Alphapapillomavirus[Alphapapillomavirus_2	1	0	59	181	0.248960
Retroviridae_Lentivirus[Human_immunodeficiency_virus_1	1	0	59	181	0.248960
Picornaviridae_Enterovirus[Enterovirus_B	28	69	32	112	0.287920
Pneumoviridae_Orthopneumovirus[Human_respiratory_syncytial_virus	57	165	3	16	0.419070
Picornaviridae_Enterovirus[Enterovirus_A	1	1	59	180	0.436720
Herpesviridae_Roseolovirus[Human_herpesvirus_6B	12	30	48	151	0.558660
Herpesviridae_Cytomegalovirus[Human_cytomegalovirus	44	126	16	55	0.627400
Adenoviridae_Mastadenovirus[Human_mastadenovirus_A	2	4	58	177	0.640550
Flaviviridae_Hepacivirus[Hepacivirus_C	3	6	57	175	0.694120
Picornaviridae_Enterovirus[Enterovirus_D	3	7	57	174	0.713240
Herpesviridae_Lymphocryptovirus[Epstein-Barr_virus	52	154	8	27	0.835730
Herpesviridae_Rhadinovirus[Human_herpesvirus_8	12	33	48	148	0.848520
Herpesviridae_Varicellovirus[Human_herpesvirus_3	36	105	24	76	0.880050
Adenoviridae_Mastadenovirus[Human_mastadenovirus_D	1	3	59	178	1.000000
Orthomyxoviridae_Betainfluenzavirus[Influenza_B_virus	1	2	59	179	1.000000

fisher_exact_tests/Vietnam_Cirrohosis_vs_cohort_seropositive_fisher_exact_tests.csv

Vietnam HCC vs cohort seropositive

Family Genus Species	Seropositive Hits HCC	Seropositive Hits Cohort	Seropositive No Hits HCC	Seropositive No Hits Cohort	p value
Herpesviridae_Simplexvirus Human_herpesvirus_1	81	66	40	54	0.065072
Herpesviridae_Varicellovirus Human_herpesvirus_3	77	64	44	56	0.117550
Herpesviridae_Lymphocryptovirus Epstein-Barr_virus	107	99	14	21	0.205620
Hepadnaviridae_Orthohepadnavirus Hepatitis_B_virus	5	1	116	119	0.212890
NA_Deltavirus Hepatitis_delta_virus	5	1	116	119	0.212890
Polyomaviridae_Betapolyomavirus Human_polyomavirus_2	3	0	118	120	0.246890
Adenoviridae_Mastadenovirus Human_mastadenovirus_C	19	14	102	106	0.454120
Herpesviridae_Cytomegalovirus Human_cytomegalovirus	87	83	34	37	0.673310
Herpesviridae_Rhadinovirus Human_herpesvirus_8	24	21	97	99	0.741350
Picornaviridae_Enterovirus Enterovirus_D	6	4	115	116	0.748670
Adenoviridae_Mastadenovirus Human_mastadenovirus_G	1	0	120	120	1.000000
Astroviridae_Mamastrovirus Mamastrovirus_1	1	0	120	120	1.000000
Coronaviridae_Betacoronavirus Severe_acute_respiratory_syndrome-related_coronavirus	1	0	120	120	1.000000
Flaviviridae_Hepacivirus Hepacivirus_C	5	4	116	116	1.000000
Orthomyxoviridae_Alphainfluenzavirus Influenza_A_virus	2	1	119	119	1.000000
Orthomyxoviridae_Betainfluenzavirus Influenza_B_virus	2	1	119	119	1.000000
Paramyxoviridae_Respirovirus Human_respirovirus_1	1	0	120	120	1.000000

fisher_exact_tests/Vietnam_HCC_vs_cohort_seropositive_fisher_exact_tests.csv

Vietnam Hepatitis B vs cohort seropositive

Family	Genus	Species	Seropositive Hits Hepatitis B	Seropositive Hits Cohort	Seropositive No Hits Hepatitis B	Seropositive No Hits Cohort	p value
Coronaviridae	Betacoronavirus	Human_coronavirus_HKU1	4	0	56	181	0.0035572
Paramyxoviridae	Rubulavirus	Human_rubulavirus_4	3	1	57	180	0.0487390
Caliciviridae	Norovirus	Norwalk_virus	1	0	59	181	0.2489600
Coronaviridae	Alphacoronavirus	Human_coronavirus_229E	1	0	59	181	0.2489600
Herpesviridae	Simplexvirus	Human_herpesvirus_2	1	0	59	181	0.2489600
Orthomyxoviridae	Gammmainfluenzavirus	Influenza_C_virus	1	0	59	181	0.2489600
Pneumoviridae	Metapneumovirus	Human_metapneumovirus	1	0	59	181	0.2489600
Poxviridae	Parapoxvirus	Orf_virus	1	0	59	181	0.2489600
Adenoviridae	Mastadenovirus	Human_mastadenovirus_B	2	3	58	178	0.6006100
Adenoviridae	Mastadenovirus	Human_mastadenovirus_A	2	4	58	177	0.6405500
Herpesviridae	Roseolovirus	Human_herpesvirus_6B	11	31	49	150	0.8454200
Adenoviridae	Mastadenovirus	Human_mastadenovirus_D	1	3	59	178	1.0000000
Orthomyxoviridae	Alphainfluenzavirus	Influenza_A_virus	1	2	59	179	1.0000000

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_seropositive_fisher_exact_tests.csv

