

Virus discovery in schizophrenia brain tissue

29 December 2021 20:38

Background: Psychosis is associated with pre-natal and post-natal viral infections (influenza, coronaviruses, bornavirus). It is typically assumed that these infections lead to systemic inflammatory signals (e.g. cytokines) that cause neurodevelopmental alterations (e.g. reduced synaptic density) which in turn then lead to the manifestation of psychosis upon a second hit (e.g. adolescence with synaptic pruning). However, one alternative explanation is that previous viral infections either lead to latent infections that become re-activated later on (e.g. similar to VZV, EBV or HSV) or that previous viral infections indicate a propensity towards viral infections and hence favour a later viral infection of the brain. In both cases, the viral re-activation/infection would trigger a specific immune response directed at the virus-infected cells in the brain and that is what causes psychosis. A previous metagenomics small study did not confirm this hypothesis, but this study was severely underpowered (<https://pubmed.ncbi.nlm.nih.gov/29478863/>)

Hypothesis: Viral RNA is more prevalent in brain tissue from psychosis patients as compared to healthy controls.

Approach: We will use existing datasets from the PsychEncode Consortium* that contain RNA sequencing data from post-mortem brain from schizophrenia patients and controls (N~1000). We will isolate the RNA transcripts that are not aligned to the human reference genome and use bioinformatics approaches** to quantify transcripts of known virus sequences. [Maybe, we will further use bioinformatics approaches for virus discovery# to quantify transcripts of unknown viruses. Maybe, we will also quantify HERV transcripts.] We will then compare viral transcript numbers between schizophrenia patients and healthy controls, while adjusting for multiple comparisons.##

Potential Outcomes & Next Steps: We might find a subset of viruses that is over-expressed in SCZ vs. controls. We can then look into mouse models of this particular virus infection and a) look at brain phenotypes b) immune phenotypes c) psychosis-like behaviour.

*<https://www.synapse.org/#!Synapse:syn4921369/wiki/235539> Common Mind, Common Mind HBCC, BrainVEGX, LIBD_szControl

** <https://github.com/dmarron/virdetect> from <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12864-020-6483-6#Sec9>

<https://github.com/yyoshiaki/VIRTUS> from <<https://academic.oup.com/bioinformatics/article/37/10/1465/5918022#264991820>>

<https://github.com/ICBI/viGEN> from <<https://www.frontiersin.org/articles/10.3389/fmicb.2018.01172/full#h3>>

<https://odin.mdacc.tmc.edu/~xsu1/VirusSeq.html> from <https://academic.oup.com/bioinformatics/article/29/2/266/202055?login=true>

similar to metagenomic approaches but for RNA instead of DNA, does a tool for this exist?

Can we use standard differential expression calculations for this? Or is this included in some of the bioinformatics pipelines above?

Virus / NMDAR research

25 March 2022 21:27

NMDAR subunits (GRIN)

HGNC ID (gene)	Approved symbol	Approved name	Previous symbols	Aliases	Chromosome
HGNC:4584	GRIN1	glutamate ionotropic receptor NMDA type subunit 1	NMDAR1	GluN1	9q34.3
HGNC:4585	GRIN2A	glutamate ionotropic receptor NMDA type subunit 2A	NMDAR2A	GluN2A	16p13.2
HGNC:4586	GRIN2B	glutamate ionotropic receptor NMDA type subunit 2B	NMDAR2B	GluN2B	12p13.1
HGNC:4587	GRIN2C	glutamate ionotropic receptor NMDA type subunit 2C	NMDAR2C	GluN2C	17q25.1
HGNC:4588	GRIN2D	glutamate ionotropic receptor NMDA type subunit 2D	NMDAR2D	GluN2D, EB11, NR2D	19q13.33
HGNC:16767	GRIN3A	glutamate ionotropic receptor NMDA type subunit 3A		GluN3A	9q31.1
HGNC:16768	GRIN3B	glutamate ionotropic receptor NMDA type subunit 3B		GluN3B	19p13.3

Searching subunits on SZDB

From <<https://www.genenames.org/data/genegroup/#/group/1201>>

Gene	Ensembl	Genome Position	logFC ^b	Avg Expression	T Statistic ^c	P-value	FDR ^d	B coefficient ^e
GRIN1	ENSG00000176884	chr9:140032842-140063207	-0.028	7.457	-0.584	5.59e-01	7.74e-01	-6.145
GRIN2A	ENSG00000183454	chr16:9852376-10276611	0.066	8.246	1.723	8.54e-02	3.06e-01	-4.868
GRIN2B	ENSG00000150086		-	0.081	7.079	2.160	3.12e-02	1.85e-01
GRIN2C	ENSG00000161509	chr17:72838162-72857627	-0.097	4.119	-1.755	7.97e-02	2.96e-01	-4.676
GRIN2D	ENSG00000105464	chr19:48898132-48948188	-0.055	1.410	-1.119	2.63e-01	5.33e-01	-5.176
GRIN3A	ENSG00000198785	chr9:104331635-104500862	0.104	4.783	3.208	1.41e-03	4.65e-02	-1.312

From <<http://www.szdb.org/cmc2.php>>

Notes:

^aGenome -Based on hg19 genome assembly.

^blogFC -log₂(Fold Change). Fold change is the ratio of the expression value of a gene in schizophrenia cases to the expression of healthy controls. If logFC > 0, this gene is highly expressed in schizophrenia. Otherwise, this gene is highly expressed in controls.

^cT Statistic -The T Statistic is used in a T test when deciding whether we should support or reject the null hypothesis.

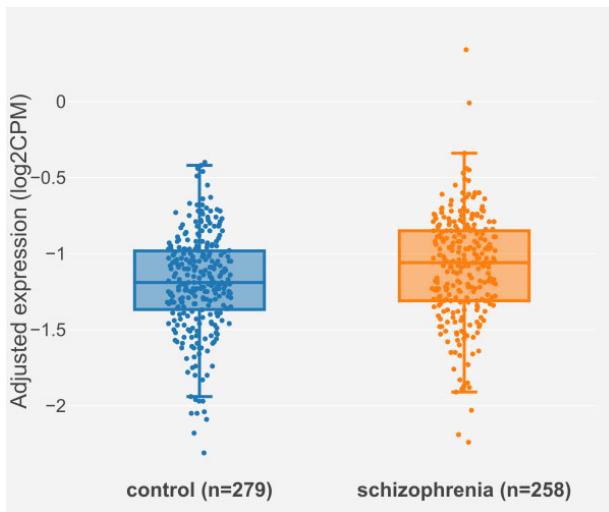
^dFDR -The false discovery rate (FDR) is a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons.

^eB -beta coefficients refer to how many standard deviations a dependent variable will change, per standard deviation coefficient increase in the predictor variable.

Note: GRIN2A was differentially methylated between schizophrenia cases and healthy controls in frontal cortex region.

From <http://www.szdb.org/gene_rank.php>

Boxplot of GRIN3A



Genomic data for GRIN3A

GRIN3A glutamate ionotropic receptor NMDA type subunit 3A [*Homo sapiens* (human)]

Studies in the knockout mouse deficient in this subunit suggest that this gene may be involved in the development of synaptic elements by modulating NMDA receptor activity. [provided by RefSeq, Jul 2008]

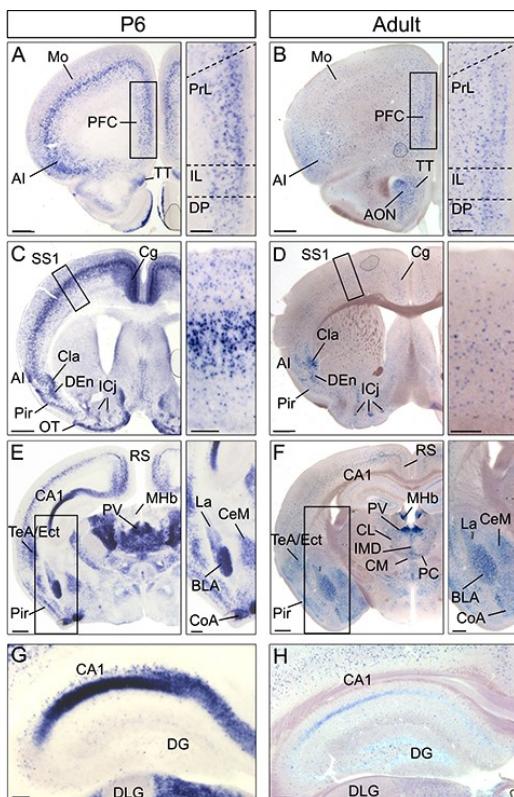
From <<https://www.ncbi.nlm.nih.gov/gene/116443>>

HIV-1 interactions ?

Temporal Dynamics and Neuronal Specificity of Grin3a Expression in the Mouse Forebrain Murillo et al. 2021

However, they have nonconventional properties that distinguish them from classical NMDARs (GluN1/GluN2 heteromers) including lower calcium permeability, reduced voltage-dependent block by magnesium, and lesser attachment to postsynaptic densities. Because of these properties, they can work as dominant-negative regulators of NMDAR-mediated plasticity and have been proposed to fine-tune the refinement of neural circuits during critical postnatal periods by inhibiting the stabilization of nonused synapses and promoting their pruning (Pachernegg et al. 2012; Perez-Otano et al. 2016).

From <<https://academic.oup.com/cercor/article/31/4/1914/6027877?login=true>>



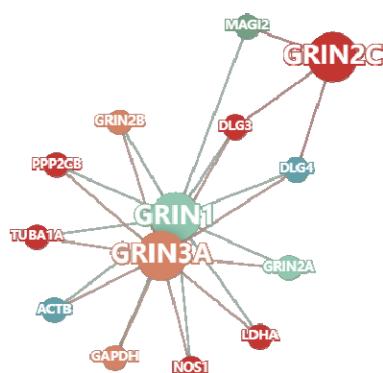
Protein-Protein Interaction

GRIN3A & GRIN1



SZDB-PPI

GRIN3A seems to have the same protein interaction as GRIN1 (GluN1) which is involved in NMDAR encephalitis, which doesn't seem to be the case for others:



SZDB-PPI

Curious about the proteins:

PPP2CB

Protein phosphatase 2A is one of the four major Ser/Thr phosphatases, and it is implicated in the negative control of cell growth and division.

TUBA1A

TUBA1A product is an alpha-tubulin that participates in the formation of microtubules

DLG3

DLG4

Encodes for PSD-95

GAPDH

an enzyme of about 37kDa that catalyzes the sixth step of glycolysis

LDHA

NOS1

Nos2 is important for protective immunity against CMV

In mice, the function of Nos2 in immunity against a number of viruses, bacteria, fungi, and parasites has been well characterized, whereas in humans the role of NOS2 has remained elusive and controversial.

[Linked to schizophrenia.](#)

ACTB

This is one of the two nonmuscle cytoskeletal actins.

Is GRIN3A suppressing spine density?

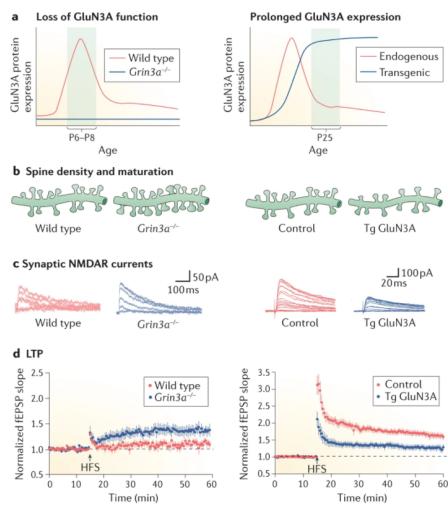
Dendritic impairments, including spine loss, are present in schizophrenia.

From <<https://www.sciencedirect.com/science/article/pii/S0304394014009203>>

GRIN3A deletion causes increased spine density:

Figure 2: GluN3A-NMDARs in synapse plasticity and maturation.

