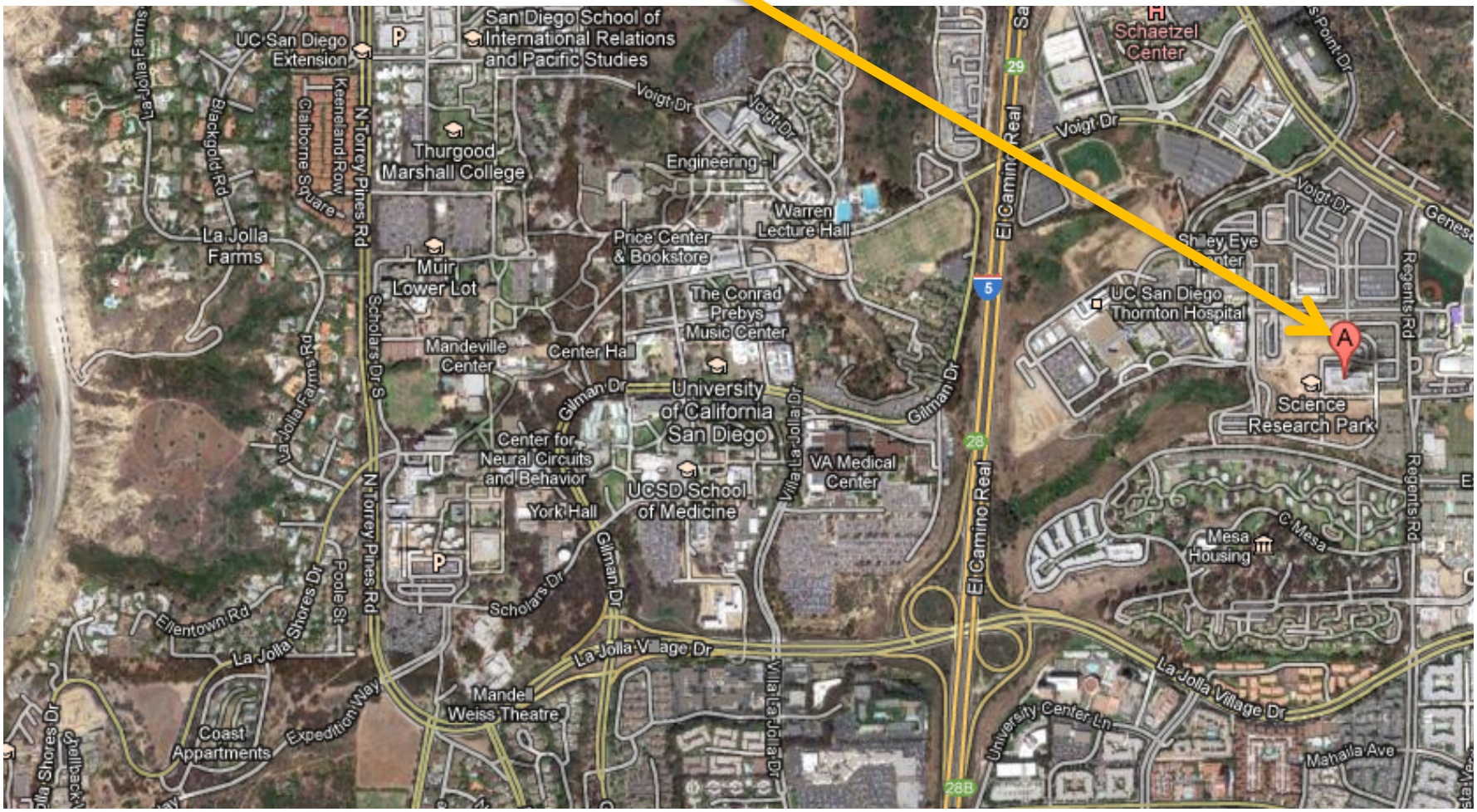
Immunoinformatics resources for the understanding of immunological information

A case study in personalized cancer immunotherapy

Bjoern Peters

La Jolla Institute for Allergy and Immunology

La Jolla Institute for Allergy and Immunology (LIAI)



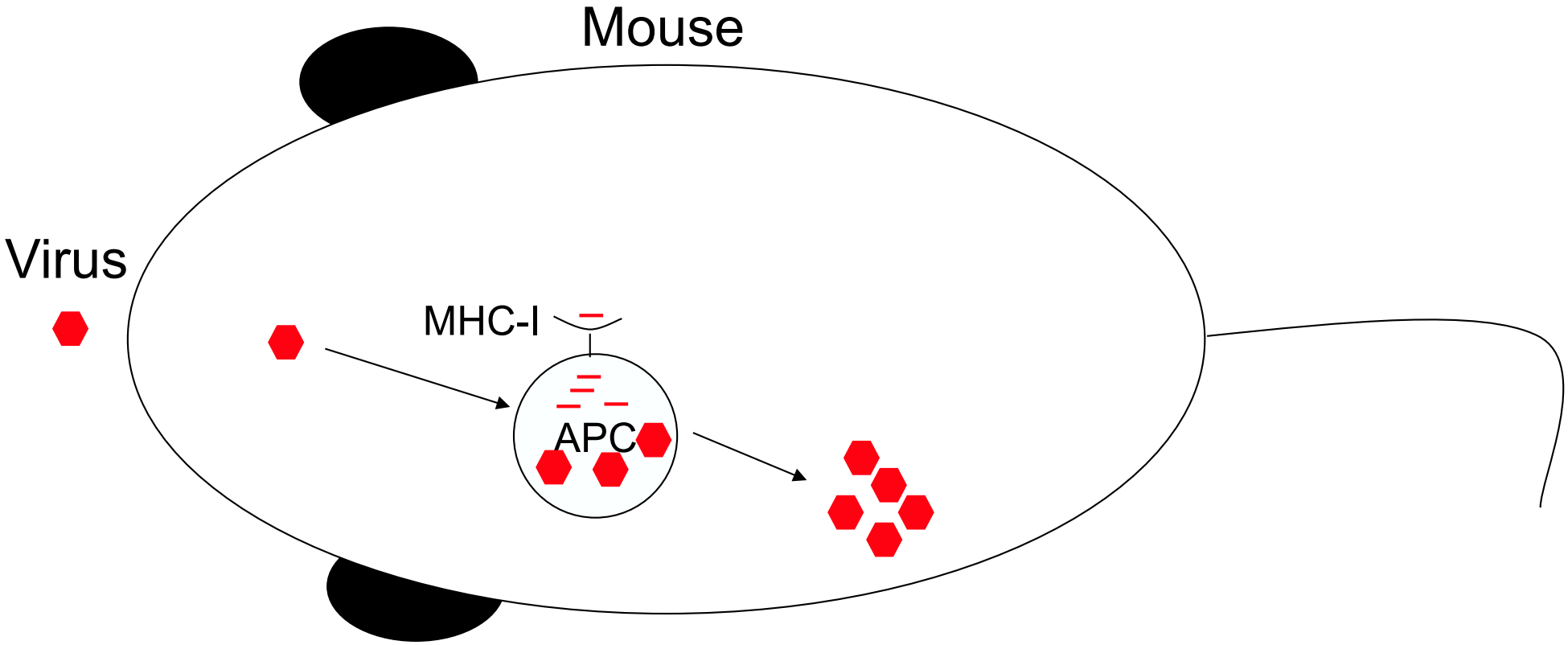
Overview

- Part I - Lecture: Biological Background
  - T cell immune responses target non-self entities
  - Cancer cells bear somatic mutations
  - Cancer immunotherapy aims to target immune responses to cancer cells
- Part II – Lecture: Bioinformatic guided approaches
  - Sequencing approaches identify tumor specific somatic mutations
  - HLA binding predictions can identify which of these will be immunogenic
- Part III – Hands on session: Design a personalized cancer vaccine

