

The Elizabeth H.
and James S. McDonnell III

**McDONNELL
GENOME INSTITUTE**
at Washington University



**Washington
University in St. Louis**

SCHOOL OF MEDICINE

PMBIO Module 08

Immune. Immunogenomics analysis

Malachi Griffith, Obi Griffith, Zachary Skidmore, Huiming Xia
Introduction to bioinformatics for DNA and RNA sequence
analysis (IBDR01)

29 October - 2 November, 2018
Glasgow



Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.



The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

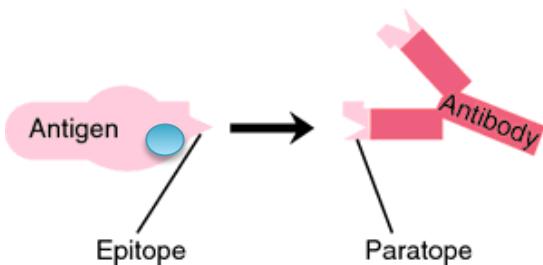
No additional restrictions — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

Learning objectives of module 08: Immune

- Key concepts: Immunogenomics, adaptive immunity, immunotherapies, personalized cancer vaccines, MHC binding prediction, neoantigen identification and prioritization
- Use results from the previous sections (germline variants, somatic variants, gene/transcript expression estimates, variant allele fractions, clonality estimates, etc.) to design a personalized cancer vaccine for our hypothetical patient
- Become familiar with the pvactools software for neoantigen identification, prioritization, selection, and DNA vector design

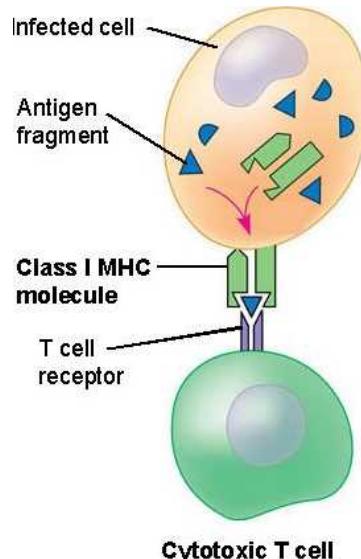
Some basic immune terminology

- Antigen: *any substance that can initiate an immune response within the body*
- Epitope: *also known as antigenic determinant, is the part of an antigen that is recognized by the immune system. i.e. bound by the receptor on an antibody, B cell, or T cell*



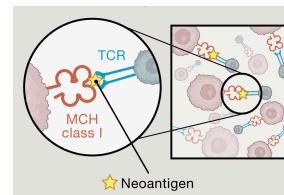
Jasreet Hundal

- MHC Class I molecules: *display fragments of proteins from within the cell to cytotoxic T cells*

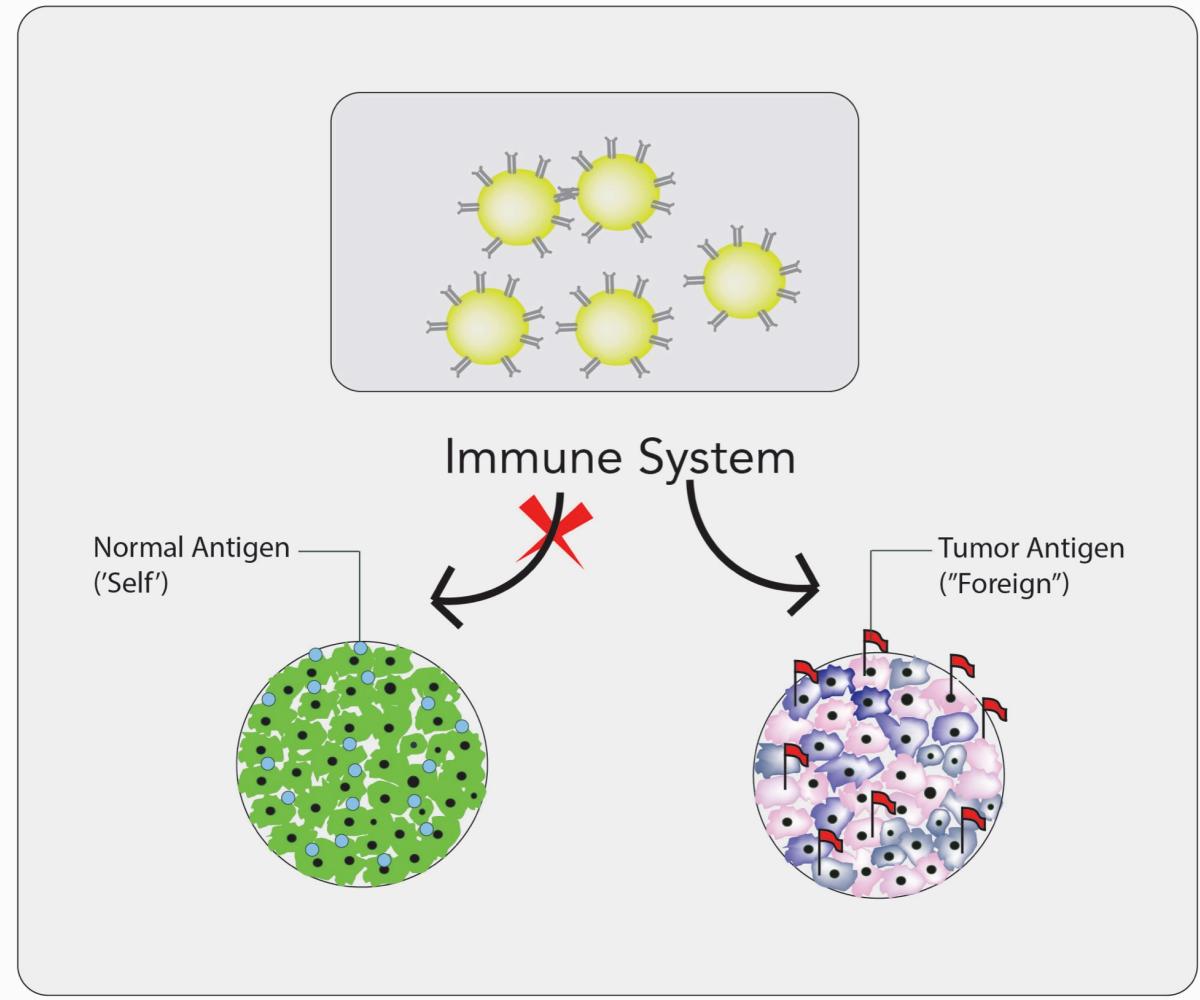


A fragment of foreign protein (antigen) inside the cell associates with an MHC molecule and is transported to the cell surface.

The combination of MHC molecule and antigen is recognized by a T cell, alerting it to the infection.



The goal of one class of cancer immunotherapy is to leverage the immune system's ability to attack non-self or "foreign" cells (recognized by their neoepitopes)

