2020 Presented at Translational Systems Immunology Keystone Symposium, January 28 – February 1, 2018; Snowbird, UT

Mono-allelic mass spectrometry improves the prediction of endogenously processed and presented MHC Class I epitopes

Michael Rooney[1], Tim O'Donnell[2], Alex Rubinsteyn[2], Dewi Harjanto[1], Sonia Ting[1], Tyler Colson[1], Amanda Creech[1], John Sauld[1], Yusuf Nasrullah[1], Robert Oslund[1], Scott Goulding[1], Theresa Addona[1], Dominik Barthelme[1], Jennifer Abelin[1]

[1]Neon Therapeutics, Cambridge, MA; [2]Icahn School of Medicine at Mt. Sinai, New York, NY

Abstract

Neoantigens are somatically mutated protein sequences presented on MHC Class I or II molecules. Mounting evidence suggests neoantigens potentiate anti-tumor immune responses during checkpoint blockade, and there is great interest in directly targeting neoantigens therapeutically. Vaccines and other neoantigen-based approaches must rely on personalized *in silico* epitope prediction given the multitude of MHC alleles in the human population (each with its own peptide binding repertoire). The accurate identification of neoantigens is critical in designing effective therapies. The current prediction paradigm is based almost entirely on machine learning algorithms (e.g. NetMHC) trained on *in vitro* p:MHC binding affinity assays.

Mass spectrometry (MS), which can be used to directly identify peptides that are processed and presented on actual cells, is an orthogonal approach to define the rules of the MHC ligandome. Recent work has used these data to better-characterize the contributions of gene expression and proteasomal processing to MHC presentation.

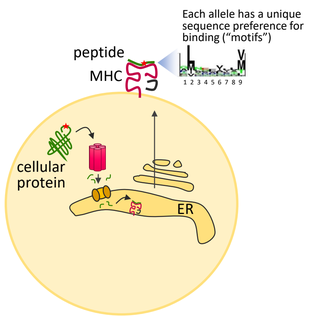
In particular, it has been shown that MS analysis of engineered cell lines that express a single MHC Class I allele (rather than the full complement of 2 HLA-A alleles, 2 HLA-B alleles, and 2 HLA-C alleles) can rapidly resolve allele-specific binding motifs and be used to train predictors that significantly out-perform NetMHC.

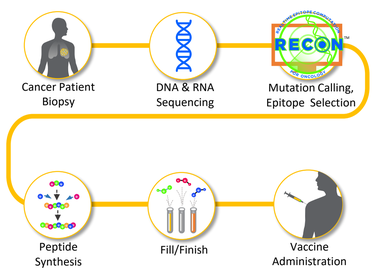
Building on this work, we have developed MHC presentation predictors for diverse individual MHC Class I alleles using MS data. These algorithms substantially out-predict standard NetMHC-based approaches and perform well on all peptide lengths (8-12) and all canonical Class I loci (A, B, and C). Meanwhile, continued analysis of diverse MS datasets has provided new insights into the nature of proteasomal processing and how it varies across cell types and tissues. These algorithms and observations are being applied to Neon Therapeutics' neoantigen prioritization pipeline, RECON™, which supports its clinical studies.

Background: Peptide presentation on MHC

Neoantigens are mutated peptides presented on MHC

Neon is developing personalized neoantigen-targeting therapies





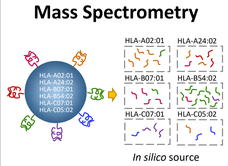
Experimental approach: Mono-allelic MS

Mono-allelic MS data provide a rapid, unbiased, and clean approach for defining peptide-binding motifs across diverse MHC alleles

Biochemical p:MHC binding assay

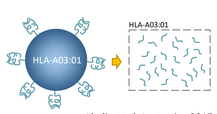
- Basis for NetMHC
- Slow / low-throughput
- No insights on processing

Multi-allelic Mass Spectrometry



- High-throughput
- Ability to learn processing rules (*e.g.* cleavage)
- ***Key challenge:*** Requires *in silico* imputation to assign peptides to alleles

Mono-allelic Mass Spectrometry



- Clean and unbiased approach for allele-specific algorithms
- Can rapidly and systematically fill allele coverage gaps
- Possible to leverage allele-specific peptide length preferences