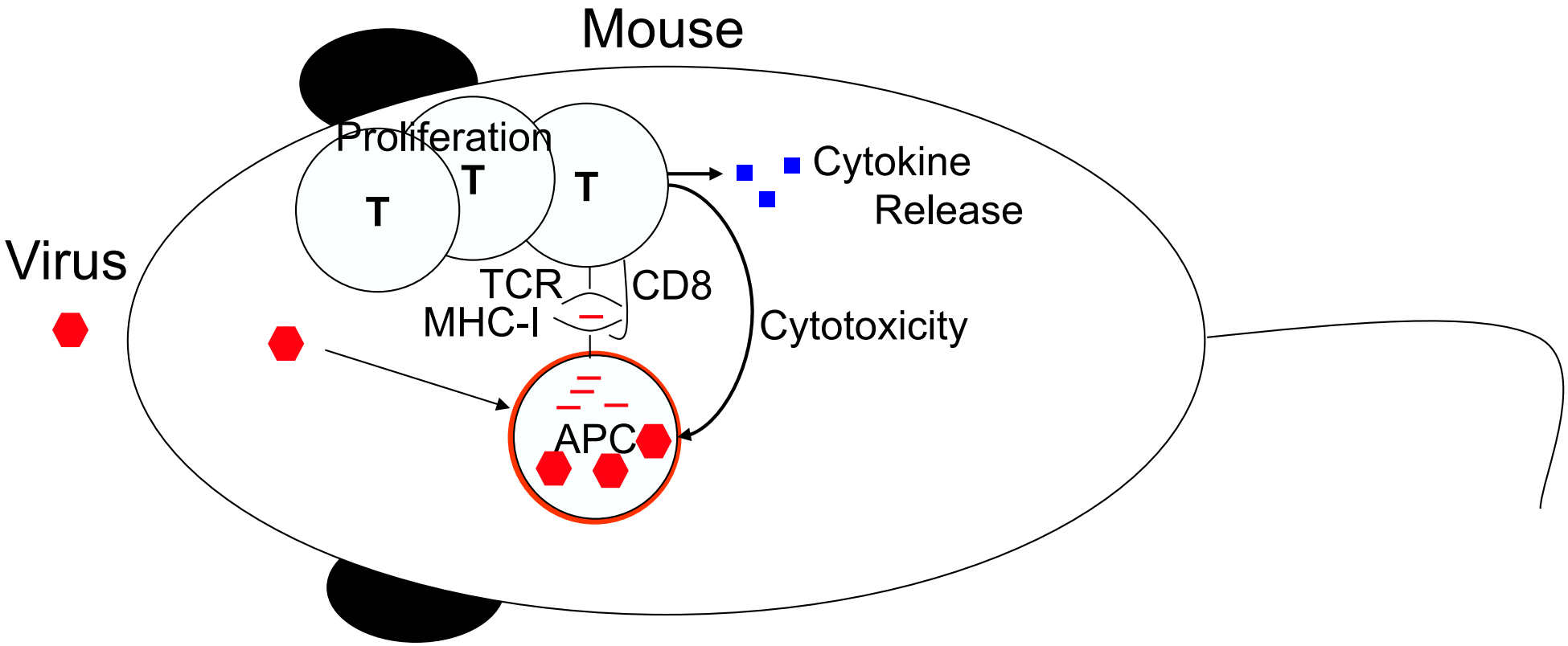
HLA molecules as sensors of non-self

HLA = Human MHC molecules

CD8^{+} T cell epitopes in viral infection

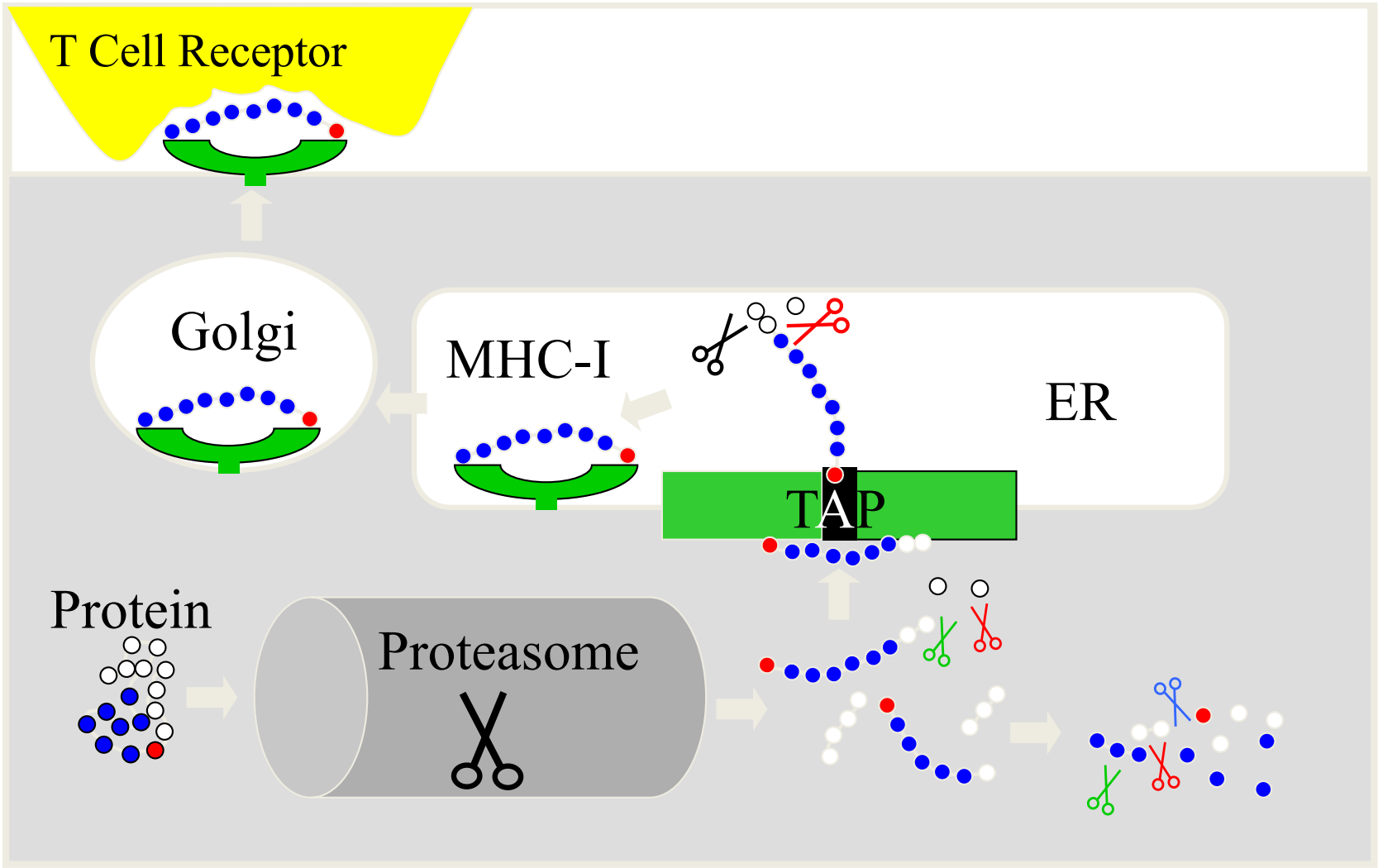


CD8$^+$ T cell epitopes in viral infection



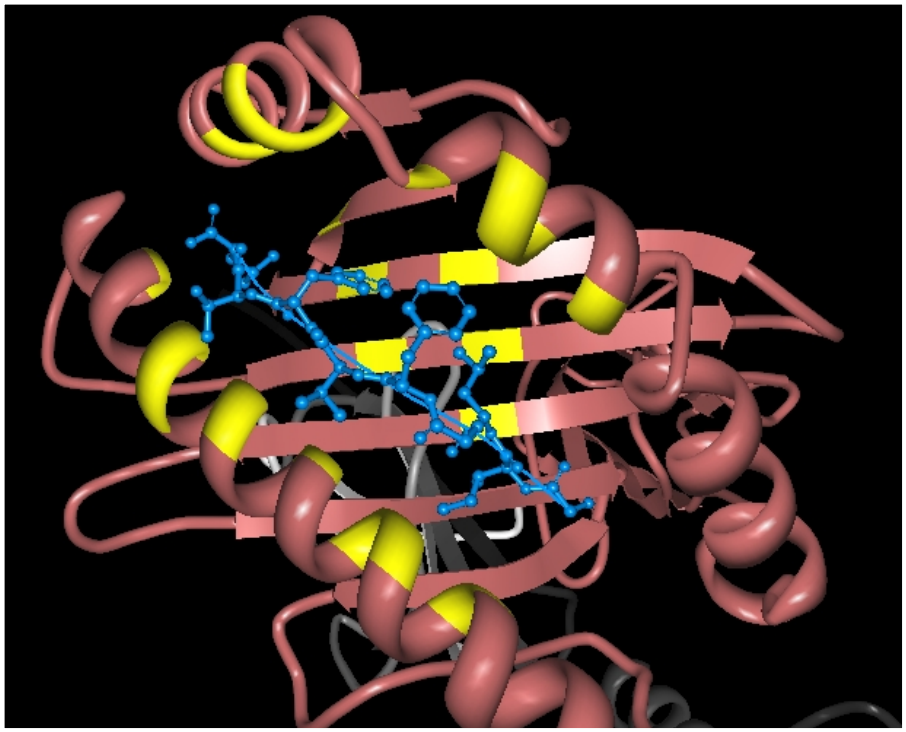
- How do peptides get loaded on MHC molecules?
- How do T cells distinguish self- from non-self peptides?