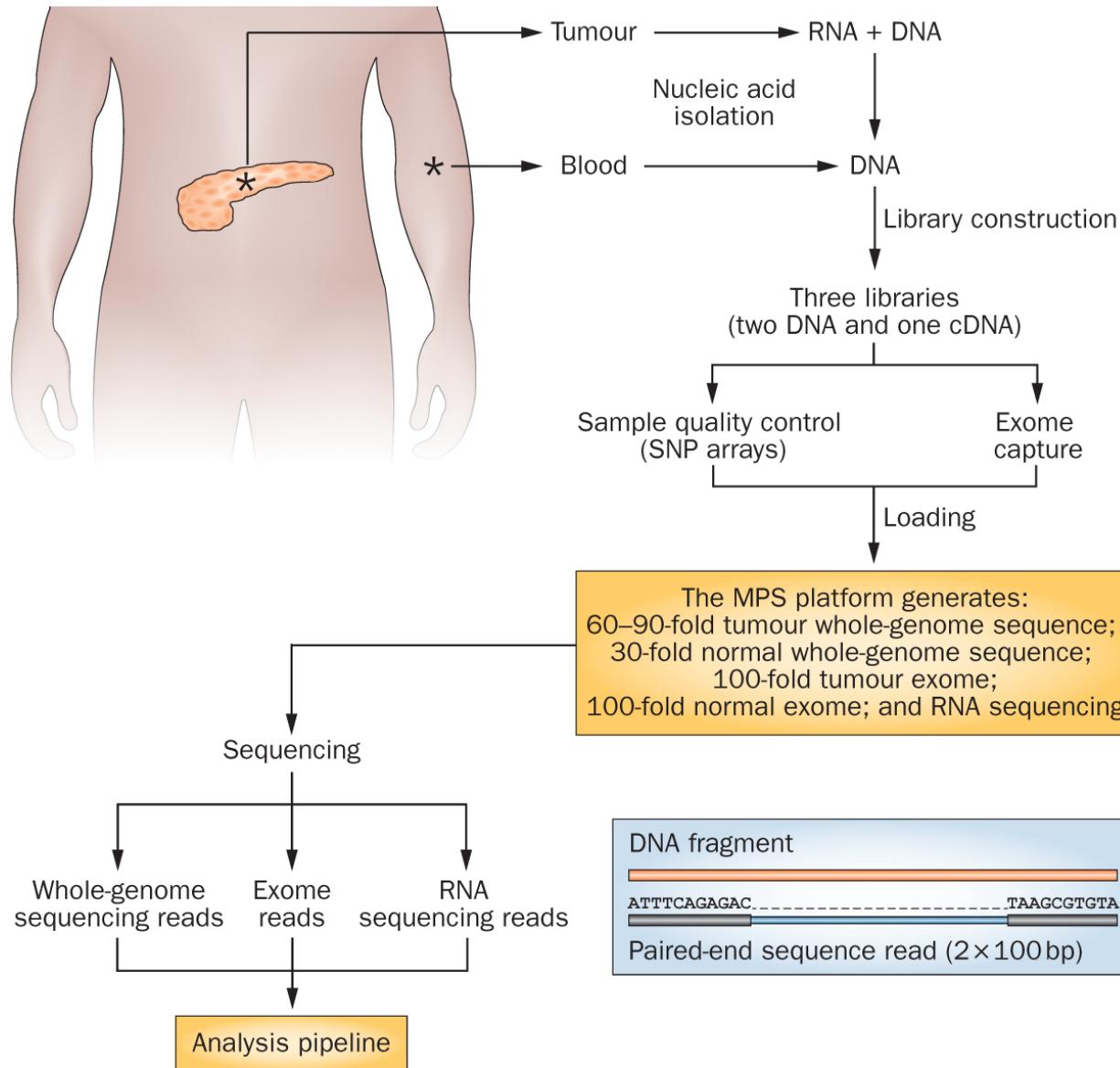


Personalized cancer vaccines and checkpoint inhibition therapies (e.g. Pembrolizumab) are two examples of this type of cancer immunotherapy

Neoepitopes are relevant to a specific aspect of the immune system

- Tumor-specific epitopes ('**Neoepitopes**') are unique to each patient
- These unique tumor epitopes result from **somatic changes in the tumor genome** (i.e. **non-synonymous substitutions, insertions, deletions, splice site alterations, fusions, etc.**)
- Historically, **experimental techniques** have not been capable of rapidly and systematically identifying neoepitopes ...

Advances in sequencing technology have changed this



Neoepitopes are peptides resulting from somatic mutations that induce an immune response*

(i)

wildtype sequence

AAG CTA CCA GAG CCA TGT CCT TCA ACG

KLPEPCPST

439.03 nM

mutant sequence

AAG CTA CCA GAG CCA TGT CCT TCA ATG

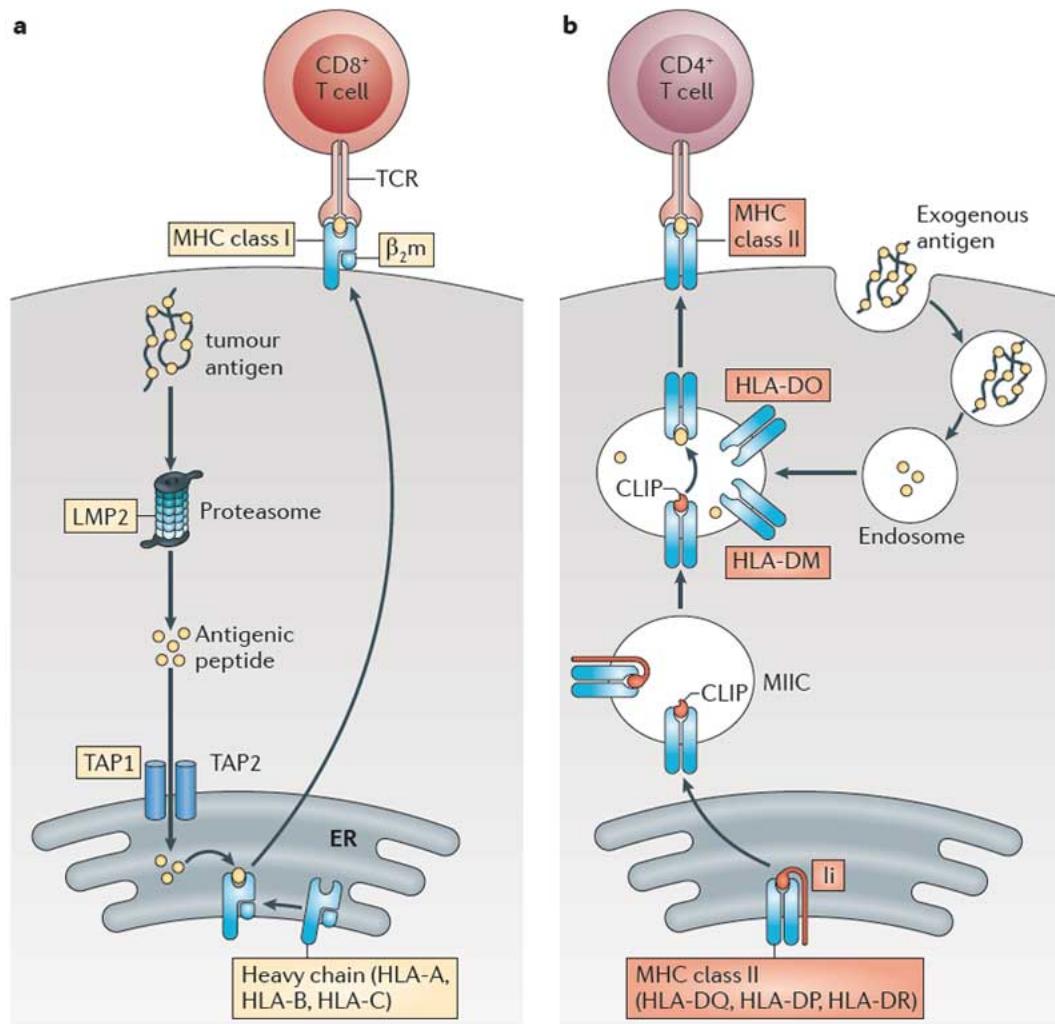
KLPEPCPSM

119.19 nM

Single base substitution of cytosine (C) to thymine (T), results in a predicted amino acid change of threonine (thr) to methionine (met).

* These are *predicted* neoepitopes. We rarely *know* if they are immunogenic...

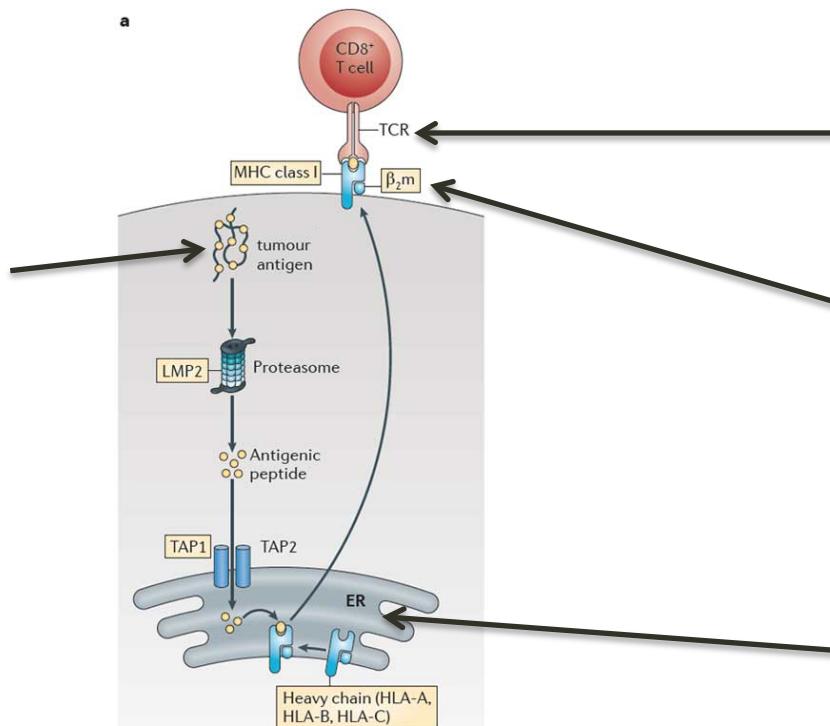
Overview of T cell recognition of neoepitopes (presented by MHC Class I or II proteins)



Two major classes of T cell that recognize epitopes presented by two classes of MHC molecule