Benefits of mono-allelic approach

Mono-allelic MS uncovers MHC-binding peptides that are poorly scored by NetMHCpan but biochemically validate as strong binders

		Peptide	netMHCpan 2.8 (nM)	measured (nM)
A*01:01	1	FLWPEAFLY	409.9	1.3
	2	YVPEHWEYY	551.9	11.1
	3	KLDLLEQEY	466.7	70.6
	4	TTDGYLLRL	792.1	78.9
	5	DTEFPNFKY	344.0	80.8
	6	ATDGKVLLW	532.5	191.4
B*51:01	1	DGLLRVLTV	11221.3	0.2
	2	DAPLNIRSI	17149.1	6.5
	3	DGLRDLPSI	12708.1	12.6
	4	DGYVVKETI	13414.5	13.4
	5	DGRLVINRV	18358.8	13.5
	6	DAYPQRIKF	12680.6	31.0
	7	IIYPTPKVV	11654.4	38.0
	8	VPPLVPKVL	11306.6	56.3
	9	TPESKIRFV	9336.0	84.8
	10	VALLVGEKV	9509.3	143.6
	11	YIIEREPLI	16193.4	1791.8

		Peptide	netMHCpan 2.8 (nM)	measured (nM)
A*29:02	1	FYPELKLAY	396.3	0.2
	2	HFLDRHLVF	825.4	1.6
	3	YLPALKVEY	113.3	2.6
	4	LPLDIQIFY	359.2	4.6
	5	ALSDLALHF	217.6	6.0
	6	GLDDKLLHY	157.5	21.3
	7	GLPDLVAKY	123.9	29.7
B*54:01	1	YPVYQPVGP	8015.0	1.7
	2	YGFVNYVTA	5011.5	6.0
	3	FVEILIPVA	6043.1	12.2
	4	FPKETFEGP	8478.9	16.4
	5	SAPVNFISA	5762.2	16.6

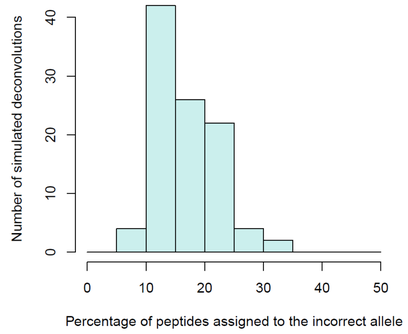
Abelin et al. *Immunity.* 2017.

Simulations of multi-allelic data show that allele ambiguity is nontrivial to resolve

Procedure

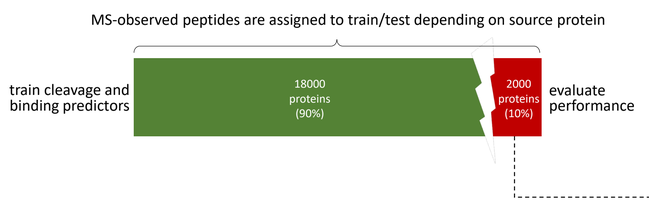
1. Generate a random six-allele patient HLA genotype (2 alleles each of HLA-A, B, and C; sampling at U.S. allele frequencies)
2. For each allele, sample 500 peptides from relevant mono-allelic experiment
3. Combine the peptides to create mock 3000-peptide multi-allelic data set
4. Assign each peptide to allele that yields best NetMHCpan %-rank score
5. Determine percentage of peptides incorrectly assigned by NetMHCpan
6. Repeat process 100 times