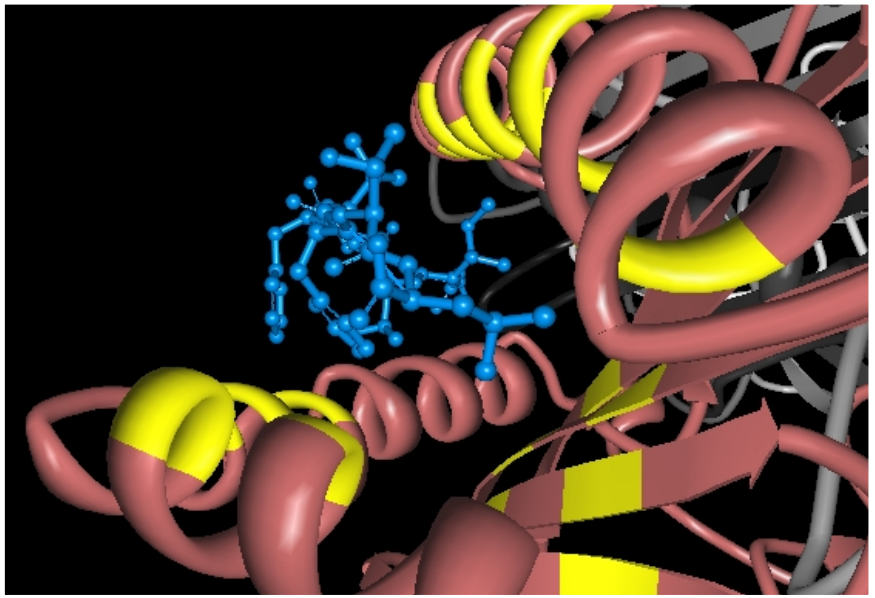
MHC I - Antigen processing and presentation pathway



Peters et al, J Mol Biol 2002, Bioinformatics 2003, J Immunol.2003; CMLS 2005 ; Assarson, J Immunol 2007

MHC:peptide binding mode

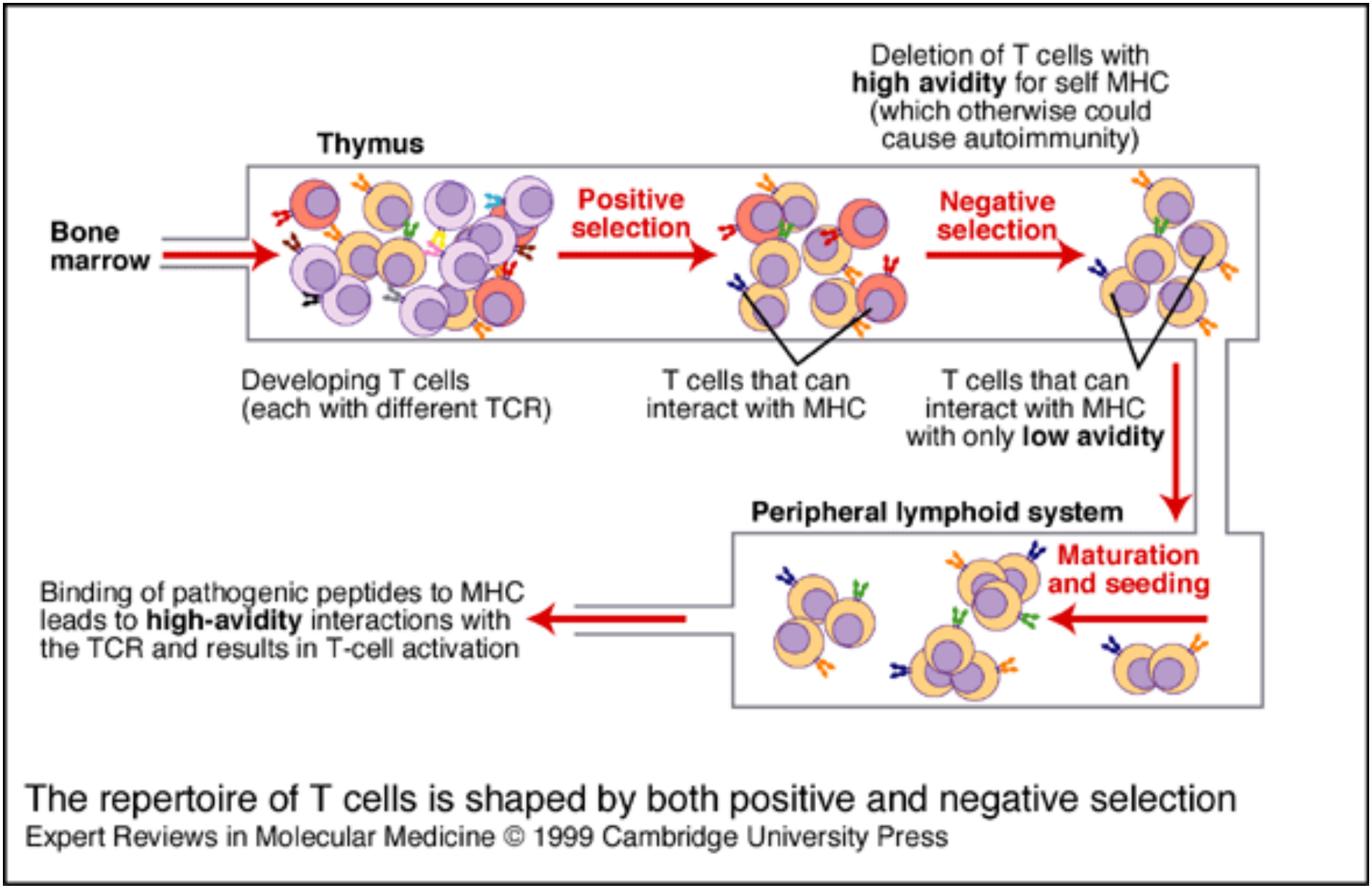




- Each human has 6 types of MHC molecules (alleles)
- >3000 alleles are known
- Distinct binding specificities → individual epitope repertoire

