|  |  |  |
| --- | --- | --- |
| **MHC I Pathway**   * All nucleated cells * 8-10mers (closed groove as bonds are made with C- & N-termini hence less sensitive to AA sequences in core of oligomer), >95% self-peptides (surveys interior) * Conjugation in the endoplasmic reticulum * Tapasin transports high-affinity pMHC-Is to cell surface * Immunogenicity prediction   + [IEDB T-cell class I tool](https://nextgen-tools.iedb.org/pipeline) seems like the best option: integrates all steps + API | | |
| **Domain (Input)** | **Tool/Resource** | **Notes** |
| Pathogen proteomes (n/a) | [ViralZone](https://viralzone.expasy.org/678) | * Manually curated, links to fasta files on Uniprot which are annotated using experimental data (~4/5 scores) * 129 common human viruses including adenovirus, rotavirus, coxsackie etc |
| RefSeq | * 52M viral proteins, variable evidence (eg some hypothetical, predicted from other baltimore groups) |
| Uniprot | * ~110,000 viral proteins with evidence of existence at protein level |
| Immunoproteosome cleavage prediction (foreign proteomes)   * Unlike ER, cleaves AAs at C-terminus | [NetChop – 3.1](https://services.healthtech.dtu.dk/services/NetChop-3.1/). no API, runs locally as a linux program (need to use BASH). IEDB-recommended | * NetChop is a predictor of proteasomal processing based upon a neural network. NetCTL is a predictor of T cell epitopes along a protein sequence. It also employs a neural network architecture. NetCTLpan is an update to the original NetCTL server that allows for prediction of CTL epitope with restriction to any MHC molecules of known protein sequence. * each AA given binary indicator of whether or not cleavage occurs immediately afterward (based on user-determined threshold, default = 0.5, increasing this increases specificity and decreases sensitivity) * Does not model PCPS but spliced peptides minimally increase the immunoproteome (<5%) * The Netchop 3.0 version has two different network methods that can be used for prediction. C-term 3.0 and 20S 3.0. C-term 3.0 network is trained with a database consisting of 1260 publicly available MHC class I ligands (using only C-terminal cleavage site of the ligands). 20S network is trained with in vitro degradation data published in Toes, et al. and Emmerich et al. C-term 3.0 network performs best in predicting the boundaries of CTL epitopes. |
| [iPCPS](http://imed.med.ucm.es/Tools/pcps/). (webserver only) | * each AA given binary indicator of whether or not cleavage occurs immediately afterward (based on user-determined threshold) * Does not model PCPS but spliced peptides minimally increase the immunoproteome (<5%) * immunoproteasome trained on 553 CD8 T cell epitopes and their flanking regions. Three models respectively based on training sets of 12, 8, 6 -meric oligopeptides, models se=90,91,76%, sp=41,54,71% * 844 unique virus-specific CD8+ T cell epitopes and their source proteins from IEDB: PCPS was clearly superior to NetChop in term of sensitivity (0.89 vs. 0.79) but somewhat inferior with regard to specificity (0.55 vs. 0.60). Judging by the Mathew’s Correlation Coefficient, PCPS predictions were overall superior to those provided by NetChop (0.46 vs. 0.39). We tuned the PCPS web server to predict CD8+ T cell epitopes and predicted the entire SARS-CoV-2 epitope space1. |
| [PCleavage](http://www.imtech.res.in/raghava/pcleavage/) | * ROC AUC for NetChop > Pcleavage, MCC 0.54 and 0.43 on in vitro and major histocompatibility complex ligand data |
| input into MHC class I molecule epitope binding prediction model/database (oligopeptides)   * Initially HLA-allele agnostic | [NetMHCpan EL - 4.1](https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/) IEDB recommended | * parameters: input \*.fasta, peptide length, HLA alleles (A, B, C, or HLA supertype), can set a threshold value to filter outputs * Affinity given as IC50 (in nM) value for each 9mer, sequentially progressing the reading frame by 1AA. this is done for each user-determined HLA allele * Training dataset had 85,217 entries in total, with ligand length ranging from 8 to 15 aa. The ligands originated from 14,797 source Ags and were restricted by 55 unique HLA molecules2 |
| NetMHCpan BA 4.1 | * Trained using BA only, not MS |
| [ViPR](https://www.bv-brc.org/): predicted/confirmed BCR/TCR epitopes. Both viruses and bacteria, continually maintained. | * Could potentially also use IEDB data which is similarly labelled |
| input into COBALT (or other BLAST tool) | Aligned sequences |  |
| Has entire process been done before?   * <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3808449/> | | |