From: [Emerging roles of GluN3-containing NMDA receptors in the CNS](#)



<https://www.nature.com/articles/nrn.2016.92>

Cytomegalovirus, CMV

Human Cytomegalovirus (HCMV) induces Human Endogenous Retrovirus (HERV) transcription

From <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3842806/>>

Virus NMDA sequence similarity

29 March 2022 13:53

<https://www.nature.com/articles/s41598-021-96233-7>

<https://www.uniprot.org/uniprot/?query=taxonomy%3AViruses+organism%3Ahuman&sort=score>

NMDAR subunit & virus common sequences (NO GAP):

Using ~3,000 'Reviewed' viruses :

<https://www.uniprot.org/uniprot/?query=taxonomy:viruses%20organism:human&fil=reviewed%3Ayes&sort=score>

The 7 NMDAR subunits :

<https://www.uniprot.org/uniprot/?query=nmdar&fil=organism%3A%22Homo+sapiens+%28Human%29+%5B9606%5D%22&sort=score>

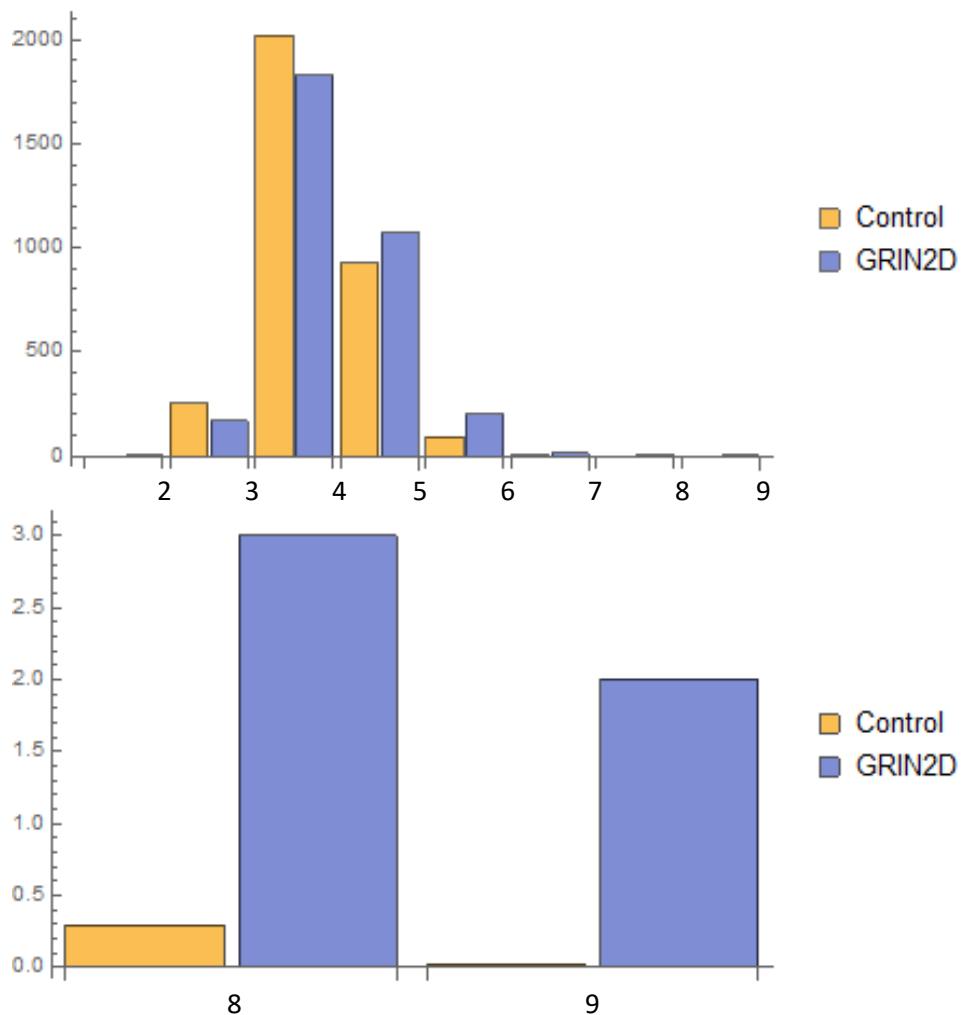
The longest common sequence for each subunit was:

{GRIN2D: 9,
GRIN3A: 6,
GRIN3B: 7,
GRIN1: 7,
GRIN2A: 7,
GRIN2B: 8,
GRIN2C: 8}

Testing Permutations

Average of 200 GRIN2D permutations

{2., 1.205}, {3., 252.335}, {4., 2017.77}, {5., 926.465}, {6., 90.595}, {7., 6.31}, {8., 0.295}, {9., 0.025}



Next using ~1.6M, including 'Unreviewed' viruses :

$\left\{ \begin{array}{l} \text{GRIN2D: 9,} \\ \text{GRIN3A: 7,} \\ \text{GRIN3B: 8,} \\ \text{GRIN1: 7,} \\ \text{GRIN2A: 8,} \\ \text{GRIN2B: 9,} \\ \text{GRIN2C: 8} \end{array} \right\}$

NMDAR subunit & virus common sequences (WITH GAP):

Test example using ~10% of GRIN2D sub-sequences (against 'verified viruses') :

Subsequence of length 12 with 3 or less non-matching AAs

AAA AAAA AAAAAA

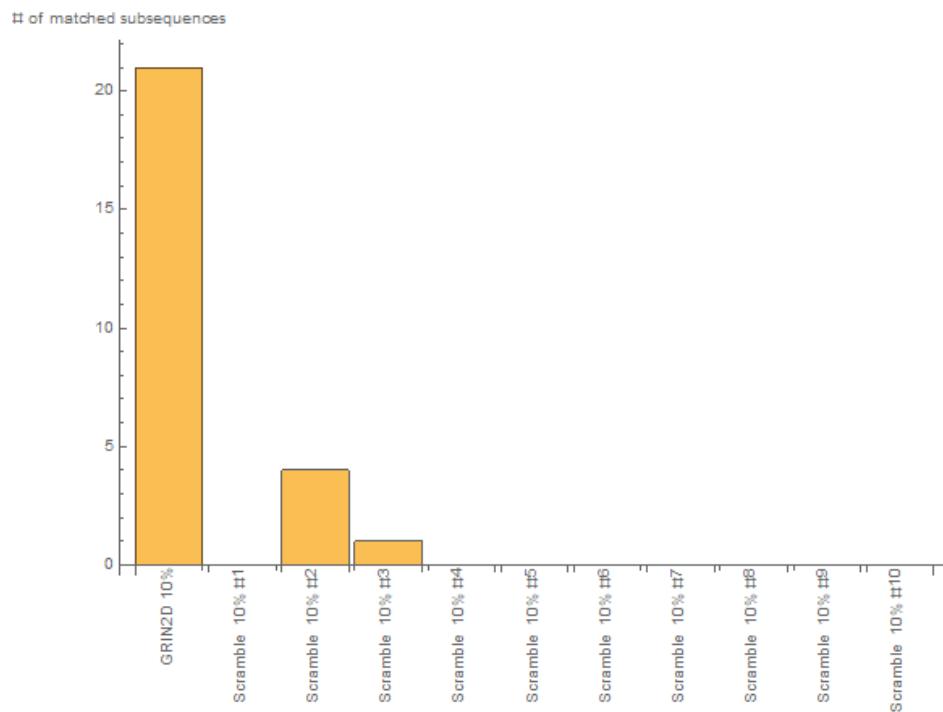
A A B A A A A A B A A

Naturally occurring?

2218 sub-sequences of 3295 viruses

7,308,310 virus subsequence of length 10 checked against each GRIN2D subsequence.

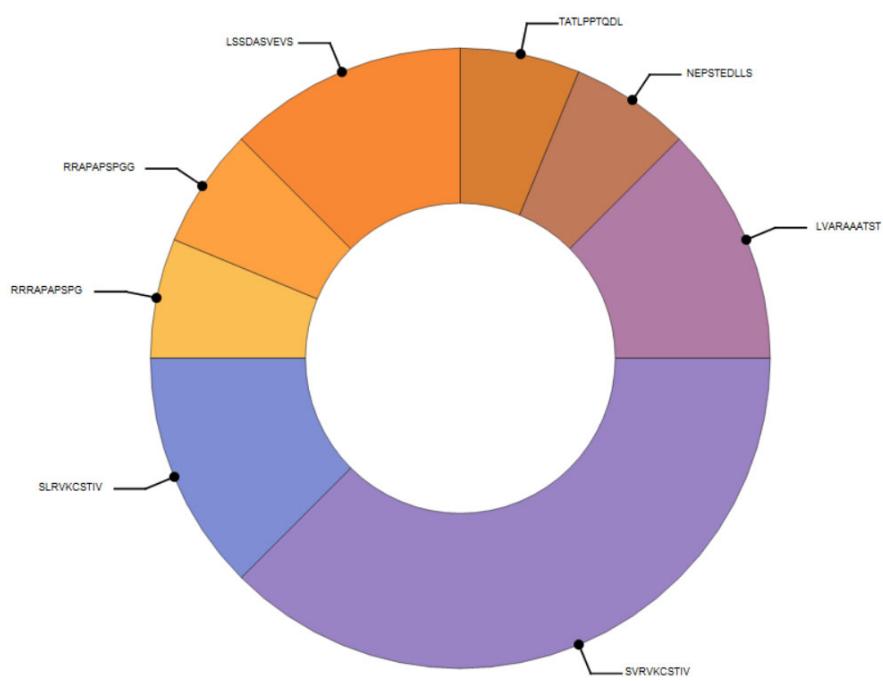
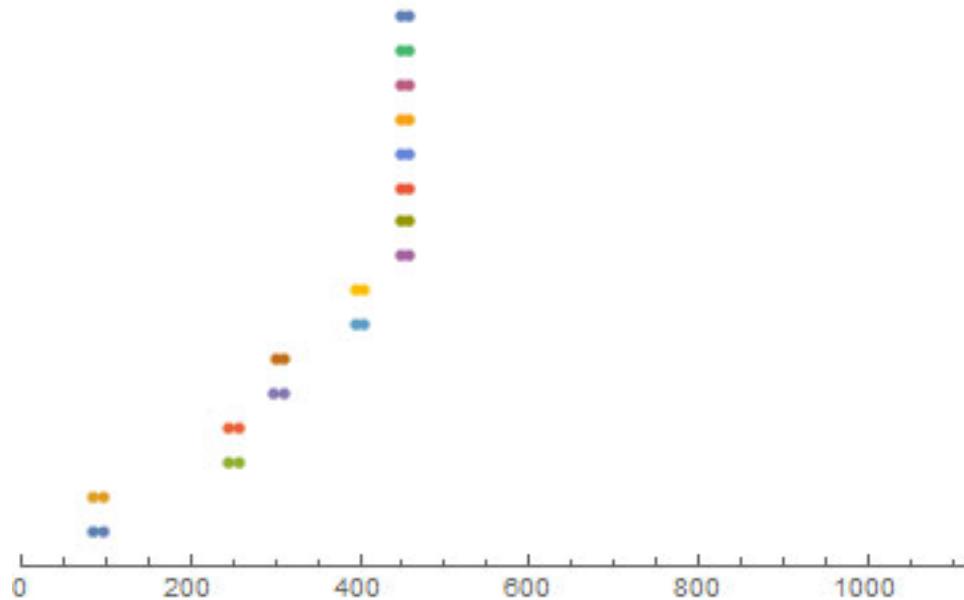
Checked against 10% of GRIN2D & scrambled permutations

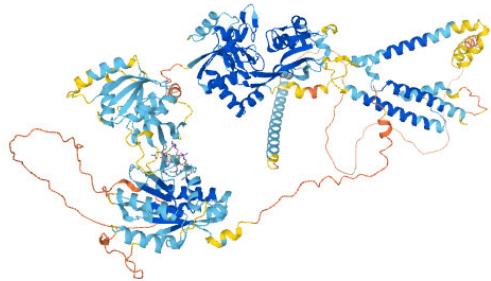


Example: GRIN3A

Position on the GRIN3A subunit with sequence of 10AA







<https://alphafold.ebi.ac.uk/entry/Q8TCU5>

Virus matches:

15 unique viruses

Human adenovirus C serotype	{"RRRAPAPSPG", "TRRSPAPSPG", }, {{"RRAPAPSPGG"}, "RRSPAPSPGA", }
Human adenovirus F serotype	"LSSDASVEVS", "LSSDADVTVS"
Human adenovirus F serotype	{"LSSDASVEVS"}, "LSSDADVTVS"
Epstein-Barr virus	"TATLPPTQDL",, "TANLPSTQDL"
Variola virus	"NEPSTEDLLS", "NLPSTQDLLS"
Human cytomegalovirus	"LVARAATST", "LVARAVATAT"
Human cytomegalovirus	"LVARAATST", "LVARAVATAT"
Human coronavirus	{"SVRVKCSTIV"}, "SIRVKGSTIV"
Human coronavirus	{"SVRVKCSTIV"}, "SIRVKGSTIV"
Human coronavirus	{"SVRVKCSTIV"}, "SIRVKGSTIV",
Human coronavirus	{"SLRVKCSTIV"}, "SIRVKGSTIV"
Human coronavirus	{"SVRVKCSTIV"}, "SIRVKGSTIV"
Human coronavirus	{"SLRVKCSTIV"}, "SIRVKGSTIV"

Virus common sub sequences results

10 April 2022 20:08

Results:



ViralComm
onSubseq...

