

# BGGN 213

## Cancer Genomics & Immunoinformatics

Lecture 18

Barry Grant  
UC San Diego

<http://thegrantlab.org/bggn213>

# Today's Menu

## Cancer Genomics

Brief review of cancer fundamentals,  
What is cancer and what causes it?

## Mining Cancer Genomic Data

**Hands-on analysis** to identify genomic changes in different cancers and identify new targets for therapy

## Towards personalized cancer treatments

Recap on how the immune system normally detects cancer cells and how we can predict mutations that can be recognized by T cells

## Cancer Immunoinformatics

**Hands-on analysis** to design personalized cancer vaccines

# What is Cancer?

“Cancer is a name given to a collection of related diseases, where some of the body’s cells begin to divide without stopping and spread into surrounding tissue”

Source: <https://www.cancer.gov>

It is estimated that cancer will strike 40% of people at some point in their lifetime with frequently devastating effects.

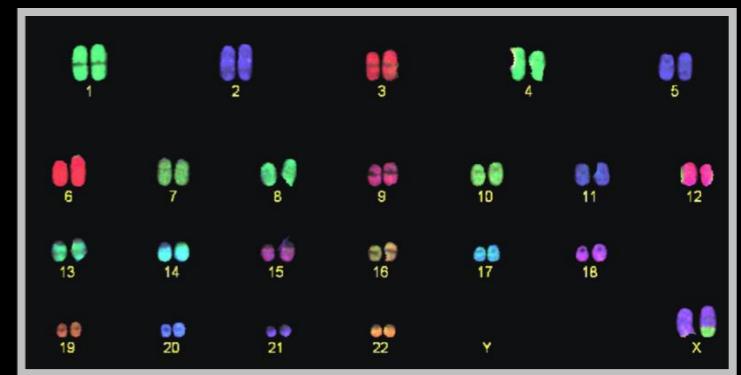
# What is Cancer?

“Cancer is a name given to a collection of related diseases, where some of the body’s cells begin to divide without stopping and spread into surrounding tissue”

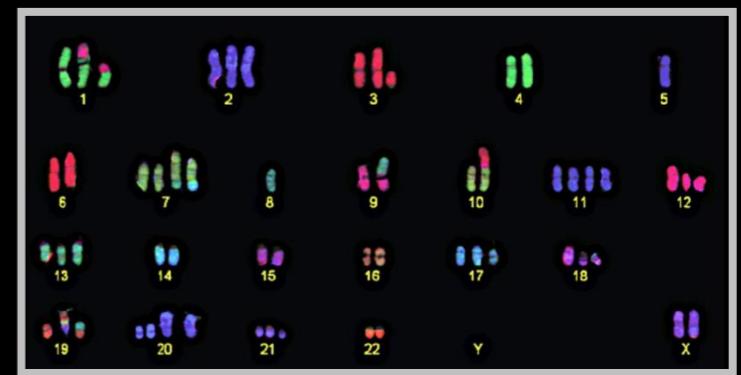
Source: <https://www.cancer.gov>

# Cancer is a disease of the Genome

- Caused by changes to genes that control the way our cells function, especially how they **grow and divide**.
- A major challenge in treating cancer is that every tumor is different: Each person's cancer has a unique combination of genetic changes (both “driver” & “passenger”).
- As the cancer continues to grow, additional changes will occur.