X-Ray Structure: Madden, Cell 1993.
Viewer: Beaver and Ponomarenko, Immunome Research, 2007

Self –reactive T cells are deleted during maturation

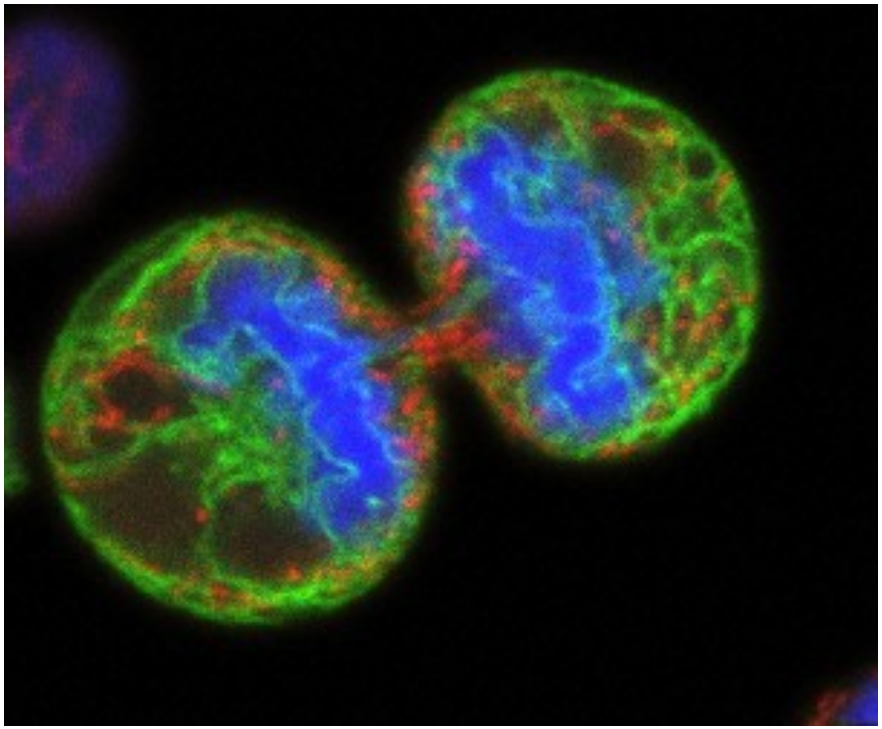


The repertoire of T cells is shaped by both positive and negative selection

Background: Cancer

What is cancer?

- All cancers derive from single cells that have acquired the characteristics of continually dividing in an unrestrained manner and invading surrounding tissues.
- Cancer cells behave in this abnormal manner because of changes in the DNA sequence of key genes, which are known as cancer genes. Therefore all cancers are genetic diseases.



Human melanoma cell undergoing cell division
Credit: Paul Smith & Rachel Errington, Wellcome Images

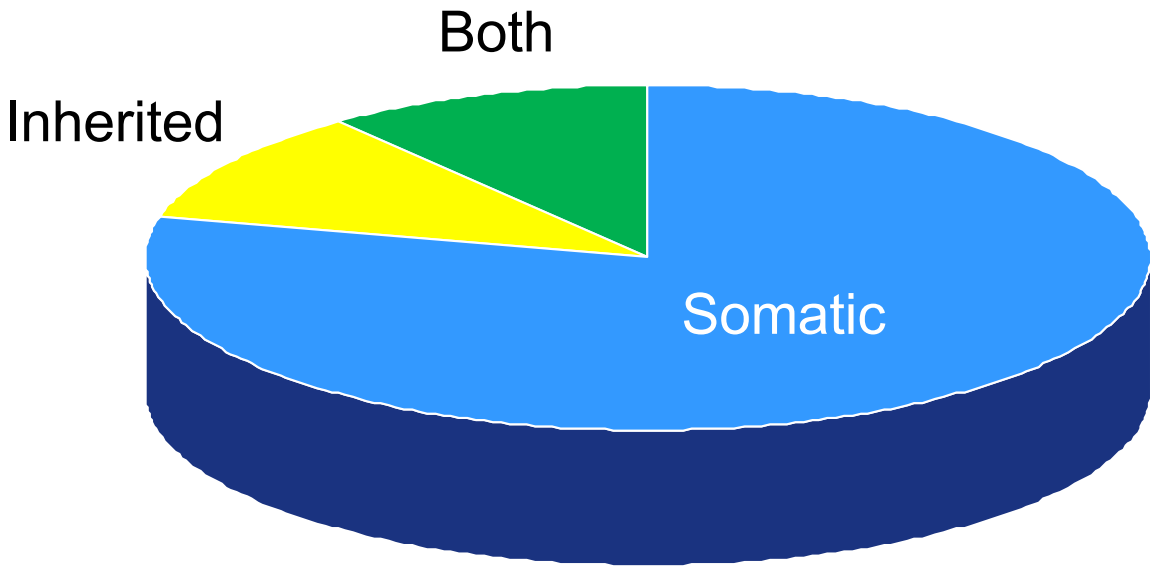
What is a mutation?

- **Germline mutation**
  - A change in the DNA sequence that can be inherited from either parent
- **Somatic mutation**
  - A change in the DNA sequence in cells other than sperm or egg
  - The mutation is present in the cancer cell and its offspring, but not in the patient's healthy cells

Mutations & cancer genes

- Cancer genes are causally implicated in *oncogenesis*
- Mutations in cancer genes can occur somatically or can be inherited.
- Mutations in some cancer genes can be inherited from parents, in which case they are present in every cell of the body. Such people are at a higher risk of developing cancer.
- Somatic mutations can occur in any of the cells of the body except the germ cells (sperm and egg) and therefore are not passed on to children.

Importance of somatic DNA changes in human cancer



Only 5 –10% of cancer cases have a clear hereditary component, e.g. *BRCA1* and *BRCA2* in breast cancer

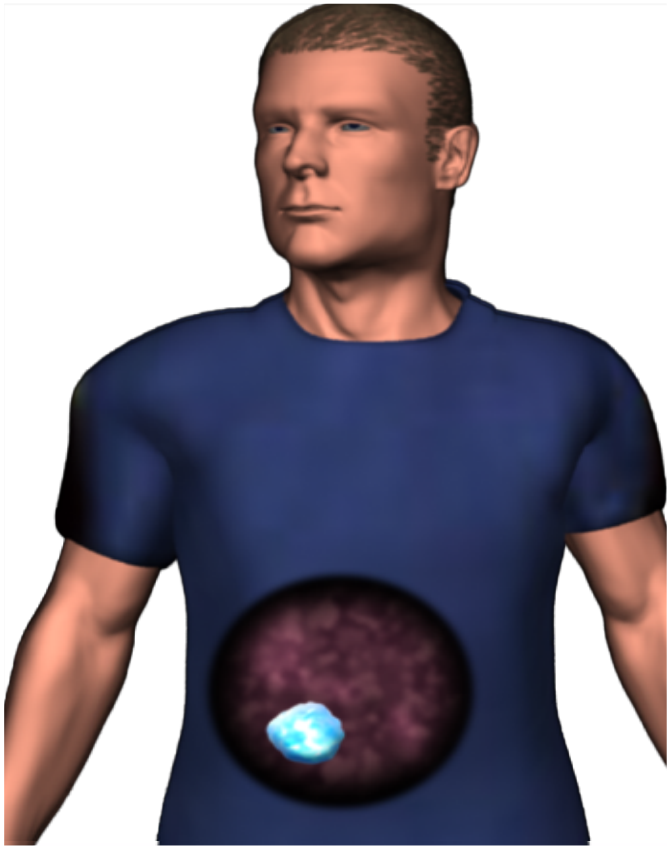
Even in those cases where susceptibility is clearly inherited, somatic changes are required for cancer to develop