- GRIN3A
- GRIN2D
- Random protein
 - o BRD4



ViralComm
onSubseq...
- All 7 subunits

BRD4 turns out to be heavily involved in virus interaction:

Involvement of Brd4 in different steps of the papillomavirus life cycle

From <<https://www.sciencedirect.com/science/article/pii/S0168170216305913>>

The EBNA1 Protein of Epstein-Barr Virus Functionally Interacts with Brd4

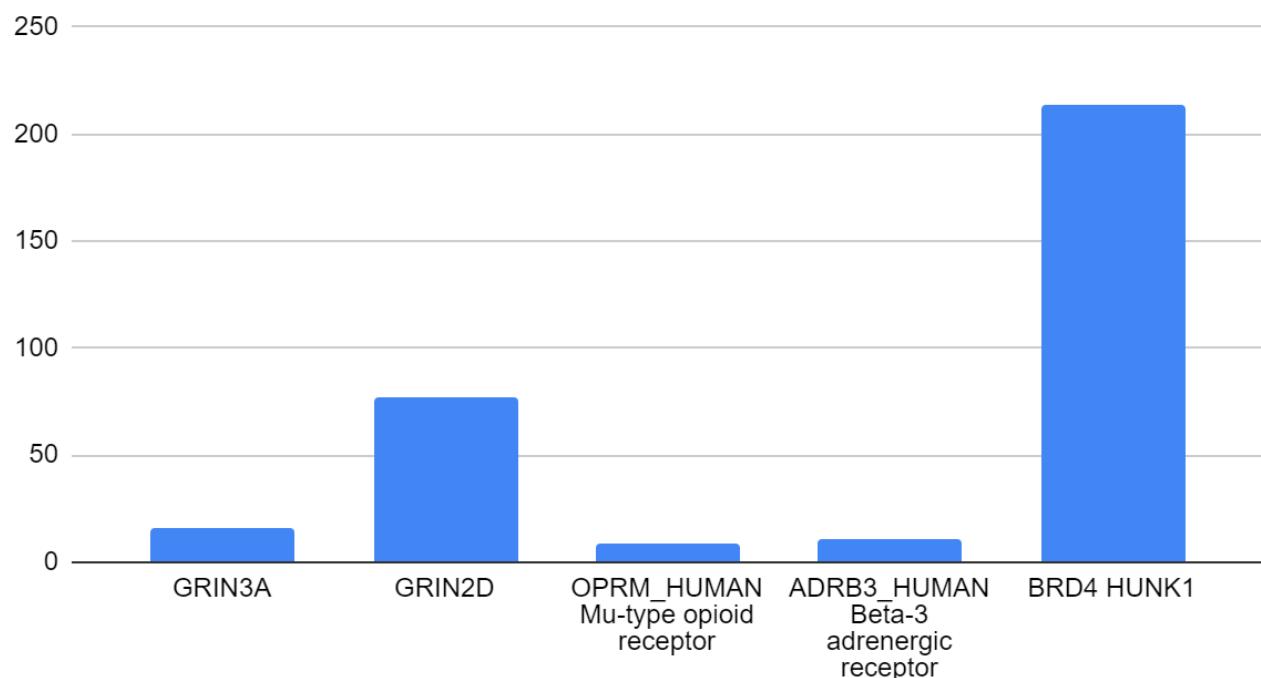
From <<https://journals.asm.org/doi/full/10.1128/JVI.01680-08>>

Cross-reactive epitope Coxsackie and GAD relevant for diabetes mellitus type 1

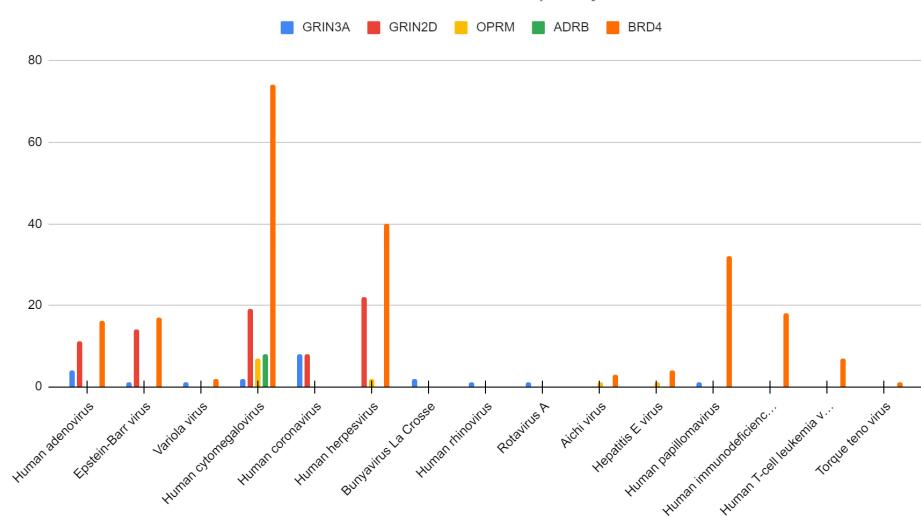
<https://www.sciencedirect.com/science/article/pii/S0009898121000097?via%3Dihub#b0095>

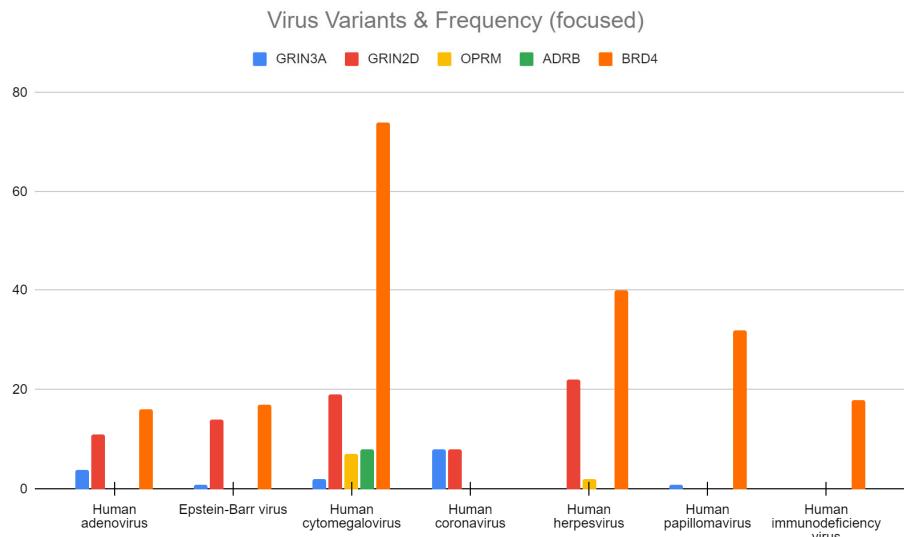
Test viral subsequences matched to alpha and beta cells, particularly the glicucon (alpha) / insulin (beta) cycle? There is a strong interaction with autoimmune for beta cells not alpha.

Viral matches



Virus Variants & Frequency



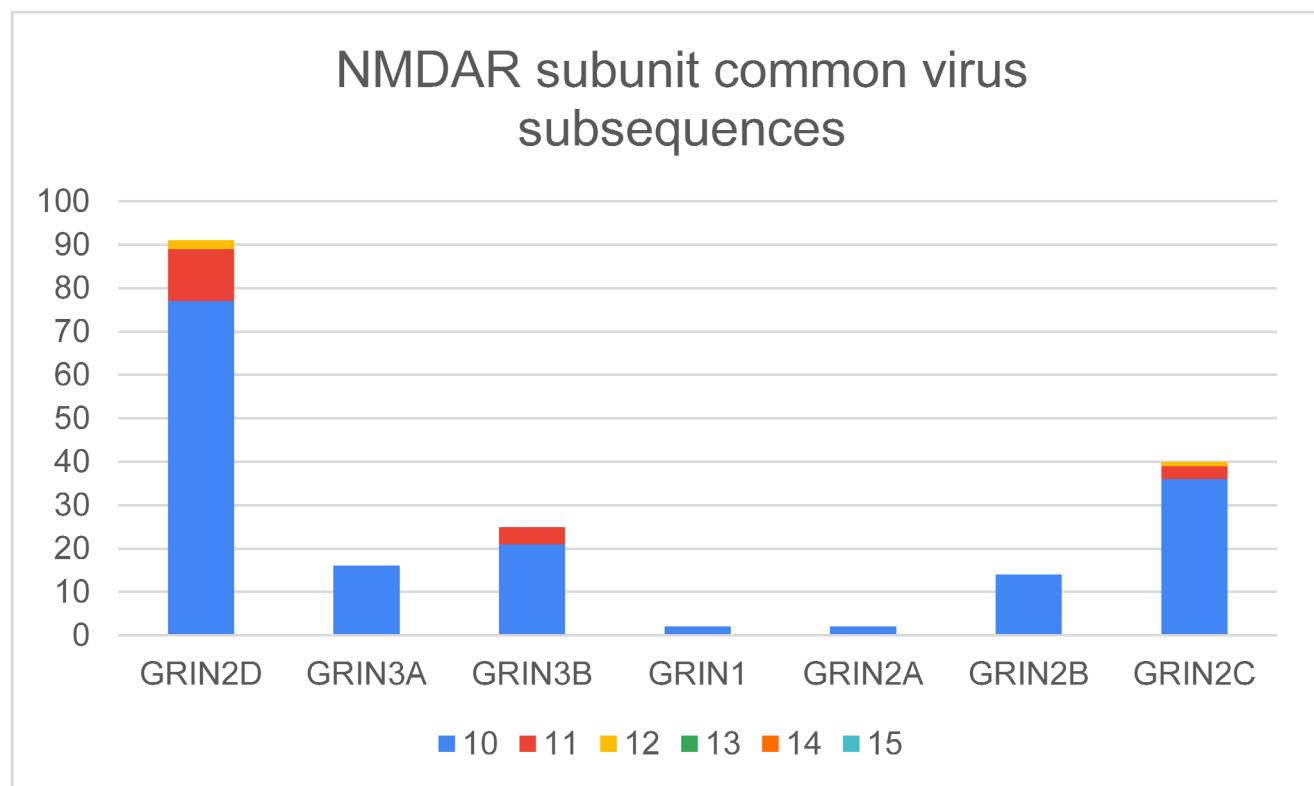


To do:

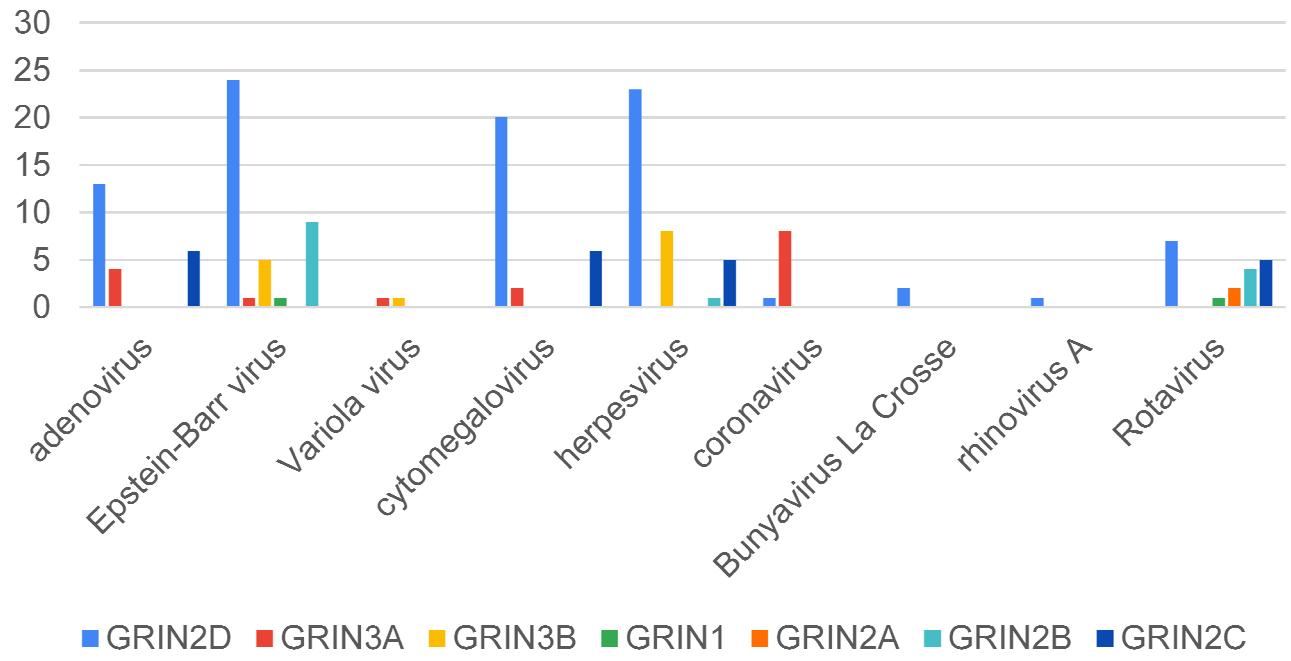
- GRIN1
- GRIN2A
- GRIN2B
- GRIN2C
- GRIN3B
- Random protein (not involved in virus interaction)