13 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
Alphapapillomavirus_6	Q80958	13	8	47	173	0.0001765
Epstein-Barr_virus	A0A0U3AZS7	40	70	20	111	0.0001817
Influenza_B_virus	P11102	39	69	21	112	0.0003247
Wesselsbron_virus	YP_002922020	10	5	50	176	0.0005000
Human_respiratory_syncytial_virus	Q3MK18	44	88	16	93	0.0009431
Human_cytomegalovirus	A0A0G2TJX8	7	2	53	179	0.0010418
Human_mastadenovirus_B	AP_000553	31	51	29	130	0.0015097
Human_respiratory_syncytial_virus	C1K8X5	49	108	11	73	0.0017529
Human_mastadenovirus_A	P36717	34	61	26	120	0.0022004
Human_respiratory_syncytial_virus	Q38182	46	99	14	82	0.0024274
Influenza_B_virus	NP_056661	36	67	24	114	0.0024778
Human_respiratory_syncytial_virus	Q3MK39	44	92	16	89	0.0025815
Human_respiratory_syncytial_virus	I1V4V3	43	89	17	92	0.0027000
Venezuelan_equine_encephalitis_virus	P36330	7	3	53	178	0.0027850
SARS_coronavirus	NP_828851	13	13	47	168	0.0033817
Human_mastadenovirus_F	P19416	13	13	47	168	0.0033817
Enterovirus_D	E7FM39	13	13	47	168	0.0033817
Gammapapillomavirus_15	YP_006393284	6	2	54	179	0.0035382
Human_immunodeficiency_virus_1	AF119819_GAG	6	2	54	179	0.0035382
Human_immunodeficiency_virus_1	HQ385479_GAG	6	2	54	179	0.0035382
Human_mastadenovirus_B	YP_002213818	4	0	56	181	0.0035572
Human_immunodeficiency_virus_1	AF286239_NEF	4	0	56	181	0.0035572
Human_immunodeficiency_virus_1	DQ085872_NEF	4	0	56	181	0.0035572
Human_mastadenovirus_D	B9A5C5	4	0	56	181	0.0035572
Epstein-Barr_virus	E2E7U4	34	64	26	117	0.0041155
Rhinovirus_B	G5D5D9	59	154	1	27	0.0042156
Human_respiratory_syncytial_virus	W8CJB3	42	88	18	93	0.0044816
Influenza_A_virus	A4U7B0	9	7	51	174	0.0055302
Human_respiratory_syncytial_virus	R9TD96	28	49	32	132	0.0064827
Human_respiratory_syncytial_virus	B0LS42	43	93	17	88	0.0067962
Human_respiratory_syncytial_virus	C1ILP1	21	32	39	149	0.0069459
Human_respiratory_syncytial_virus	O91941	38	78	22	103	0.0073941
Influenza_A_virus	A4U6V2	8	6	52	175	0.0081105
Human_mastadenovirus_B	Q2KRX1	11	11	49	170	0.0081184
Human_respiratory_syncytial_virus	X4YAY5	19	28	41	153	0.0084585
Human_immunodeficiency_virus_1	AB286851_GAG	6	3	54	178	0.0085311
Human_mastadenovirus_G	YP_213980	32	61	28	120	0.0090822
Human_immunodeficiency_virus_1	AY008715_TAT	15	19	45	162	0.0091542
Human_immunodeficiency_virus_1	AY588970_TAT	15	19	45	162	0.0091542
Human_enterovirus	YP_009259673	20	30	40	151	0.0093852
Human_herpesvirus_1	A0A0F7GT05	17	24	43	157	0.0099402
Human_respiratory_syncytial_virus	A0A0A0VAE3	36	73	24	108	0.0106880
Human_respiratory_syncytial_virus	Q50G60	41	88	19	93	0.0107920
Human_respiratory_syncytial_virus	E3W9K0	28	51	32	130	0.0109310
Human_respiratory_syncytial_virus	G8EIT1	38	79	22	102	0.0109970
Human_respiratory_syncytial_virus	A0A068B4W3	18	26	42	155	0.0112550
Human_respiratory_syncytial_virus	E3WHY5	29	53	31	128	0.0114090
Tanapox_virus	NP_073464	5	2	55	179	0.0114700
Human_immunodeficiency_virus_1	FN392874_GAG	5	2	55	179	0.0114700
Human_immunodeficiency_virus_1	EU697904_GAG	5	2	55	179	0.0114700
Human_immunodeficiency_virus_2	AB731742_ENV	5	2	55	179	0.0114700
Hepatitis_B_virus	Q69602	5	2	55	179	0.0114700
Epstein-Barr_virus	A0A052YR54	5	2	55	179	0.0114700
Human_immunodeficiency_virus_1	EU697908_TAT	13	16	47	165	0.0118240
Human_mastadenovirus_C	AP_000511	48	113	12	68	0.0118770
Human_respiratory_syncytial_virus	A0A0B5HKD1	29	54	31	127	0.0119290
Human_respiratory_syncytial_virus	W5RNM1	30	56	30	125	0.0123860
Human_respiratory_syncytial_virus	B7U384	30	56	30	125	0.0123860
Human_respiratory_syncytial_virus	B2Z2Y7	10	10	50	171	0.0126180
Human_respiratory_syncytial_virus	Q59616	21	34	39	147	0.0128200
Human_respiratory_syncytial_virus	Q80R93	25	44	35	137	0.0132610
Epstein-Barr_virus	A0A0U3U7S7	53	131	7	50	0.0134910
Epstein-Barr_virus	R4QWP9	53	132	7	49	0.0137530
Mupapillomavirus_1	NP_597739	8	7	52	174	0.0139270
Human_immunodeficiency_virus_1	EF091932_TAT	14	18	46	163	0.0142190
Influenza_B_virus	P12442	12	14	48	167	0.0143830
Gammapapillomavirus_11	YP_004934015	4	1	56	180	0.0144240
Rotavirus_B	Q86108	4	1	56	180	0.0144240
Epstein-Barr_virus	G3CKS1	4	1	56	180	0.0144240
Human_mastadenovirus_A	T1UEZ7	4	1	56	180	0.0144240
Epstein-Barr_virus	A0A0X8W7I3	4	1	56	180	0.0144240
Epstein-Barr_virus	YP_401659	3	0	57	181	0.0148530
Gammapapillomavirus_11	YP_006393305	3	0	57	181	0.0148530
Rotavirus_A	Q77Q91	3	0	57	181	0.0148530
Marburg_marburgvirus	Q1PD54	3	0	57	181	0.0148530
Human_cytomegalovirus	A0A0G2UCG9	3	0	57	181	0.0148530
Human_cytomegalovirus	A8T767	3	0	57	181	0.0148530
Human_mastadenovirus_C	Q64829	3	0	57	181	0.0148530
BK_polyomavirus	D2KKI4	3	0	57	181	0.0148530
Hepatitis_B_virus	U5JTY1	3	0	57	181	0.0148530
Human_herpesvirus_2	A0A109R0V4	3	0	57	181	0.0148530
Rhinovirus_A	P04936	34	69	26	112	0.0157070
Human_respiratory_syncytial_virus	A0A075BF85	34	70	26	111	0.0165540
Human_respiratory_syncytial_virus	R4WVH0	28	52	32	129	0.0170860
Human_herpesvirus_2	YP_009137175	6	4	54	177	0.0171500
Alphapapillomavirus_6	P36798	6	4	54	177	0.0171500
Human_herpesvirus_1	F8RGE9	6	4	54	177	0.0171500
Torque_teno_mini_virus_9	NP_054655	10	11	50	170	0.0173120
Epstein-Barr_virus	E2E7X3	23	40	37	141	0.0174180
Human_respiratory_syncytial_virus	Q9IH88	27	51	33	130	0.0177570
Human_respiratory_syncytial_virus	A0A0B5HI44	28	53	32	128	0.0177710
Human_respirovirus_3	P06162	12	15	48	166	0.0180000
Human_herpesvirus_1	I3TC88	12	15	48	166	0.0180000
Human_respiratory_syncytial_virus	A0A068B4Y1	24	42	36	139	0.0186120
Enterovirus_A	NP_042242	24	43	36	138	0.0196950
Human_immunodeficiency_virus_1	AY173951_NEF	9	9	51	172	0.0196980
Human_mastadenovirus_A	T1UEY7	9	9	51	172	0.0196980
Cowpox_virus	O72759	15	21	45	160	0.0199270
Human_respiratory_syncytial_virus	H2EIR7	30	59	30	122	0.0203360
Human_immunodeficiency_virus_1	U42720_POL	7	6	53	175	0.0205680