Examples of mutations

Sequence 1	Sequence 2	Type
ACTCGTTAGGCA	ACTCCTTAGGCA	Substitution
ACTCGTTAGGCA	ACTCGGCA	Deletion
ACTCGTTAGGCA	ACTCGTTATCAGGCA	Insertion
ACTCGTTAGGCA	ACTTTGCAGGCA	Inversion
ACTCGTTAGGCA	ACTCGTTAGTTAGGCA	Duplication

Cancer progression

Mutations in multiple cancer genes are required for the development and progression of a single cancer



Benign Tumour

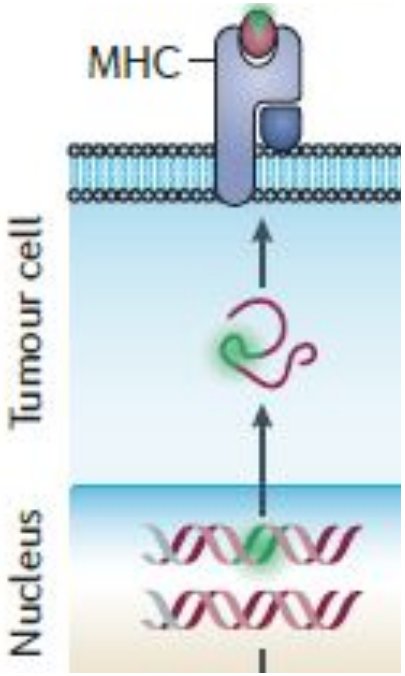
In situ cancer

Invasive cancer

Metastatic cancer

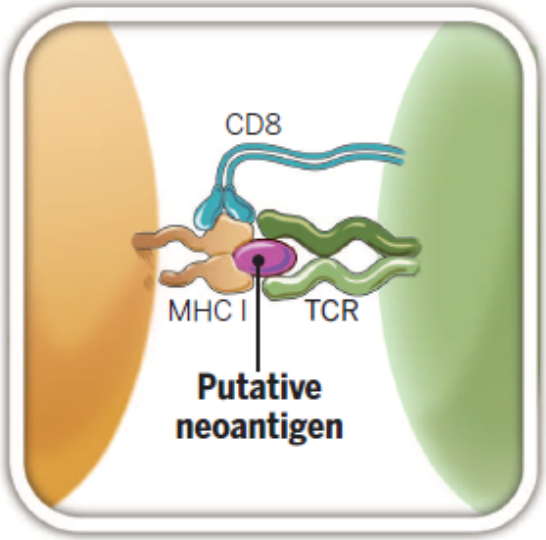
Neoepitopes (Neoantigens)

- Cancers genomes accumulate mutations
- Mutations in coding regions are translated in mutated protein sequences
- Mutated peptides can be presented as epitopes on MHC to T cells



Neoepitopes are presumably recognized by tumor-infiltrating lymphocytes **(TILs)**

Neoepitopes are highly tumor-specific!



Coulie et al, Nat Rev Cancer. 2014 Feb;14(2):135-46
Schumacher & Schreiber, Science. 2015 Apr 3;348(6230):69-74

NCT

Cancer Immunotherapy

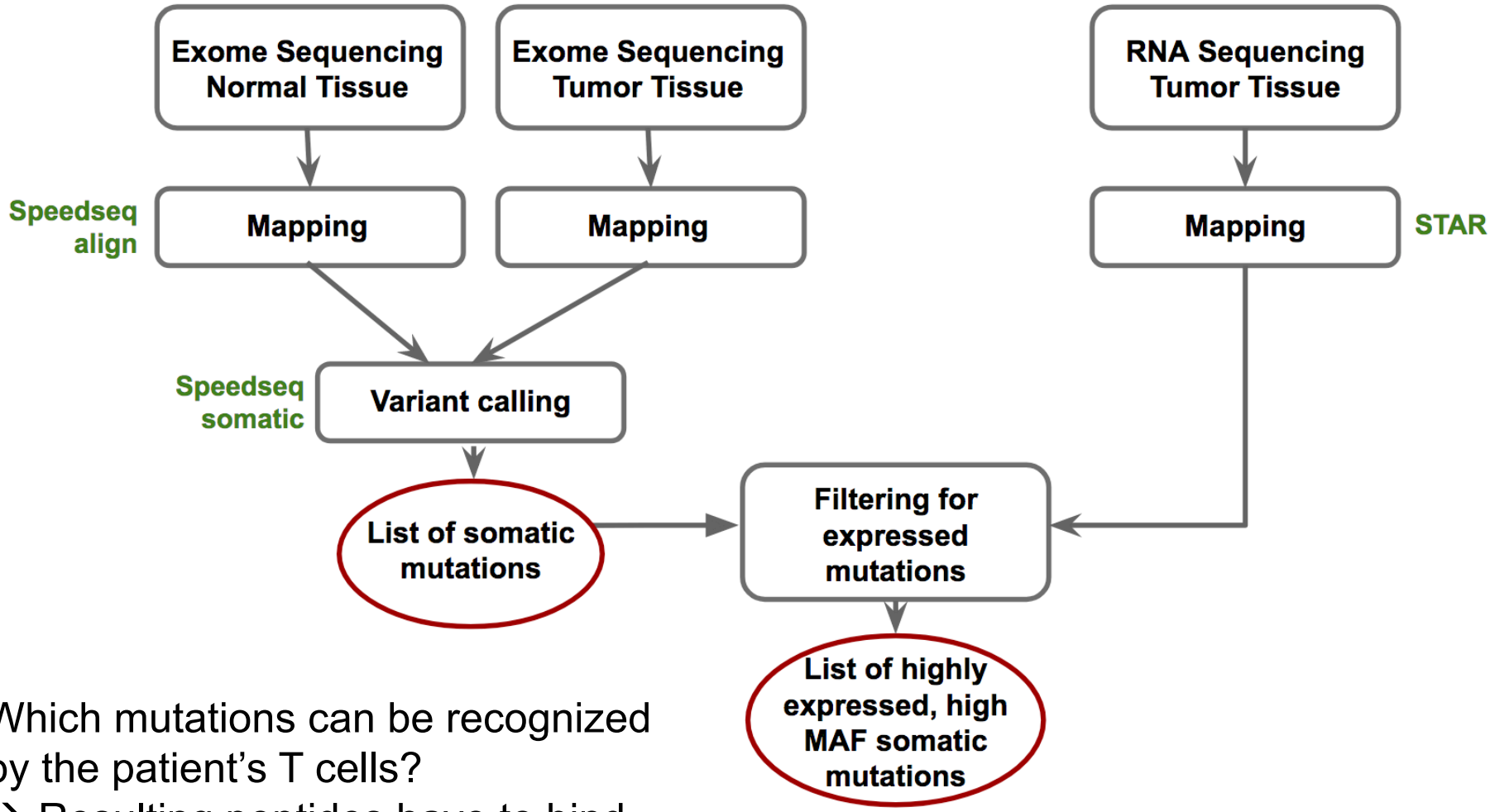
- Vaccination: Introduce or boost an immune response against a specific target (antigen)
- Cancer cells contain non-self antigens that *could* be recognized by T cells, but presence of cancer means this mechanism has failed, typically by the tumor suppressing immune responses
- Checkpoint blockade treatments: Block immune suppressive mechanisms to boost T cell immune responses against cancer cells.
- Problem: Checkpoint blockade is unspecific, and will also boost unwanted autoimmune responses
- Personalized Cancer Immunotherapy: Boost anti-tumor response with vaccine containing peptides corresponding to cancer mutations that can be recognized by T cells.