Updated results

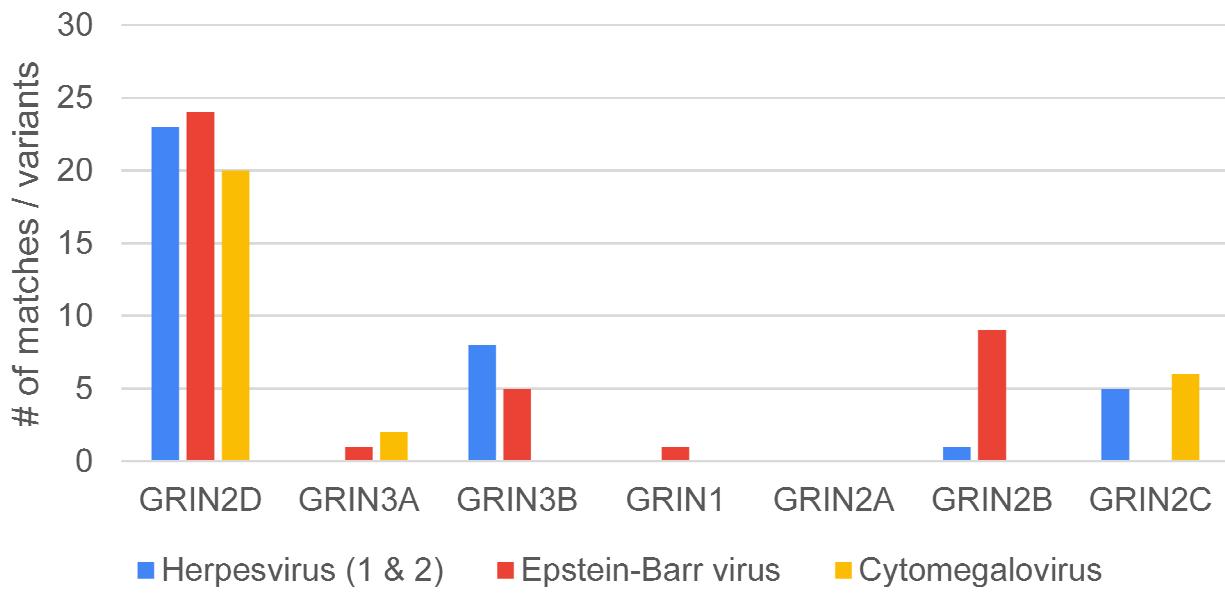
NMDAR subunits

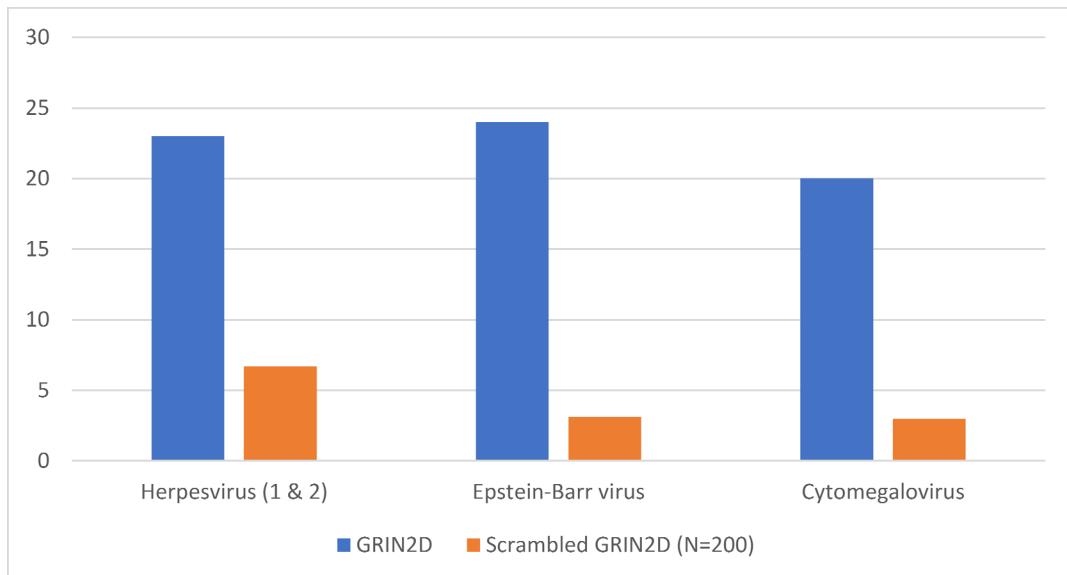


Virus variants



Viruses with links to schizophrenia / psychosis





SD ~ 2

MAX ~ 12,26,9

Diabetes type 1

20 April 2022 19:07

Cross-reactive epitope Coxsackie and GAD relevant for diabetes mellitus type 1

<https://www.sciencedirect.com/science/article/pii/S0009898121000097?via%3Dihub#b0095>

in genetically predisposed subjects, a particular condition in which chronic local inflammation occurs through the persistence of the infecting virus in pancreatic tissue and the activation of autoimmunity by molecular mimicry, bystander activation or both.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997365/>

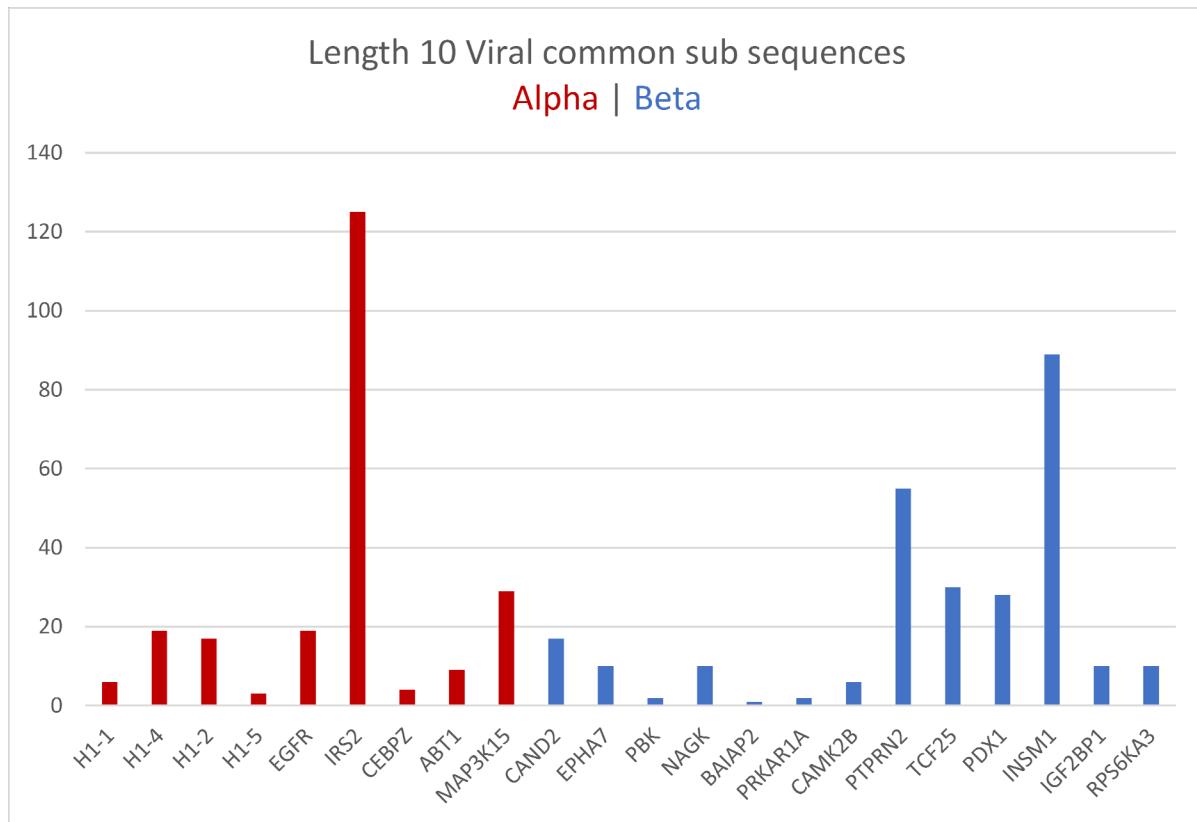
Table 2

Proteins with different expression levels in alpha and beta cells.

Leading Protein ID	Gene	Protein Names	β/α (Log2)	p-value
IPI00228830	Ube3c	Ubiquitin-protein ligase E3C	2.0634	0.0051
IPI00420577	Cand2	Cullin-associated NEDD8-dissociated protein 2;Cullin-associated and neddylation-dissociated protein 2;p120 CAND2;TBP-interacting protein TIP120B;TBP-interacting protein of 120 kDa B	2.7923	0.0003
IPI00124993	Epha7	Ephrin type-A receptor 7;Tyrosine-protein kinase receptor EHK-3	2.0891	0.0047
IPI00467350	Pbk	Lymphokine-activated killer T-cell-originated protein kinase	2.286	0.0022
IPI00136625	Nagk	N-acetyl-D-glucosamine kinase;GlcNAc kinase	2.3244	0.0019
IPI00222731	Baiap2	Brain-specific angiogenesis inhibitor 1-associated protein 2;Insulin receptor tyrosine kinase 53 kDa substrate	2.5636	0.0007
IPI00119575	Prkar1a	Putative uncharacterized protein;cAMP-dependent protein kinase type I-alpha regulatory subunit	4.081	0.0000
IPI00649296	Camk2b	Calcium/calmodulin-dependent protein kinase II	4.3953	0.0000
IPI00465657	Map3k15	Mitogen-activated protein kinase kinase kinase 15;MAPK/ERK kinase kinase 15	-3.3046	0.0000
IPI00471226	Abt1	Activator of basal transcription 1	-1.9123	0.0097
IPI00928012	Ptpn2	Protein tyrosine phosphatase, receptor type, N polypeptide 2;Receptor-type tyrosine-protein phosphatase N2;PTP	2.3397	0.0018

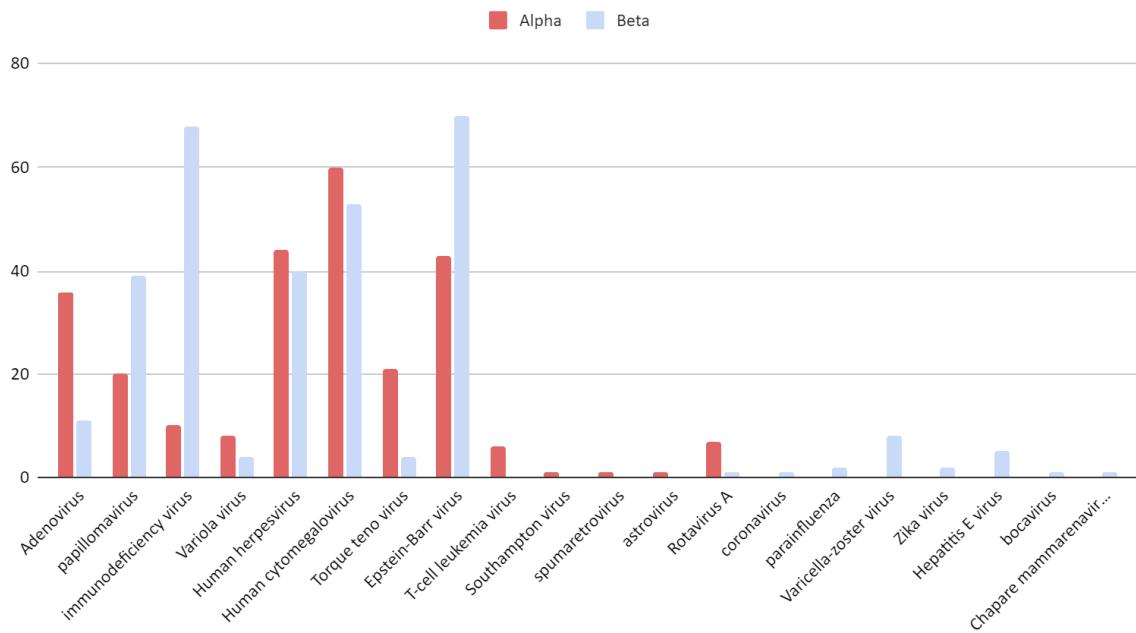
		IA-2beta;Protein tyrosine phosphatase-NP		
IPI00752710	Cebpz	CCAAT/enhancer-binding protein zeta;CCAAT-box-binding transcription factor	-2.0338	0.0064
IPI00674690	Tcf25	Transcription factor 25;Nuclear localized protein 1;Nuclear localized protein-1 isoform d	2.3231	0.0019
IPI00923679	Irs2	Insulin receptor substrate 2;4PS	-2.1428	0.0044
IPI00222731	Baiap2	Brain-specific angiogenesis inhibitor 1-associated protein 2;Insulin receptor tyrosine kinase 53 kDa substrate;Insulin receptor substrate p53;Insulin receptor substrate protein of 53 kDa;Brain-specific angiogenesis inhibitor 1-associated protein 2, isoform CRA_c	2.5636	0.0007
IPI00132557	Pdx1	Pancreas/duodenum homeobox protein 1;Insulin promoter factor 1;Islet/duodenum homeobox 1;Somatostatin-transactivating factor 1	2.6903	0.0004
IPI00471240	Insm1	Insulinoma-associated protein 1;Zinc finger protein IA-1	3.9295	0.0000
IPI00134310	Ins1	Insulin-1;Insulin-1 B chain;Insulin-1 A chain	4.4525	0.0000
IPI00134311	Ins2	Insulin-2;Insulin-2 B chain;Insulin-2 A chain	5.1812	0.0000
IPI00886028	Igf2	Insulin-like growth factor II;Multiplication-stimulating polypeptide;IGF-II;Preptin	6.1304	0.0000
IPI00135645	Gcg	Glucagon	-4.2082	0.0000
IPI00121190	Egfr	Epidermal growth factor receptor	-2.0428	0.0062
IPI00230133	Hist1h1b	Histone H1.5;H1 VAR.5;H1b	-2.0679	0.0057
IPI00223713	Hist1h1c	Histone H1.2;H1 VAR.1;H1c	-2.2984	0.0024
IPI00223714	Hist1h1e	Histone H1.4;H1 VAR.2;H1e	-2.4348	0.0014
IPI00228616	Hist1h1a	Histone H1.1;H1 VAR.3	-2.4669	0.0013
IPI00114333	Rps6ka3	Ribosomal protein S6 kinase alpha-3	1.8178	0.0117