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
Sindbis_virus	O93006	53	22	68	98	0.0000253
Human_respirovirus_1	P36354	32	14	89	106	0.0049711
Hepatitis_delta_virus	Q760F6	17	4	104	116	0.0049988
Human_immunodeficiency_virus_1	AJ302647_GAG	8	0	113	120	0.0069331
Hepatitis_B_virus	E3Q0K8	8	0	113	120	0.0069331
Hepatitis_B_virus	K7QG62	12	2	109	118	0.0105660
Human_respiratory_syncytial_virus	B2Z2N1	7	0	114	120	0.0142920
Hepatitis_B_virus	U3MTJ7	15	4	106	116	0.0148470
Human_polyomavirus_2	Q8JV81	9	1	112	119	0.0189960
Human_respiratory_syncytial_virus	A0A173DTM0	11	2	110	118	0.0190950
Human_cytomegalovirus	A0A0A7CGS9	14	4	107	116	0.0249950
Human_respiratory_syncytial_virus	A0A0K1H410	14	4	107	116	0.0249950
Human_herpesvirus_3	NP_040161	24	11	97	109	0.0272250
Hepatitis_B_virus	Q4R1S2	6	0	115	120	0.0293290
Hepatitis_B_virus	A0A142MJR5	12	3	109	117	0.0297690
Hepatitis_B_virus	A0A142MK62	12	3	109	117	0.0297690
Human_respiratory_syncytial_virus	S5M3T2	52	35	69	85	0.0317960
Human_cytomegalovirus	YP_081532	10	2	111	118	0.0339850
Epstein-Barr_virus	A0A0C7SX58	10	2	111	118	0.0339850
Human_respiratory_syncytial_virus	A0A0U4AV18	10	2	111	118	0.0339850
Human_cytomegalovirus	C8BKZ5	10	2	111	118	0.0339850
Epstein-Barr_virus	A0A0C7TX23	55	38	66	82	0.0342150
Human_respiratory_syncytial_virus	A0A0A7NQZ4	22	10	99	110	0.0355820
Tanapox_virus	NP_073456	8	1	113	119	0.0356130
Hepatitis_delta_virus	L0CV01	8	1	113	119	0.0356130
Human_cytomegalovirus	Q91D82	45	29	76	91	0.0359940
Human_herpesvirus_1	G8H888	38	23	83	97	0.0376960
Hepatitis_B_virus	D2X485	13	4	108	116	0.0413640
Human_respiratory_syncytial_virus	I1V4U9	41	26	80	94	0.0438480
Human_respiratory_syncytial_virus	Q6ECJ5	92	77	29	43	0.0495120
Hepatitis_B_virus	U3MSN1	11	3	110	117	0.0503700
Influenza_A_virus	Q67352	25	13	96	107	0.0509210
Human_respiratory_syncytial_virus	U6BZQ0	108	96	13	24	0.0509210
Human_respiratory_syncytial_virus	X2C1V9	61	45	60	75	0.0517230
Human_herpesvirus_3	Q0Z9V5	21	10	100	110	0.0530710
Hepatitis_delta_virus	A0A0N6Y551	14	5	107	115	0.0534070
Human_cytomegalovirus	D2W9P1	49	34	72	86	0.0576630
Human_respiratory_syncytial_virus	NP_056858	5	0	116	120	0.0599230
Human_herpesvirus_6B	NP_050276	5	0	116	120	0.0599230
Hepatitis_delta_virus	P0C6M6	5	0	116	120	0.0599230
Influenza_A_virus	Q6XSW4	5	0	116	120	0.0599230
Human_herpesvirus_8	A0A024A4F8	5	0	116	120	0.0599230
Hepatitis_B_virus	A0A142MJH0	5	0	116	120	0.0599230
Human_cytomegalovirus	A0A140GX97	5	0	116	120	0.0599230
Hepatitis_B_virus	I3VPK9	5	0	116	120	0.0599230
Human_mastadenovirus_A	D3JI54	5	0	116	120	0.0599230
Human_cytomegalovirus	Q6SWZ1	40	26	81	94	0.0601460
Human_respiratory_syncytial_virus	E9NW86	52	37	69	83	0.0617030
Human_respiratory_syncytial_virus	K4G8J7	100	87	21	33	0.0650660
Human_respirovirus_1	P26590	15	6	106	114	0.0655750
Yaba_monkey_tumor_virus	NP_938294	7	1	114	119	0.0658030
Hepatitis_delta_virus	V5TAS2	7	1	114	119	0.0658030
Hepatitis_delta_virus	E9KP24	7	1	114	119	0.0658030
Human_cytomegalovirus	V9PSV6	7	1	114	119	0.0658030
Hepatitis_B_virus	B6D725	7	1	114	119	0.0658030
Human_respiratory_syncytial_virus	Q9IUC5	7	1	114	119	0.0658030
Human_herpesvirus_1	A0A1C3J7A4	7	1	114	119	0.0658030
Hepatitis_delta_virus	A1VP99	12	4	109	116	0.0672040
Human_rubulavirus_4	C4NTR1	19	9	102	111	0.0690000
Gammapapillomavirus_7	YP_004169296	29	17	92	103	0.0707530
Human_mastadenovirus_C	F2YRV5	29	17	92	103	0.0707530
Epstein-Barr_virus	A0A0U3THE3	46	32	75	88	0.0734240
Epstein-Barr_virus	Q07286	30	18	91	102	0.0754690
Cowpox_virus	A0A0K2YUL6	16	7	105	113	0.0776480
Human_cytomegalovirus	Q6RXH9	10	3	111	117	0.0836160
Hepatitis_delta_virus	A0A0N6YRG4	10	3	111	117	0.0836160
Human_parainfluenza_2_virus	P23056	13	5	108	115	0.0839250
Human_herpesvirus_3	G9IX77	26	15	95	105	0.0857370
Human_respiratory_syncytial_virus	E9NW32	81	67	40	53	0.0861720
Human_herpesvirus_1	G8H8J0	78	64	43	56	0.0894470
Human_respiratory_syncytial_virus	A0A0B5HHB9	59	45	62	75	0.0911740
Human_respiratory_syncytial_virus	X2C1W3	59	45	62	75	0.0911740
Human_respiratory_syncytial_virus	F8U5U7	28	17	93	103	0.0977710
Epstein-Barr_virus	K9UT27	14	6	107	114	0.1002100
Tanapox_virus	NP_073481	8	2	113	118	0.1020800
Human_mastadenovirus_C	AP_000500	8	2	113	118	0.1020800
Human_polyomavirus_2	Q91NN6	8	2	113	118	0.1020800
Epstein-Barr_virus	A0A0B6VQ04	8	2	113	118	0.1020800
Hepatitis_B_virus	B5U704	8	2	113	118	0.1020800
Human_respiratory_syncytial_virus	W5RPX3	8	2	113	118	0.1020800
Epstein-Barr_virus	A0A0C7TYV8	19	10	102	110	0.1121200
Human_respiratory_syncytial_virus	U3LVQ0	80	67	41	53	0.1139600
Orf_virus	FIAXJ9	15	7	106	113	0.1158800
Epstein-Barr_virus	A0A075FFC3	78	65	43	55	0.1164600
Torque_teno_mid1_virus_1	YP_001109584	25	15	96	105	0.1185800
Human_betaherpesvirus_6A	Q77Z83	6	1	115	119	0.1195500
Human_herpesvirus_1	A0A0F7CWU8	6	1	115	119	0.1195500
Human_respiratory_syncytial_virus	A0A0A0RBN0	6	1	115	119	0.1195500
Hepatitis_B_virus	A0A0G3VT08	6	1	115	119	0.1195500
Hepatitis_delta_virus	Q5Y989	6	1	115	119	0.1195500
Epstein-Barr_virus	Q90688	6	1	115	119	0.1195500
Hepatitis_B_virus	U5XKD1	6	1	115	119	0.1195500
Human_respiratory_syncytial_virus	X2C1R4	6	1	115	119	0.1195500
Human_respiratory_syncytial_virus	A9UF88	6	1	115	119	0.1195500
Hepatitis_delta_virus	E2J7V7	6	1	115	119	0.1195500
Hepatitis_B_virus	A8UNY0	6	1	115	119	0.1195500
Human_respiratory_syncytial_virus	G3K4S1	6	1	115	119	0.1195500
Human_respiratory_syncytial_virus	G9M581	6	1	115	119	0.1195500
Influenza_A_virus	YP_308852	4	0	117	120	0.1218900
Orf_virus	NP_957889	4	0	117	120	0.1218900