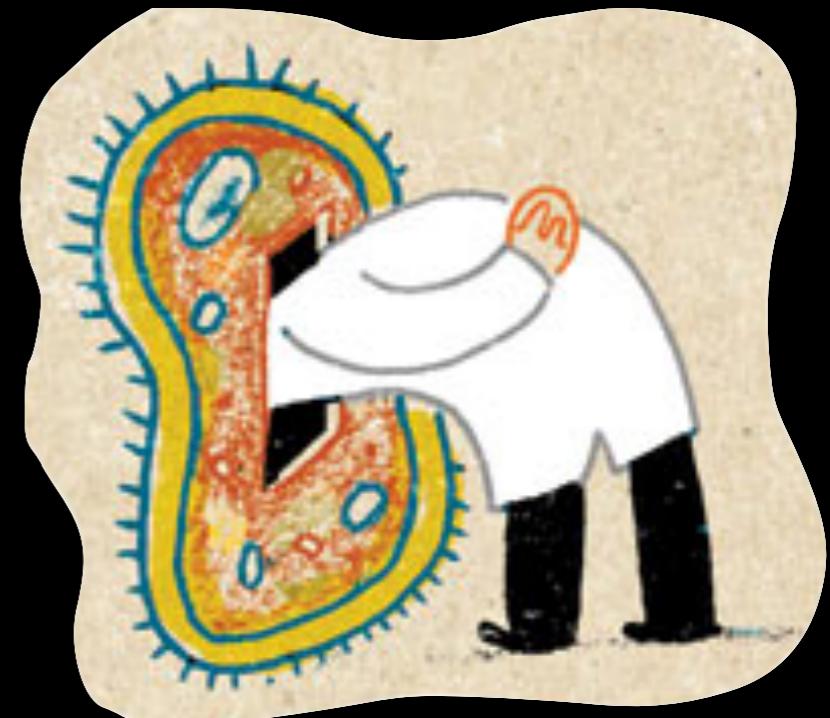
Healthy 46 chromosomes



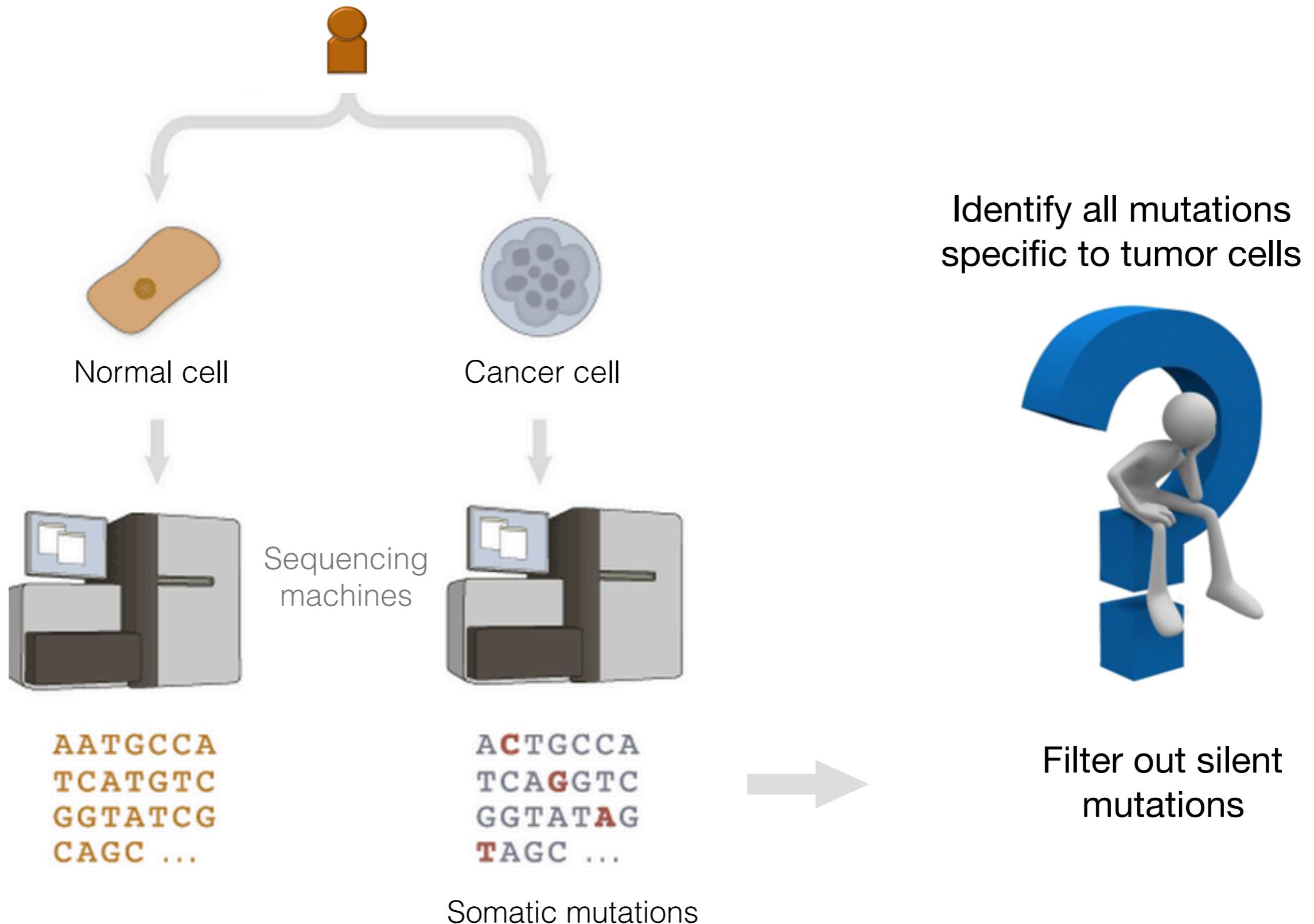
Example cancer 59 chromosomes

# Goals of Cancer Genome Research

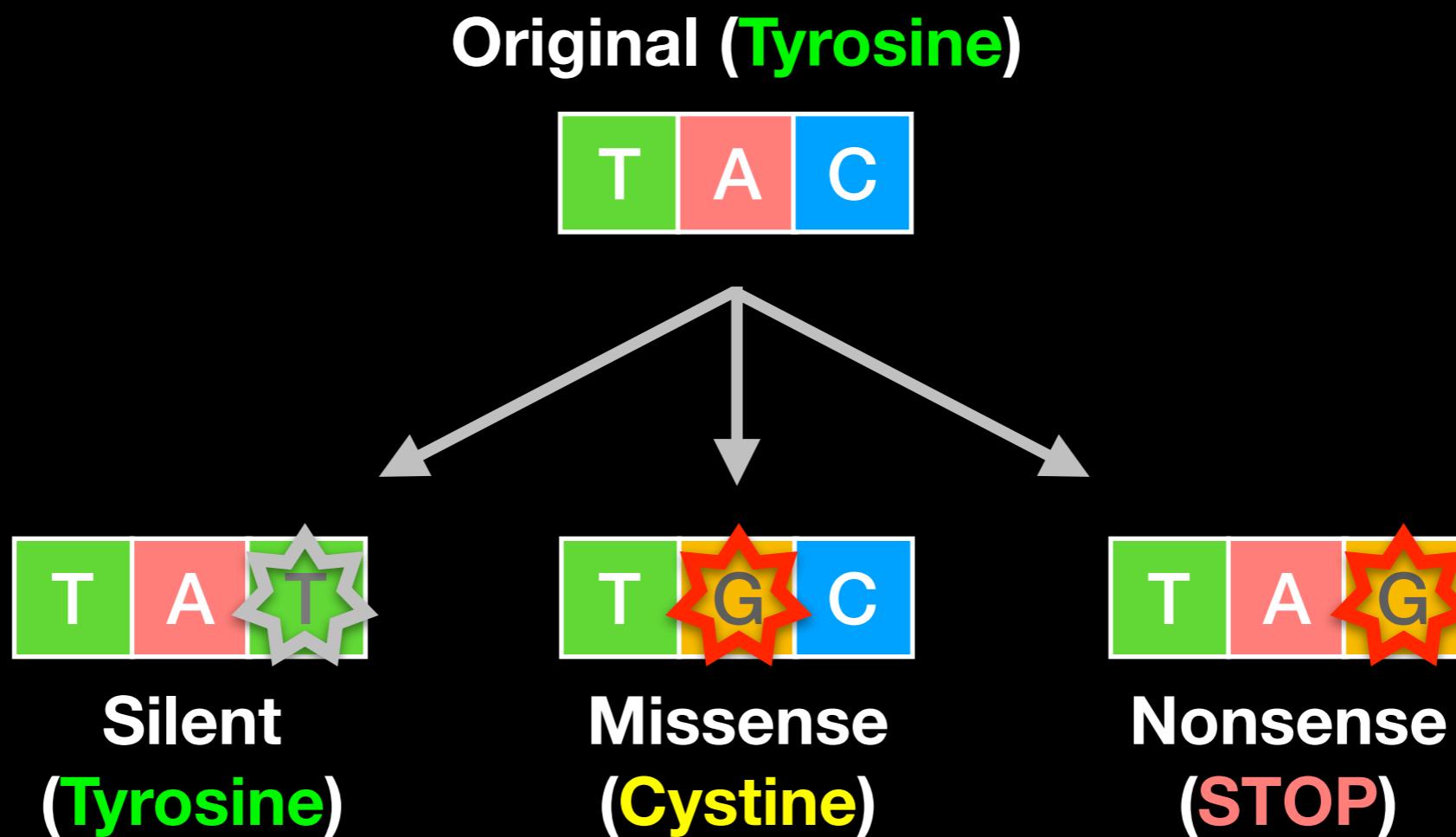
- Identify changes in the genomes of tumors that drive cancer progression
- Identify new targets for therapy
- Select drugs based on the genomics of the tumor
- Provide early cancer detection and treatment response monitoring
- Utilize cancer specific mutations to derive neoantigen immunotherapy approaches



# Finding Cancer Associated Mutations



# Mutations detected: Point mutations



# Mutations detected: Indels

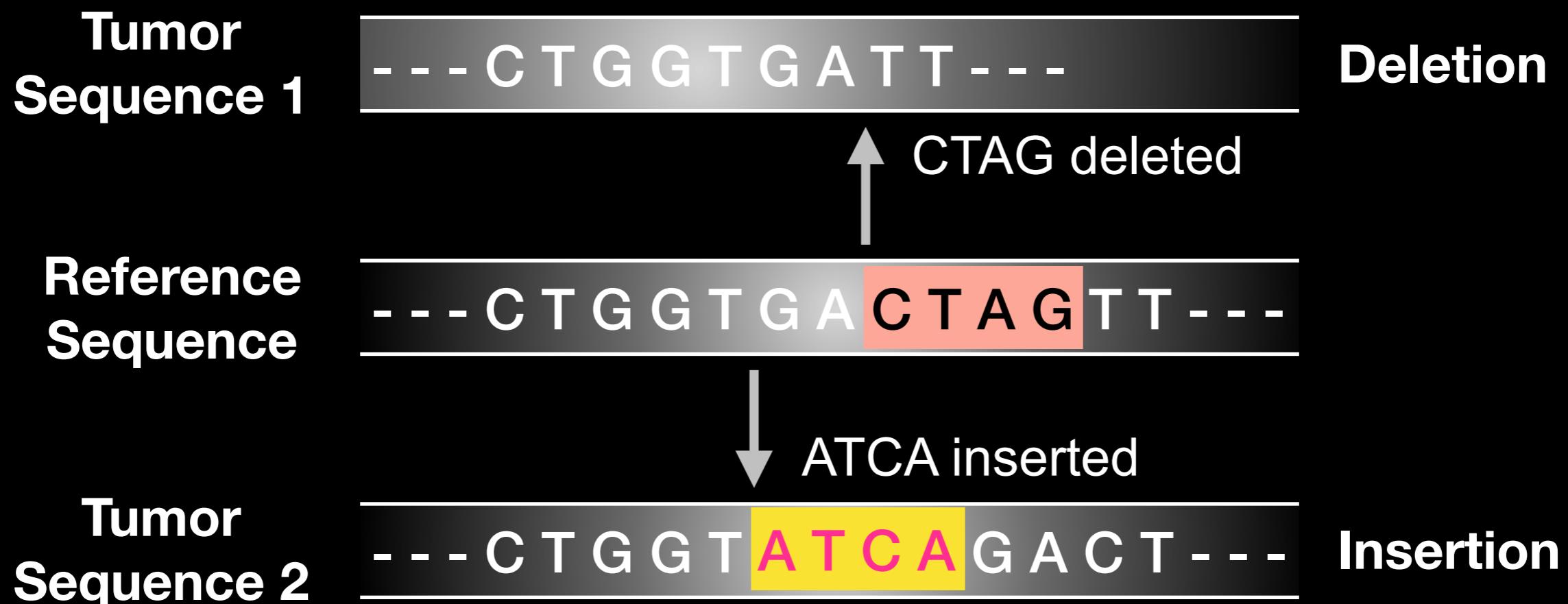
Reference  
Sequence

--- C T G G T G A C T A G T T ---

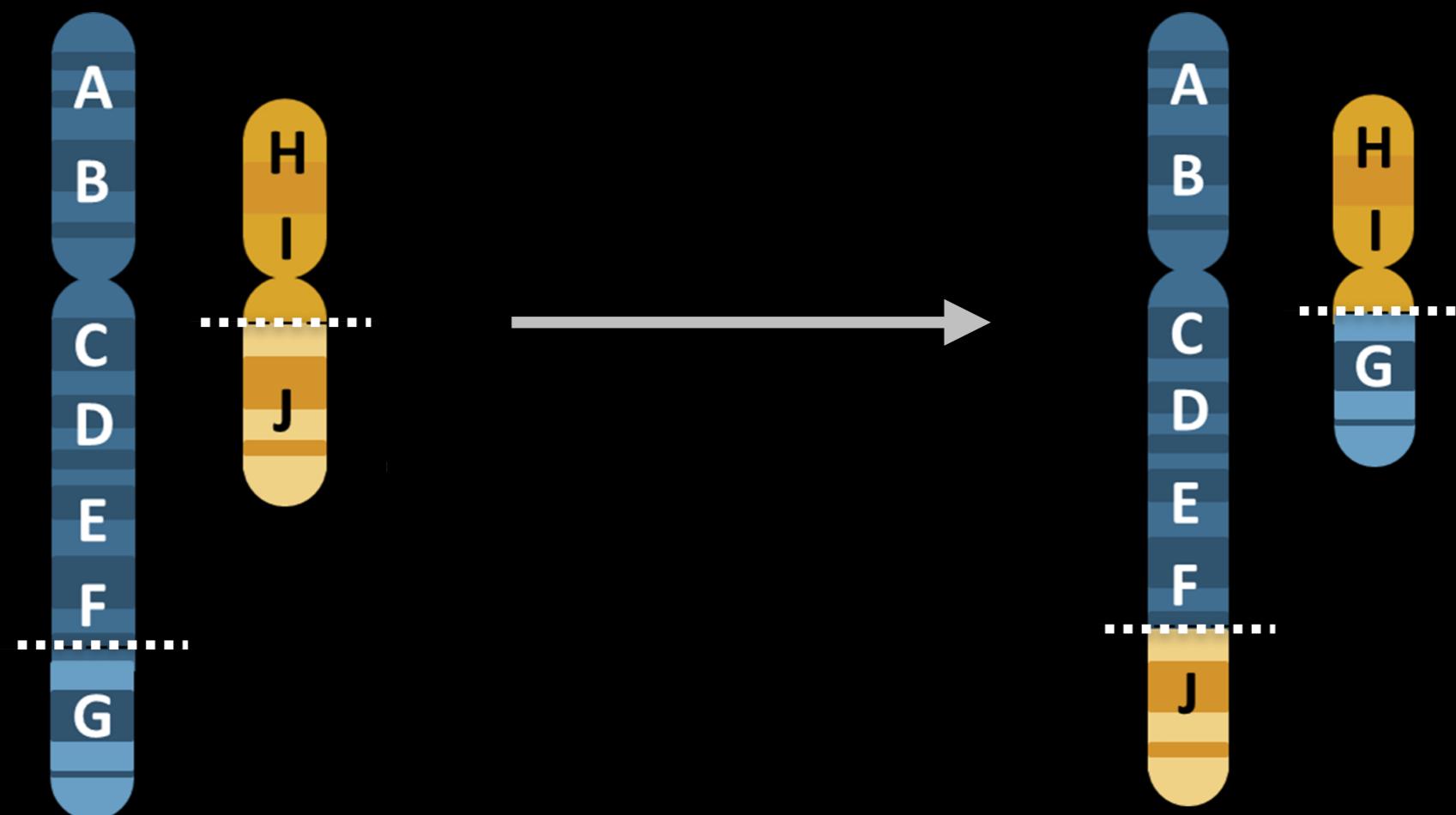
# Mutations detected: Indels

The diagram illustrates a deletion mutation. On the left, 'Tumor Sequence 1' is shown as a grey box containing the sequence:   
- - - C T G G T G A T T - - -  
On the right, 'Deletion' is written. Below it, a white arrow points upwards from the sequence, indicating the position of the deletion. The deleted sequence 'CTAG' is highlighted in a red box. On the far left, 'Reference Sequence' is shown as a grey box containing the sequence:   
- - - C T G G T G A C T A G T T - - -

