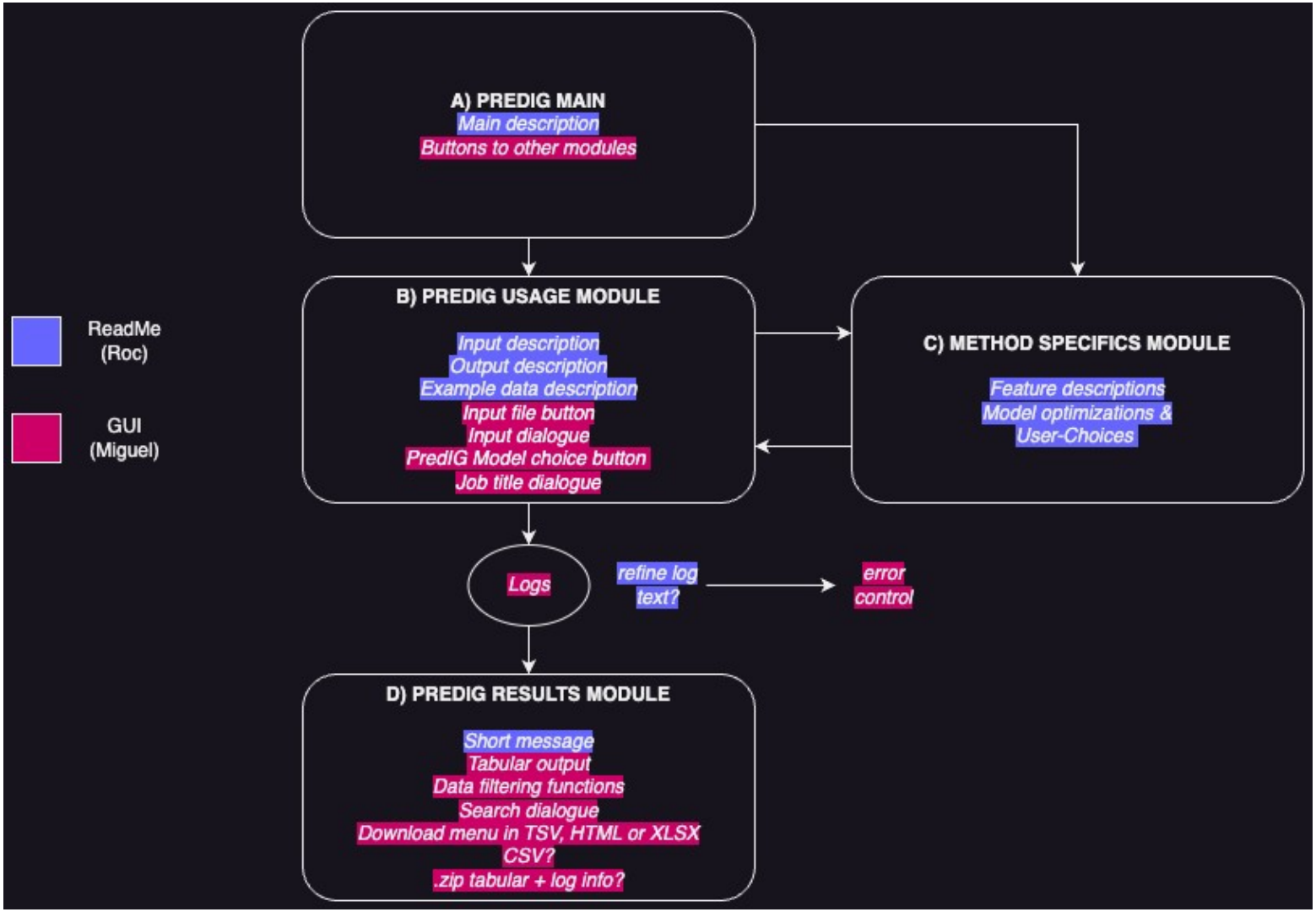
PredIG Server - Structure & Readmes



A) Main

PredIG is a Predictor of CD8+ T-cell epitope ImmunoGenicity given paired peptides and HLA-I alleles.