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
Eastern_equine_encephalitis_virus	YP_006491233	28	5	32	176	0.0000e+00
Influenza_A_virus	P03448	25	0	35	181	0.0000e+00
Hepatitis_B_virus	D6QW13	28	0	32	181	0.0000e+00
Human_cytomegalovirus	A0A0G2TP82	28	1	32	180	0.0000e+00
West_Nile_virus	A0A076E7A3	28	4	32	177	0.0000e+00
Human_associated_cyclovirus_8	YP_009047065	21	0	39	181	1.0000e-14
Variola_virus	NP_042071	22	1	38	180	3.0000e-14
Human_immunodeficiency_virus_1	FN392876_TAT	19	0	41	181	2.8000e-13
West_Nile_virus	A0A067ZS69	21	3	39	178	1.1000e-11
Human_cytomegalovirus	P16847	16	0	44	181	4.0210e-11
Human_mastadenovirus_B	YP_002213767	19	2	41	179	4.0560e-11
Human_immunodeficiency_virus_1	EU697906_GAG	18	2	42	179	1.9343e-10
Human_herpesvirus_2	A0A0K0KN18	15	0	45	181	2.0195e-10
Hepatitis_delta_virus	Q70E10	15	0	45	181	2.0195e-10
Human_immunodeficiency_virus_2	JN863898_ENV	14	0	46	181	9.9656e-10
Human_herpesvirus_8	G3KH28	14	0	46	181	9.9656e-10
Hepatitis_B_virus	I0DG61	14	0	46	181	9.9656e-10
Human_herpesvirus_2	A0A0Y0QNP6	14	0	46	181	9.9656e-10
Hepatitis_B_virus	U5JQH2	14	0	46	181	9.9656e-10
Hepatitis_B_virus	A0A1B1ABA1	15	1	45	180	2.6280e-09
Human_herpesvirus_1	A0A110AV91	19	5	41	176	4.5691e-09
Epstein-Barr_virus	YP_401693	11	0	49	181	1.0826e-07
Influenza_A_virus	A4GXG3	11	0	49	181	1.0826e-07
Cosavirus_E	YP_002956072	12	1	48	180	2.4182e-07
Human_coronavirus_229E	NP_073556	16	5	44	176	2.9369e-07
Betapapillomavirus_1	P50770	13	2	47	179	3.3166e-07
Human_immunodeficiency_virus_1	EU581825_VIF	10	0	50	181	5.0016e-07
Hepatitis_B_virus	C1K1Z4	10	0	50	181	5.0016e-07
Aichi_virus	Q8B6B2	10	0	50	181	5.0016e-07
Human_cytomegalovirus	Q6RXD9	10	0	50	181	5.0016e-07
Human_herpesvirus_8	YP_001129378	12	2	48	179	1.3644e-06
Human_betaherpesvirus_6A	P24443	9	0	51	181	2.2752e-06
Zika_virus	A0A166XWL4	9	0	51	181	2.2752e-06
Zaire_ebolavirus	NP_066244	8	0	52	181	1.0195e-05
Human_immunodeficiency_virus_1	AY900572_VPR	8	0	52	181	1.0195e-05
Human_immunodeficiency_virus_1	GQ372987_VIF	8	0	52	181	1.0195e-05
Human_herpesvirus_1	A0A0X8EA33	8	0	52	181	1.0195e-05
Alphapapillomavirus_11	Q82007	12	4	48	177	1.7696e-05
Influenza_A_virus	P03476	7	0	53	181	4.5011e-05
Cowpox_virus	G0XW74	7	0	53	181	4.5011e-05
Human_immunodeficiency_virus_2	KC693505_POL	10	3	50	178	7.403e-05
Hepacivirus_C	YP_001469630	9	2	51	179	8.0494e-05
Rio_Bravo_virus	NP_620044	9	2	51	179	8.0494e-05
Human_mastadenovirus_C	E7DY96	9	2	51	179	8.0494e-05
Human_immunodeficiency_virus_1	AF457051_NEF	12	6	48	175	0.00011678
Human_immunodeficiency_virus_1	EU735355_ENV	6	0	54	181	0.00019588
Human_immunodeficiency_virus_1	AY173951_VPU	6	0	54	181	0.00019588
Influenza_A_virus	Q64968	6	0	54	181	0.00019588
Epstein-Barr_virus	P03235	6	0	54	181	0.00019588
Cowpox_virus	U5TB51	6	0	54	181	0.00019588
Cowpox_virus	G0XW68	6	0	54	181	0.00019588
Mayaro_virus	NP_579968	14	10	46	171	0.00023279
Alphapapillomavirus_8	P36727	9	3	51	178	0.00025811
Hepatitis_B_virus	F2X840	9	3	51	178	0.00025811
Human_immunodeficiency_virus_1	U88826_POL	7	1	53	180	0.00028872
Human_immunodeficiency_virus_1	AY423387_POL	8	2	52	179	0.00029475
Human_mastadenovirus_D	YP_003038612	16	14	44	167	0.00041252
Hepatitis_B_virus	A0A075BZB0	9	4	51	177	0.0006723
Colorado_tick_fever_virus	Q9ENK8	5	0	55	181	0.00084051
Alphapapillomavirus_6	Q02515	5	0	55	181	0.00084051
Human_cytomegalovirus	A0A0G2TC35	5	0	55	181	0.00084051
Hepatitis_B_virus	F5COM6	5	0	55	181	0.00084051
Hepatitis_B_virus	X2FVR0	5	0	55	181	0.00084051
Rosavirus_A	YP_009028557	8	3	52	178	0.00086611
Human_immunodeficiency_virus_2	U05357_ENV	8	3	52	178	0.00086611
Human_immunodeficiency_virus_2	U27200_GAG	7	2	53	179	0.0010418
Human_immunodeficiency_virus_1	AF289550_VPU	6	1	54	180	0.0011011
Human_cytomegalovirus	A0A126NBR7	6	1	54	180	0.0011011
Human_cytomegalovirus	YP_081561	30	49	30	132	0.0014621
Mamastrovirus_1	YP_009275018	16	17	44	164	0.001834
Human_cytomegalovirus	A0A0G2TD75	8	4	52	177	0.0020821
HMO_Astrovirus_A	YP_003275951	11	9	49	172	0.0025267
Human_herpesvirus_6B	NP_050212	7	3	53	178	0.002785
Rotavirus_A	Q9E780	7	3	53	178	0.002785
Human_immunodeficiency_virus_2	L07625_POL	27	43	33	138	0.0028826
Human_immunodeficiency_virus_1	L20571_ENV	6	2	54	179	0.0035382
Alphapapillomavirus_4	P22162	6	2	54	179	0.0035382
Human_herpesvirus_2	A0A109R110	6	2	54	179	0.0035382
Bunyamwera_orthobunyavirus	A0A0U2GK49	6	2	54	179	0.0035382
Jugra_virus	YP_009344969	4	0	56	181	0.0035572
Epstein-Barr_virus	A0A0S2YRL2	4	0	56	181	0.0035572
Human_immunodeficiency_virus_1	AF492624_GAG	4	0	56	181	0.0035572
Alphapapillomavirus_7	P24830	4	0	56	181	0.0035572
Influenza_A_virus	Q92929	4	0	56	181	0.0035572
Human_cytomegalovirus	Q52MJ2	4	0	56	181	0.0035572
Epstein-Barr_virus	E2ESI7	4	0	56	181	0.0035572
Epstein-Barr_virus	A0A0C7T5F6	4	0	56	181	0.0035572
Epstein-Barr_virus	Q9WGA9	4	0	56	181	0.0035572
Human_herpesvirus_2	A0A109R057	4	0	56	181	0.0035572
Human_betaherpesvirus_7	Q89530	4	0	56	181	0.0035572
Human_herpesvirus_6B	NP_050255	5	1	55	180	0.0040636
Orf_virus	F1AXE0	5	1	55	180	0.0040636
Human_immunodeficiency_virus_1	DQ845387_POL	9	7	51	174	0.0055302
Human_cytomegalovirus	A0A0G2U480	11	10	49	171	0.0060485
Gammapapillomavirus_7	YP_004169293	7	4	53	177	0.0061431
Human_rubulavirus_4	A0A097F113	14	17	46	164	0.0077441
Enterovirus_D	NP_040760	29	53	31	128	0.011409
Human_mastadenovirus_D	YP_003038629	5	2	55	179	0.01147
Human_herpesvirus_2	A0A0Y0QNI6	5	2	55	179	0.01147
Hepatitis_delta_virus	Q217C3	5	2	55	179	0.01147