From <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997365/table/pone-0095194-t002/?report=objectonly>>

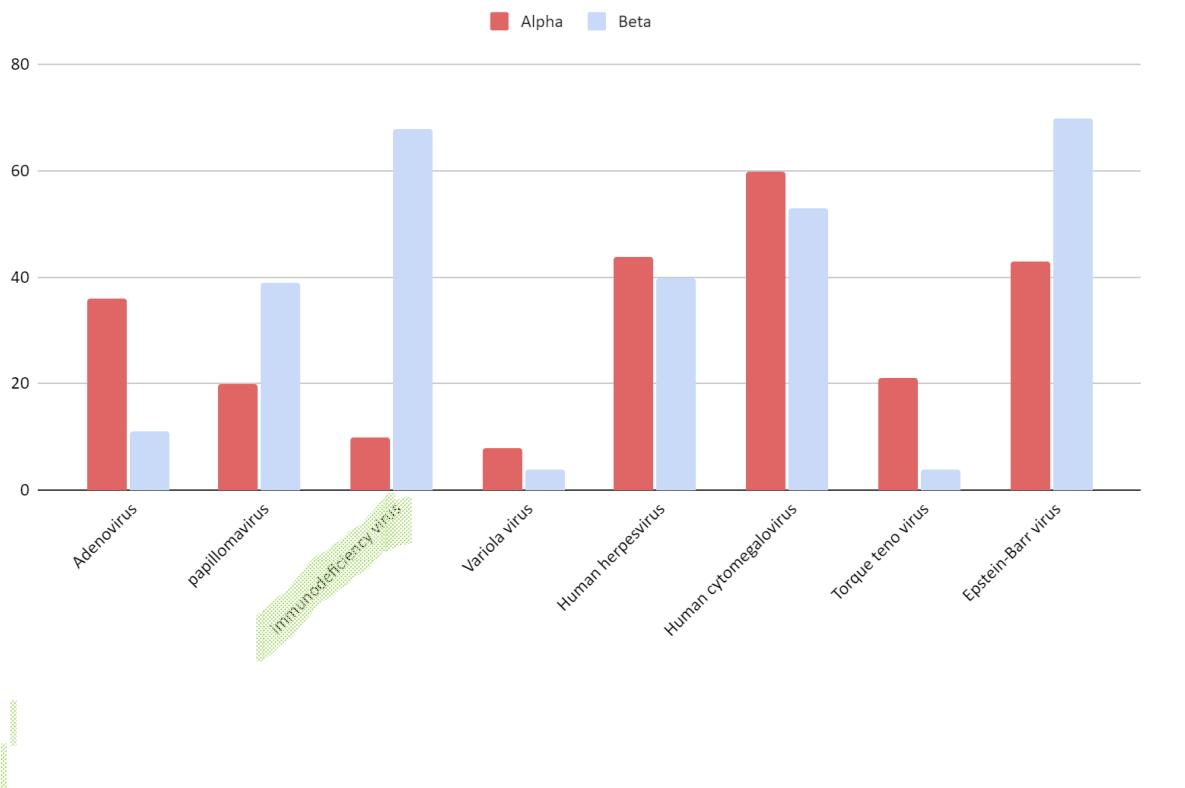


High matches are associated with Insulinoma?

Virus variants: differentially expressed proteins in Alpha and Beta cells



Virus variants: differentially expressed proteins in Alpha and Beta cells



Genome of Coxsackie Virus

<https://www.frontiersin.org/articles/10.3389/fmicb.2019.01001/full#h4>

Cell (differenti- ally expressed)	Coxsacki- e ID	Protein	Sequenc- e length	coxsackie Position	Coxsackie sequence	Protein ID	Protein sequence	Protein position	Coxsackie viral sequence / genome
Beta	1	UBE3C	10	414	GREGPW QSTL	1	GRIGPL QSTL	657	AXS68447.1 polyprotein [Coxsackiev irus A4]
Alpha	-	-	-	-	-	-	-	-	-

Antibody prediction

T cell prediction

Diabetes type 1 : T cells and B cells... Oh my.

09 May 2022 18:23

GAD65

Out[225]= Virus

```
Out[226]= {{HLA-DQA1*03:01-DQB1*03:02, VKILPEVK, 0.178732, 0.0220406},  
           {HLA-DQA1*03:01-DQB1*03:02, KILPEVKE, 0.138507, 0.04851},  
           {HLA-DQA1*03:01-DQB1*03:02, ILPEVKEK, 0.0931147, 0.0355584},  
           {HLA-DQA1*03:01-DQB1*03:02, LPEVKEKH, 0.0334235, 0.0330668}}}
```

Out[227]= Protein

```
Out[228]= {{HLA-DQA1*03:01-DQB1*03:02, FKMFPEVK, 0.163576, 0.00732368},  
           {HLA-DQA1*03:01-DQB1*03:02, KMFPEVKE, 0.1216, 0.0151587},  
           {HLA-DQA1*03:01-DQB1*03:02, MFPEVKEK, 0.0917334, 0.0158245},  
           {HLA-DQA1*03:01-DQB1*03:02, FPEVKEKG, 0.0497132, 0.0279057}}}
```

There is a sequence homology that involves 7–9 amino acid residues of which 6 in particular correspond to the identical “PEVKEK” fragment in the proteins hGAD65 and CVB4 and the “PEVKTK” fragment differs by a single residue in the hGAD67 ([Table 1](#)). As a

INS

Out[233]= Virus

```
Out[234]= {{HLA-DQA1*03:01-DQB1*03:02, LHGSPPGA, 0.0423898, 0.00429962}}}
```

Out[235]= Protein

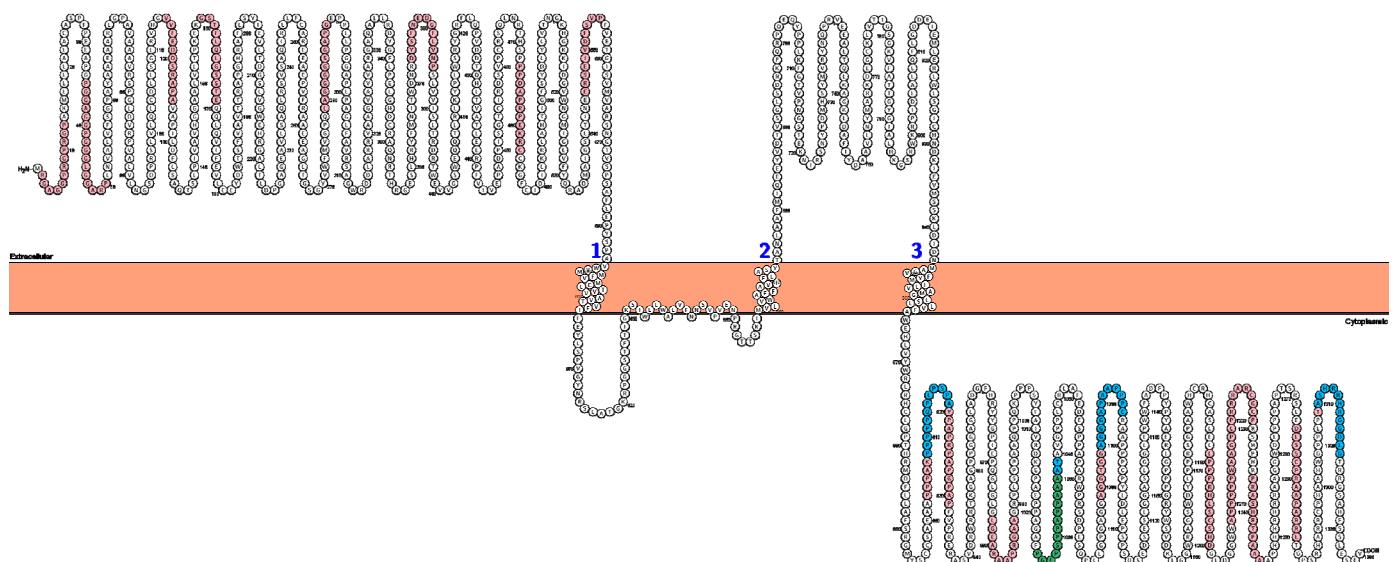
```
Out[237]= {{HLA-DQA1*03:01-DQB1*03:02, LGGGPPGA, 0.0525819, 0.0183631}}}
```

A 'broken' subsequence with identical matches ~3

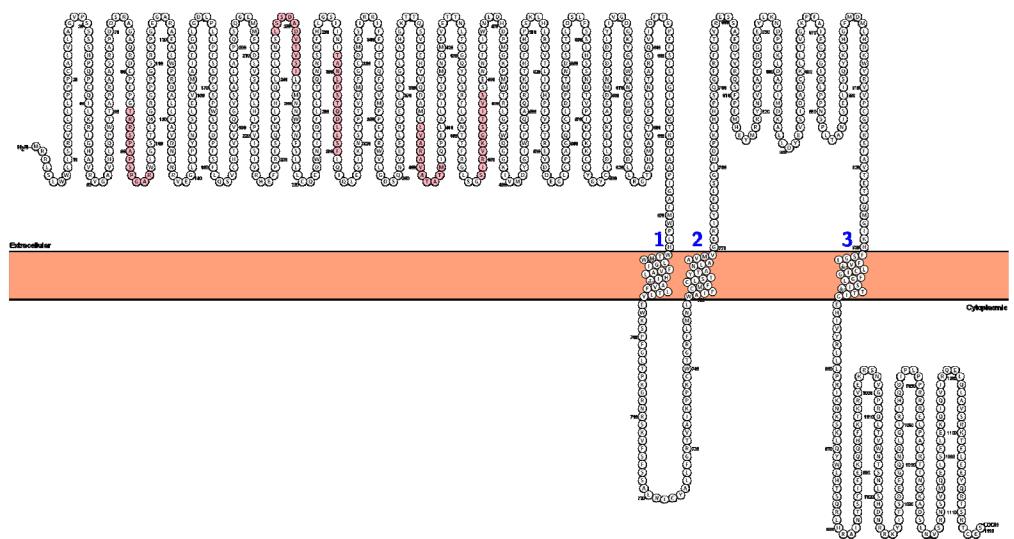
Protter plots

26 April 2022 17:09

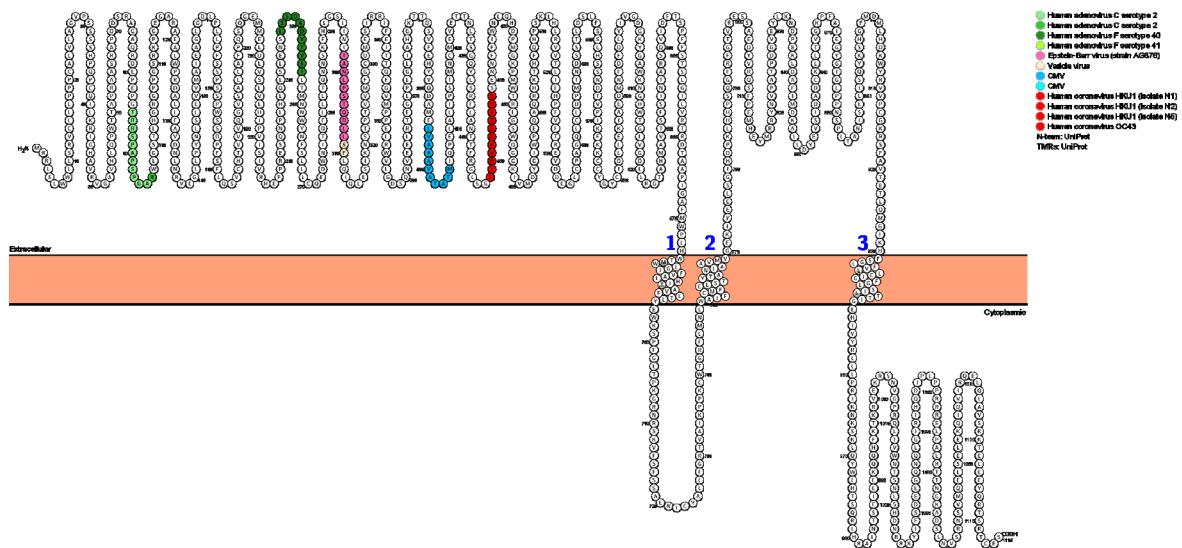
GRIN2D



GRIN3A



Virus ID positions:



[Protter - interactive protein feature visualization \(ethz.ch\)](#)

Compiled data

05 May 2022 14:02

Updated library 1.02

Diabetes Type 1

- Coxsackievirus proteome
 - A10
 - A21
 - B4 (which is linked to N1)
- Alpha & Beta cell proteins (differentially expressed)
- Autoantigens (Human sequences)
 - INS / INS1 (= Preproinsulin ?)
 - IGF2
 - Glutamic decarboxylase 65 (GAD65)
 - GAD67
 - PTPRN ICA3, ICA512
 - INSM2 (= IA-2)
 - PTPRN2
 - Peripherin
 - Peripherin-2
 - Carboxypeptidase E
 - HSP60
 - IGRP
 - Ganglioside
 - GM2A
 - GDAP1

(from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312399>)

HLA Sequences

From <[https://www.ebi.ac.uk/ipd/imgt/hla/alleles/?query=and\(contains\(previous_nomenclature,%22DR%22\),contains\(previous_nomenclature,%223%2A%22\)\)&next=gANdcQBYCAAAEhMQTAxODcwcQFhLg](https://www.ebi.ac.uk/ipd/imgt/hla/alleles/?query=and(contains(previous_nomenclature,%22DR%22),contains(previous_nomenclature,%223%2A%22))&next=gANdcQBYCAAAEhMQTAxODcwcQFhLg)>

- DQA1*03010101 *
- DQB1*0302 *
- DQB1030102
- DQB103010101
- DR3 & DR4 - couldn't find specific mention so compiled a selection of:
 - DRB3
 - DRB4

From <<https://www.sciencedirect.com/science/article/pii/S0303720718301886?via%3Dihub>>

NMDAR

HLA Sequences

- Linked with Schizophrenia

From <<https://onlinelibrary.wiley.com/doi/full/10.1111/iji.12507>>

- DRB1*10:01:01
- Pre-loaded on NetHMC
 - A*24:02:01
 - B*37:01:01
 - C*06:02:01
- DQB1*05:01:01 (Remaining)

NeuroPeptides

- Human & Mouse from NeuroPep

- A list of active peptides from [Neuropeptides.nl](#)
- Enzymes

Virus database

- Basic Virus database (N = ~4000 ; Unique = ~60)
- Specific virus database (Reduced Duplications)
 - [Human viruses table ~ ViralZone \(expasy.org\)](#)
 - Comes in pieces - either whole or selective viruses?