Rates of misassignment in 100 simulated deconvolutions

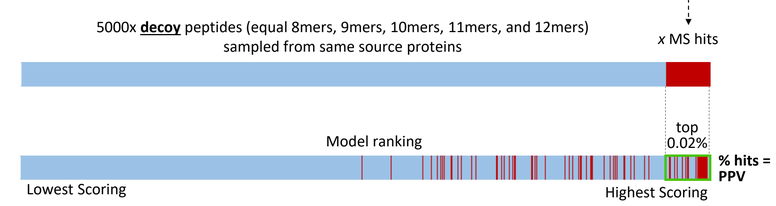


Benchmarking for MHC Class I predictor

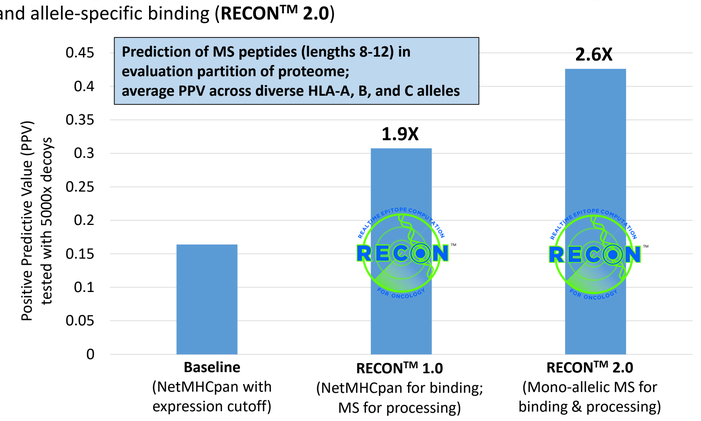
Model training and evaluation are conducted on non-overlapping source proteins



Evaluation approach employs a 5000:1 excess of decoys to true binders

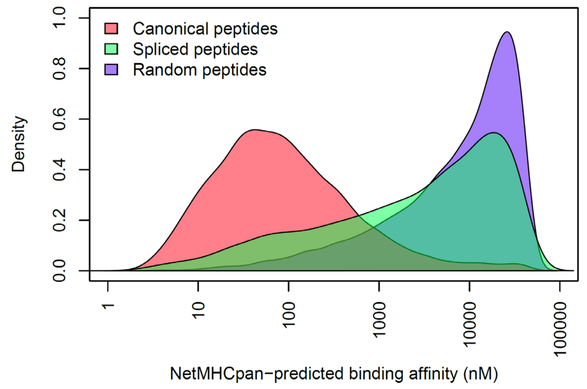


MS drives significant prediction improvement both in terms of processing (**RECON™ 1.0**) and allele-specific binding (**RECON™ 2.0**)

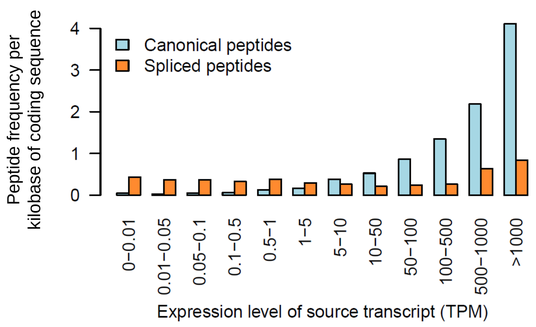


New insights into peptide processing

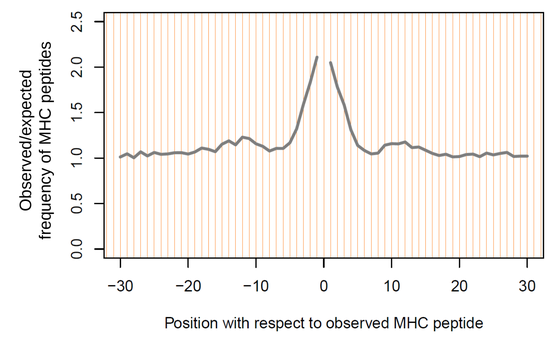
Spliced peptides (Liepe et al. Science. 2016) exhibit predicted binding strengths similar to random peptides



Spliced peptides (Liepe et al. Science. 2016) are equally sampled from strongly and weakly expressed source genes



Clusters of MHC Class I peptides within source genes ("hot spots": Müller et al. *Frontiers in Immunology*. 2017) tend to be no wider than 10 amino acids



Conclusions

- Mono-allelic profiling of MHC-bound peptides by MS is an effective approach for defining allele-specific binding motifs that have been missed in previous low-throughput studies or that are difficult to extract from multi-allelic data
- MHC Class I presentation predictors trained on mono-allelic MS data outperform NetMHCpan-based predictors, forming the basis of RECON™ 2.0, the bioinformatics engine that will select and rank epitopes in future Neon trials
- Similar efforts for underway for building predictors for MHC Class II
- Spliced peptides do not appear to obey expected patterns of MHC binding potential and expression level
- MHC Class I peptides tend to be clustered in their source proteins, but such "hotspots" tend to be small in size (<10AA) and do not improve prediction

Acknowledgments

- The authors acknowledge the financial support of Neon Therapeutics and many productive conversations with the extended Neon research team

NEON THERAPEUTICS