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_protein_fisher_exact_tests.csv

14 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.

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
Human_herpesvirus_1	Q09I76	8	2	52	179	0.0002947
Human_respirovirus_3	NP_599250	5	2	55	179	0.0114700
Human_respiratory_syncytial_virus	R9TC20	20	34	40	147	0.0309140
Rhinovirus_A	P04936	4	2	56	179	0.0351440
Rhinovirus_B	G5D5D9	18	30	42	151	0.0387960
Severe_acute_respiratory_syndrome-related_coronavirus	Q0Q468	3	1	57	180	0.0487390
Human_respirovirus_3	P06162	3	1	57	180	0.0487390
Rhinovirus_A	Q82122	17	29	43	152	0.0564660
Human_respiratory_syncytial_virus	A0A1C9V2Q5	2	0	58	181	0.0612030
Human_herpesvirus_1	A0A181ZFG9	2	0	58	181	0.0612030
Epstein-Barr_virus	A0A0C7TX23	2	0	58	181	0.0612030
Rhinovirus_A	P23008	7	8	53	173	0.0614590
Rhinovirus_B	Q82081	21	41	39	140	0.0629690
Enterovirus_C	P22055	4	3	56	178	0.0667090
Human_respiratory_syncytial_virus	Q0QX19	24	49	36	132	0.0742460
Human_respiratory_syncytial_virus	Q38I82	9	13	51	168	0.0755300
Human_respiratory_syncytial_virus	Q0QXF9	9	13	51	168	0.0755300
Human_herpesvirus_1	P10220	7	9	53	172	0.0796650
Epstein-Barr_virus	YP_001129441	8	11	52	170	0.0946040
Human_respiratory_syncytial_virus	Q4KV07	23	48	37	133	0.1019600
Rhinovirus_C	YP_001552411	4	4	56	177	0.1087300
Human_herpesvirus_3	Q0Q8T5	43	108	17	73	0.1232800
Enterovirus_B	Q66790	9	14	51	167	0.1260800
Epstein-Barr_virus	A0A0U3TY20	16	31	44	150	0.1319300
Enterovirus_C	Q7T7P4	7	10	53	171	0.1424800
Enterovirus_C	P36290	5	6	55	175	0.1475000
Human_respiratory_syncytial_virus	Q50G60	5	6	55	175	0.1475000
Rhinovirus_B	NP_041009	2	1	58	180	0.1539000
Human_herpesvirus_1	G8H888	2	1	58	180	0.1539000
Rhinovirus_A	NP_042288	2	3	57	178	0.1652800
Human_respiratory_syncytial_virus	A0A059TE74	21	47	39	134	0.1884200
Epstein-Barr_virus	A0A0C7TXA7	12	23	48	158	0.2036300
Epstein-Barr_virus	A0A075FFC3	12	23	48	158	0.2036300
Human_respiratory_syncytial_virus	A0A075LY43	9	16	51	165	0.2202100
Human_respiratory_syncytial_virus	U3LLK0	4	5	56	176	0.2320500
Human_cytomegalovirus	A0A0R5ECB9	9	17	51	164	0.2348600
Enterovirus_C	C6EN06	6	10	54	171	0.2381200
Venezuelan_equine_encephalitis_virus	NP_040822	1	0	59	181	0.2489600
Dugbe_orthonairovirus	NP_690576	1	0	59	181	0.2489600
Human_betaherpesvirus_7	YP_073752	1	0	59	181	0.2489600
Epstein-Barr_virus	YP_401657	1	0	59	181	0.2489600
Hepacivirus_C	YP_001469633	1	0	59	181	0.2489600
Human_cytomegalovirus	P06434	1	0	59	181	0.2489600
Enterovirus_A	Q66479	1	0	59	181	0.2489600
Human_betaherpesvirus_6A	Q06093	1	0	59	181	0.2489600
Human_respirovirus_1	A0A1B2CW80	1	0	59	181	0.2489600
Human_herpesvirus_1	A0A089X4K7	1	0	59	181	0.2489600
Epstein-Barr_virus	A0A0C7TMS7	1	0	59	181	0.2489600
Human_respiratory_syncytial_virus	A0A0A0RD50	1	0	59	181	0.2489600
Enterovirus_B	S6DP08	1	0	59	181	0.2489600
Human_respiratory_syncytial_virus	Q6ECI9	1	0	59	181	0.2489600
Human_herpesvirus_1	A0A1C3J7Q4	1	0	59	181	0.2489600
Human_parainfluenza_2_virus	A0A1V0E1U9	1	0	59	181	0.2489600
Epstein-Barr_virus	B5LU11	1	0	59	181	0.2489600
Human_herpesvirus_8	Q9DUM3	1	0	59	181	0.2489600
Human_cytomegalovirus	A0A0G2TD98	1	0	59	181	0.2489600
Enterovirus_C	NP_041277	2	2	58	179	0.2590700
Enterovirus_B	P29813	2	2	58	179	0.2590700
Severe_acute_respiratory_syndrome-related_coronavirus	Q3I5J5	2	2	58	179	0.2590700
Respiratory_syncytial_virus	Q18NW7	2	2	58	179	0.2590700
Human_cytomegalovirus	A0A0G2TXR2	2	2	58	179	0.2590700
Enterovirus_B	Q9YLG5	4	6	56	175	0.2714800
Human_respiratory_syncytial_virus	A0A075BF70	4	6	56	175	0.2714800
Human_respiratory_syncytial_virus	A0A075BF97	11	22	49	159	0.2777700
Human_respiratory_syncytial_virus	A0A0F7JHI5	13	29	47	152	0.3298900
Human_respiratory_syncytial_virus	Q4KJU7	22	54	38	127	0.3391800
Human_herpesvirus_6B	NP_050192	5	9	55	172	0.3461900
Human_respiratory_syncytial_virus	E9NWA9	3	4	57	177	0.3700700
Epstein-Barr_virus	A0A0A7DK99	11	24	49	157	0.3973700
Human_herpesvirus_6B	NP_050266	3	5	57	176	0.4148100
Enterovirus_D	Q68T42	7	14	53	167	0.4271900
Human_enterovirus	YP_009259673	1	1	59	180	0.4367200
Influenza_B_virus	P12442	1	1	59	180	0.4367200
Severe_acute_respiratory_syndrome-related_coronavirus	P0C6W2	1	1	59	180	0.4367200
Epstein-Barr_virus	H9C5V9	1	1	59	180	0.4367200
Human_respiratory_syncytial_virus	A7YLT5	1	1	59	180	0.4367200
Human_herpesvirus_1	A0A0X8EA26	1	1	59	180	0.4367200
Human_cytomegalovirus	Q91D82	1	1	59	180	0.4367200
Human_herpesvirus_1	A0A0X8EA29	1	1	59	180	0.4367200
Epstein-Barr_virus	A0A052YQE3	1	1	59	180	0.4367200
Enterovirus_B	Q05ED4	1	1	59	180	0.4367200
Epstein-Barr_virus	H2E7N5	1	1	59	180	0.4367200
Hepacivirus_C	A7XB67	1	1	59	180	0.4367200
Human_respiratory_syncytial_virus	B0LS43	7	15	53	166	0.4424600
Human_cytomegalovirus	YP_081561	4	7	56	174	0.4735100
Enterovirus_B	Q66474	4	7	56	174	0.4735100
Human_respiratory_syncytial_virus	F8U5T5	4	8	56	173	0.4996800
Rhinovirus_A	B9V447	5	10	55	171	0.5366200
Human_respiratory_syncytial_virus	A0A0A0VAE3	6	14	54	167	0.5930200
Human_herpesvirus_8	YP_001129431	2	3	58	178	0.6006100
Hepacivirus_C	O92530	2	3	58	178	0.6006100
Human_herpesvirus_1	A0A0X8E9W2	2	3	58	178	0.6006100
Hepacivirus_C	O92531	2	4	58	177	0.6405500
Human_herpesvirus_8	A0A0N9SCV5	2	4	58	177	0.6405500
Human_herpesvirus_8	Q76SB0	2	4	58	177	0.6405500
Human_herpesvirus_8	A0A0N9RMS1	9	23	51	158	0.6631900
Hepacivirus_C	Q68801	3	6	57	175	0.6941200
Human_respiratory_syncytial_virus	A0A075BF87	3	6	57	175	0.6941200
Epstein-Barr_virus	A0A125T0X3	3	7	57	174	0.7132400
Human_respiratory_syncytial_virus	A0A068B4C9	4	9	56	172	0.7414700