# Mutations detected: Indels



# Mutations detected: Translocations



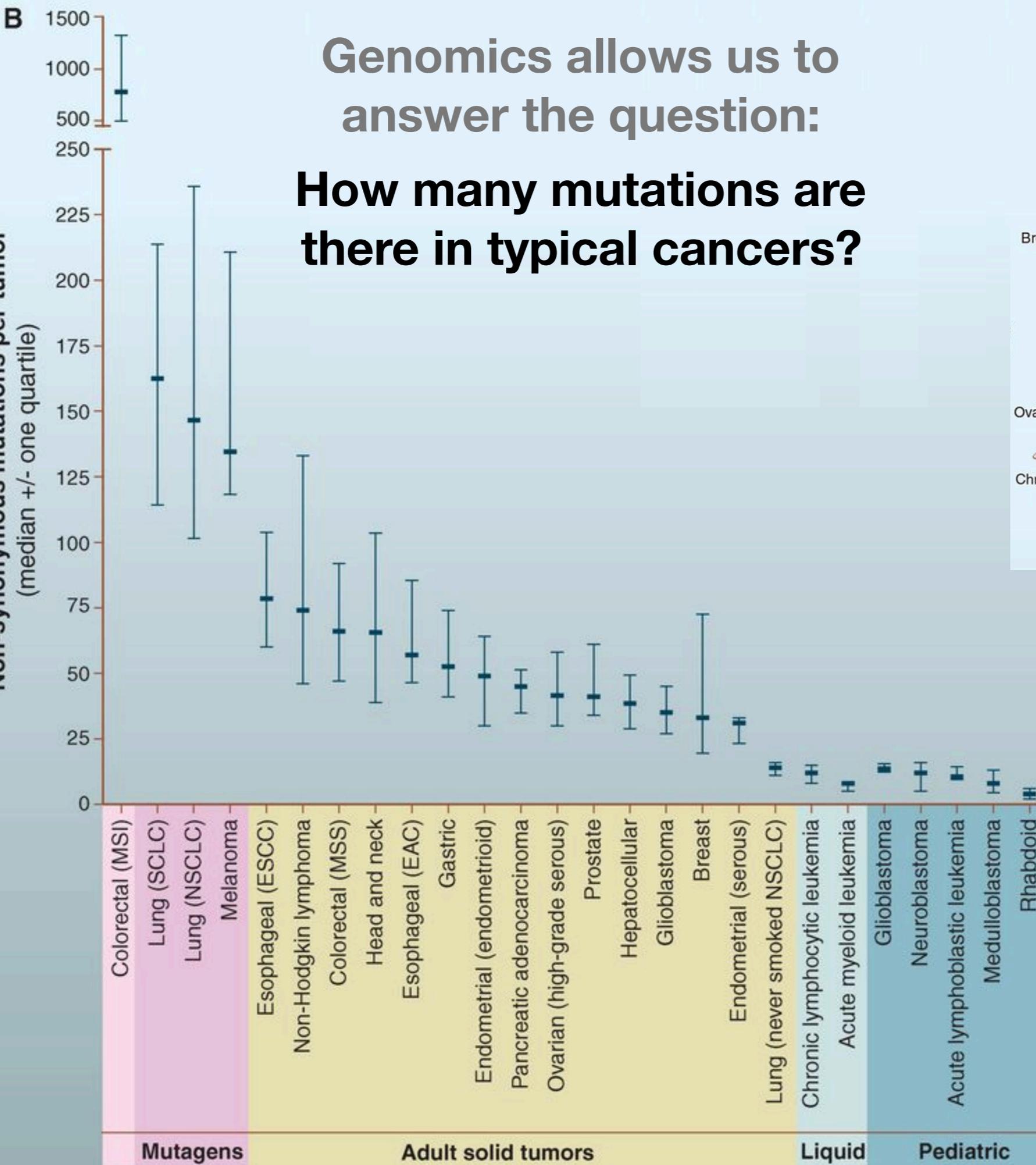
(e.g. Philadelphia chromosome (Ph) and found in over 90% of CML patients)

# What can go wrong in cancer genomes?

Type of change	Some common technology to study changes
DNA mutations	WGS, WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array, SNP array, WGS
DNA methylation	Methylation array, RRBS, WGBS
mRNA expression changes	mRNA expression array, RNA-seq
miRNA expression changes	miRNA expression array, miRNA-seq
<i>Protein expression</i>	Protein arrays, mass spectrometry

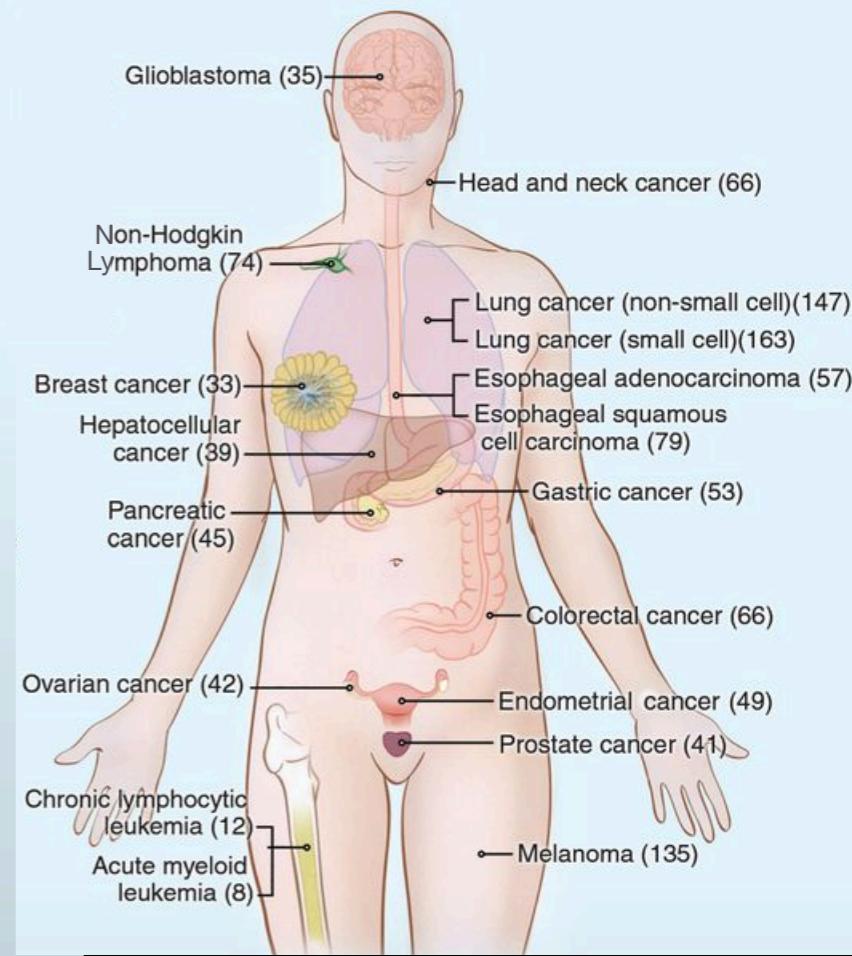
WGS = whole genome sequencing, WXS = whole exome sequencing

RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing



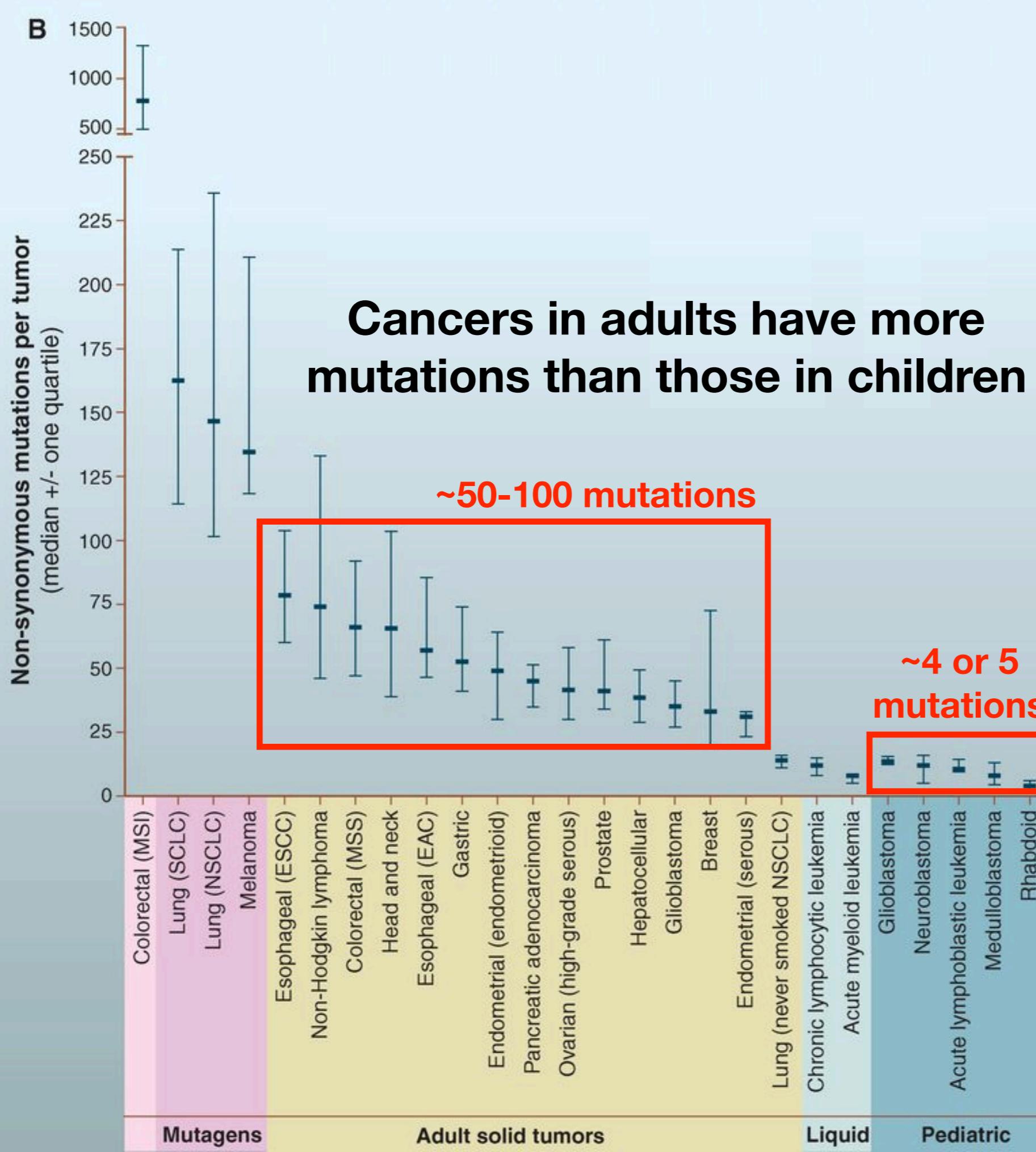
# Genomics allows us to answer the question:

## How many mutations are there in typical cancers?

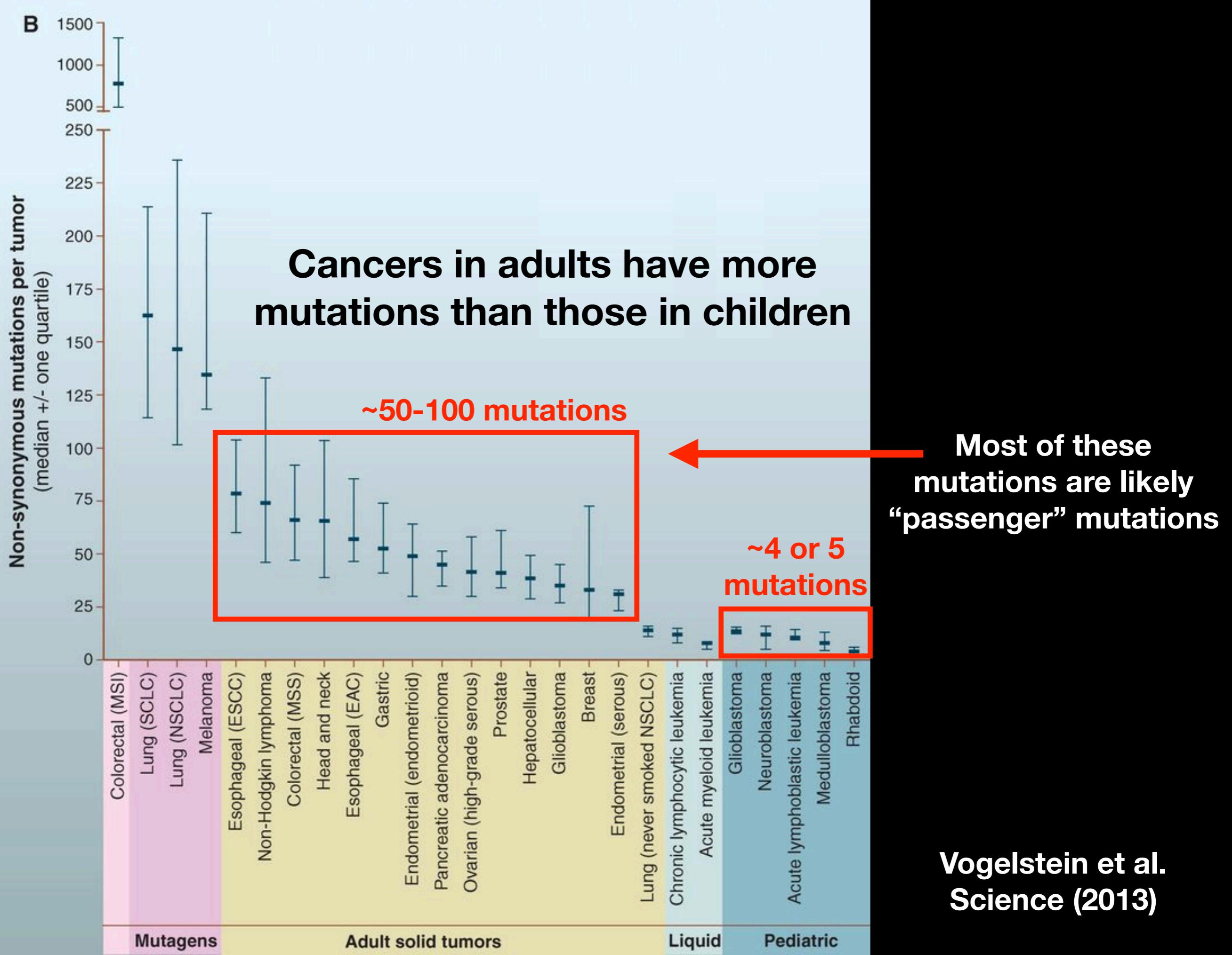


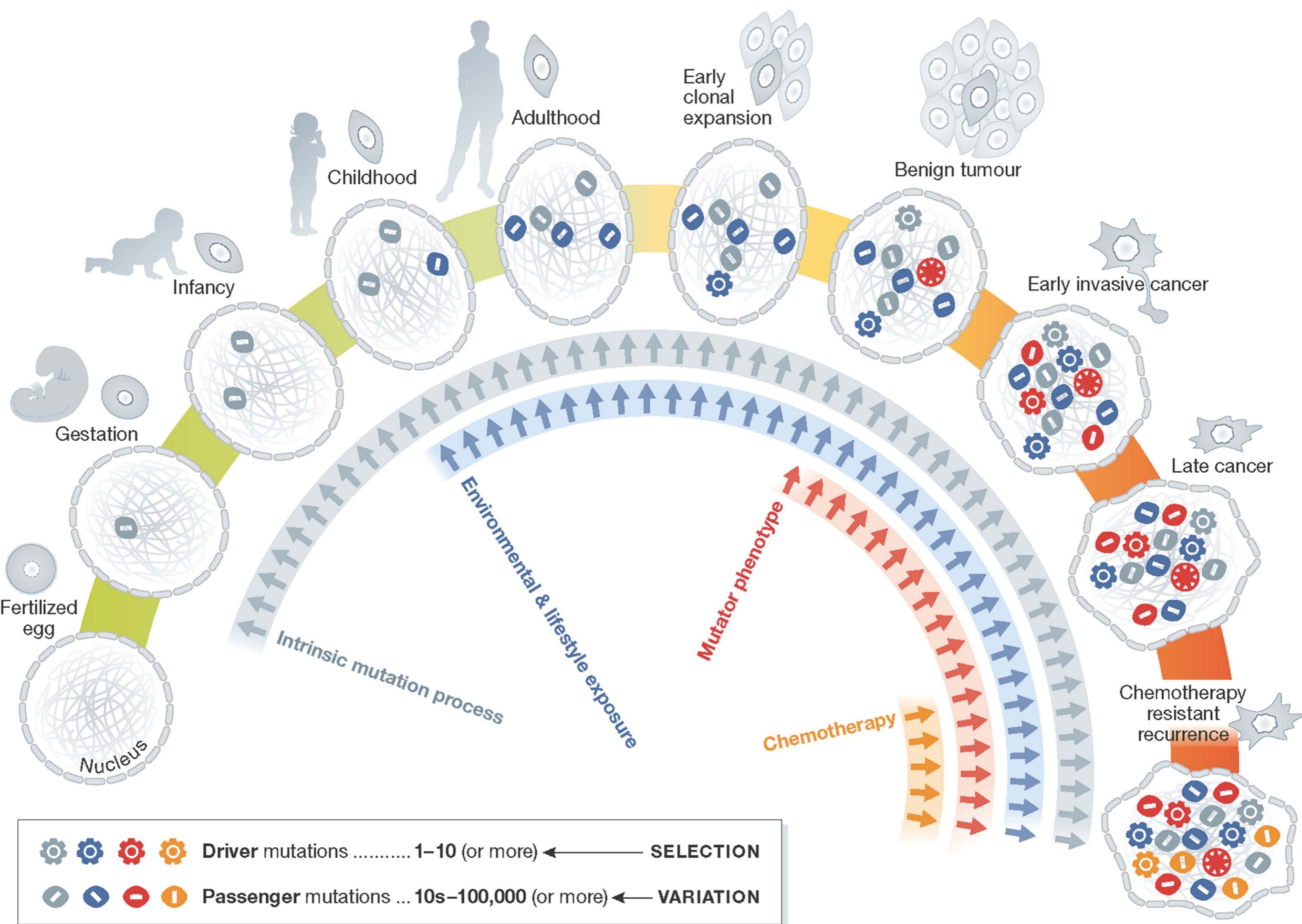
Number of somatic mutations in representative human cancers, detected by genome-wide sequencing studies

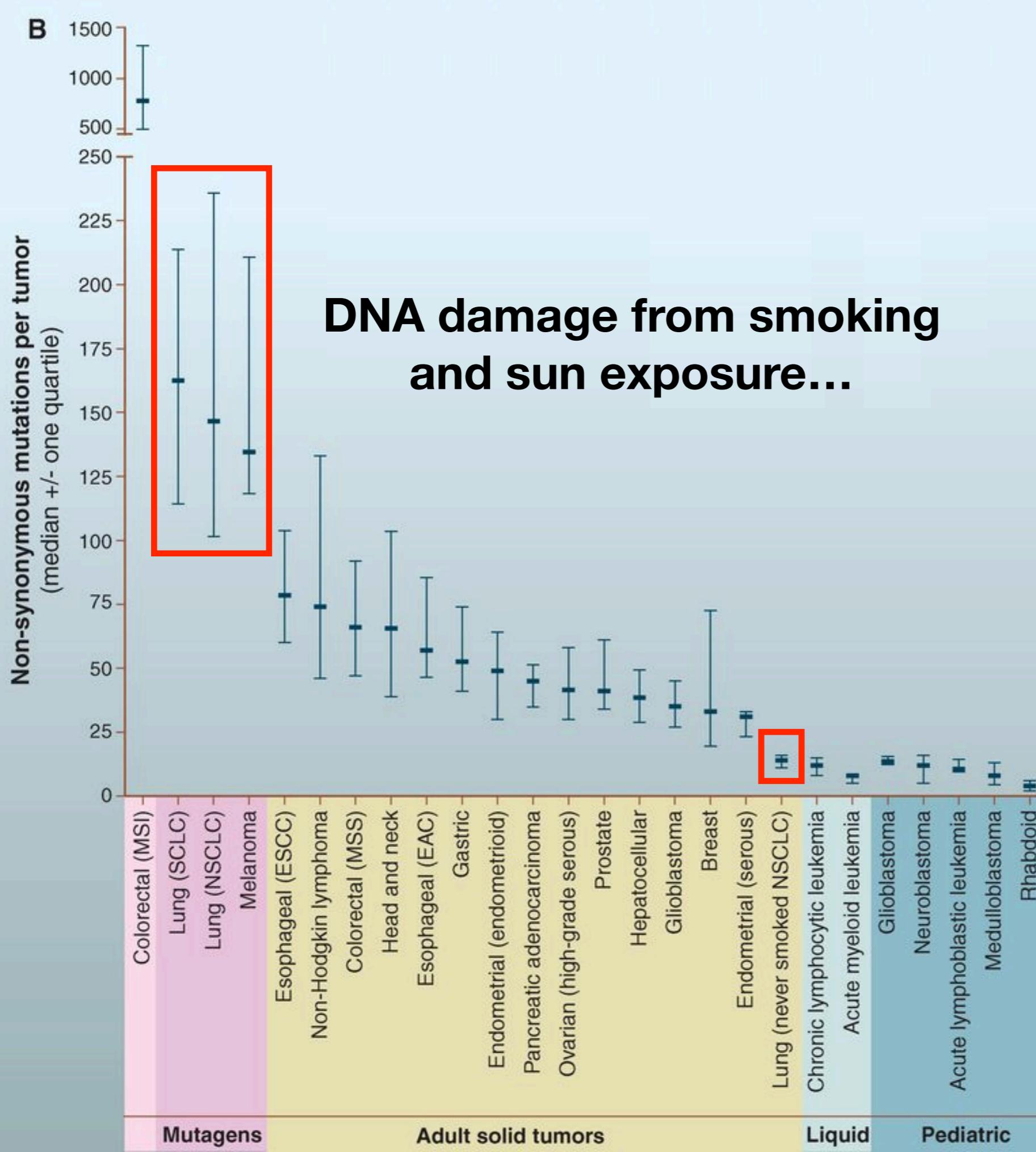
Vogelstein et al.  
Science (2013)