|  |  |  |
| --- | --- | --- |
| **MHC II Pathway**   * Antigen-presenting cells only * 13-25mers (open binding groove .: protein can slide and bind non-canonically, much more sensitive to AAs in oligomer core P1, 4, 6, 9), extracellular/foreign peptides (surveys exterior). More difficult to predict compared to MHC-I as binding more promiscuous * Conjugation in the late endolysosome * HLA-DM transports high-affinity pMHC-IIs to cell surface | | |
| **Domain (Input)** | **Tool/Resource** | **Notes** |
| Pathogen proteomes (n/a) | [ViralZone](https://viralzone.expasy.org/678) | * Manually curated, links to fasta files on Uniprot which are annotated using experimental data (~4/5 scores) * 129 common human viruses including adenovirus, rotavirus, coxsackie etc |
| RefSeq | * 52M viral proteins, variable evidence (eg some hypothetical, predicted from other baltimore groups) |
| Uniprot | * ~110,000 viral proteins with evidence of existence at protein level |
| Endolysosome cleavage (protein) | Can use immuno-proteasome cleavage predictions as constitutive expression by all cells during inflammation under the control of IFN-gamma [citation] |  |
| input into MHC class II molecule epitope binding prediction model/database (oligopeptides) | [NetMHCIIpan 4.0](https://services.healthtech.dtu.dk/services/NetMHCIIpan-4.0/) | * parameters: input \*.fasta, peptide length, HLA alleles (25 HLA-DR alleles, 20 HLA-DQ, 9 HLA-DP), can set a threshold value to travel along ROC curve * affinity given as IC50 (in nM) value for each 9mer, sequentially progressing the reading frame by 1AA. this is done for each user-determined HLA allele |
| MixMHC2pred2 | * 627,013 unique MHC-II ligands identified by mass spectrometry * Outperforms NetMHCIIpan overall (ROC AUC), especially when considering non-canonical binding of epitopes, which more closely match peptidomic data * Kullback-Leibler divergence (KLD) much higher for NetMHCIIpan vs MixMHC2pred against reference proteomics (in vitro) data |
| others: MARIA (major histocompatibility complex analysis with recurrent integrated architecture), MHCnuggets |  |
| input into COBALT (or other BLAST tool) | Aligned sequences |  |

Scenarios? (not mutually exclusive)

* Cell-mediated
  + Neuronal MHC-I presentation of self-peptide (may/not be a cell-surface protein)
    - Leads to cytotoxicity due to an already autoreactive CD8+ cell due to prior infection by mimicking pathogen
    - Primes a CD8+ cell to become autoreactive, needs:
      * Sufficient antigen presentation, often only achieved by APCs, or
      * The necessary cytokines (?) in an inflammatory milieu to obviate CD4+ help, or
      * CD4+ help (requires antecedent MHC-II presentation)
* Humoral