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
Epstein-Barr_virus	YP_001129463	12	2	109	118	0.010566
Epstein-Barr_virus	A0A0F6QKC9	21	10	100	110	0.053071
Epstein-Barr_virus	A0A0U3TY67	20	11	101	109	0.122820
Enterovirus_B	Q66575	9	3	112	117	0.135910
Enterovirus_B	K7DW09	14	7	107	113	0.169480
Human_respiratory_syncytial_virus	E9NW64	3	0	118	120	0.246890
Epstein-Barr_virus	A0A0U3U757	3	0	118	120	0.246890
Human_herpesvirus_3	Q0Q8T5	80	71	41	49	0.288370
Epstein-Barr_virus	Q3KST1	7	3	114	117	0.333370
Hepacivirus_C	YP_001469632	4	1	117	119	0.369780
Human_respiratory_syncytial_virus	A0A0B5HBX9	4	1	117	119	0.369780
Human_respiratory_syncytial_virus	A0A0A7NRR5	22	16	99	104	0.377210
Hepacivirus_C	YP_001469634	5	2	116	118	0.446230
Epstein-Barr_virus	A0A0C7TXA7	20	15	101	105	0.465150
Influenza_A_virus	YP_308871	2	0	119	120	0.497930
Hepacivirus_C	Q00269	2	0	119	120	0.497930
Human_respiratory_syncytial_virus	Q8V218	2	0	119	120	0.497930
Human_cytomegalovirus	J7EHC9	2	0	119	120	0.497930
Human_coronavirus_HKU1	S5ZBQ7	2	0	119	120	0.497930
Human_herpesvirus_8	A0A0N9SCV1	6	3	115	117	0.499490
Epstein-Barr_virus	A0A075FDV6	10	7	111	113	0.616180
Norwalk_virus	P54634	3	1	118	119	0.621870
Human_herpesvirus_8	O40947	12	9	109	111	0.648830
Rhinovirus_B	Q82081	33	29	88	91	0.658860
Epstein-Barr_virus	YP_001129465	4	2	117	118	0.683560
Human_mastadenovirus_A	P36717	4	2	117	118	0.683560
Human_respiratory_syncytial_virus	O91941	4	2	117	118	0.683560
Human_herpesvirus_8	A0A0N7GF08	4	2	117	118	0.683560
Human_betaherpesvirus_6A	NP_042902	5	3	116	117	0.721940
Human_herpesvirus_6B	NP_050266	5	3	116	117	0.721940
Hepacivirus_C	Q81487	5	3	116	117	0.721940
Hepacivirus_C	Q68798	5	3	116	117	0.721940
Human_herpesvirus_6B	NP_050228	40	37	81	83	0.782640
Human_respiratory_syncytial_virus	A0A0A0VAE3	11	9	110	111	0.816050
Human_herpesvirus_8	A0A0N7GEY6	14	12	107	108	0.835980
Human_herpesvirus_8	A0A0N9RMS1	17	15	104	105	0.849790
Human_cytomegalovirus	YP_081607	1	0	120	120	1.000000
Human_herpesvirus_3	NP_040134	4	3	117	117	1.000000
Tanapox_virus	NP_073461	1	0	120	120	1.000000
Human_cytomegalovirus	YP_081561	6	5	115	115	1.000000
Epstein-Barr_virus	YP_001129441	10	9	111	111	1.000000
Enterovirus_B	P21404	1	0	120	120	1.000000
Enterovirus_C	P03302	4	3	117	117	1.000000
Hepacivirus_C	Q81495	3	2	118	118	1.000000
Enterovirus_B	P13900	4	3	117	117	1.000000
Hepacivirus_C	Q68749	2	1	119	119	1.000000
Enterovirus_A	Q9QF31	1	0	120	120	1.000000
Hepacivirus_C	Q68801	5	4	116	116	1.000000
Enterovirus_C	P36290	6	5	115	115	1.000000
Hepacivirus_C	O92532	3	2	118	118	1.000000
Hepacivirus_C	P26661	1	0	120	120	1.000000
Hepacivirus_C	P29846	4	3	117	117	1.000000
Hepacivirus_C	Q5I2N3	2	1	119	119	1.000000
Hepacivirus_C	Q913D4	3	2	118	118	1.000000
Hepatitis_delta_virus	A0A190LCH0	1	0	120	120	1.000000
Sindbis_virus	Q8QQZ3	1	0	120	120	1.000000
West_Nile_virus	A0A067ZR50	1	0	120	120	1.000000
Hepatitis_delta_virus	A0A190LCP5	1	0	120	120	1.000000
Human_respirovirus_3	A0A1C9FM37	1	0	120	120	1.000000
Human_respiratory_syncytial_virus	A0A075BF65	1	0	120	120	1.000000
Human_herpesvirus_1	A0A0X8E9V5	1	0	120	120	1.000000
Hepatitis_delta_virus	E2J7V6	1	0	120	120	1.000000
Human_respiratory_syncytial_virus	C1K8Y2	1	0	120	120	1.000000
Human_respiratory_syncytial_virus	A0A068B4C9	7	6	114	114	1.000000
Enterovirus_C	Q7T7P4	9	8	112	112	1.000000
Human_respiratory_syncytial_virus	Q0QXK9	2	1	119	119	1.000000
Human_herpesvirus_8	A0A0N7GEZ7	1	0	120	120	1.000000
Human_cytomegalovirus	R45D58	1	0	120	120	1.000000
Rhinovirus_A	A5GZE6	1	0	120	120	1.000000
Human_herpesvirus_8	A0A0N9S3L8	2	1	119	119	1.000000
Human_respiratory_syncytial_virus	C1K8X6	2	1	119	119	1.000000
Epstein-Barr_virus	A0A0C7TL19	1	0	120	120	1.000000
Human_herpesvirus_8	Q9DUN0	3	2	118	118	1.000000
Epstein-Barr_virus	A0A075FFC3	18	17	103	103	1.000000