Briefly, PredIG score consists of a probability from 0 to 1, being 1 the maximum likelihood for epitope immunogenicity. This score can be used to rank candidates for prioritization approaches or for classification using adaptable thresholds.

Please, refer to the Usage module for specifications of input format and data requirements; refer to Method Specifics module for a detailed description on feature space, model optimizations and user-choices.

B) Usage Module

Input

PredIG requires as data points pairs of epitope and presenting HLA-I allele.

For the file input, data points should be provided in a CSV with two columns named "Epitope" and "HLA_allele". CSV should be separated by commas.

For the dialogue input, upload a list of epitope and HLA-I alleles in two separated columns entitled Epitope and HLA_allele, with commas as separators.

A job title can be assigned in the specific box.

Please, load the example data for further clarifications.

Output
PredIG provides flexible tabular outputs including HTML, TSV and EXCEL.

The HTML output page can also be filtered and searched upon.

The results include the 14 features predicted within PredIG as well as an immunogenicity likelihood score termed PredIG score.

Briefly, PredIG score consists of a probability from 0 to 1, being 1 the maximum likelihood for epitope immunogenicity. This score can be used to rank candidates for prioritization approaches or to classify them using adaptable thresholds.

Please refer to Method Specifics section for a detailed description on feature space, model optimizations and user-choices.

C) Method Specifics Module

PredIG is a Predictor of CD8+ T-cell epitope ImmunoGenicity trained on the largest T-cell assay dataset including reactivity (cytokine release assay) and binding (MHC-tetramer) experiments.

As data points, the model uses individual pairs of epitope and presenting HLA-I allele, labelled as immunogenically positive or negative according to its ground truth experimental source. The presenting HLA-I allele was refined from different HLA-ontologies and prediction sources.

Epitope and HLA-I allele are used to build a feature space using bioinformatic predictors that comprises 7 antigenic predictions covering antigen processing, transport and presentation and 7 physicochemical characteristics describing the entire epitope and its central residues specifically.

PredIG takes advantage of a feature space that comprehensively describes the epitope pathway to T-cell recognition and uses XGBoost and explainable AI techniques to deliver interpretable predictions. A careful assessment of the imbalanced nature of T-cell assay data (ie much more expected negatives than positive cases) is built in the algorithm to precisely adapt to cancer and pathogen data.

Briefly, PredIG score consists of a probability from 0 to 1, being 1 the maximum likelihood for epitope immunogenicity. This score can be used to rank candidates for prioritization approaches or to classify them using adaptable thresholds.

Feature table and descriptions

Feature Label	Feature Name	Prediction Source	Predicted Process	Scoring Range (Interpretation)	Reference
NOAH_score	NOAH	NOAH	HLA-I peptide binding (structural)	Likelihood for binding probability from negative to positive, being negative best. <-1 Binders <-5 Strong Binders	Aguilar-Gurrieri, C. et al. High immunogenic VLP-based vaccines elicit new T cell specificities against melanoma neoantigens in mice. J Immunother Cancer 9, (2021).
netcleave	NetCleave	NetCleave2.0	C-terminal Antigen Processing	Probability score for C-terminal processing by the proteasome. From 0 to 1, being 1 best. >= 0.6 Processed peptides. >= 0.8 Optimally processed peptides.	Farriol-Duran, R., Vallejo-Vallés, M., Amengual-Rigo, P., Floor, M. & Guallar, V. NetCleave: An Open-Source Algorithm for Predicting C-Terminal Antigen Processing for MHC-I and MHC-II. in Computational Vaccine Design (ed. Reche, P. A.) 211–226 (Springer US, 2023). doi:10.1007/978-1-0716-3239-0_15.
mhcflurry_affinity	MHCflurry Affinity	MHCflurry2.0	HLA-I peptide	Percentile rank for HLA-I	O'Donnell, T. J.,