I compiled some data:

Diabetes Type 1

- Coxsackievirus proteome
 - A10
 - A21
 - B4 (which is linked to N1)
- Alpha & Beta cell proteins (differentially expressed) I would suggest to go with a list of known autoantigens, e.g.
glutamic acid decarboxylase 65 & 67
proinsulin
carboxypeptidase H
insulinoma antigen-2 (IA-2)
IA-2 β (also termed phogrin or ICA512)

(from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312399>)

- Also, there are some known HLA associations that we can use to double check our results
DR4/DQ8 or DR3/DQ2 in 90% of patients
DQ8 (DQA1*0301, DQB1*0302) raises by factor 11

From <<https://www.sciencedirect.com/science/article/pii/S0303720718301886?via%3Dihub>>

=> We would be confident in our algorithm if it showed for example Coxsackie B4 and GAD-65 share an epitope that is preferentially presented by the MHC II complex with HLA alleles DR4/DQ8 or DR3/DQ2

NeuroPeptides

- Human & Mouse from NeuroPep
- A list of active peptides from [Neuropeptides.nl](#)
 - need to find the AA sequences The genes encoding the peptides could help with that (e.g. PENK, POMC)
 - In addition (this can be for later, but I am putting it here for later reference), it would be great to include enzymes that process neuropeptides. Candidates are
 - Endopeptidases: cleaving of large precursor into smaller peptides
 - prohormone convertase 1 (also named prohormone convertase 3)
 - prohormone convertase 2
 - prohormone/proprotein convertase 5A (also known as prohormone/proprotein convertase 6A)

From <<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/mas.20079>>

- Carboxypeptidases: removal C-terminal residues
 - Carboxypeptidase H (also called carboxypeptidase E)
 - Carboxypeptidase D
 - Metallocarboxypeptidase
- Non-conventional processing

- endothelin converting enzyme-2
- carboxypeptidase A5
- carboxypeptidase A6
- number of extracellular peptidases

From <<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/mas.20079>>
- Posttranslational modifications
 - peptidyl- α -amidating monooxygenase
 - tyrosylprotein sulfotransferase
 - And other enzymes for amidation, acetylation, phosphorylation, sulfation, and glycosylation

From <<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/mas.20079>>

Original virus database

- Should we the one that George sent us?

NMDAR subunits

2 thoughts:

1. For the final analysis, it would be great to have a protein database of neuronal receptors/surface proteins, I am sure someone must have assembled some databases already. Or maybe a uniprot query will do with neurotransmitter receptor activity AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"
From <<https://www.uniprot.org/uniprot/P02708>>
2. I would expect the neurotransmitter receptors to be more important for B cell immunity, rather than for T cell immunity (but maybe I am wrong about this). For B cell immunity, George suggested to basically do what you have been doing so far, but migrate it to blast.

Sars Covid 2 proteome (thought it could be interesting to add) definitely, great!!

Is there something missing? I'm just not sure how to approach the narcolepsy idea

<https://www.nature.com/articles/s41467-021-22637-8ls> this the paper we were referencing? From what I understand they are doing effectively what we want to do and actually show that there is a link between HLA-DQB1*0602, Narcolepsy & a T cell epitope common between H1N1 & primarily a protein POMT1. Does this already show the idea or did we want to explore B cells in this scenario? Sorry I may have misunderstood.

Wow, that is cool! Indeed, this paper does what we wanted to do as a positive control! We should definitely include the trio H1N1, POMT1 and HLA-DQB1*0602 as a positive control for our T cell pipeline. As I understand their paper, they use simple blast plus in-vivo experiments and not a specific bioinformatic T cell approach as George suggested, and it would be good to see whether our approach gives converging results.

RESULTS

Significant type 1 diabetes associations were observed at all class I HLA loci. After accounting for LD with HLA class II, the most significantly type 1 diabetes-associated alleles were B*5701 (odds ratio 0.19; $P = 4 \times 10^{-11}$) and B*3906 (10.31; $P = 4 \times 10^{-10}$). Other significantly type 1 diabetes-associated alleles included A*2402, A*0201, B*1801, and C*0501 (predisposing) and A*1101, A*3201, A*6601, B*0702, B*4403, B*3502, C*1601, and C*0401 (protective). Some alleles, notably B*3906, appear to modulate the risk of all DRB1-DQA1-DQB1 haplotypes on which they reside, suggesting a class I effect that is independent of class II. Other class I type 1 diabetes associations appear to be specific to individual class II haplotypes. Some apparent associations (e.g., C*1601) could be attributed to strong LD to another class I susceptibility locus (B*4403).

From <<https://diabetesjournals.org/diabetes/article/59/11/2972/15961/HLA-Class-I-and-Genetic-Susceptibility-to-Type-1>>

Cut using:

<https://services.healthtech.dtu.dk/service.php?NetChop-3.1>

Then run for HMC ii using

<https://bertmhc.privacy.nlehd.de/>

Do we need to make peptides based on cleavage sites - apparently HMC ii unlike HMC i have a sliding window and can allow peptides to wiggle.

SCZ MHC

10 May 2022 17:20

MHCi

• Review:

- HLA_A*101 (Swedish study 'strongest evidence'* Re-confirmed by UK study)
- HLA_C*0701
- HLA_B*0801 (Swedish study 'strongest evidence'* Re-confirmed by UK study)

GRIN2D gave strong and weak binding matches

Searching for exact 6 AA matches against the original virus database for potential strain matches:

```
Out[139]TableForm
sp|B8XTP8|POLG_COSAA Genom polyprotein OS-Cosavirus A (isolate Human/Pakistan/0553/-) OX-1554483 PE=3 SV=1
sp|Q3KSP5|LF2_EBV Protein LF2 OS-Epstein-Barr virus (strain GD1) OX-10376 GN-LF2 PE=1 SV=1
sp|F7V998|U150A_HCMW Uncharacterized protein UL150A OS-Human cytomegalovirus (strain Merlin) OX-295027 GN-UL150A PE=4 SV=1
sp|P16847|UL28_HCMW Uncharacterized protein UL28 OS-Human cytomegalovirus (strain AD169) OX-10360 GN-UL28 PE=3 SV=1
sp|C0H677|UL298_HCMW Protein UL298 OS-Human cytomegalovirus (strain AD169) OX-10360 GN-UL298 PE=1 SV=1
sp|Q6SWA3|UL298_HCMW Protein UL298/28 OS-Human cytomegalovirus (strain Merlin) OX-295027 GN-UL298 PE=3 SV=2
```

```
Out[137]{{EKGSTF}, {APGHRA}, {LERLWL}, {PRERAS}, {PRERAS}}
```

Test binding

• EBV

- Strain GD1
 - <https://www.uniprot.org/proteomes/UP000274863>
 - Unreviewed
 - High completeness

• CMV

- Strain Merlin
- Strain AD169

+ Cosavirus A ?

Pos	MHC	Peptide	Core Of Gp Gl Ip Il	Icore	Identity	Score_El	#Rank_El	Score_BA	#Rank_BA	Aff(nM)	BindLevel
<hr/>											
<hr/>											
<hr/>											
933	HLA-B*08:01	VPRERASV VPRERA-SV	0 0 0 6 1	VPRERASV sp_015399_NMDE4	0.4988200	0.224	0.451084	0.494	379.61	< SB	
<hr/>											
<hr/>											
<hr/>											
75	HLA-B*08:01	PRERASGV -PRERASGV	0 0 0 0 1	PRERASGV sp_P16847_UL28_-0.0166820	5.915	0.169024	8.849	8851.61			
75	HLA-B*08:01	KPRERASG KPRER-ASG	0 0 0 5 1	KPRERASG sp_P16847_UL28_-0.0066390	10.011	0.158502	9.098	8998.58			
74	HLA-B*08:01	PKPRERAS -PKPRERAS	0 0 0 0 1	PKPRERAS sp_P16847_UL28_-0.0009160	25.719	0.057703	41.967	26780.90			
76	HLA-B*08:01	PRERASGV -PRERASGV	0 0 0 0 1	PRERASGV sp_P16847_UL28_-0.0166820	5.915	0.169024	8.849	8851.61			

Large virus db and re-search 2D SB

Virus Database re-try

Pos	MHC	Peptide	Core Of Gp Gl Ip Il	Icore	Identity	Score_El	#Rank_El	Score_BA	#Rank_BA	Aff(nM)	BindLevel
<hr/>											
<hr/>											
<hr/>											
82	HLA-B*08:01	LDVRPVAL -LDVRPVAL	0 0 0 0 1	LDVRPVAL sp_015399_NMDE4 0.1463950	1.129	0.186230	6.452	6666.24	< WB		
83	HLA-B*08:01	DVRPVAL DVRPVAL-V	0 0 0 7 1	DVRPVALV sp_015399_NMDE4 0.0202270	4.089	0.123601	14.333	13127.21	WB		
84	HLA-C*07:01	VRPVALVL VR-PVALVL	0 0 0 2 1	VRPVALVL sp_015399_NMDE4 0.1177950	0.267	0.344228	0.380	1206.29	< SB		

Rubella Virus

Pos	MHC	Peptide	Core Of Gp Gl Ip Il	Icore	Identity	Score_El	#Rank_El	Score_BA	#Rank_BA	Aff(nM)	BindLevel
<hr/>											
<hr/>											
<hr/>											
909	HLA-B*08:01	KTVRPVAL KT-VRPVAL	0 0 0 2 1	KTVRPVAL sp_P08563_2_POL 0.0887200	1.827	0.205753	5.193	5396.88	< WB		
910	HLA-B*08:01	TRPVALP TVR-PVALP	0 0 0 3 1	TRPVALP sp_P08563_2_POL 0.0005780	31.000	0.057416	42.204	26864.20	WB		
911	HLA-B*08:01	VRPVALPR VR-PVALPR	0 0 0 3 1	VRPVALPR sp_P08563_2_POL 0.0001800	47.545	0.0330301	69.306	34975.05	WB		
909	HLA-C*07:01	KTVRPVAL KTV-VRPVAL	0 0 0 3 1	KTVRPVAL sp_P08563_2_POL 0.0018840	5.643	0.090320	12.247	18815.54			
910	HLA-C*07:01	TRPVALP TVR-PVALP	0 0 0 3 1	TRPVALP sp_P08563_2_POL 0.0000100	50.000	0.030427	50.165	35974.48			
911	HLA-C*07:01	VRPVALPR VR-PVALPR	0 0 0 2 1	VRPVALPR sp_P08563_2_POL 0.0000990	9.549	0.065074	26.785	24278.02			
909	HLA-A*01:01	KTVRPVAL KT-VRPV-AL	0 0 0 6 1	KTVRPVAL sp_P08563_2_POL 0.0008240	20.930	0.035500	41.551	34053.10			
911	HLA-A*01:01	VRPVALPR VR-PVALPR	0 0 0 0 1	VRPVALPR sp_P08563_2_POL 0.0000420	74.231	0.015551	84.012	42256.79			
910	HLA-A*01:01	TRPVALP TVRPVA-LP	0 0 0 6 1	TRPVALP sp_P07566_1_POL 0.0001430	50.323	0.021583	69.517	39586.95			

Identical subsequence	Position	Allele	Epitope				Binding (WB-Weak bind)		Brain region (\$)	
VRPVAL	82	HLA-B*08:01	LDVRPVAL	NMDE4	2D	1.129	6.452	WB	SC >>	1.03
VRPVAL	909	HLA-B*08:01	KTVRPVAL	Rubella Virus		1.827	5.193	WB		0.41