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Positive Controls** | | | | |
| **Disease** | **Infective agent & protein** | **HLA class II associations** | **Self protein** | **Common sequence (if known)** |
| **T1DM** | Coxsackie B4 P2C nonstructural protein (PEVKEK AA sequence)3 | (in format DRB1–DQA1–DQB1) **Risk**: 04:05-03:01-03:02, 04:01-03:01-03:02, 03:01-05:01-02:01, 04:02-03:01-03:02. For fine-mapping see ref. 4 implicating DQB1 (incl. 02:01, 02:02, 03:02, 03:04, 03:05) & DRB1 (incl. 04:01, 04:09) alleles. Binding to DRB1\*03:01 observed in ref. 5.  **Protective**:  13:03-05:01-03:01, 11:04-05:01-03:01, 15:01-01:02-06:02,  07:01-02:01-03:03,  14:01-01:01-05:03 see ref. 6, 7. | PEVKEK/PEVKTK sequences (GAD65/67 respectively)8 | PEVKXK |
| Enteroviruses, conserved VP1 protein (ALTAV sequence)9 | Receptor-type tyrosine-protein phosphatase-like N9 | ALTAV |
| Rotavirus VP7 (serotype G3, human strain P) | tyrosine phosphatase-like insulinoma Ag 2 (IA2), GAD6510 | Longest contiguous sequence is 3aa but region of similarity aa16-49, see ref. 10 |
| **MS** | EBV nuclear antigen 1 (EBNA1)11 | **Risk:** HLA-DRB1\*15:01 (primary risk allele)12, supported by fine mapping13 which also implicates \*03:01, \*13:03, \*04:04, \*04:01, and \*14:01.  Protective: DRB1\*01 (01:01), DRB1\*1114 | Myelin basic protein (MBP), anoctamin 2 (ANO2), α-crystallin B chain (CRYAB), β-synuclein (β-SYN), glial cell adhesion molecule (GCAM) (great review11) |  |
| EBNA1 aa386-405 | glial cell adhesion molecule (GlialCAM) aa370-389 | PPRXP (see fig. 3 and results subsection ‘GlialCAM phosphorylation enables mimicry’ in ref. 15) |
| EBNA1 aa411-426 | Myelin basic protein (MBP) aa85-9916,17 |  |
| EBNA1 aa431-440 | anoctamin 2 (ANO2) aa140-149 | PGXIEXGP (see fig. 2 ref. 18) |
| EBNA1 aa392-411 | α-crystallin B chain (CRYAB)19 |  |
| **Coeliac Disease** | Gliadin three-way mimicry (adenovirus type 12, HCV, Campylobacter jejuni, Giardia lamblia, enterovirus, rotavirus all implicated)20 | DQ2.5, DQ2, DQ821 | type 2 tTG (EC 2.3.2.13) (8 isoenzymes named based on location), the type present in the skin, which is the target of autoantibodies in dermatitis herpetiformis = type 3 tTG; type 6 tTG, which has been identified as the target of autoantibodies in patients with diseases of the central nervous system (e.g., ataxia). Elio Tonutti, Nicola Bizzaro, in Autoantibodies (Third Edition), 2014 |  |
| Interesting hypothesis (gliadin fundamental to complex formation potentiating TH activation of autoreactive BL22) | |  | |
| **Narcolepsy-cataplexy (Narcolepsy type 1)** | Influenza A/H1N1 AS03 adjuvanted vaccine23 (Pandemrix), specifically peptides within the A/California/7/2009 H1N1 (A/H1N1pdm09) virus antigen: haemagglutinin (HA) aa275-287/neuraminidase (NA)24 | **Risk:** DQB1\*06:0225  Protective:  DQB1\*06:03, DQB1\*05:01, DQB1\*06:09 DQB1\*0226 | hypocretin (HCRT)1 aa56-68 and HCRT2 aa87-99 | CASSQETQGRNYGYTF  (study difficult to understand27, explainer of method [here](https://www.mblbio.com/bio/g/product/allergy-Immunology/pickup/mhctetramer-feature.html)) |
| **Rheumatic Fever** | *S. pyogenes* (GABHS) m-protein | Class II in general but no specific allele28 | Type 4 collagen |  |
| a streptococcal M protein N-terminus domain has been shown to bind to the CB3 region in collagen type IV. This binding seems to initiate an antibody response to the collagen and result in ground substance inflammation. These antibodies do not cross-react with M proteins, and we believe that no failure of immune system and, possibly, no molecular mimicry occur in rheumatic fever. This alternative hypothesis shares similarity with collagen involvement in both Goodpasture syndrome and Alport syndrome29 | |  | |