Vogelstein et al.  
Science (2013)

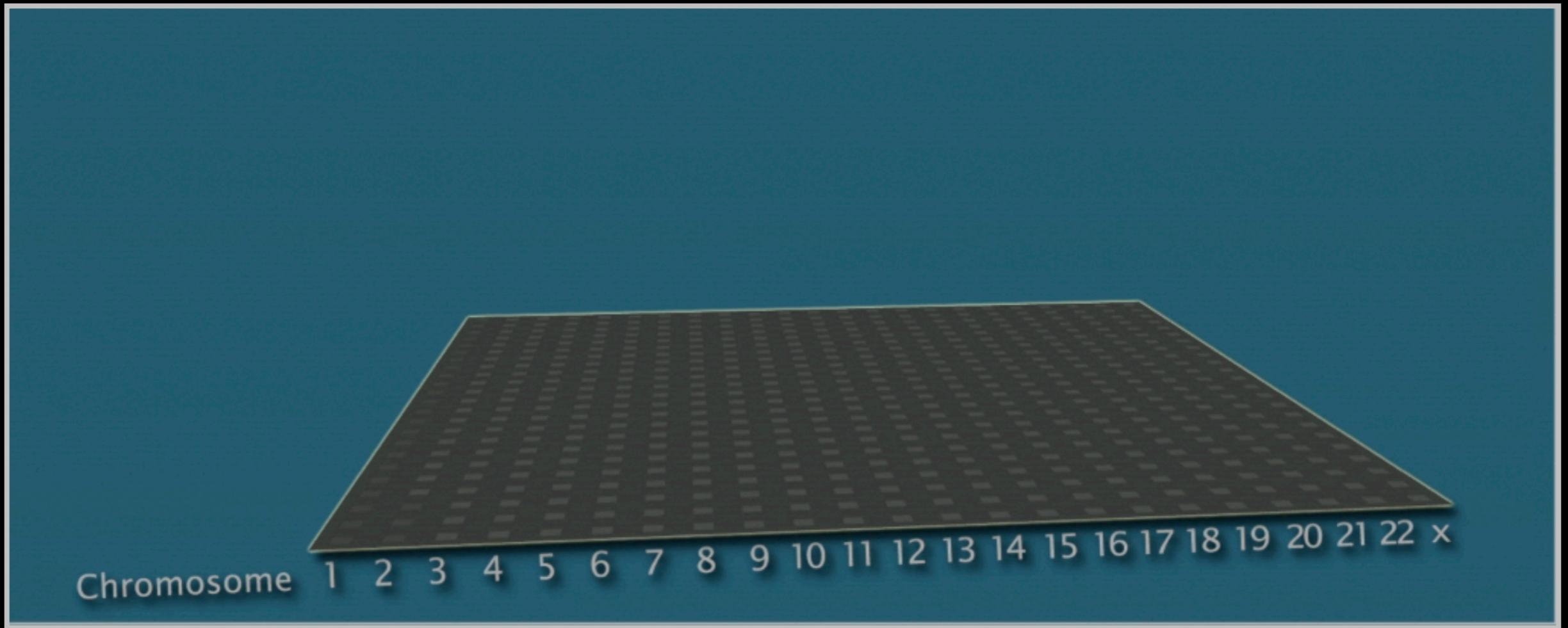






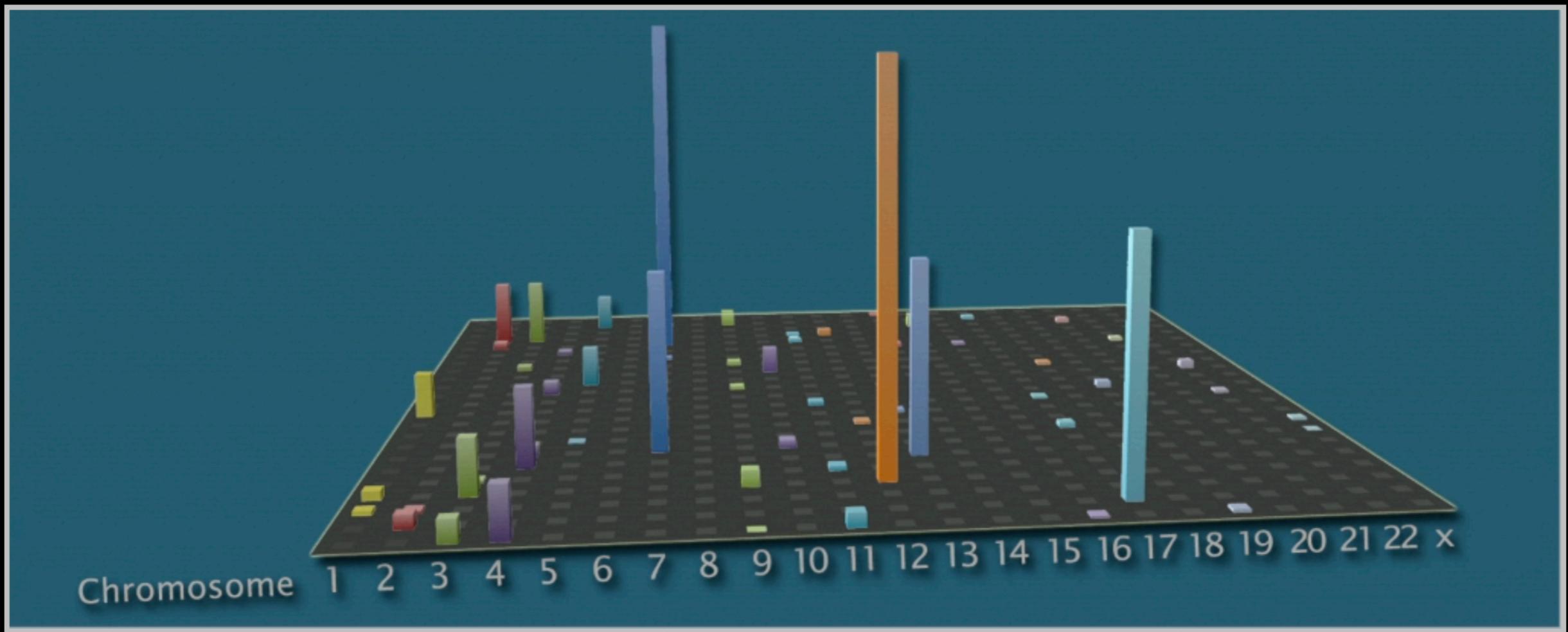
Vogelstein et al.  
Science (2013)

**Genomic approaches can identify the genes most commonly mutated in cancer**



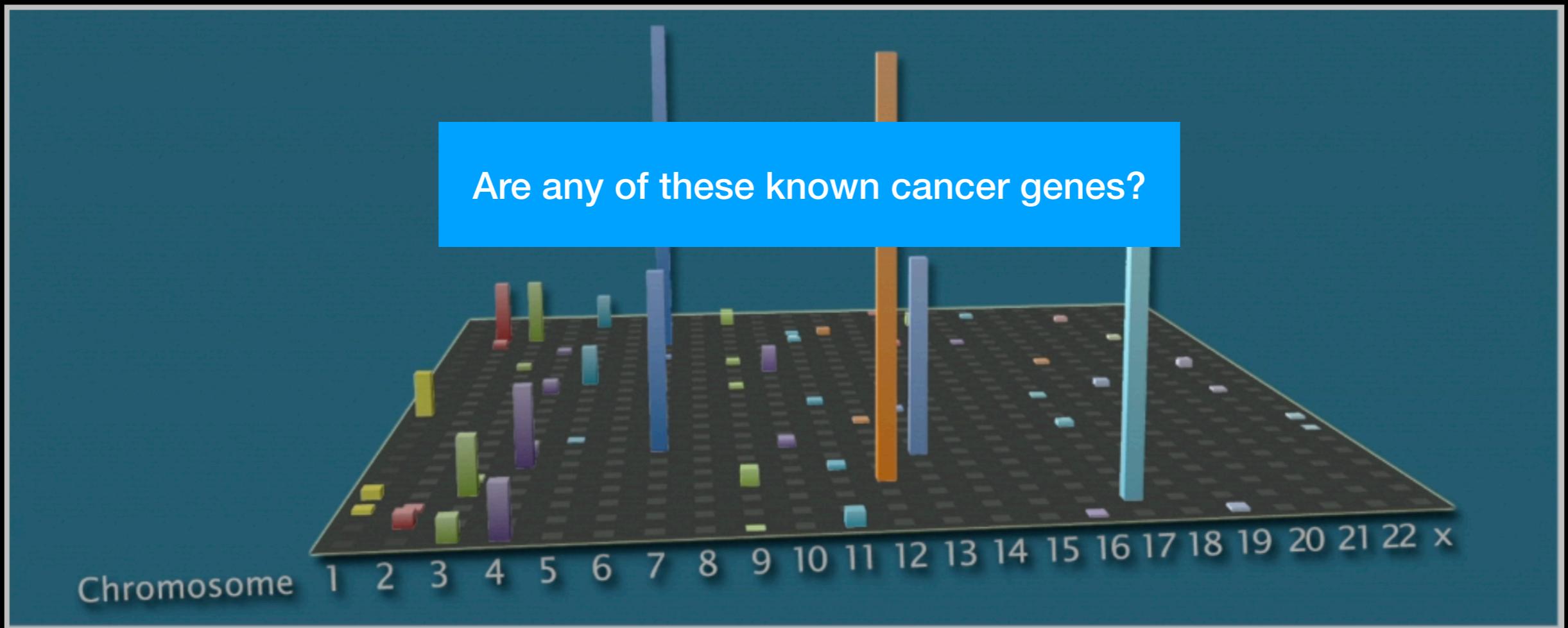
Arrange all genes in a matrix, ordered by chromosomes