Sequence	Seq_No.	Length	Hydropathy_Index
L D V R P V A L	10	8	1.03

I,V,L	Most Hydrophobic
F,C,M,A	More Hydrophobic
G,T,S,W,Y,P	
K,H,N,Q,D,E	More Hydrophilic
R	Most Hydrophilic

From https://www.peptide2.com/peptide_generator2.php

All subunits
GRIN1 & GRIN2A again have significantly less matches

SCZ protein list 01

- glia maturation factor beta (GMF-β)
- the brain-derived neurotrophic factor (BDNF)
- the 115-kDa isoform of the Rab3 GTPase-activating protein catalytic subunit (RAB3GAP1)
From <<https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00885/full>>

- Synaptophysin
From <<https://www.nature.com/articles/s41380-018-0041-5>>

- PSD-95
From <<https://www.sciencedirect.com/science/article/pii/S0278584617306589>>

CYFIP1

From <<https://www.nature.com/articles/s41467-019-11203-y>>

NCAM1

From <<https://www.sciencedirect.com/science/article/pii/S2666379122001148>>

Identical Epitope	Subunit Pos	HLA allele	epitope	Protein / Virus		% Rank_EL	% Rank_BA	Binding		Brain region (\$)	Hydropathy index
LERLWL	817	HLA-B*08:01	EMLERLW L	NMDE4	2D	0.198	0.256	SB	SC >>	0.11	
LERLWL	210	HLA-B*08:01	VILERLWL	cytomegalovirus (strain Merlin)		0.592	1.098	WB		1.3	
VFLTLY	688	HLA-A*01:01	TAVFLTLY	NMD3A	3A	1.778	1.37	WB			
VFLTLY	616	HLA-A*01:10	PVFLTLYY	herpesvirus 8 type P		2.331	2.085				
VFLTLY	616	HLA-A*01:01	PVFLTLYY	herpesvirus 8 type P		2.869	3.645				
RARARA	50	HLA-B*08:01	RARARAAL	NMD3B	3B	0.439	0.153	SB			
RARARA	691	HLA-B*08:01	RARARAP F	herpesvirus 8 type P		2.022	0.547				
AALHLT	583	HLA-B*08:01	AALHLTAL	NMD3B	3B	1.405	2.636	WB			
AALHLT	582	HLA-B*08:01	FAALHHTA	NMD3B	3B	4.86	2.685				
AALHLT	2145	HLA-C*07:01	RAAHLHTY	Saffold virus		1.438	1.465	WB			
AITSTL	880	HLA-B*08:01	FRAITSTL	sp_Q05586_NMDZ1	1	3.298	3.175				
AITSTL	880	HLA-C*07:01	FRAITSTL	sp_Q05586_NMDZ1	1	0.171	0.144	SB		0.78	
AITSTL	94	HLA-C*07:01	TYAITSTL	herpesvirus 2 (strain HG52)		1.731	2.783	WB		0.74	
DKSIHL	128	HLA-C*07:01	YSDKSIHL	NMDZ1	1	1.515	4.067	WB			
DKSIHL	128	HLA-A*01:01	YSDKSIHL	NMDZ1	1	0.765	1.023	WB			
DKSIHL	128	HLA-B*08:01	YSDKSIHL	NMDZ1	1	1.133	9.256	WB		-0.65	
DKSIHL	130	HLA-B*08:01	DKSIHLSF	NMDZ1	1	2.613	10.313				
DKSIHL	543	HLA-B*08:01	FLDKSIHL	Variola virus		0.305	2.098	SB		0.44	
DKSIHL	543	HLA-C*07:01	FLDKSIHL	Variola virus		2.335	5.103				
DKSIHL	543	HLA-A*01:01	FLDKSIHL	Variola virus		1.951	2.94	WB			
DKSIHL	543	HLA-A*01:10	FLDKSIHL	Variola virus		2.355	3.756				
QVHPRL	356	HLA-B*08:01	QVHPRLV V	NMDE1	2A	0.538	1.895	WB		0.45	
QVHPRL	355	HLA-C*07:01	YQVHPRL V	NMDE1	2A	4.944	5.372				
QVHPRL	32	HLA-B*08:01	QVHPRLVL	Chapare mammarenavirus		0.061	0.503	SB		0.4	
VHPRLV	356	HLA-B*08:01	QVHPRLV V	NMDE1	2A	0.538	1.895	WB	HIP >	0.45	
VHPRLV	357	HLA-B*08:01	VHPRLVVI	NMDE1	2A	3.899	9.256				
VHPRLV	355	HLA-C*07:01	YQVHPRLV	NMDE1	2A	4.944	5.372				
VHPRLV	32	HLA-B*08:01	QVHPRVLV	Chapare mammarenavirus		0.061	0.503	SB		0.4	
SDKGNL	936	HLA-A*01:01	SDKGNLM Y	NMDE1	2A	0.102	0.166	SB			
SDKGNL	935	HLA-A*01:01	VSDKGNLM	NMDE1	2A	1.2	1.738	WB			
SDKGNL	934	HLA-B*08:01	MVSDKGN L	NMDE1	2A	2.198	5.95				
SDKGNL	557	HLA-B*08:01	YLSDKGNL	Guanarito mammarenavirus		0.553	2.207	WB			
SDKGNL	558	HLA-A*01:01	LSDKGNLV	Guanarito mammarenavirus		2.489	2.001				
ILKKLA	461	HLA-B*08:01	ILKKLARV	NMDE3	2C	1.476	2.714	WB	CB >	0.73	
ILKKLA	79	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain B95-8)		0.842	1.754	WB		0.95	
ILKKLA	79	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain GD1)		0.842	1.754	WB		0.95	

Sequence	Seq_No.	Length	Hydropathy_Index
V I L E R L W L 1	8	1.4	
E M L E R L W L 2	8	0.11	
F R A I T S T L 3	8	0.78	
T Y A I T S T L 4	8	0.74	
Y S D K S I H L 5	8	-0.65	
F L D K S I H L 6	8	0.44	
Q V H P R L V V 7	8	0.45	
Q V H P R L V L 8	8	0.4	
I L K K L A R V 11	8	0.73	
I L K K L A Y F 12	8	0.95	
L P L E I Q P L 14	8	0.71	
D P L E I Q P L 15	8	-0.2	
D P L E I Q P L 16	8	-0.2	
L P L E I Q P L 17	8	0.71	
D P L E I Q P L 18	8	-0.2	
L P L E I Q P L 19	8	0.71	
D P L E I Q P L 20	8	-0.2	
L P L E I Q P L 21	8	0.71	
D P L E I Q P L 22	8	-0.2	

LEIQPL	58	HLA-B*08:01	LPLEIQPL	NMDE3	2C	1.34	2.589	WB	CB >	0.71
LEIQPL	570	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain B95-8)		1.47	8.982	WB		-0.2
LEIQPL	570	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain GD1)		1.47	8.982	WB		-0.2
PLEIQP	58	HLA-B*08:01	LPLEIQPL	NMDE3	2C	1.34	2.589	WB	CB >	0.71
PLEIQP	570	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain GD1)		1.47	8.982	WB		-0.2

\$ -

Distribution of GRIN2 A,B,C,D throughout rat brain:

<https://www.sciencedirect.com/science/article/pii/S0169328X9900100X>

2C EBV CB>CTX>BrStm>C>SC>HIP>OB>STR>RET>HYP
 2A Chapare mammarenavirus HIP>CTX>SC>C>CB>STR>BrStm>RET>OB>HYP
 2D CMV + Rubella SC>IC>HYP>BrStm>RET>HIP>STR>OB>CTX>CB

Proter plots are all extracellular except GRIN1 which is in cytoplasm - still none within transmembrane sequences.

Truncating mutations in *GRIN2C*

From <https://www.nature.com/articles/tb201152>

If the immune system is attacking the subunit would it mutate to survive?

Iran SarsCovid2 epitope reactivity paper:

<https://www.frontiersin.org/articles/10.3389/fimmu.2021.705772/full>

Distribution of GRIN2 A,B,C,D throughout rat brain:

<https://www.sciencedirect.com/science/article/pii/S0169328X9900100X>

Antigens binding to class I molecules are 8–10 amino acids long

From <<https://www.miltenyi-biotec.com/GB-en/products/bla-b7-b27-antibody-anti-human-reaffinity-reo176.html#gref>>

Identical Epitope	HLA allele	epitope	Protein / Virus	Binding	Brain region (\$)
VRPVAL	HLA-B*08:01	LDVRPVAL	GRIN2D	WB	SC >>
VRPVAL	HLA-B*08:01	KTVRPVAL	Rubella Virus	WB	
LERLWL	HLA-B*08:01	EMLERLWL	GRIN2D	SB	SC >>
LERLWL	HLA-B*08:01	VILERLWL	cytomegalovirus (strain Merlin)	WB	
AITSTL	HLA-C*07:01	FRAITSTL	GRIN1	SB	
AITSTL	HLA-C*07:01	TYAITSTL	herpesvirus 2 (strain HG52)	WB	
DKSIHL	HLA-A*01:01	YSDKSIHL	GRIN1	WB	
DKSIHL	HLA-B*08:01	YSDKSIHL	GRIN1	WB	
DKSIHL	HLA-B*08:01	FLDKSIHL	Variola virus	SB	
DKSIHL	HLA-A*01:01	FLDKSIHL	Variola virus	WB	
QVHPRL	HLA-B*08:01	QVHPRLV	GRIN2A	WB	
QVHPRL	HLA-B*08:01	QVHPRVLV	Chapare mammarenavirus	SB	
VHPRLV	HLA-B*08:01	QVHPRLV	GRIN2A	WB	HIP >
VHPRLV	HLA-B*08:01	QVHPRVLV	Chapare mammarenavirus	SB	
ILKKLA	HLA-B*08:01	ILKKLARV	GRIN2C	WB	CB >
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain B95-8)	WB	
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain GD1)	WB	
LEIQPL	HLA-B*08:01	LPLEIQPL	GRIN2C	WB	CB >
LEIQPL	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain B95-8)	WB	
LEIQPL	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain GD1)	WB	
PLEIQP	HLA-B*08:01	LPLEIQPL	GRIN2C	WB	CB >
PLEIQP	HLA-B*08:01	DPLEIQPL	Epstein Bar Virus (strain GD1)	WB	

Testing 100 HLA-B alleles close to HLA-B*08:01 - NMDAR and original virus database

~ 12 000 NMDAR peptides tested

~ 60 matches with common HLA

Visually they all seem to be some form of:

PPPQ

PAPAP

...

Slightly interesting, none of the GWAS peptides above were present.

Virus matches were pretty wide spread - similar to initial pattern matching

Predominantly Rotavirus

Sequence	Seq_No.	Length	Hydropathy_Index
V I L E R L W L 1	8	1.4	
E M L E R L W L 2	8	0.11	
F R A I T S T L 3	8	0.78	
T Y A I T S T L 4	8	0.74	
Y S D K S I H L 5	8	-0.65	
F L D K S I H L 6	8	0.44	
Q V H P R L V V 7	8	0.45	
Q V H P R L V V 8	8	0.4	
Q V H P R L V V 9	8	0.45	
Q V H P R L V V 10	8	0.4	
I L K K L A R V 11	8	0.73	
I L K K L A Y F 12	8	0.95	
I L K K L A Y F 13	8	0.95	
L P L E I Q P L 14	8	0.71	
D P L E I Q P L 15	8	-0.2	
D P L E I Q P L 16	8	-0.2	
L P L E I Q P L 17	8	0.71	
D P L E I Q P L 18	8	-0.2	
L P L E I Q P L 19	8	0.71	
D P L E I Q P L 20	8	-0.2	
L P L E I Q P L 21	8	0.71	
D P L E I Q P L 22	8	-0.2	

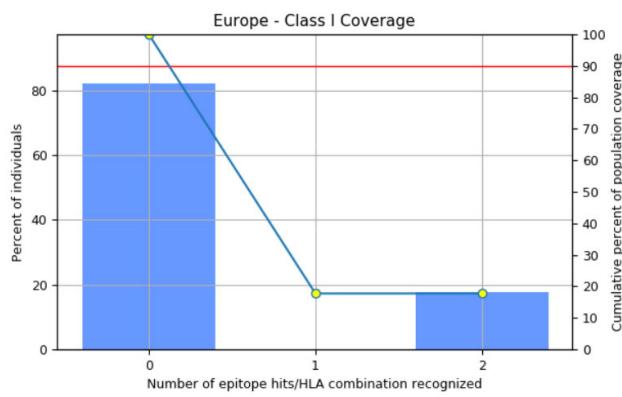
From <https://www.peptide2.com/peptide_generator2.php>

