IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

User Workshop Q&A - Written Responses - Day 1

These are the written responses for some questions captured in Zoom. The remaining questions were responded to verbally live.

Question	Answer
Do you plan to make an open API available for database searches?	Yes, we are. We are estimating to have that done in the next 12 months
Is there any information about molecular mimicry?	We capture if an epitope is studied as a 'mimotope' of something else. That can currently be searched for only under the specialized 'epitope details' search. It will be integrated in the main search in the next 3 months or so.
Is there a way to run batch search? e.g. search data for several peptides, or several CDR3 sequences.	Not currently. That is something we are working on.
What is the difference between Curated and Calculated chains (e.g., after export TCR data to .csv file, there are columns like Curated Chain 1 D Gene and Calculated Chain 1 D Gene)?	Curated data comes directly from the publication as stated by the authors. Calculated data is calculated by the IEDB based on which portions of the provided receptor sequence is most likely to be the CDR1, 2, 3 or V domain using our algorithms. In these cases, the calculated CDR3 may be trimmed relative to the curated sequence. We have a publication explaining this in more detail if you are interested.
Have the MHC binding tools been updated in the last 6 months?	Yes, tools are continuously updated. There is a changelog that tells you when that occurs.
How does the benchmark work? Is it possible to download the dataset for the benchmark?	**Links to paper:** https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007757 **Repository:** https://gitlab.com/iedb-tools/cd8-t-cell-epitope-prediction-benchmarking

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

User Workshop Q&A - Written Responses - Day 2

These are the written responses for some questions captured in Zoom. The remaining questions were responded to verbally live.

Question	Answer
Can one MHC bind to more than one peptide core?	Yes. This can be a problem for generating 3D structures for crystallization for MHC-II, which works best if all peptides bind in the same conformation. If there are multiple cores that an MHC molecule can bind to, there is a mixture of states, and the peptide can often not be resolved.
Can MHC binding affinity be assumed as a surrogate for the stability of the complex? i.e. the time the complex will last on the surface and therefore the time an epitope is being presented?	There is a very high correlation between stability (essentially measured by off-rates) and IC50 values (which are a proxy for Kd values). There are a few examples where there is divergence, but we have not found that to be a practically relevant concern, and have not found that stability based assays give any improvements on predicting epitopes.
Mass spec data also contains false matches (FDR 1 or 5%) and contaminants (e.g. class I in II sets), Could the IEDB data set be used to identify these?	Yes, and many mass spec pipelines include that as a step. We are a bit torn, as we like to know about unexpected binding peptides identified by mass spec, so we like to see the unfiltered results - which has led to some interesting discoveries like unconventional binding modes. But yes, given that false positives are a real problem in MS, it makes sense to remove predicted non-binders as a QC step.
Can you use the MHCII peptide predictions as sequences and make a DNA vaccine with that?	The MHCII peptide predictions are predictors of binding, not immunogenicity. Binding is necessary, and generally the limiting factor, but not sufficient for eliciting an immune response.
In MHCI prediction, what is very important and considered most between the score and the percentile rank?	Both are important. The percentile rank helps to enable comparisons between different alleles predicted by the same method as well as comparisons between methods.
What are some options that can be used to search for epitopes in peptides with unnatural amino acids?	The IEDB database contains experimentally characterized epitopes with modified / unnatural amino acids. We have not yet implemented any tools that deal with them, as this is not a common use case, but we have started some work in this area.

Does the netMHCpan algorithm include protein cleavage data?	No. The EL algorithm is trained on naturally eluted ligands, which implicitly includes training on cleavage patterns.
For cancer neoantigens, should we use all 15 AA frames containing mutated amino acids?	Yes, to be comprehensive, you can consider all 15-mer peptides containing the mutation, starting from (-14,0), (-13,1), (-12,2), ... ,(0,14), which covers a 29-mer stretch.
Is there any concept of protective and harmful HLA alleles based on in silico epitope prediction?	No.
What happens if I use a method that is not recommended?	The 'recommended' methods are provided to give users a simple way to immediately run what we consider (based on benchmarks) to be the most accurate prediction methods. As alluded to in other questions, a single method might not be the best performing method across all alleles. We therefore make all of the methods available and are working on ways to improve our automated benchmarks and make that information easily accessible to users.
Will you discuss Pan-specific stability of peptide:MHC-I complexes algorithm and results?	No, we have not found those predictions to provide an advantage when predicting epitopes. We might reconsider that as we do more benchmarks
How do the databases account for IC50-value differences that result from using different reference peptide/ competitors?	They don't. If the assay is performed under the right conditions, there should not be a difference in theory; in practice there can be a difference but it tends to be low.
Are the consensus method and NetMHCIIpan orthogonal to each other? If yes and there is a hybrid method combining these two methods, do you think it can improve prediction performance?	They are giving quite similar results. We are evaluating new consensus methods that would include NetMHCIIpan itself.
Besides amino acid residues which are highly immunogenic, does the immunogenicity prediction tool use another variable to work its predictions?	The class I does not. The class II tool is an ANN, so a bit of a blackbox, and it can include higher order interactions.
Is lower value = higher binder true for the MHC Class I immunogenicity tool as well?	No, for class I immunogenicity, higher = more immunogenic.
Both LYRA and SCEptRe have the green flask symbol next to them. What are the drawbacks of using them right now and when do you estimate they will be ready for production?	The flask indicates it is a 'Labs' tool, which we typically stamp onto tools that are new to the website. The label is meant to indicate that the implementation on the website may not be stable and doesn't indicate anything about the underlying algorithm. We haven't done a good job of

	re-evaluating when that logo should be removed, but we will plan to do so.
For immunogenicity prediction especially in biologics design, how important is to include MHCI in addition to MHCII?	We normally don't include class I. Given that the biologics are externally supplied (and not expressed inside a cell), they are expected to be going for the class II pathway only
Is there a way to get low reactivity data for a germline/ constant region?	Ideally, that kind of information should show up in the immunome browser, which maps reactivity rates on the protein. But antibodies might be tricky, as they are highly divergent in the VDJ region, and the mapping might not work. If you are interested, we could pursue this specific question further; immunogenicity of antibodies obviously has high relevance.
Is HIV mutation rate on key proteins like gp120 that high in order to make a vaccine almost impossible?	Not only the rate is high, but we have many diverse viruses circulating at a time. Influenza evolves are close to the same rate, but one form or a few related forms circulate around the world each year. Also, HIV integrates into the host genome (of T-cells mostly) for life. So we want a vaccine that prevents infection, and not just one that prevents disease. Polio vaccine probably does not prevent infection completely, but just reduces virus symptoms to eliminate polio disease.
Is it possible to dump the confirmed neoantigens using the IEDB interface?	Yes, if you go to the IEDB home page, use the "Specialized Searches" drop down menu (blue top header bar), choose "Epitope Details". In the Epitope search pane, towards the bottom of that Epitope search pane, use the "Epitope Related Object" section to select "The epitope is a neo-epitope of:" and then click the green "Search" button. You can then export those epitopes once the results are returned (1716 epitopes). We are currently adding this search feature to the results page (rather than the "specialized Searches" and expect this to be live in 2021.
This is a data question. Is there a link between entries in IEDB and patients/samples, similar to what was mentioned in the HIV database?	In each of the assay pages, details, and exports, there is a section on the immunization, which describes the host, their age, gender, disease states, disease stage, and MHC types present (when provided by the authors).