# Identifying genes most commonly mutated in cancer



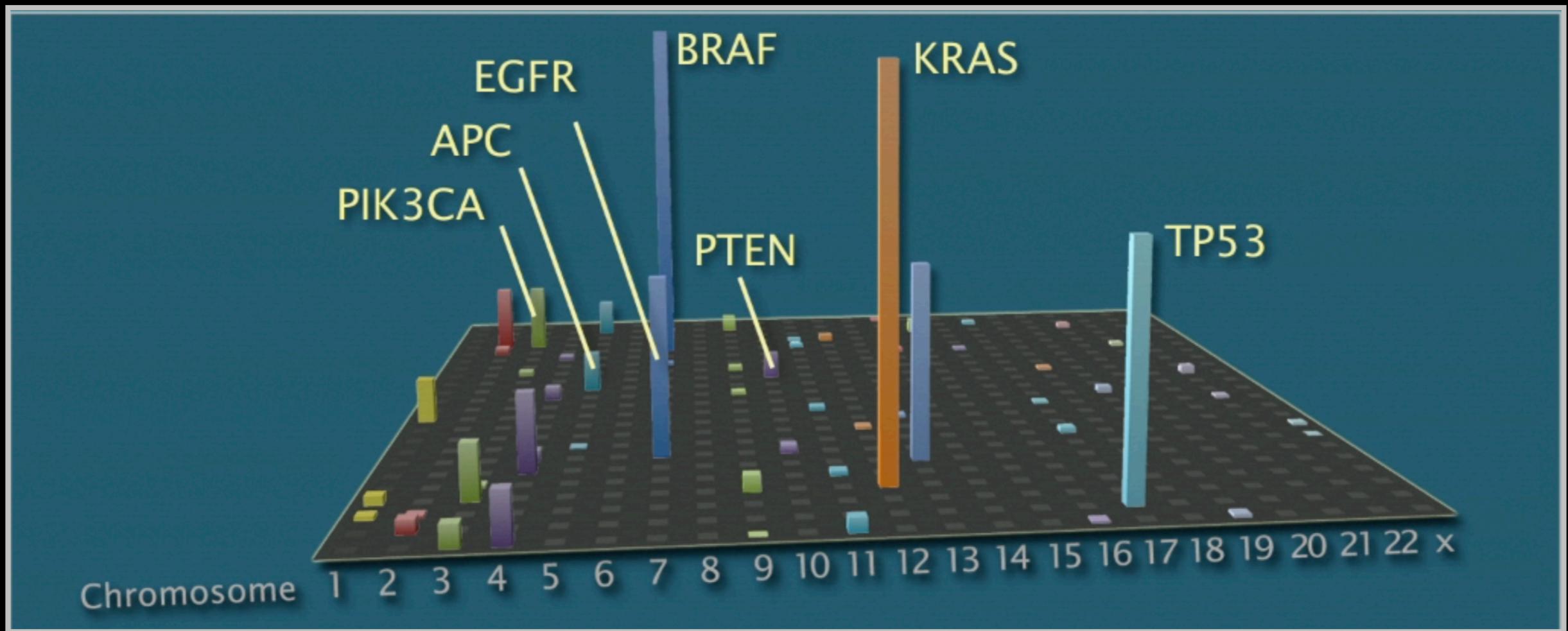
Add all data together to see which genes are most often mutated

# Identifying genes most commonly mutated in cancer



Add all data together to see which genes are most often mutated

# Identifying genes most commonly mutated in cancer



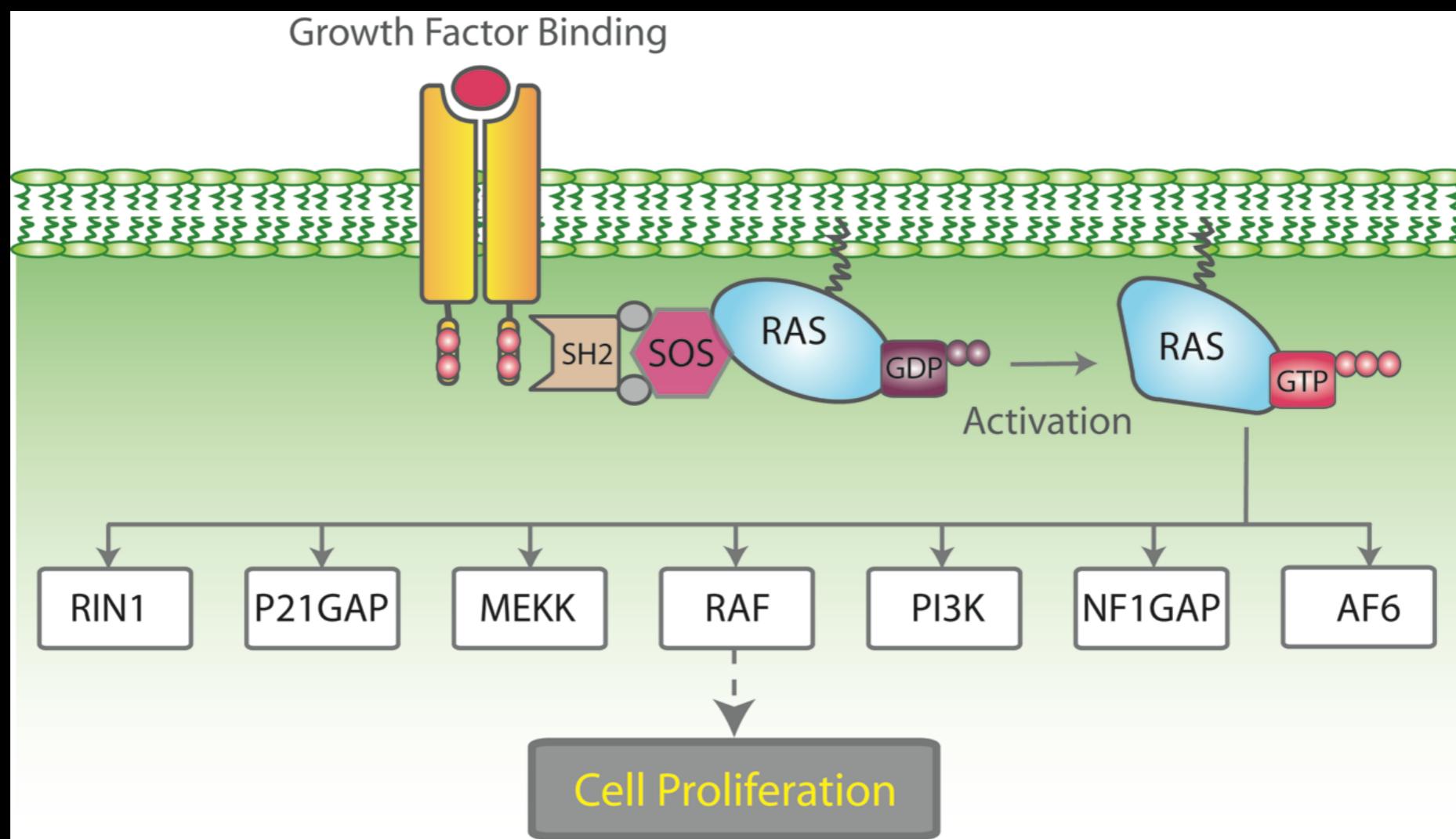
Many are famous proto-oncogenes, many others are new cancer genes!

# Three Main Types of Cancer Genes:

- **Oncogenes**, such as **Ras**, normally function to accelerate cell division and growth. They can be mutated to act like stuck gas pedals.
- **Tumor suppressor genes**, such as **p53** normal act like breaks. Mutations can cause these breaks to fail.
- **DNA repair genes**, such as **BRCA1 & 2**, normally function to fix minor damage to DNA when it replicates. When these genes are mutated, DNA damage can accumulate and lead to cancer.