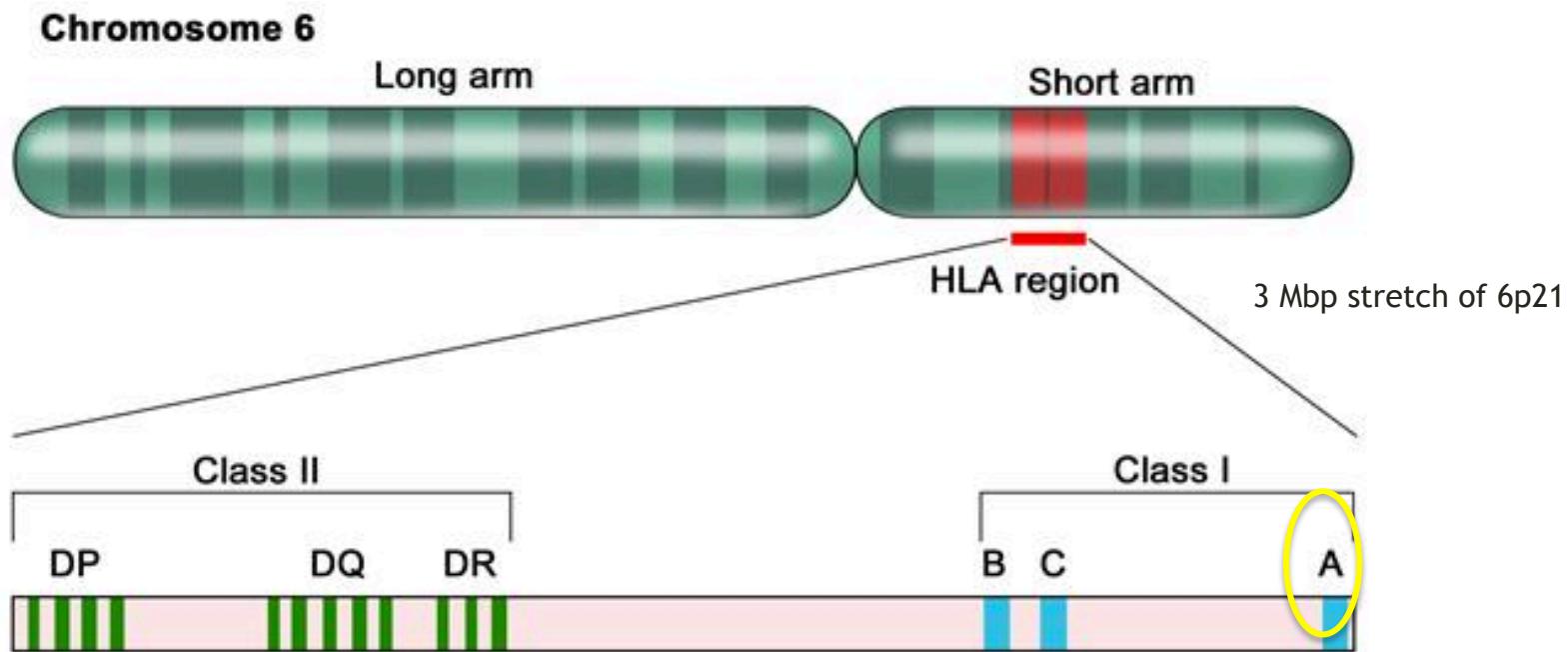
Invoking an adaptive immune response against the tumor (focus on CD8+ T cells)

1. Tumor produces a unique peptide corresponding to a somatic mutation

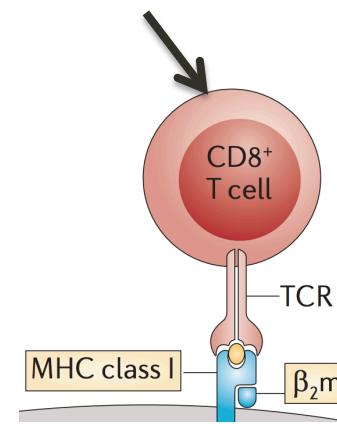


4. T cell receptor that uniquely matches the tumor peptide
3. Neoantigen peptide presented by MHC
2. Processing and presentation of the tumor specific peptide

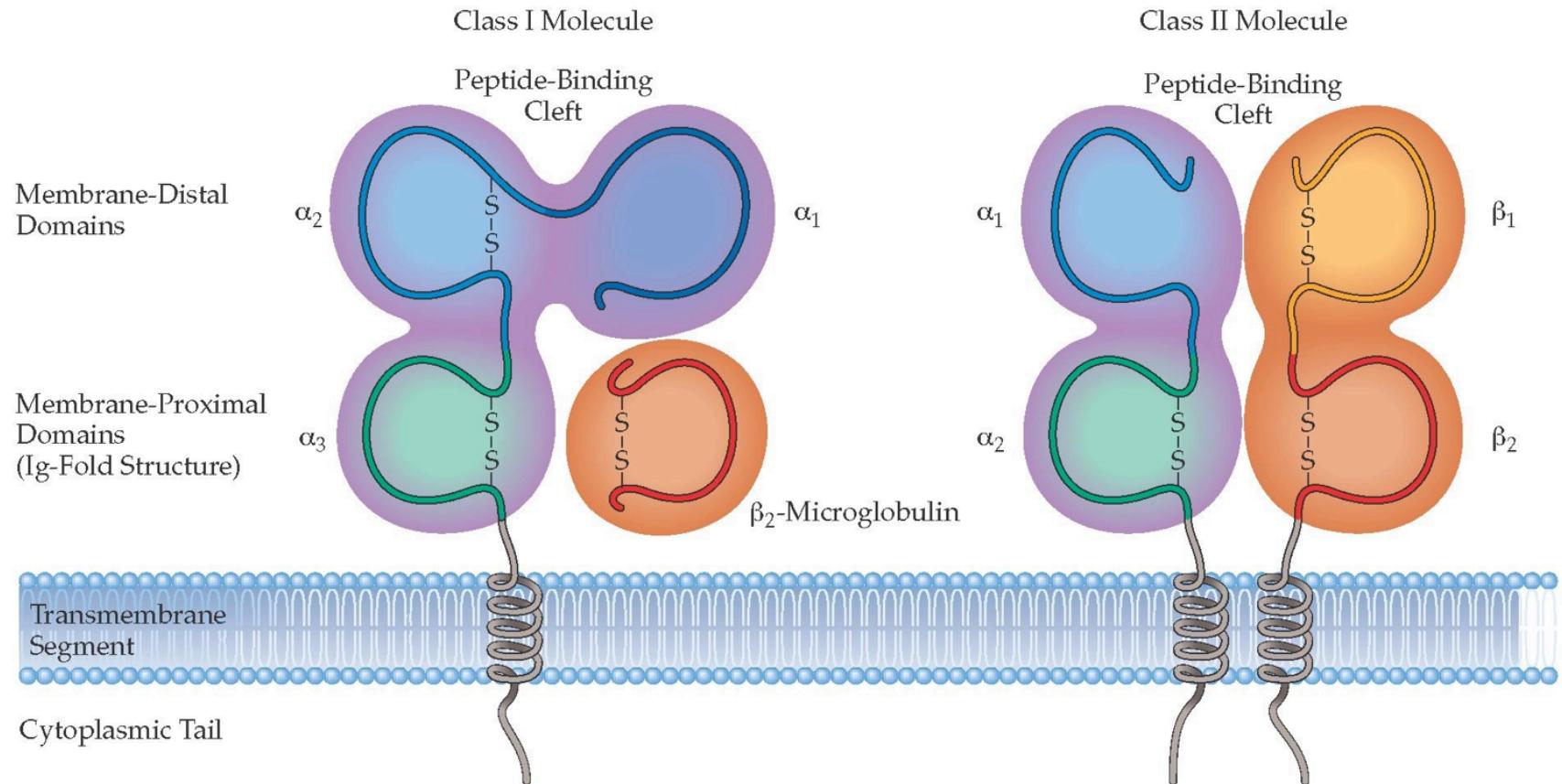
To accurately predict binding of particular peptides to MHC, we need to know which MHC alleles each patient has



- The HLA locus is the most heterogeneous region of the genome
 - >5,000 known variations of the MHC Class I sequences alone
 - e.g. HLA-A*02:101:01:02N ...
- Challenge to classical HLA allele typing
 - Polymorphisms can be outside the sequenced hypervariable region (exons 2-4 of the class I genes)
- Challenge to NGS HLA typing
 - Relies on good sequence coverage and a well annotated database of known HLA sequences
 - IMGT/HLA Database (www.ebi.ac.uk/imgt/hla/)