→ How can such a vaccine be designed?

Overview

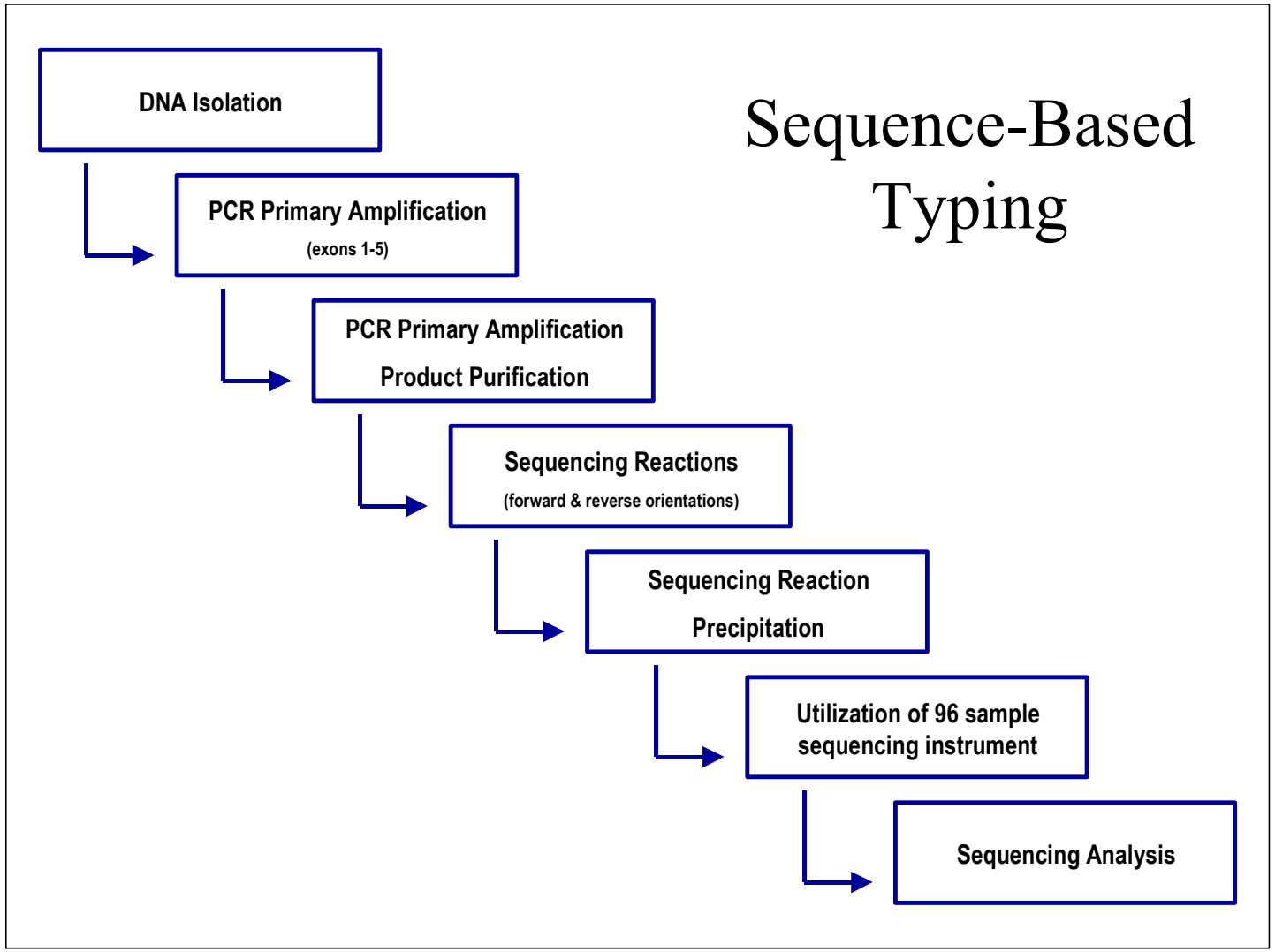
- Part I - Lecture: Biological Background
  - T cell immune responses target non-self entities
  - Cancer cells bear somatic mutations
  - Cancer immunotherapy aims to target immune responses to cancer cells
- Part II – Lecture: Bioinformatic guided approaches
  - Sequencing approaches identify tumor specific somatic mutations
  - HLA binding predictions can identify which of these will be immunogenic
- Part III – Hands on session: Design a personalized cancer vaccine

DNA and RNA sequencing identifies tumor specific somatic mutations



Which mutations can be recognized by the patient's T cells?
→ Resulting peptides have to bind HLA molecules of the patient

HLA Typing: Targeted sequencing of HLA locus

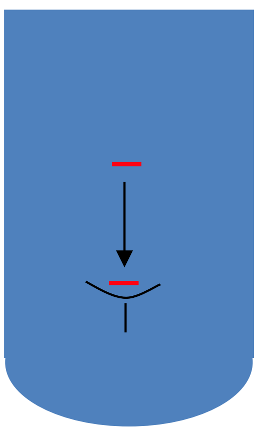


•http://www.ashi-hla.org/publicationfiles/ASHI_Quarterly/25_2_2001/highthrusbt3.htm

Measuring and predicting MHC:peptide binding

Experimental Basis: MHC Binding Assay

List of peptides with allele specific binding affinity



Sequence	IC_{50}
QIVTMFEAL	3.6
LKGPDIYKG	308
NFCNLTSAF	50,000
AQSQCRTFR	38,000
CTYAGPFGM	143
CFGNTAVAK	50,000
. . .	

$\log(IC_{50}) \sim$ Binding free Energy

low IC_{50} → high affinity

Impossible to measure all peptides

→ Predict binding peptides using machine learning

Find function F_i in $F_1, F_2, F_3, \ldots$

F_i (Sequence) ≈ Affinity

Many different approaches (ANN, SVM, HMM, LP, ...)

T cell epitope mapping

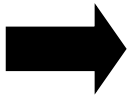
ORF 1	M G Q I V T M F E A L P H I I D E V I N I V I I V L I V I T G I K A V Y N ...
ORF 2	M G L K G P D I Y K G V Y Q F K S V E F D M S H L N L T M P N A C S A N N ...
ORF 3	M H N F C N L T S A F N K K T F D H T L M S I V S S L H L S I D G N S N Y ...
ORF 4	M S A Q S Q C R T F R G R V L D M F R T A F G G K Y M R S G W G W T G S D ...
ORF 5	M H C T Y A G P F G M S R I L L S Q E K T K F F T R R L A G T F T W T L S ...
ORF 6	M K C F G N T A V A K C N V N H D A E F C D M L R L I D Y N K A A L S K F ...
ORF 7	M L M R N H L L D L M G V P Y C N Y S K F W Y L E H A K T G E T S V P K C ...