Human_mastadenovirus_C	I3VK64	1	0	120	120	1.000000
Human_respiratory_syncytial_virus	U6BZQ0	5	4	116	116	1.000000
Hepatitis_delta_virus	Q80JX3	1	0	120	120	1.000000
Rhinovirus_A	B9V447	8	7	113	113	1.000000
Human_herpesvirus_8	A0A0N7GF19	2	1	119	119	1.000000
Human_herpesvirus_8	DOUZU1	2	1	119	119	1.000000
Human_herpesvirus_8	A0A0N9SEW3	2	1	119	119	1.000000

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
Human_respiratory_syncytial_virus	A0A0F7JH5	15	27	45	154	0.080316
Human_respirovirus_3	A0A1C9FM91	3	2	57	179	0.100210
Human_rubulavirus_4	P17241	2	1	58	180	0.153900
Human_betaherpesvirus_7	YP_073771	1	0	59	181	0.248960
Human_coronavirus_NL63	YP_003771	1	0	59	181	0.248960
Human_mastadenovirus_E	YP_068029	1	0	59	181	0.248960
Hepacivirus_C	YP_001469630	1	0	59	181	0.248960
Human_metapneumovirus	YP_012606	1	0	59	181	0.248960
Enterovirus_D	NP_040760	1	0	59	181	0.248960
Human_respiratory_syncytial_virus	G8EIZ0	1	0	59	181	0.248960
Human_coronavirus_HKU1	Q0ZJE3	1	0	59	181	0.248960
Epstein-Barr_virus	G1CS70	1	0	59	181	0.248960
Epstein-Barr_virus	A0A0C7TGZ2	1	0	59	181	0.248960
Orf_virus	A0A0F6N3A1	1	0	59	181	0.248960
Human_respiratory_syncytial_virus	X2C1V9	1	0	59	181	0.248960
Human_respiratory_syncytial_virus	G8FOX7	1	0	59	181	0.248960
Epstein-Barr_virus	Q1EH77	1	0	59	181	0.248960
Human_respiratory_syncytial_virus	C1ILQ6	2	2	58	179	0.259070
Human_respiratory_syncytial_virus	A0A075BF70	4	6	56	175	0.271480
Epstein-Barr_virus	YP_401669	9	19	51	162	0.356890
Human_respiratory_syncytial_virus	E9NWA9	3	4	57	177	0.370070
Hepatitis_C_virus_genotype_1	NP_671491	1	1	59	180	0.436720
Hepacivirus_C	P26662	1	1	59	180	0.436720
Human_herpesvirus_8	Q91LX9	1	1	59	180	0.436720
Human_herpesvirus_1	A0A0X8EA26	1	1	59	180	0.436720
Epstein-Barr_virus	A0A0S2YQE3	1	1	59	180	0.436720
Epstein-Barr_virus	I1YP21	1	1	59	180	0.436720
Human_respiratory_syncytial_virus	Q38I82	7	15	53	166	0.442460
Human_respiratory_syncytial_virus	A0A075LY43	8	17	52	164	0.463070
Human_respiratory_syncytial_virus	Q4KV07	20	51	40	130	0.513680
Human_respiratory_syncytial_virus	A0A075BF97	10	23	50	158	0.515530
Human_respiratory_syncytial_virus	G8E0C0	5	10	55	171	0.536620
Human_herpesvirus_3	Q0Q9E6	5	11	55	170	0.554330
Epstein-Barr_virus	YP_001129465	2	4	58	177	0.640550
Human_herpesvirus_4_type_2	YP_001129464	31	86	29	95	0.655240
Human_herpesvirus_3	_W5RSJ9	9	22	51	159	0.656360
Human_respiratory_syncytial_virus	A0A059VDK9	3	6	57	175	0.694120
Epstein-Barr_virus	A0A125TOX3	3	7	57	174	0.713240
Human_herpesvirus_6B	NP_050192	4	10	56	171	0.753230
Human_herpesvirus_8	A0A0N9SNQ9	4	10	56	171	0.753230
Human_respiratory_syncytial_virus	B0LS43	6	16	54	165	0.797840
Human_cytomegalovirus	A0A0R5ECB9	7	19	53	162	0.811880
Human_respiratory_syncytial_virus	R9TC20	14	40	46	141	0.859140
Enterovirus_C	NP_041277	1	3	59	178	1.000000
Hepatitis_C_virus_genotype_7	YP_009272536	1	3	59	178	1.000000
Norwalk_virus	_P54634	1	3	59	178	1.000000
Hepacivirus_C	Q68749	1	2	59	179	1.000000
Rhinovirus_A	P23008	4	11	56	170	1.000000
Hepacivirus_C	Q91936	1	3	59	178	1.000000
Human_respiratory_syncytial_virus	Q4KUU7	19	57	41	124	1.000000
Human_respiratory_syncytial_virus	Q50G60	3	8	57	173	1.000000
Human_herpesvirus_1	F8RFJ2	1	2	59	179	1.000000
Human_respiratory_syncytial_virus	X2C165	1	3	59	178	1.000000
Human_respiratory_syncytial_virus	A0A075BFA4	1	3	59	178	1.000000
Human_herpesvirus_1	I3TCA8	1	2	59	179	1.000000
Epstein-Barr_virus	A0A127AWN9	1	2	59	179	1.000000
Human_respiratory_syncytial_virus	F8UST5	3	9	57	172	1.000000
Epstein-Barr_virus	H6VXA3	1	2	59	179	1.000000
Human_herpesvirus_8	A0A0N9SF30	1	3	59	178	1.000000
Human_herpesvirus_8	A0A0N9S3L8	1	2	59	179	1.000000
Human_respiratory_syncytial_virus	C1K8X6	1	2	59	179	1.000000
Human_respiratory_syncytial_virus	Q50G63	1	2	59	179	1.000000
Human_respiratory_syncytial_virus	A0A0X8CW15	3	9	57	172	1.000000