<https://www.nature.com/articles/s41569-019-0258-2>

**Preliminary Results (Netchop-epitopes vs. selected proteins)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | T1DM | | |
|  | DR/DP/DQ Molecule Haplotype | Number of 5AA matches against viral dataset | |
| Targeted Proteins | Control Proteins |
| Risk | "DQA1\_05\_01\_\_DQB1\_02\_01" |  |  |
| "DQA1\_03\_01\_\_DQB1\_03\_02" | AVVNN (G5EGJ5 IDA-1/P10190 – HHV-1) |  |
| "DRB1\_04\_01" |  | APANY (Q9NQA5 TRPV-5/O56278 – Human betaherpes virus 7) |
| "DRB1\_04\_09" | EHLRD (G5EGJ5 IDA-1/P10211 – HHV-1) |  |
| Protective | "DRB1\_13\_03", |  |  |
| "DRB1\_15\_01", |  |  |
| "DQA1\_01\_01\_\_DQB1\_05\_03", |  |  |
| "DQA1\_02\_01\_\_DQB1\_03\_03" |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | MS | | |
|  | DR/DP/DQ Molecule Haplotype | Number of 5AA matches against viral dataset | |
| Targeted Proteins | Control Proteins |
| Risk | "DRB1\_15\_01", |  |  |
| "DRB1\_03\_01", |  |  |
| "DRB1\_13\_03", |  |  |
| "DRB1\_04\_01", |  | EDWSD (P08617 Hep A genome polyprotein/ Q6ZP01 RMB44) |
| "DRB1\_04\_09", |  |  |
| "DRB1\_14\_01", | TKPSY (P03520 P Vesicular stomatitis Indiana virus /Q14CZ8 GlialCAM) |  |
| Protective | "DRB1\_01\_01", |  |  |
| "DRB1\_11\_01", |  |  |
| "DRB1\_11\_02", |  |  |
| "DRB1\_11\_03", |  |  |
| "DRB1\_11\_04", |  | FKNGV (A0A058Z – small HSP/ P04217 A1BG) |
| "DRB1\_11\_05" |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | NT1 | | |
|  | DR/DP/DQ Molecule Haplotype | Number of 5AA matches against viral dataset | |
| Targeted Proteins | Control Proteins |
| Risk | "DQA1\_01\_01\_\_DQB1\_06\_02" |  | EEEEE (P0C6U8 SARS-CoV replicase /P13727) |
| Protective | DQA1\_01\_01\_\_DQB1\_05\_01 |  |  |

1. Gomez-Perosanz, M., Ras-Carmona, A., Lafuente, E. M. & Reche, P. A. Identification of CD8+ T cell epitopes through proteasome cleavage site predictions. *BMC Bioinformatics* **21**, 484 (2020).

2. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology* **199**, 3360–3368 (2017).

3. Dotta, F. *et al.* Coxsackie B4 virus infection of β cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. *Proceedings of the National Academy of Sciences* **104**, 5115–5120 (2007).

4. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* **47**, 898–905 (2015).

5. Vreugdenhil, G. R. *et al.* Molecular mimicry in diabetes mellitus: the homologous domain in coxsackie B virus protein 2C and islet autoantigen GAD65 is highly conserved in the coxsackie B-like enteroviruses and binds to the diabetes associated HLA-DR3 molecule. *Diabetologia* **41**, 40–46 (1998).

6. Noble, J. A. & Valdes, A. M. Genetics of the HLA Region in the Prediction of Type 1 Diabetes. *Curr Diab Rep* **11**, 533–542 (2011).

7. Erlich, H. *et al.* HLA DR-DQ Haplotypes and Genotypes and Type 1 Diabetes Risk: Analysis of the Type 1 Diabetes Genetics Consortium Families. *Diabetes* **57**, 1084–1092 (2008).

8. Kaufman, D. L. *et al.* Autoimmunity to two forms of glutamate decarboxylase in insulin-dependent diabetes mellitus. *J Clin Invest* **89**, 283–292 (1992).