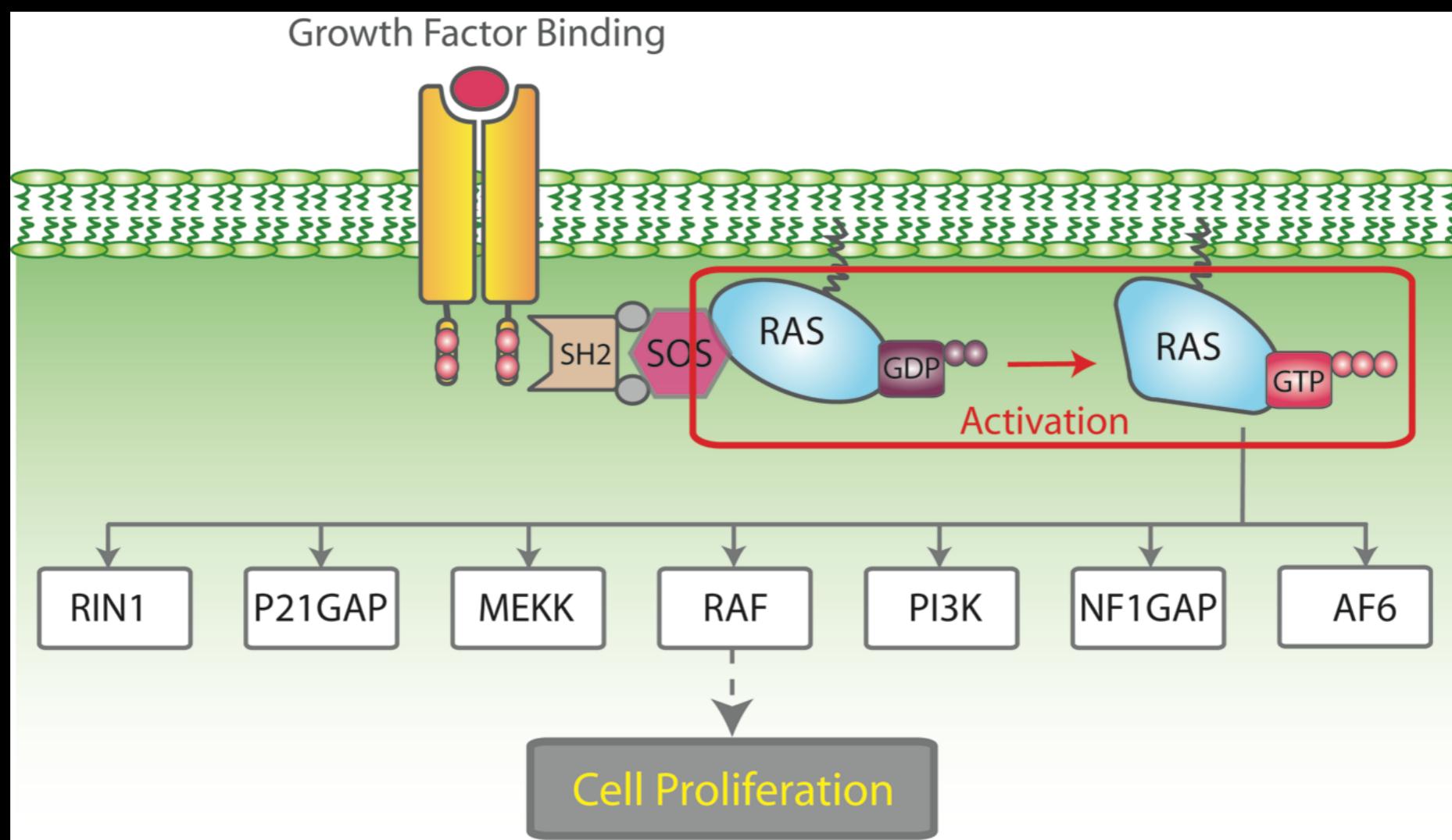
# Cell growth and survival genes

Many participate in signaling pathways that promote cell proliferation  
(E.G. EGFR, Ras, BRAF, MEK etc.)

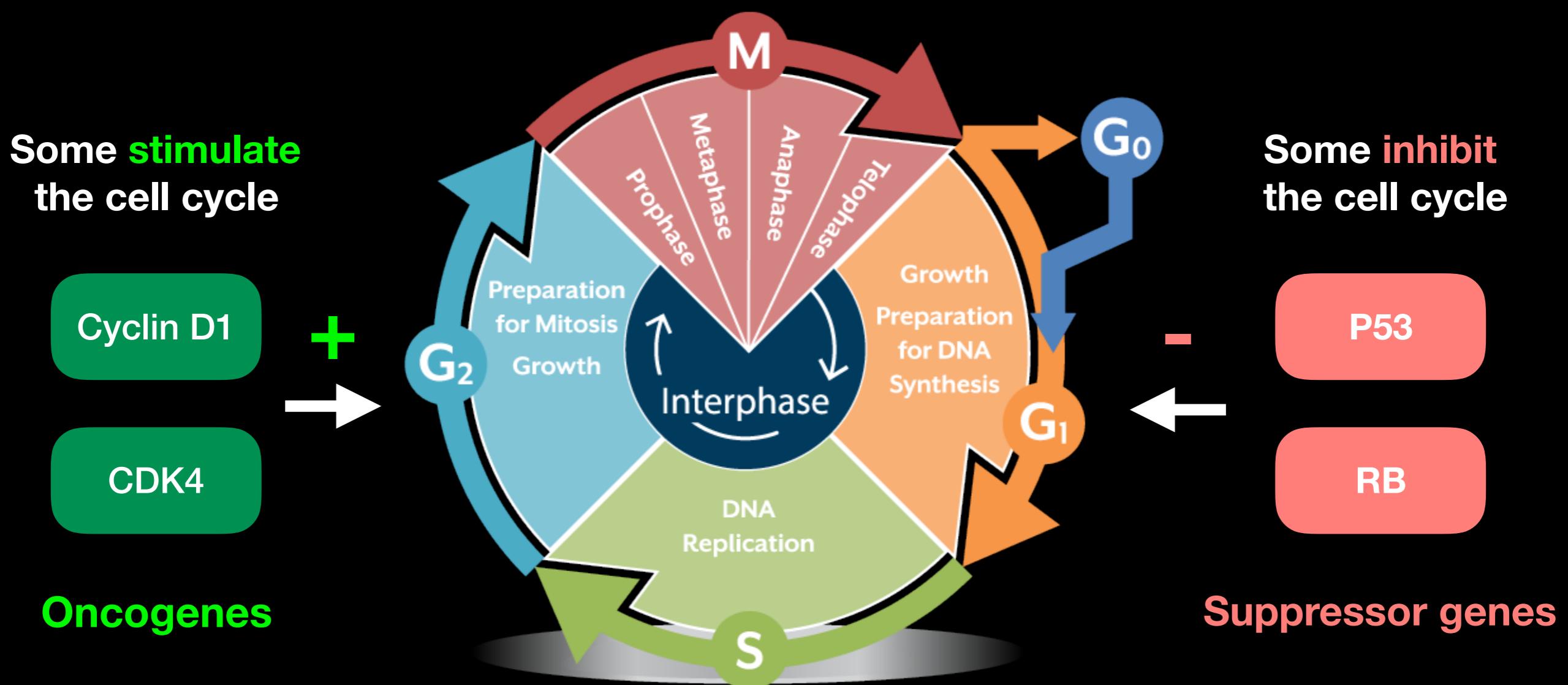


# Cell growth and survival genes

Many participate in signaling pathways that promote cell proliferation  
(E.G. EGFR, Ras, BRAF, MEK etc.)



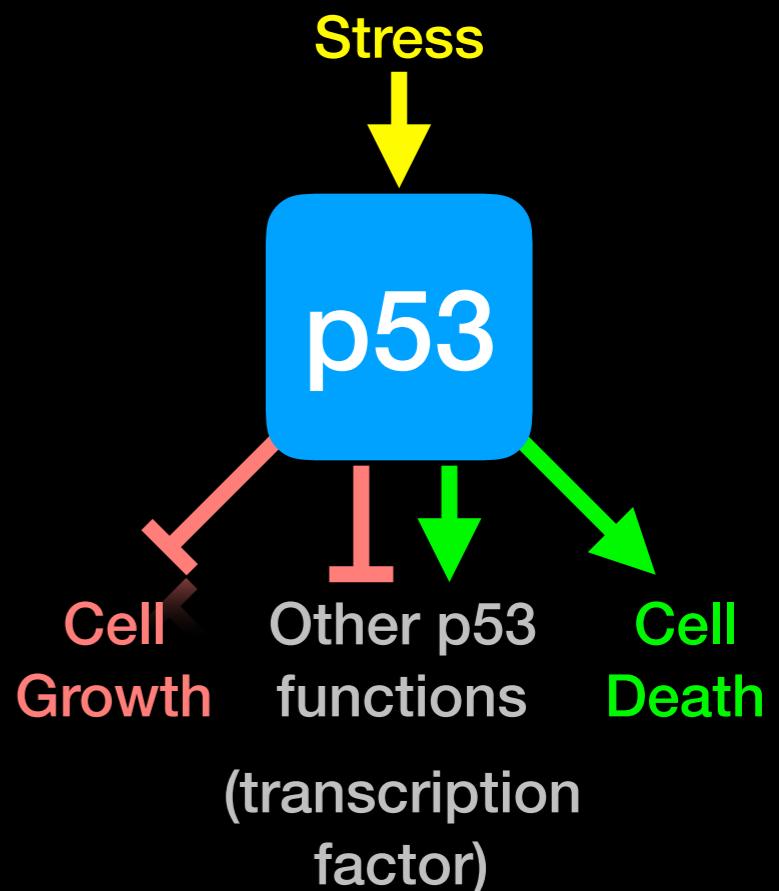
# Regulators of Cell Cycle and Cell Death



# p53 Regulates Cell Division

Probably the most famous cancer gene that is mutated in about half of all tumors. Often called the '*guardian of the genome*'

- p53 normally shuts down cell division when a cell is stressed (e.g. by DNA damage)
- When DNA is damaged, p53 activates genes that stop cell growth or trigger the cell to die.
- Thus, p53 guards against changes to cells that might lead to tumor formation.
- It appears necessary to inactivate p53 to develop many forms of cancer.