<https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#jobs::PEP-FOLD3.A13933215542078>



poster003

Identical Epitope	HLA allele	epitope	Protein / Virus	Binding	Brain region (\$)	Epitope cut score	Epitope cut (CTL)	EL score	Binding	Population coverage (Europe)		Average hit	pc90
										Coverage (%)			
VRPVAL	HLA-B*08:01	LD VRPVAL	GRIN2D	WB	SC >>	0.9941	DVRPVALVL	0.47411	SB	17.74	-	0.35	0.24
VRPVAL	HLA-B*08:01	KT VRPVAL	Rubella Virus	WB						-	-	-	-
LERLWL	HLA-B*08:01	EM LERLWL	GRIN2D	SB	SC >>	0.79	LERLWLSGI	0.002805	■				
LERLWL	HLA-B*08:01	V LERLWL	cytomegalovirus (strain Merlin)	WB									
AITSTL	HLA-C*07:01	FRA ITSTL	GRIN1	SB		1.0042	TFRAITSTL	0.02034	WB	25.11	-	0.5	0.27
AITSTL	HLA-C*07:01	TY AITSTL	herpesvirus 2 (strain HG52)	WB									
DKSIHL	HLA-A*01:01	YSD DKSIHL	GRIN1	WB						0.03	■	25.67	
DKSIHL	HLA-B*08:01	YSD DKSIHL	GRIN1	WB		1.1898	SDKSIHLSF	0.098	WB	17.74	-	-	
DKSIHL	HLA-B*08:01	F DKSIHL	Variola virus	SB									
DKSIHL	HLA-A*01:01	F DKSIHL	Variola virus	WB									
QVHPRL	HLA-B*08:01	QV HPRLVV	GRIN2A	WB		0.9933	QVHPRLVVI	0.703359	SB				
QVHPRL	HLA-B*08:01	QV HPRLVL	Chapare mammarenavirus	SB									
VHPRLV	HLA-B*08:01	QV HPRLVV	GRIN2A	WB	HIP >								
VHPRLV	HLA-B*08:01	QV HPRLVL	Chapare mammarenavirus	SB									
ILKKLA	HLA-B*08:01	ILKKLARV	GRIN2C	WB	CB >	1.5315	ILKKLARVV	0.275215	WB	17.74	-	0.35	0.24
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain B95-8)	WB									
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain GD1)	WB									
LEIQPL	HLA-B*08:01	L LEIQPL	GRIN2C	WB	CB >								
LEIQPL	HLA-B*08:01	D LEIQPL	Epstein Bar Virus (strain B95-8)	WB									
LEIQPL	HLA-B*08:01	D LEIQPL	Epstein Bar Virus (strain GD1)	WB									
PLEIQP	HLA-B*08:01	L PLEIQP	GRIN2C	WB	CB >	0.3853	DL PLEIQP	0.131063	WB	17.74	-	-	-
PLEIQP	HLA-B*08:01	D PLEIQP	Epstein Bar Virus (strain GD1)	WB									

<http://tools.iedb.org/main/>

T Cell:

<http://tools.iedb.org/main/tcell/><http://tools.iedb.org/mhci/result/>
<http://tools.iedb.org/netchop/table/>

B Cell:

<http://tools.iedb.org/main/bcell/><http://tools.iedb.org/ellipro/result/predict/>

Pipeline

Protein

NetCTL(Protein)

NetMHCi(% , length = 8, 9)

Partition(% (Strong and weak binders) , length = 6)

Virus

Matches(%)

NetCTL(% (Matching virus proteins))

NetMHCi(% , length = 8, 9)

Identical Epitope	HLA allele	epitope	Protein / Virus	Binding	Brain region (\$)	Epitope cut score	Epitope cut (CTL)	EL score	Binding	Population coverage (Europe)		Average hit	pc90
										Coverage (%)			
VRPVAL	HLA-B*08:01	LDVRPVAL	GRIN2D	WB	SC >>	0.9941	DVRPVALVL	0.47411	SB	17.74	-	0.35	0.24
VRPVAL	HLA-B*08:01	KTVRPVAL	Rubella Virus	WB						-	-	-	-
AITSTL	HLA-C*07:01	FRAITSTL	GRIN1	SB		1.0042	TFRAITSTL	0.02034	WB	25.11	-	0.5	0.27
AITSTL	HLA-C*07:01	TYAITSTL	herpesvirus 2 (strain HG52)	WB						-	-	-	-
DKSIHL	HLA-B*08:01	YSDKSIHL	GRIN1	WB		1.1898	SDKSIHLSF	0.098	WB	17.74	-	-	-
DKSIHL	HLA-B*08:01	FLDKSIHL	Variola virus	SB						-	-	-	-
QVHPRL	HLA-B*08:01	QVHPRLVV	GRIN2A	WB	HIP >	0.9933	QVHPRLVVI	0.703359	SB	17.74	-	-	-
QVHPRL	HLA-B*08:01	QVHPRLVL	Chapare mammarenavirus	SB						-	-	-	-
VHPRLV	HLA-B*08:01	QVHPRLVV	GRIN2A	WB	HIP >	0.9933	QVHPRLVVI	0.703359	SB	17.74	-	-	-
VHPRLV	HLA-B*08:01	QVHPRLVL	Chapare mammarenavirus	SB						-	-	-	-
ILKKLA	HLA-B*08:01	ILKKLARV	GRIN2C	WB	CB >	1.5315	ILKKLARVV	0.275215	WB	17.74	-	0.35	0.24
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain B95-8)	WB						-	-	-	-
ILKKLA	HLA-B*08:01	ILKKLAYF	Epstein Bar Virus (strain GD1)	WB						-	-	-	-

Blast+ user manual

<https://www.ncbi.nlm.nih.gov/books/NBK279690/>

Class I Immunogenicity

Allele: HLA-B0801

Masking: custom

Masked variables: [2, 5, 9]

Peptide	Length	Score
DVRPVALVL	9	0.05723
QVHPRLVVI	9	0.04786
DLPLEIQPL	9	0.00628
TFRAITSTL	9	-0.02423
LERLWLSGI	9	-0.12462
ILKKLARVV	9	-0.18237
SDKSIHLSF	9	-0.31204

Class I Immunogenicity

Masking: default

Masked variables: [1, 2, 'cterm']

Peptide	Length	Score
TFRAITSTL	9	0.10537
DLPLEIQPL	9	0.10378
QVHPRLVVI	9	0.09826
DVRPVALVL	9	0.09743
LERLWLSGI	9	0.09108
SDKSIHLSF	9	-0.18244
ILKKLARVV	9	-0.19317

Class I Immunogenicity

Masking: default

Masked variables: [1, 2, 'cterm']

Peptide	Length	Score
ILKKLAYF	8	-0.26409
ILKKLAYF	8	-0.26409
YSDKSIHL	8	-0.21832
FLDKSIHL	8	-0.21832
ILKKLARV	8	-0.21729
FRAITSTL	8	0.06145
TYAITSTL	8	0.06145
QVHPRLVV	8	0.07414
QVHPRLVL	8	0.07414
QVHPRLVV	8	0.07414

QVHPRILVL	8	0.07414
LDVRPVAL	8	0.12656
KTVRPVAL	8	0.12656

From <<http://tools.iedb.org/immunogenicity/result/>>

----- Searching subunits Genome Position on SZDB for variation at epitopes position?

Gene	Ensembl	Genome Position	logFCs	Avg Expression	T Statistic	P-value	FDRd	B coefficient
GRIN1	ENSG00000176884	chr9:140032842-140063207	-0.028	7.457	-0.584	5.59e-01	7.74e-01	-6.145
GRIN2A	ENSG00000183454	chr16:9852376-10276611	0.066	8.246	1.723	8.54e-02	3.06e-01	-4.868
GRIN2B	ENSG00000150086	-	0.081	7.079	2.160	3.12e-02	1.85e-01	-4.061
GRIN2C	ENSG00000161509	chr17:72838162-72857627	-0.097	4.119	-1.755	7.97e-02	2.96e-01	-4.676
GRIN2D	ENSG00000105464	chr19:48898132-48948188	-0.055	1.410	-1.119	2.63e-01	5.33e-01	-5.176
GRIN3A	ENSG00000198785	chr9:104331635-104500862	0.104	4.783	3.208	1.41e-03	4.65e-02	-1.312

From <<http://www.szdb.org/cmc2.php>>

----- Neuropeptides

One match of 7 identical AA - Cut match - Binding match (check)

NPID	NP04921		
Name	ProSAAS		
Organism	Homo sapiens		
NCBI Taxa	9606		
ID			
Tissue	Expressed in brain and pancreas.		
Specificity			
Family	ProSAAS		
UniProt ID	PCSK1_HUMAN		
Length	227		
Modification	NA		
Gene Ontology	GO ID	GO Term	Definition
	GO:0005615	Cellular Component	extracellular space
	GO:0005794	Cellular Component	Golgi apparatus
	GO:0030141	Cellular Component	secretory granule
	GO:0005802	Cellular Component	trans-Golgi network
	GO:0004866	Molecular Function	endopeptidase inhibitor activity
	GO:0005102	Molecular Function	receptor binding
	GO:0004867	Molecular Function	serine-type endopeptidase inhibitor activity
	GO:0010951	Biological Process	negative regulation of endopeptidase activity
	GO:0007218	Biological Process	neuropeptide signaling pathway
	GO:0016486	Biological Process	peptide hormone processing
	GO:0009409	Biological Process	response to cold
	GO:0002021	Biological Process	response to dietary excess
Sequence	ARPVKEPRL _[10] SAASPLAET _[20] GAPRRFRRSV _[30] PRGEAAGAVQ _[40] EL _[41] ALARALAH _[50] L _[51] EAERQERARA _[60] EAQEAEQQ _[70] RVLAQLLRVW _[80] GAPRNSDPAL _[90] GLDDDPDAPA _[100] AQLARALLRA _[110] RLDPAA _[120] AAQ _[120] LVPAPVPAAA _[130] LRPRPPVYDD _[140] GPAGPDAEEA _[150] GDETPDVDP _[160] LLRYLLGRIL _[170] AGSADSEGV _[180] APRRLRRAAD _[190] HDVGSELPP _[200] GVLGALLRVK _[210] RLETPAPQVP _[220] ARRLLPP		

From <http://isyslab.info/NeuroPep/search_info?pepNum=NP04921>

B cell analysis

No.	Chain	Start	End	Peptide	Number of residues	Score	3D structure
1	A	852	950	LLVAMGLSLLVFAWEHLVYWRRLHCLGPTHMRDFLLAFSRGMYSCCSAEAAPPAKPPPPPQPLPSPAYPAPRPAPGPAPFVPRERASVDRWRTKGAG	99	0.844	View
2	A	1273	1334	EDLSSCPRAAPARRLTGPSRHRRCPHAAHWGPLPTASHRRHRGGDLGTRRGSAHFSSLES	62	0.823	View
3	A	1171	1269	PRSGPAWHCRHCASLELLPPPRHLSCHDGLDGGWWAPPPWAAGPLPQQARCGCPRSHPRPRASHRTPAAAAPHHHRRRAAGGWDLPPPAPTS	99	0.768	View
4	A	1017	1108	AEPPAGAFPGFPSPAPPAAAATVGPPLCRLAFADESPAPARWPRSDPESQPLLPGAGGGAGGTGGAGGGAPAAPPCRAAPPPCPYLDL	92	0.761	View
5	A	976	1001	GLGEARAAPRGAAGRPLSPPAACQQPQ	26	0.735	View
6	A	587	668	VMMFVMCLTVAVTVFIFELSPVGYNRSLATGKRPGGSTFTIGKSIWLLWALVFNNNSPVENPRGTTSKIMVLVWAFFAVI	82	0.73	View
7	A	1	117	MRGAGGPRGPRGPAKMLLLALACASPPEEAPPGGAGGGGGLGGARPLNVALFSGPAYAAEAARLGPAAAVRSPGLD VRPVAL VLNGSDPRSLVQLCDLLSGLRVHGVVF	117	0.711	View
8	A	327	365	VARGAQALLRDYGFLPELGHDCRAQRTHRGESELHYFM	39	0.664	View
9	A	120	141	DSRAPAVAPILDLFLSAQTSLPI	22	0.56	View
10	A	223	235	LDPGAGEAVLSAQ	13	0.555	View
11	A	155	160	KEKGST	6	0.545	View
12	A	1120	1127	SLGGASLG	8	0.538	View
13	A	294	298	LPGGA	5	0.51	View

Identical Epitope	HLA allele	Protein / Virus	Binding	Epitope cut score	Epitope cut (CTL)	EL score	Binding
VRPVAL	HLA-B*08:01	GRIN2D	WB	0.9941	DVRPVALVL	0.47411	SB
VRPVAL	HLA-B*08:01	Rubella Virus	WB				
AITSTL	HLA-C*07:01	GRIN1	SB	1.0042	TFRAITSTL	0.02034	WB
AITSTL	HLA-C*07:01	herpesvirus 2 (strain HG52)	WB				
DKSIHL	HLA-B*08:01	GRIN1	WB	1.1898	SDKSIHLSF	0.098	WB
DKSIHL	HLA-B*08:01	Variola virus	SB				
QVHPRL	HLA-B*08:01	GRIN2A	WB	0.9933	QVHPRLVVI	0.703359	SB
QVHPRL	HLA-B*08:01	Chapare mammarenavirus	SB				
VHPRLV	HLA-B*08:01	GRIN2A	WB	0.9933	QVHPRLVVI	0.703359	SB
VHPRLV	HLA-B*08:01	Chapare mammarenavirus	SB				
ILKKLA	HLA-B*08:01	GRIN2C	WB	1.5315	ILKKLARVV	0.275215	WB
ILKKLA	HLA-B*08:01	Epstein Bar Virus (strain B95-8)	WB				
ILKKLA	HLA-B*08:01	Epstein Bar Virus (strain GD1)	WB				