fisher_exact_tests/Vietnam_Hepatitis_B_vs_cohort_polyclonal_fisher_exact_tests.csv

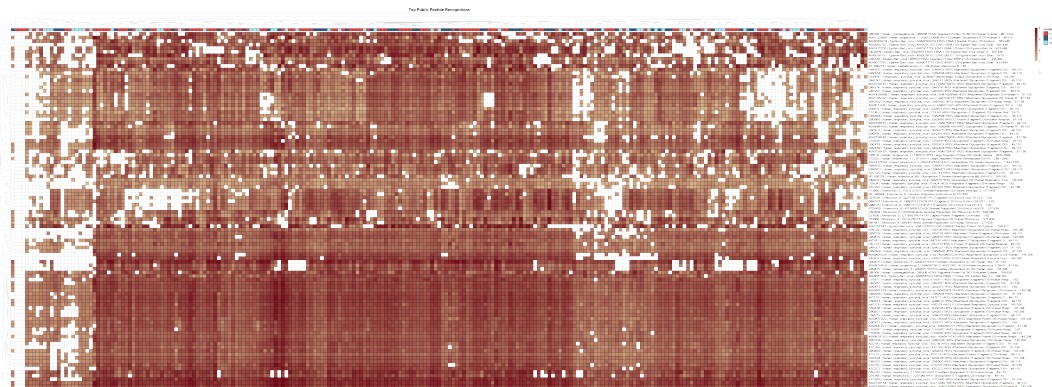
15 Summary Heatmaps: Top Public

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

16 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*.

17 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.

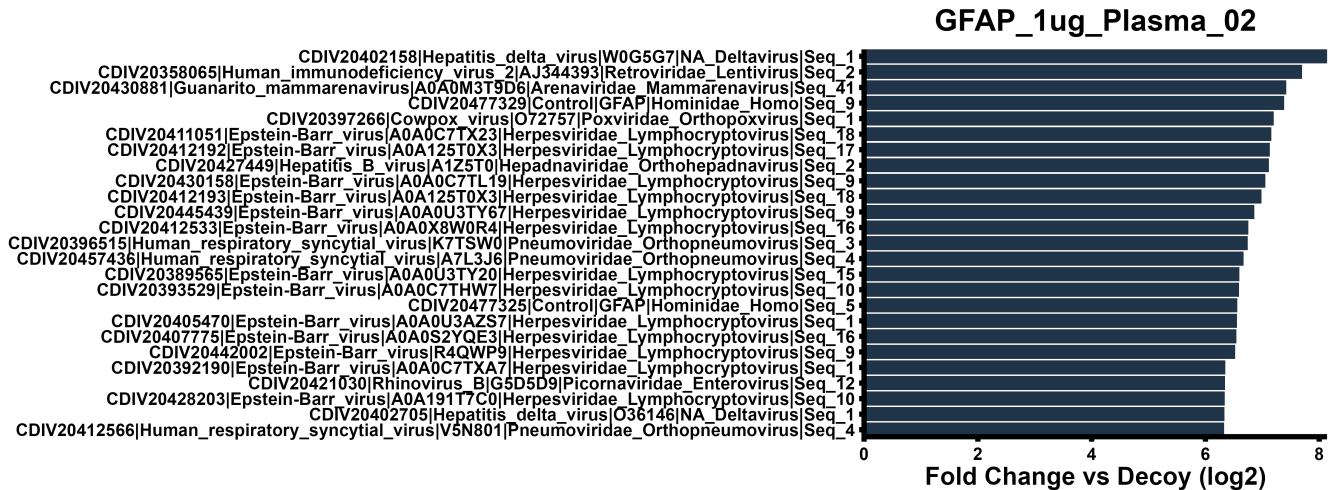
Heatmaps have been created for all proteins containing a peptide represented in the top_public_peptides file. Note: as some proteins will typically include two or more peptides in the top variants list there are often fewer than 100 heatmap files made.

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.

18 Individual Sample Barplots

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

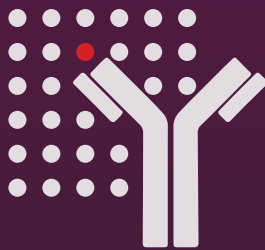
These barplots are created from the *edgeR_enriched_log2foldchg.csv* file and represent the top x25 enriched fold change signals versus 'beadsonly' decoy controls for each individual sample. If fewer than x25 significant hits are available, fewer peptides are plotted. An example of one of these images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



GFAP_1ug_Plasma_02_enriched_foldchange_barplot.png

19 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. Schubert, R. et al. Pan-viral serology implicates enteroviruses in acute flaccid myelitis. *Nat. Med.* 25(11): 1748–1752. (2019).
4. Mohan, D. et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* 13, 1958–1978 (2018).



THE AUTOANTIBODY SEROMICS COMPANY

CDI Labs Canada

MaRS Centre - West Tower
661 University Ave
Suite 1385
Toronto, ON
CANADA M5G 0B7

Sales

Email: info@cdilabs.ca

Support

Email: support@cdilabs.ca

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

© **CDI Labs Canada.** All rights reserved.

CDI Labs Canada and VirScan™ are trademarks of CDI Labs Canada.