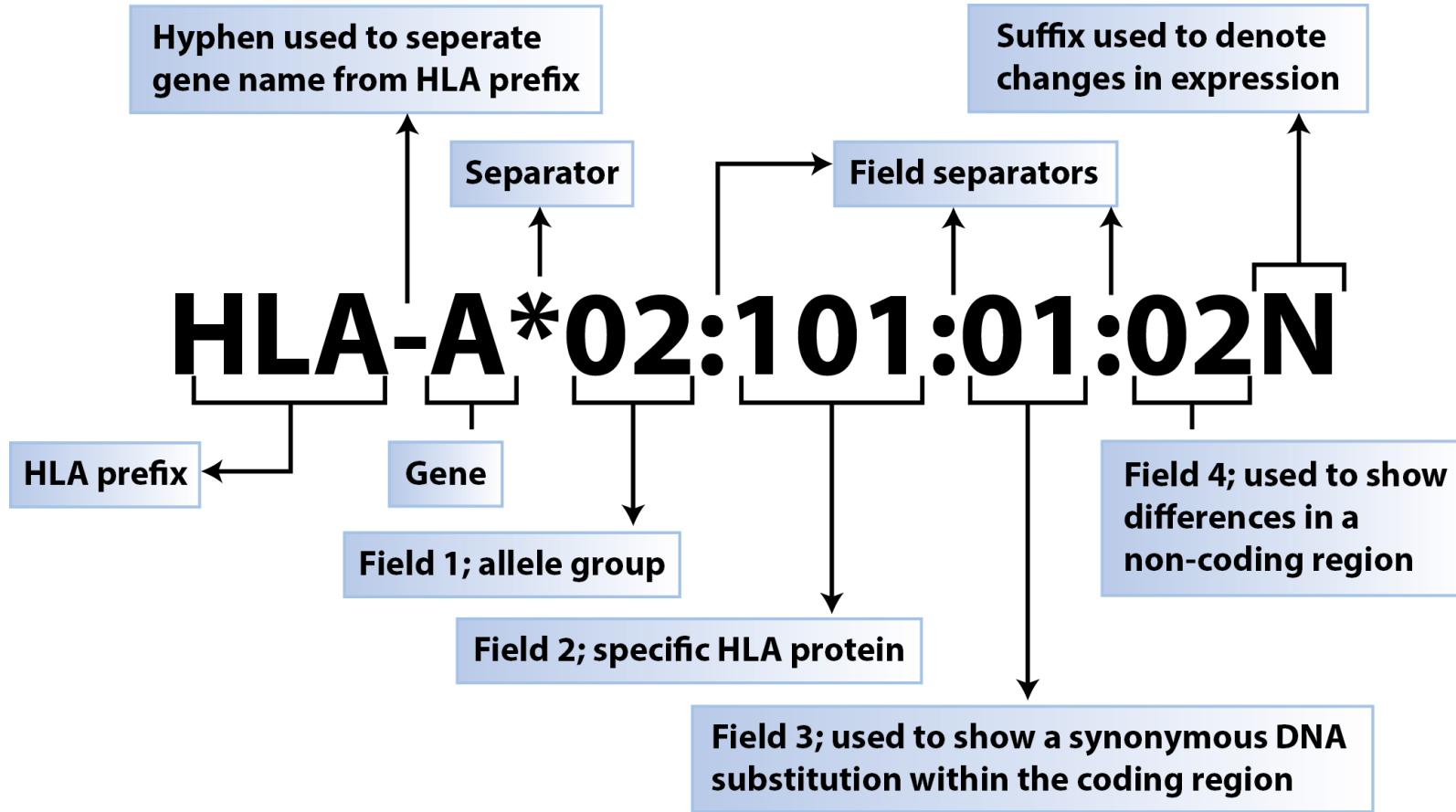
Class I and class II MHC structures



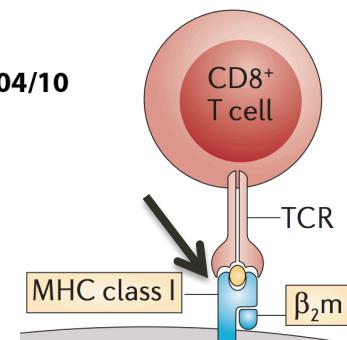
HLA class II pairings

	Alpha	Beta
HLA-DM	HLA-DMA	HLA-DMB
HLA-DO	HLA-DOA	HLA-DOB
HLA-DP	HLA-DPA1	HLA-DPB1
HLA-DQ	HLA-DQA1, HLA-DQA2	HLA-DQB1, HLA-DQB2
HLA-DR	HLA-DRA	HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5

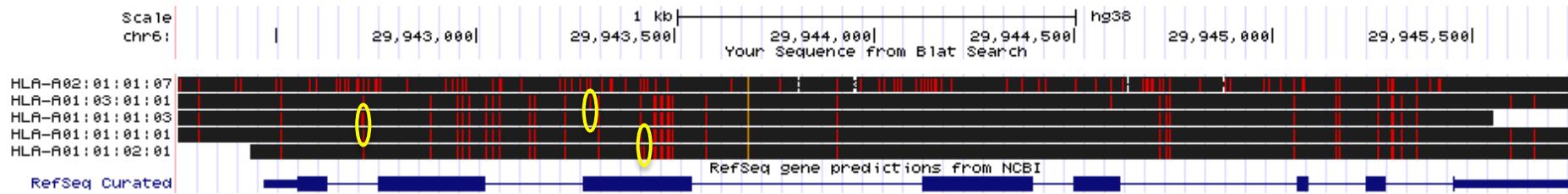
Conventions for HLA allele naming



(c) SGE March 04/10

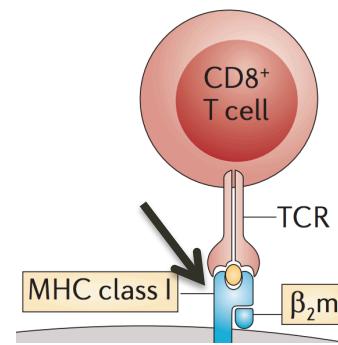


DNA alignment view of some example HLA sequences

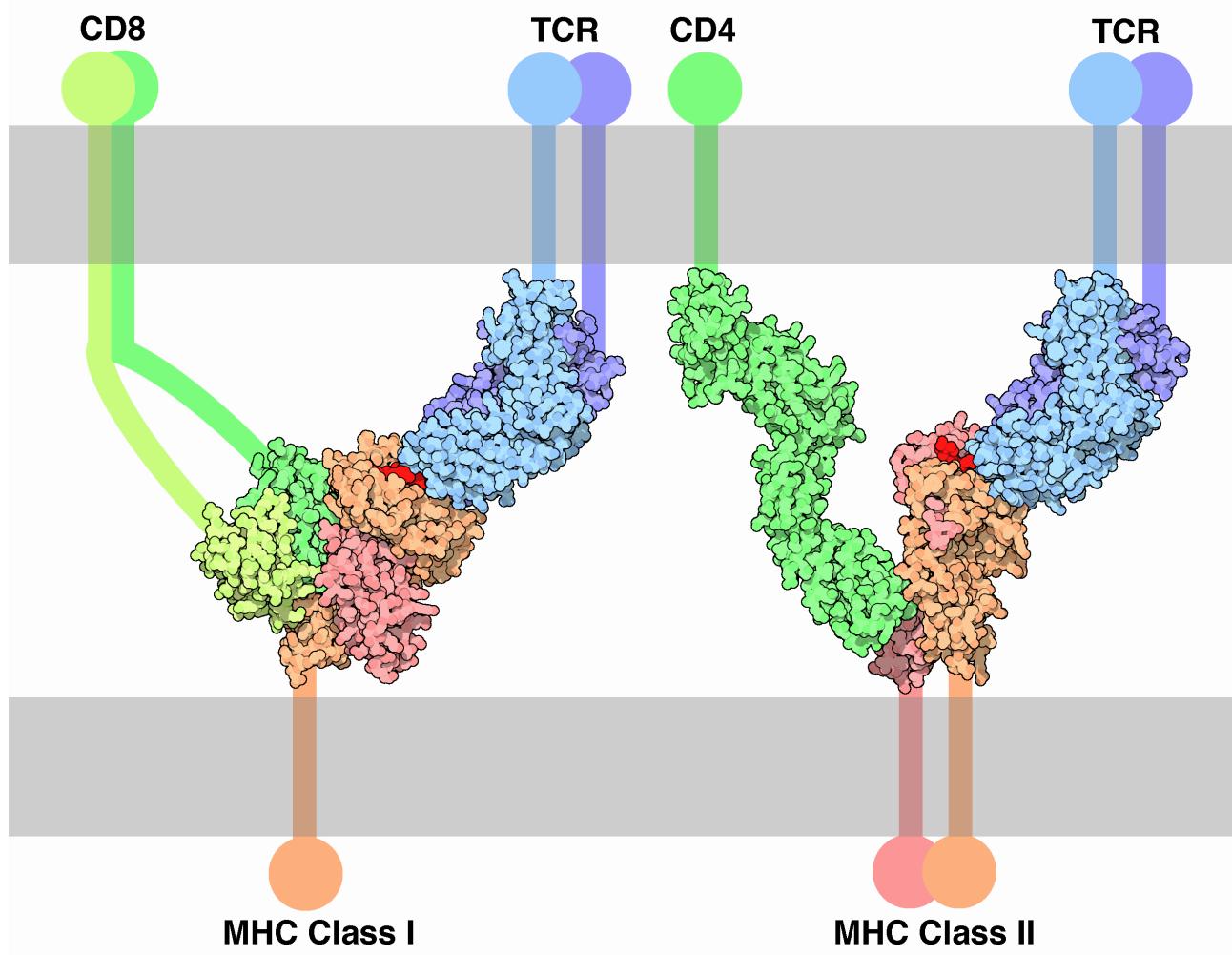


Sequence comparison between ~3500 bp genomic sequences of five HLA-A sequences:

- HLA-A*01:01:01:01
- HLA-A*01:01:01:03 – differences in non-coding regions
- HLA-A*01:01:02:01 – synonymous changes in coding regions
- HLA-A*01:03:01:01 – some protein coding changes
- HLA-A*02:01:01:07 – many protein coding changes (different allele group)

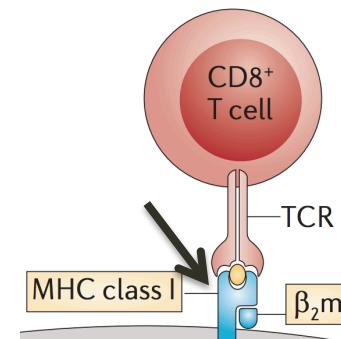


3d view of MHC complex



The three dimensional structure is what really matters

<http://pdb101.rcsb.org/motm/63>

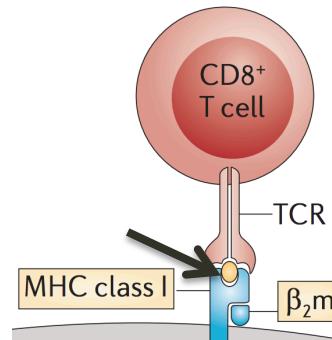


MHC / HLA resources ...

- IMGT. The international ImMunoGeneTics information system (<http://www.imgt.org>)
 - Genome, transcript, protein, and structure data
 - Immunoglobulins/antibodies, T cell receptors (TCRs), major histocompatibility complex (MHC) molecules, etc.
- The subset of IMGT that deals with sequences of the human MHC (HLA) is hosted separately at EBI:
 - [The IPD-IMGT/HLA Database](#)

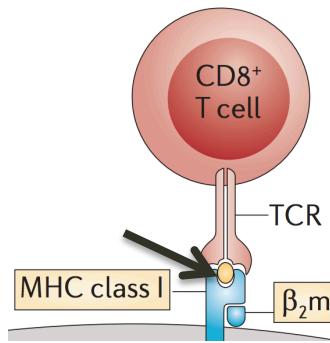
Peptide-MHC binding

- Major components of neoepitope identification:
 - Predicting amino acid changes caused by somatic variation in tumor DNA
 - Predicting whether a short peptide sequence (8-12 amino acids) containing the somatic event will be strongly bound by one of the patient's MHC molecules
 - For class I -> two alleles of HLA-A, HLA-B, and HLA-C
 - Tools: Athlates, HLA-Miner, etc.
 - Consider all peptide lengths vs. all HLA alleles vs. all registers for the mutation
 - Tools: pVACtools, OpenVax, etc.
 - Many more factors ...



Peptide-MHC binding

- There are many existing algorithms that attempt to model the peptide-MHC complex
 - NetMHC, NetMHCpan, MHCflurry, PickPocket, SMM, SMMPMBEC, NetMHCIIpan, NN-align, SMM-align, etc.
 - Trained on *relatively* small libraries of peptides with *in vitro* binding measurements for a relatively small set of MHC alleles
 - Algorithms try to generalize the sequence/binding relationship to any sequence and MHC allele.
- Note. Binding doesn't mean presentation, and presentation doesn't mean T cell recognition or cell death.



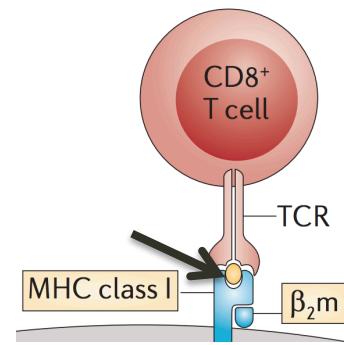
Many algorithms, varied performance

MHC Class I binding algorithms:

Algorithm	mean AUC (range)
NetMHC 3.0 (ANN)	0.90 (0.74-0.98)
SMM	0.91 (0.73-0.98)
SMMPPMBEC	0.91 (0.74-0.98)
NetMHCpan	0.94 (0.57-1.0)
PickPocket	0.895 (range unavailable)

MHC Class II binding algorithms

Algorithm	mean AUC (range)
SMM-align	0.78 (0.66-0.83)
NN-align	0.80 (0.62-0.86)
NetMHCIIPan	0.85 (0.63-0.93)



Peptide-MHC binding resources

- IEDB analysis resource
- Others you will hear about today

MHC-I Binding Predictions ←

Prediction Method Version 2013-02-22 [Older versions]

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. ([Browse for sequences in NCBI](#))

```
>LCMV Armstrong, Protein GP  
MGQIVTMFEALPHIDEVINIVIILVITGIKAVYNFATCGIFALISFLLAGRSCGM  
YGLKGPDIFYKGVYOFKSVEFDMSHLNLTPNACSAHHYISMGTSGLELTFTNDSII  
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNNGITIQYNLTFSDA  
QSQQSQCRTFRGRVLDMFRTAF-GGKYMRSGWGTGSDGKTTWCSQTSYQYLIIQNRTWE  
NHCTYAGPFGMSRLSQEKTKFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE  
LKCFCNTAVAKCNVNHDAEFCDMRLRIDYNAKALSKFKDEVESALHLFKTTVNSLISDQ  
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSPVKCWLVTNGSYLNETHFSQIEQEA  
DNMITEMLRKDYIKRQGSTPLALMDLMFSTSAYLVSIFLHLVKIPTHRHKGSCP  
HRLTNKGICSGAFKVPGVKTVWKRR
```

Or select file containing sequence(s) Choose File No file chosen

Choose sequence format auto detect format

Choose a Prediction Method

Prediction Method IEDB recommended Help on prediction method selections

Sp Consensus netMHCpan ANN SMMPPMBEC SMM CombLib_Sidney2008 PickPocket netMHCcons netMHCstabpan

MHC source species

Show only frequently occurring alleles: Select MHC allele(s) Select HLA allele reference set:

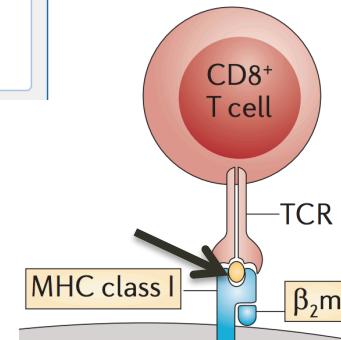
Upload allele file [?](#)

Epitope Analysis Resource

T Cell Epitope Prediction [?](#)
Scan an antigen sequence for amino acid patterns indicative of:
MHC I Binding
MHC II Binding
MHC I Processing (Proteasome,TAP)
MHC I Immunogenicity

B Cell Epitope Prediction [?](#)
Predict linear B cell epitopes using:
Antigen Sequence Properties
Predict discontinuous B cell epitopes using antigen structure via:
Discotope
ElliPro