9. Härkönen, T., Lankinen, H., Davydova, B., Hovi, T. & Roivainen, M. Enterovirus infection can induce immune responses that cross-react with β-cell autoantigen tyrosine phosphatase IA-2/IAR. *Journal of Medical Virology* **66**, 340–350 (2002).

10. Honeyman, M. C., Stone, N. L., Falk, B. A., Nepom, G. & Harrison, L. C. Evidence for Molecular Mimicry between Human T Cell Epitopes in Rotavirus and Pancreatic Islet Autoantigens. *The Journal of Immunology* **184**, 2204–2210 (2010).

11. Soldan, S. S. & Lieberman, P. M. Epstein–Barr virus and multiple sclerosis. *Nat Rev Microbiol* **21**, 51–64 (2023).

12. Cook, S. Multiple sclerosis and viruses. *Mult Scler* **3**, 388–389 (1997).

13. Patsopoulos, N. A. *et al.* Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLOS Genetics* **9**, e1003926 (2013).

14. Mamedov, A. *et al.* Protective Allele for Multiple Sclerosis HLA-DRB1\*01:01 Provides Kinetic Discrimination of Myelin and Exogenous Antigenic Peptides. *Frontiers in Immunology* **10**, (2020).

15. Lanz, T. V. *et al.* Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* **603**, 321–327 (2022).

16. Wucherpfennig, K. W. & Strominger, J. L. Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80**, 695–705 (1995).

17. Lünemann, J. D. *et al.* EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-γ and IL-2. *Journal of Experimental Medicine* **205**, 1763–1773 (2008).

18. Tengvall, K. *et al.* Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proceedings of the National Academy of Sciences* **116**, 16955–16960 (2019).

19. Hecker, M. *et al.* High-Density Peptide Microarray Analysis of IgG Autoantibody Reactivities in Serum and Cerebrospinal Fluid of Multiple Sclerosis Patients \*. *Molecular & Cellular Proteomics* **15**, 1360–1380 (2016).

20. Vazquez, D. S., Schilbert, H. M. & Dodero, V. I. Molecular and Structural Parallels between Gluten Pathogenic Peptides and Bacterial-Derived Proteins by Bioinformatics Analysis. *Int J Mol Sci* **22**, 9278 (2021).

21. Martina, S. *et al.* Genetic susceptibilty and celiac disease: what role do HLA haplotypes play? *Acta Biomed* **89**, 17–21 (2018).

22. Sollid, L. M., Molberg, Ø., McADAM, S. & Lundin, K. E. A. Autoantibodies in coeliac disease: tissue transglutaminase—guilt by association? *Gut* **41**, 851–852 (1997).

23. Boquete Castro, A. & Cortés Mejía, J. M. Adjuvanted AH1N1 influenza vaccine precipitating the appearance of narcolepsy. *Vacunas (English Edition)* **22**, 119–124 (2021).

24. Buonocore, S. M. & van der Most, R. G. Narcolepsy and H1N1 influenza immunology a decade later: What have we learned? *Front Immunol* **13**, 902840 (2022).

25. Mignot, E. *et al.* Complex HLA-DR and -DQ Interactions Confer Risk of Narcolepsy-Cataplexy in Three Ethnic Groups. *Am J Hum Genet* **68**, 686–699 (2001).

26. Tafti, M. *et al.* DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep* **37**, 19–25 (2014).

27. Luo, G. *et al.* Autoimmunity to hypocretin and molecular mimicry to flu in type 1 narcolepsy. *Proc Natl Acad Sci U S A* **115**, E12323–E12332 (2018).

28. Bryant, P. A., Robins-Browne, R., Carapetis, J. R. & Curtis, N. Some of the People, Some of the Time. *Circulation* **119**, 742–753 (2009).

29. Tandon, R. *et al.* Revisiting the pathogenesis of rheumatic fever and carditis. *Nat Rev Cardiol* **10**, 171–177 (2013).