_percentile	Percentile		binding (binding assays + MS)	binding, normalized between HLA-I alleles. < 2 Binders < 0.5 Strong Binders	Rubinsteyn, A. & Laserson, U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. Cell Systems 11, 42-48.e7 (2020).
mhcflurry_affinity	MHCflurry Affinity	MHCflurry2.0	HLA-I peptide binding (binding assays + MS)	Percentile rank for HLA-I binding, NOT normalized between HLA-I alleles. <= 500 nM Binders <= 50 nM Strong Binders	
mhcflurry_processing_score	MHCflurry Antigen Processing Score	MHCflurry2.0	HLA-I peptide processing (MS)	Processing likelihood, from 0 to 1, being 1 best.	
mhcflurry_presentation_score	MHCflurry Antigen Presentation Score	MHCflurry2.0	HLA-I peptide binding (binding assays + MS)	Presentation likelihood, from 0 to 1, being 1 best. Integrates binding and processing predictions.	
TAP	TAP	NetCTLpan2.1	Peptide Transport to ER by TAP transporter.	Transport likelihood. From 0 to 1, being 1 best. Trained on MS data.	Stranzl, T., Larsen, M. V., Lundegaard, C. & Nielsen, M. NetCTLpan: Pan-specific MHC class I pathway epitope predictions. Immunogenetics 62, 357–368 (2010).
Physicochemical Features					

Full Epitope Calculated for the entire epitope sequence.					
hydroph_peptide	Epitope Hydrophobicity	Peptides R package	Hydrophobicity based on KyteDoolittle Scale.	The hydrophobicity index is calculated adding the hydrophobicity of individual amino acids and dividing this value by the length of the sequence. Highly expected transmembrane peptides generally have higher hydrophobicity values than 0.5 using KyteDoolittle scale.	Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: A Package for Data Mining of Antimicrobial Peptides. The R Journal 7, 4 (2015).
mw_peptide	Epitope Molecular Weight (bulkiness)	Peptides R package	Molecular Weight. Proxy for amino acid bulkiness.	The molecular weight is the sum of the masses of each atom constituting a molecule. The molecular weight is directly related to the length of the amino acid sequence and is expressed in daltons (Da).	
charge_peptide	Epitope Net Charge	Peptides R package	Net electric charge.	The net sum of the charges of each of the	

				amino acids comprised in the peptide.	
stab_peptide	Epitope Stability	Peptides R package	Peptide Stability in solution.	This index predicts the stability of a protein based on its amino acid composition.	
TCR Contact Region Calculated for the central residues of the epitope. These are reported to interact directly with the TCR CDR loops. Includes amino acids from position 4 to -2 of the epitope sequence.					
hydroph_tcr_contact	TCR Contact Region Hydrophobicity	Peptides R package	Hydrophobicity based on KyteDoolittle Scale.	Hydrophobicity calculated as above against the central residues of the epitope.	Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: A Package for Data Mining of Antimicrobial Peptides. The R Journal 7, 4 (2015).
mw_tcr_contact	TCR Contact Region Molecular Weight	Peptides R package	Molecular Weight. Proxy for amino acid bulkiness.	Molecular Weight calculated as above against the central residues of the epitope.	
charge_tcr_contact	TCR Contact Region Net Charge	Peptides R package	Net electric charge.	The net sum of the charges of each of the amino acids comprised in the central region of the peptide.	

Model Optimizations

PredIG uses XGBoost tunning to optimize its performance and adapt to different data sources. The main goal was to optimize the model to foster the scoring of epitopes in extreme class imbalance conditions where few immunogenic candidates are expected among many immune silent epitopes. Thus, we provide different models for the user to select the expected class imbalance in their target data.

Model Name	**Class Imbalance**	**SPW Value**	**Target Data**
PredIG-SPW-X	X	spw-x	Pathogen
PredIG-SPW-Y	Y	spw-y	Cancer

Please refer to the publication for an in-depth description of the class imbalance effects in PredIG predictions.

Citation

D) Results Module

Find the results of your PredIG prediction including a PredIG score (termed "predig") and all 14 features of the model calculated per epitope - HLA-I allele pair. Use the download menu below to select your output format.

Results table