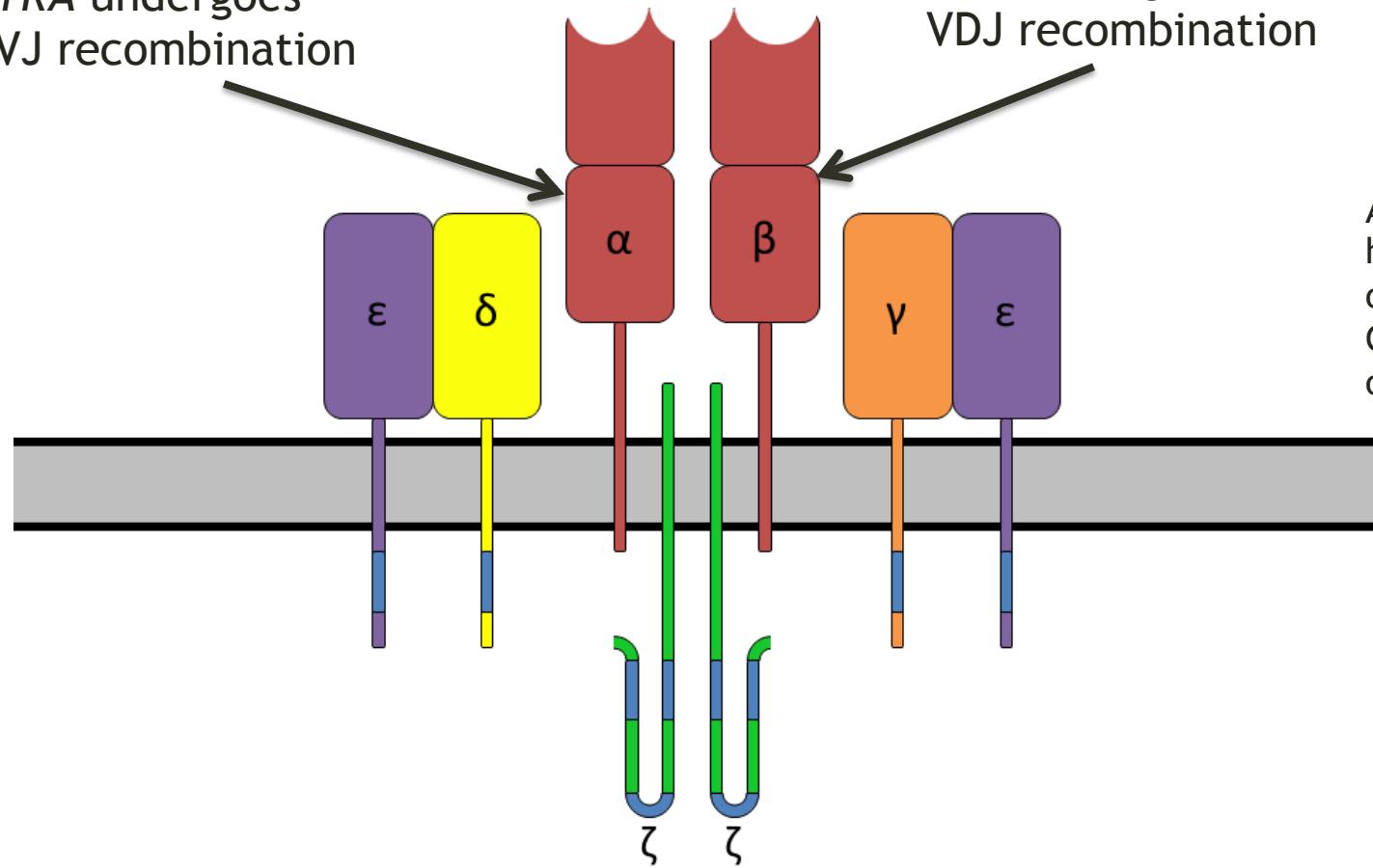
Epitope Analysis Tools [?](#)
Analyze epitope sets of:
Population Coverage
Conservation Across Antigens
Clusters with Similar Sequences



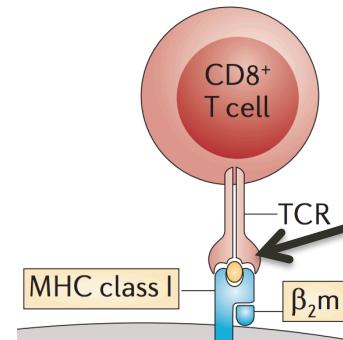
Structure of T cell receptor

TRA undergoes VJ recombination

TRB undergoes VDJ recombination



Alpha/Beta form a heterodimer (95% of T cells) or Gamma/Delta (5% of T cells)



Generation of T cell receptor (TCR) diversity

Germline configuration:



(1) D to J recombination



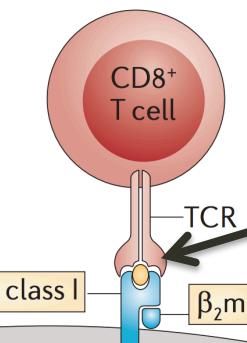
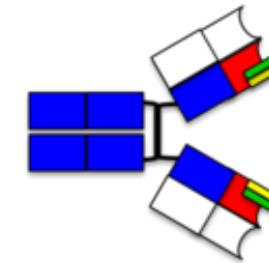
(2) V to DJ recombination



(3) Transcription & splicing



(4) Translation & assembly

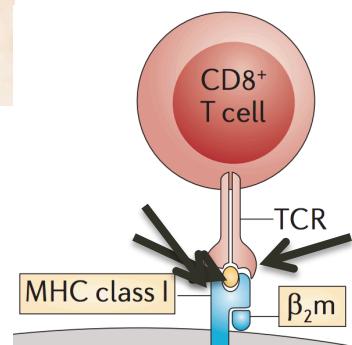
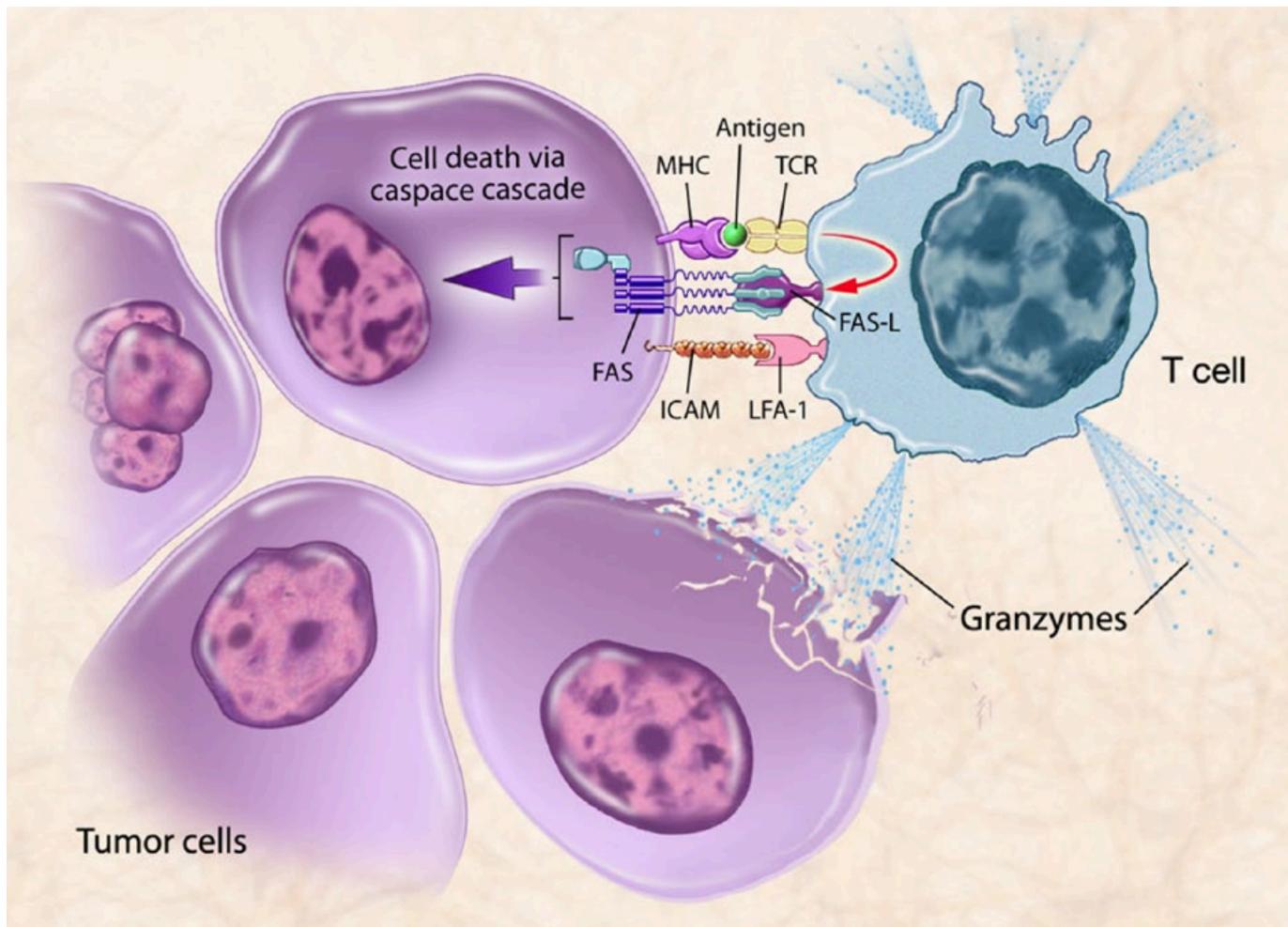