Calculate scoring matrix from affinities

Machine learning PSSM = Minimize the difference between predicted and measured binding affinities by varying the matrix values

N peptides with measured binding affinities

log (IC50)	Peptide
0.50	FQPQNGSFI
0.72	ISVANKIYM
2.37	RVYEALYYV
3.42	FQPQSGQFI
3.46	LYEKVKSQL
4.07	FKSVEFDMS
4.18	FQPQNGQFH
4.24	VLMLPVWFL
4.39	YMTLGQVVF
4.40	EDVKNAVGV
4.90	VFYEQMKRF
	...



	HLA A*0201								
	1	2	3	4	5	6	7	8	9
A	-0.3	0.8	-0.3	-0.3	-0.2	-0.3	0.0	0.0	-0.9
C	0.2	0.9	0.0	0.3	-0.5	-0.1	0.1	0.2	0.4
D	0.8	0.9	-0.4	-0.3	0.3	0.2	0.4	0.3	0.6
E	0.6	-0.4	0.7	-0.2	0.1	-0.4	-0.2	-0.2	-0.5
F	-1.3	0.5	-0.5	0.1	-0.1	0.0	-0.3	-0.4	-0.8
G	-0.2	0.1	0.3	-0.1	0.0	0.4	0.3	-0.1	0.2
H	1.1	0.9	-0.1	0.4	0.1	0.2	0.0	0.2	0.8
I	-0.4	-0.7	-0.4	0.1	-0.1	-0.4	-0.5	0.5	-1.4
K	-0.3	0.0	1.1	0.1	0.1	0.6	0.9	0.2	0.9
L	0.0	-1.9	-0.4	-0.2	0.0	-0.2	0.0	-0.1	-1.1
M	-0.7	-1.2	-0.7	0.2	-0.6	0.0	0.0	0.0	-0.8
N	-0.1	0.3	0.1	-0.3	-0.1	-0.3	0.0	0.2	0.7
P	1.2	0.5	0.6	-0.3	0.4	0.0	-0.4	-0.5	0.7
Q	0.4	-1.1	0.0	-0.1	0.4	-0.2	-0.3	0.2	0.7
R	-0.2	0.9	1.0	0.3	0.1	0.4	0.7	0.0	0.9
S	-0.3	0.1	0.1	-0.4	0.1	0.3	-0.2	-0.1	0.2
T	-0.2	-0.5	0.1	0.4	0.1	-0.5	0.2	0.0	-0.1
V	-0.1	-0.9	-0.1	0.2	0.0	-0.3	0.1	0.1	-1.9
W	0.0	0.7	-0.5	-0.2	-0.1	0.2	-0.3	-0.1	0.4
Y	-0.3	0.2	-0.6	0.2	0.0	0.4	-0.4	-0.3	0.8

Offset: 4.3

Predictions available as webserver

- Immune Epitope Database (IEDB) Analysis resource
- http://tools.iedb.org/mhci/

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)	
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
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele Length — Upload allele file
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	
	Submit Reset

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)	>Region 1 SPLPSQAMLDLMLSPDD >Region 2 DPGPDEAPWMPEAAPPV
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
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: ☑ Select MHC allele(s) Select HLA allele reference set: ☐	Allele Length Upload allele file
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	
	Submit Reset

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)	>Region 1 SPLPSQAMLDLMLSPDD >Region 2 DPGPDEAPWMPEAAPPV
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

Specify what to make binding predictions for

MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele — Length — *Upload allele file*

HLA-A*01:01
HLA-A*02:01
HLA-A*02:06
HLA-A*03:01
HLA-A*11:01
HLA-A*23:01
HLA-A*24:02
HLA-A*25:01
HLA-A*26:01
HLA-A*29:02
HLA-A*30:01
HLA-A*30:02
HLA-A*31:01
HLA-A*32:01

ecify Output

Sort peptides by	
Show	
Output format	
Email address (optional)	

Submit Reset

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)	>Region 1 SPLPSQAMLDLMLSPDD >Region 2 DPGPDEAPWMPEAAPPV
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

Specify what to make binding predictions for

MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*02:01 — Length: 9 — Upload allele file

Sp

Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	

Length options: 8, 9, 10, 11, 12, 13, 14, All lengths

Submit Reset

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)	>Region 1 SPLPSQAMLDLMLSPDD >Region 2 DPGPDEAPWMPEAAPPV
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
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: ☑ Select MHC allele(s) Select HLA allele reference set: ☐	Allele: HLA-A*02:01 — Length: 9 Upload allele file
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	
	Submit Reset

MHC-I Binding Predictions

Loading... please wait.

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	Reg 1	SPLPSQAMLDLMLSPDD
2	Reg 2	DPGPDEAPWMPEAAPPV

Prediction method: IEDB recommended | Low percentile_rank = good binders

Download result

Citations

Check to expand the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank
HLA-A*02:01	2	9	17	9	WMPEAAPPV	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	8	16	9	MLDLMLSPD	Consensus (ann/comblib_sidney2008/smm)	2.9
HLA-A*02:01	1	7	15	9	AMLDLMLSP	Consensus (ann/comblib_sidney2008/smm)	4.0
HLA-A*02:01	1	5	13	9	SQAMLDLML	Consensus (ann/comblib_sidney2008/smm)	7.7
HLA-A*02:01	1	6	14	9	QAMLDLMLS	Consensus (ann/comblib_sidney2008/smm)	26.0
HLA-A*02:01	2	5	13	9	DEAPWMPEA	Consensus (ann/comblib_sidney2008/smm)	32.0
HLA-A*02:01	1	1	9	9	SPLPSQAML	Consensus (ann/comblib_sidney2008/smm)	33.0
HLA-A*02:01	1	3	11	9	LPSQAMLDL	Consensus (ann/comblib_sidney2008/smm)	39.0
HLA-A*02:01	1	4	12	9	PSQAMLDLM	Consensus (ann/comblib_sidney2008/smm)	43.0

Evaluating binding predictions

- Percentile rank < 0.5% = high affinity binder
- Percentile rank 0.5%-1% = intermediate binder
- Percentile rank 1% - 2% = low affinity binder
- Percentile rank 2% - 5% = borderline
- Percentile rank >5% is a non-binder

Overview

- Part I - Lecture: Biological Background
  - T cell immune responses target non-self entities
  - Cancer cells bear somatic mutations
  - Cancer immunotherapy aims to target immune responses to cancer cells
- Part II – Lecture: Bioinformatic guided approaches
  - Sequencing approaches identify tumor specific somatic mutations
  - HLA binding predictions can identify which of these will be immunogenic
- Part III – Hands on session: Design a personalized cancer vaccine

Input data from actual patient

>P53_HUMAN Cellular tumor antigen p53 - Healthy Tissue

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD

>P53_HUMAN Cellular tumor antigen p53 - Tumor Tissue

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMLDLMLSPDDIEQWFTEDPGP
DEAPWMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFVHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEV

HLA typing results:
HLA-A*02:01, HLA-A*68:01
HLA-B*07:02 , HLA-B*35:01

Steps

- Step 1: Identify sequence regions that contain all 9-mer peptides that are only found in the tumor
- Step 2: Run HLA binding prediction to identify 9-mer peptides in the sequence regions unique to the tumor that can be presented to T cells
- Step 3: Select the top peptide for each HLA allele
- Step 4: What is the un-mutated form of the chosen peptides in the patient? What is their MHC binding affinity?
- Step 5: Are the peptides really specific for the tumor? Examine this using NCBI BLAST
- Step 6: Decide: Which peptide would you choose?