T cell receptor specificity and repertoire analysis resources

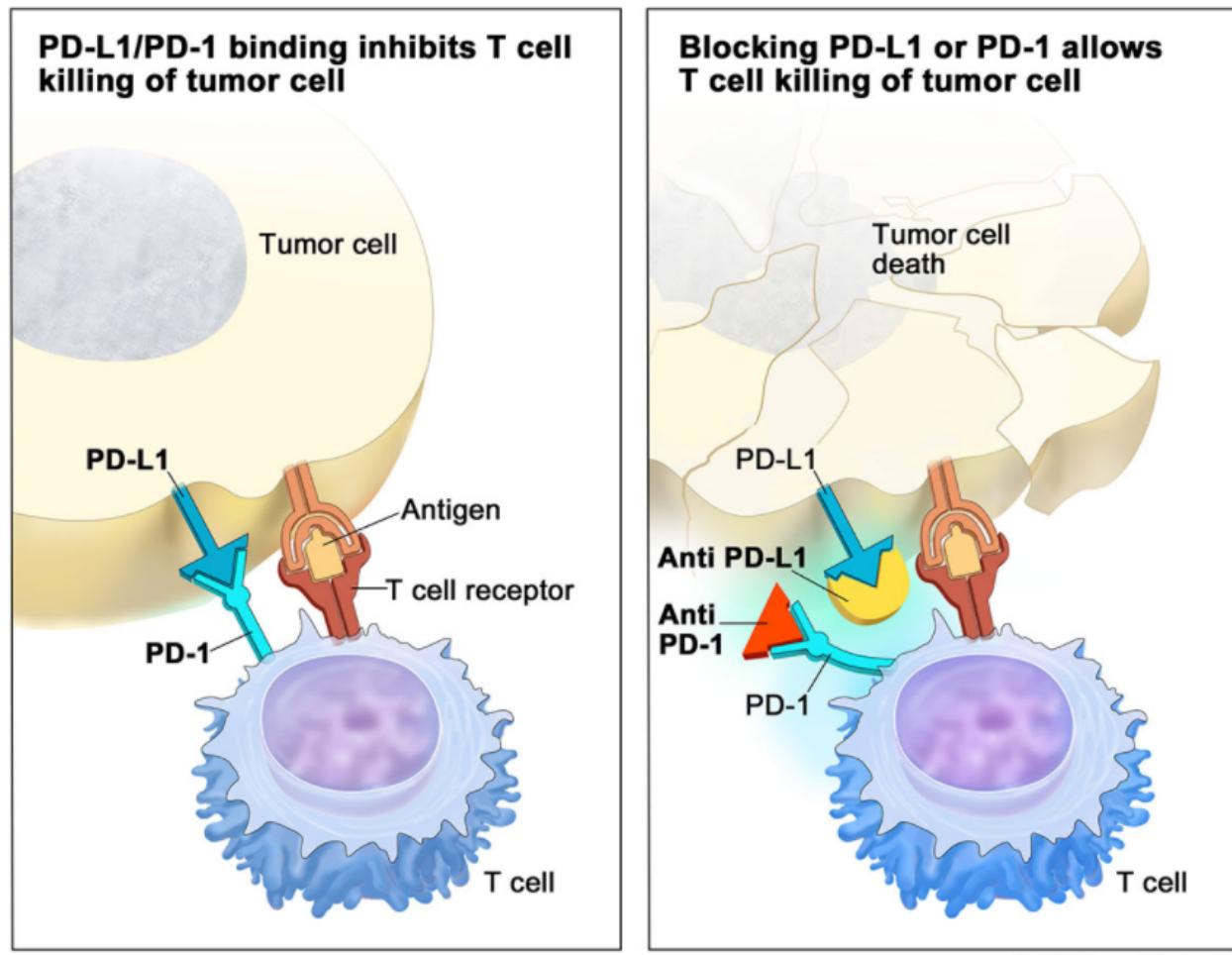
- T cell receptor sequences and annotations can be obtained from [IMGT/LIGM-DB](#)
 - e.g. V, D, and J segments of the gene TRB
 - Sequences are also stored in GenBank
- [MIXCR](#)
 - Analysis of T- and B- cell receptor repertoire sequencing data
 - Alignment of sequence reads to reference TCR sequences to determine T cell clonotypes
- [VDJdb](#)
 - A curated database of T-cell receptor sequences with known antigen specificity

Count	Freq	CDR3 Nucleotide	CDR3 Amino Acid	V	D	J
2690	0.05	TGTGCCAGTAACCCAACAGGGC	CASNPTGPDNSPLHF	TRBV28	TRBD1	TRBJ1-6
834	0.02	TGTGCCTGTGACCCTTGTACTG	CACDPLYWGRSTGTDKLIF	TRDV2	TRDD3	TRDJ1

T cell mediated cell death



Many tumors express checkpoint molecules to suppress this immune cell death signaling

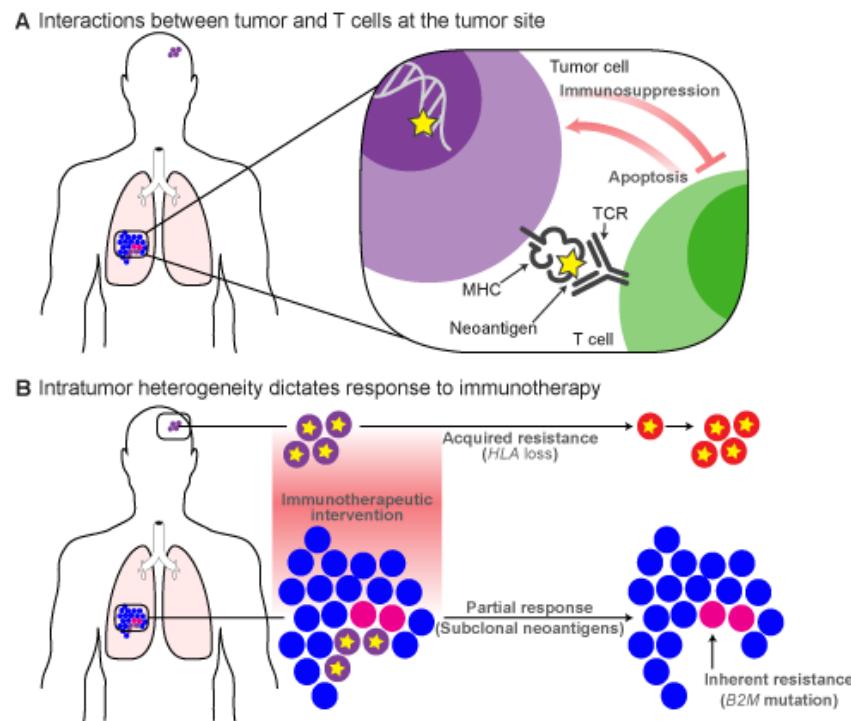


© 2015 Terese Winslow LLC
All rights reserved.

Checkpoint blockade/inhibitors such as pembrolizumab intercept these immunosuppressive signals and allow T cell mediated death to resume

There are many other diverse mechanisms of resistance to immune response

- Loss of the tumor variant
 - deletion of that genomic region
 - outgrowth of a sub-clonal population that does not contain the neoepitope variant
- Interfere with correct processing of the neoepitope
- Reduced expression or loss of the MHC molecule that presents the neoepitope
 - Or important co-factors such as B2M
- Interference with signaling cascade that leads to apoptosis
- As with all cancer treatment modalities, mechanisms of resistance will be diverse
 - Inherent/acquired, intrinsic/extrinsic, ephemeral/sustained, direct/indirect ...