backup

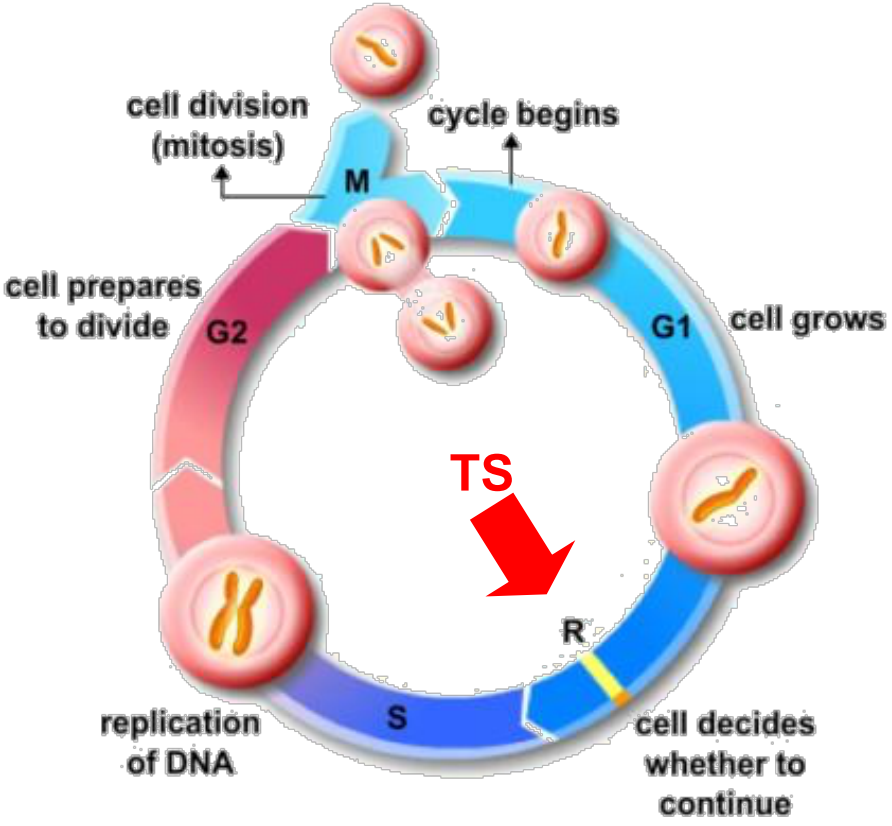
Cancer genes

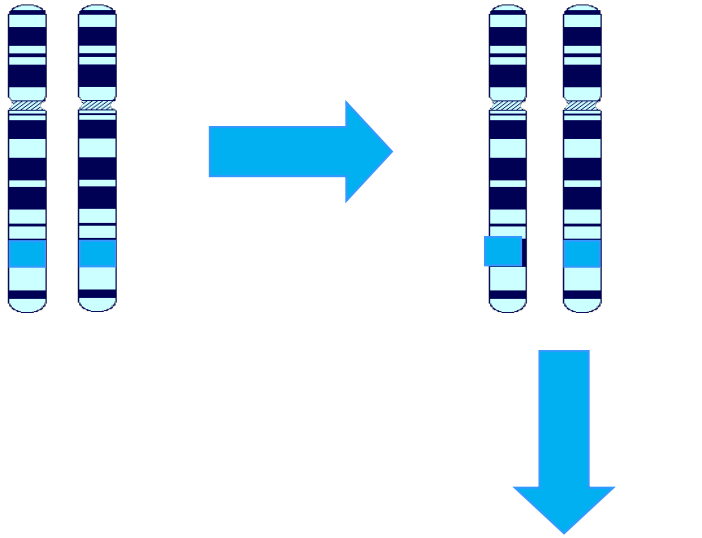
- There are two types of cancer genes:
  - **Tumour suppressor genes**
  - **Oncogenes**
- To date, we know of approximately 400 somatic "cancer genes" * but there are almost certainly more to be found
- COSMIC is a catalogue of somatic mutations found in cancer genes in human tumours and is available at: http://www.sanger.ac.uk/genetics/CGP/cosmic/

*(COSMIC v47release. July 2010)

Tumour suppressor gene

These genes normally function to PREVENT cell growth/division

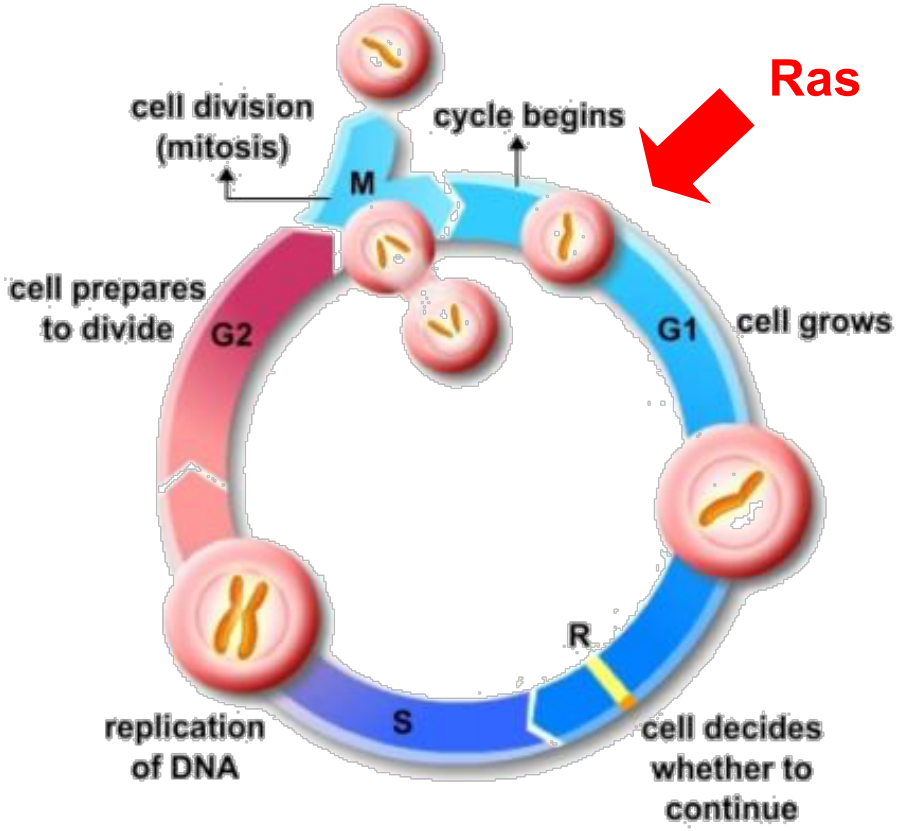


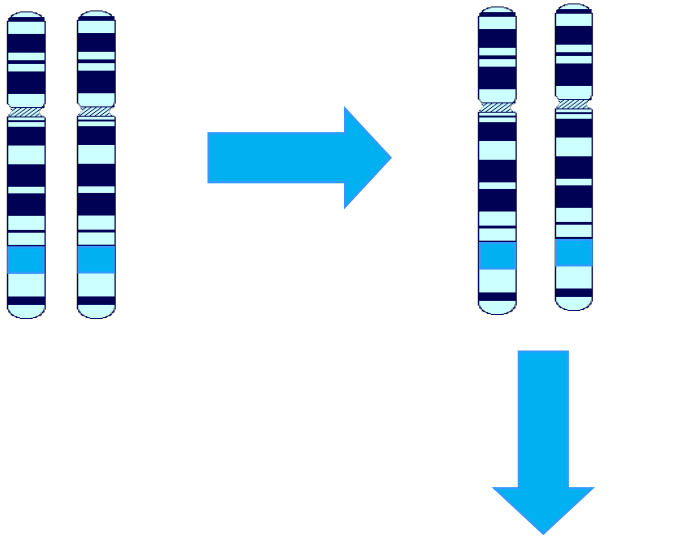


Cancer

Oncogene

Genes which normally function to PROMOTE cell growth/division in a controlled manner





Cancer