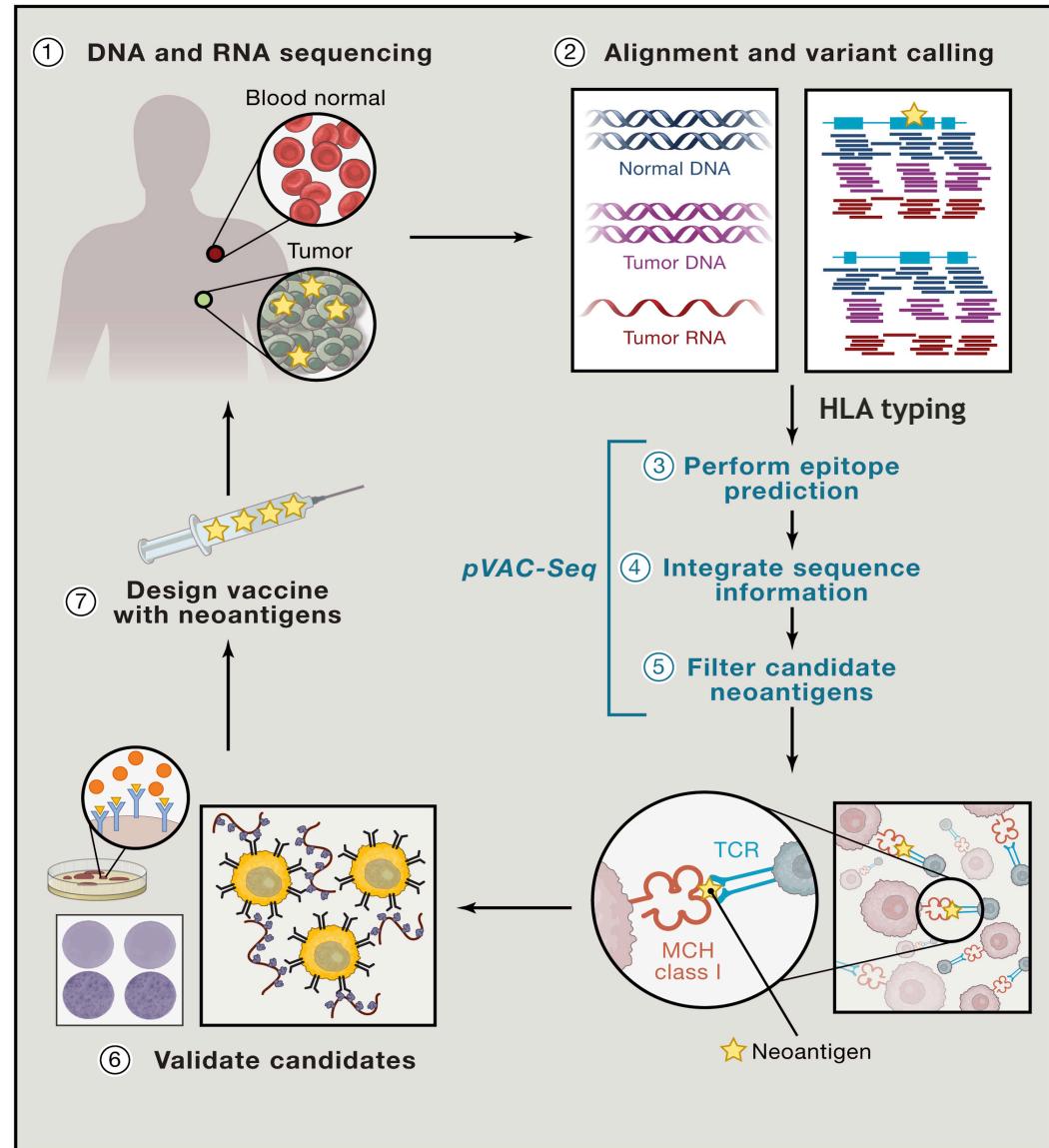


Identifying and characterizing neoepitopes is one critical element of this area of research

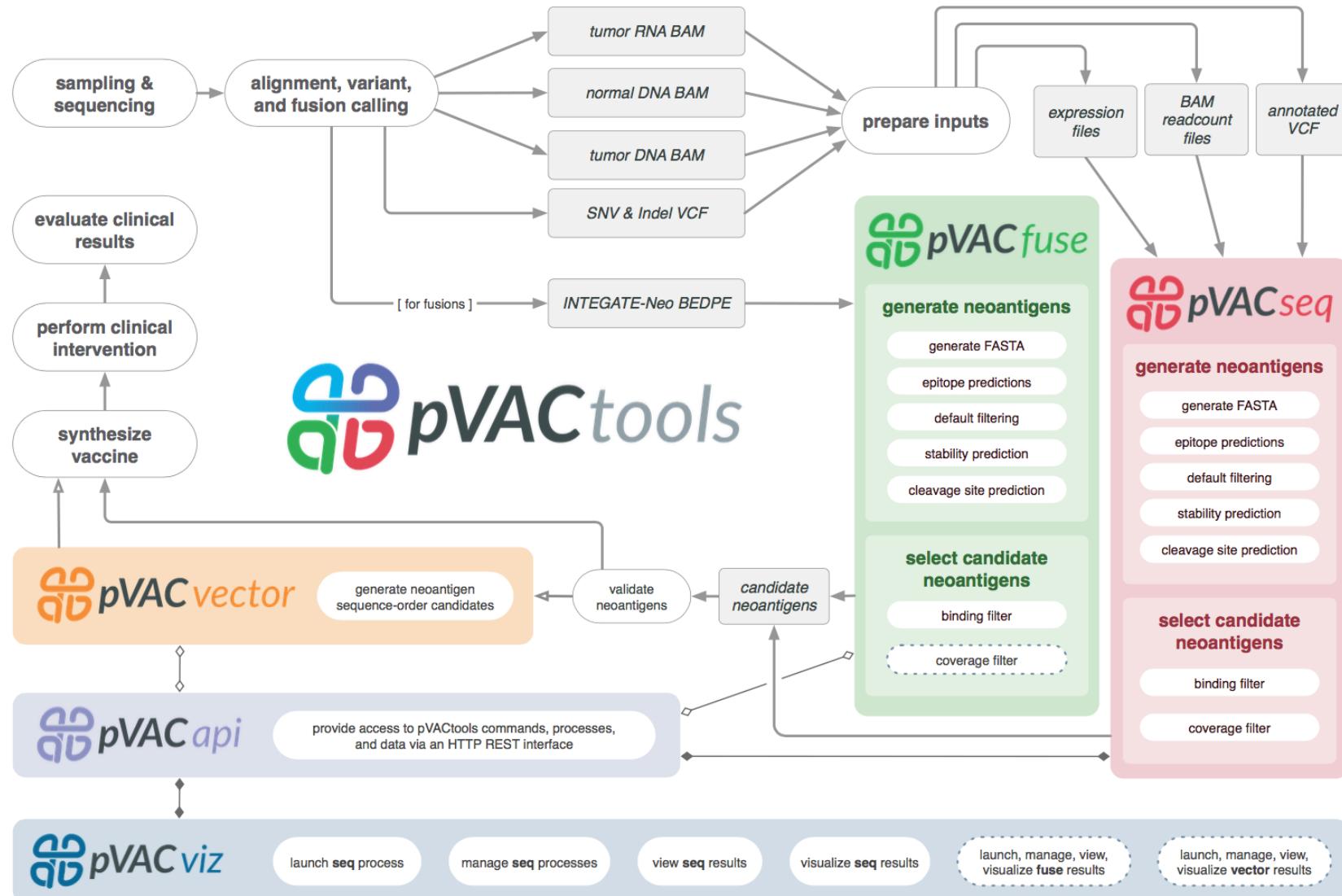
For understanding response to checkpoint inhibitors and for designing personalized cancer vaccines that direct an immune attack

An example neoepitope characterization workflow

Jasreet Hundal & Katie Campbell
Shirley X. Liu, Elaine R. Mardis.
Applications of Immunogenomics
to Cancer. Cell Press. 2017.



pVACtools (pvactools.org) is one of several neoepitope identification/prioritization pipelines



Example result from pVACtools

	A	B	C	D	E	F	G	H	I	J
1	Chr	Pos	Reference	Variant	Transcript	Gene Name	Variant Type	Mutation	Protein Position	
2	19	48197614	G	A	ENST00000396720	GLTSCR1	missense	V/I	843	
3	12	25398283	C	A	ENST00000256078	KRAS	missense	G/V	12	
4	7	6647640	C	T	ENST00000344417	C7orf26	missense	A/V	400	
5										
6	HLA Allele	Peptide Length	Sub-peptide Position	Mutation Position	MT Epitope Seq	WT Epitope Seq	Best MT Score Method	Best MT Score	WT Score	WT/MT Fold Change
7	HLA-B*07:02	9	11	1	IPPPASNPA	VPPPASNPA	SMMPMBEC	226.03	305.6	1.352
8	HLA-A*03:01	9	7	5	VVGAVGVGK	VVGAGGVGK	NetMHCpan	168.9	246.7	1.461
9	HLA-B*07:02	10	7	5	SPHAVLPPGF	SPHAALPPGF	NetMHCpan	66	49	0.742
10										
11	Tumor DNA Depth	Tumor DNA VAF	Tumor RNA Depth	Tumor RNA VAF	Normal Depth	Normal VAF	Gene Expression			
12	163	7.36	4	50	88	0	6.13549			
13	165	6.02	44	15.91	101	0	11.9842			
14	475	4.62	17	5.88	312	0	17.2583			
15										
16	Median MT Score	Median WT Score	Median Fold Change	NetMHC WT Score	NetMHC MT Score	NetMHCpan WT Score	NetMHCpan MT Score			
17	260.18	340.99	1.311	803.06	680.13	1635.3	1540.5			
18	230.87	292.66	1.268	277.44	202.62	246.7	168.9			
19	200.28	201.66	1.007	59.68	81.88	49	66			
20										
21	PickPock et WT Score	PickPock et MT Score	SMM WT Score	SMM MT Score	SMMPMBEC WT Score	SMMPMBEC MT Score				
22	340.99	260.18	335.65	251.13	305.6	226.03				
23	792.98	605.05	338.69	267.79	292.66	230.87				
24	573.19	573.19	203.62	230.05	201.66	200.28				
25										

Results that validate the immunogenicity of predicted neoepitopes in patients are still lacking

Summary of useful resources (questions?)

Databases

- [IPD-IMGT/HLA Database](#) (EBI)
- [Immune Epitope Database, IEDB](#) (La Jolla IAI)
- [IMGT database](#) (LIGM, Montpellier, France)

Tools

- [IEDB tools](#) - binding predictors
- [pVACtools](#) - vaccine design
- [OpenVax](#) - vaccine design
- [MIXCR](#) - TCR repertoire analysis