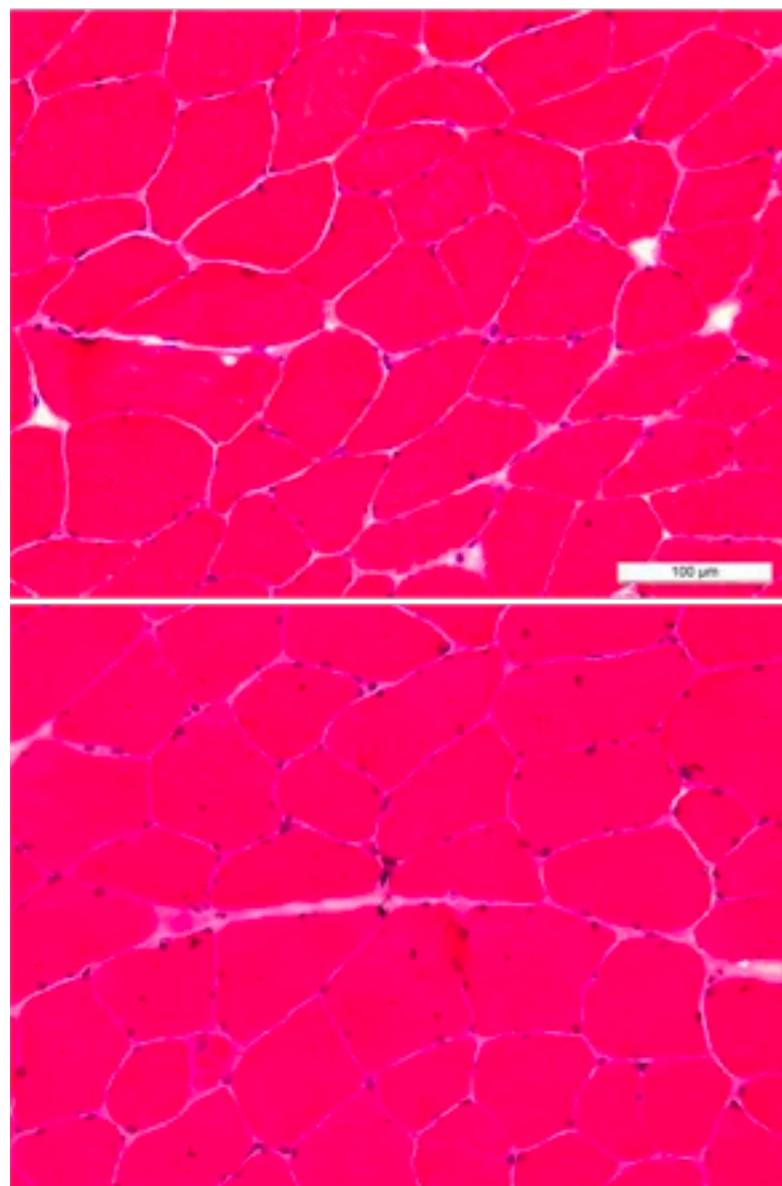


# Hands-on time!

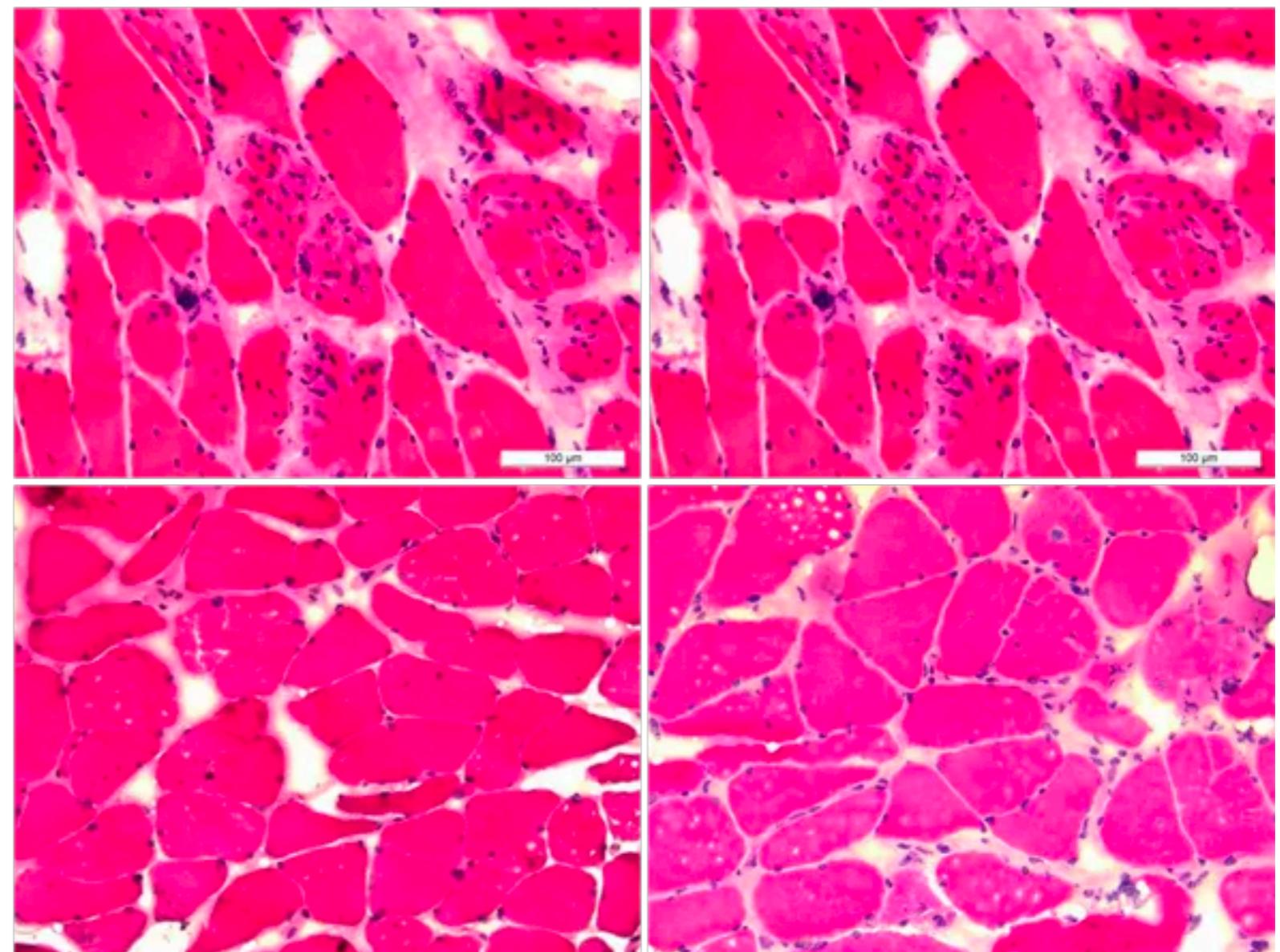
[https://bioboot.github.io/bggn213\\_F19/lectures/#18](https://bioboot.github.io/bggn213_F19/lectures/#18)

**Part 1 Only Please**

## Control



## Pancreatic Cancer



Representative H&E micrographs of rectus abdominis biopsies are displayed for two patients without cancer (*left*) and four patients with pancreatic cancer (*right*)

# Today's Menu

Next Up:

Cancer Genomics

Brief review of cancer fundamentals,  
What is cancer and what causes it?

Mining Cancer  
Genomic Data

**Hands-on analysis** to identify genomic changes in different cancers and identify new targets for therapy

Towards personalized  
cancer treatments

Recap on how the immune system normally detects cancer cells and how we can predict mutations that can be recognized by T cells

Cancer  
Immunoinformatics

**Hands-on analysis** to design personalized cancer vaccines

# Towards personalized cancer treatments using Immunoinformatics

**Bjoern Peters**

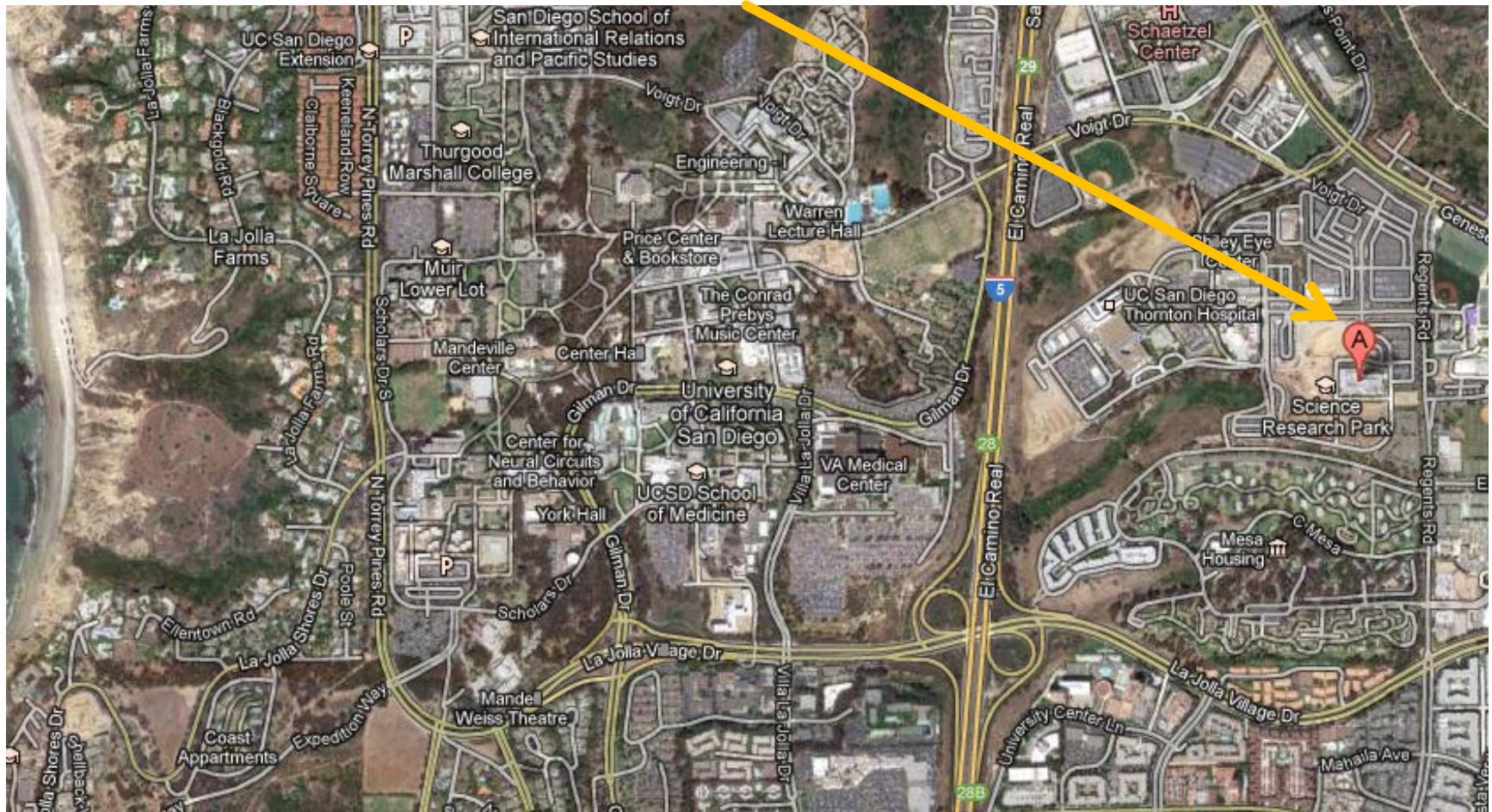
**Zeynep Kosaloglu-Yalcin**

La Jolla Institute for Immunology

**La Jolla  
Institute**  
FOR IMMUNOLOGY

Life Without Disease.®

# La Jolla Institute for Immunology (LJI)



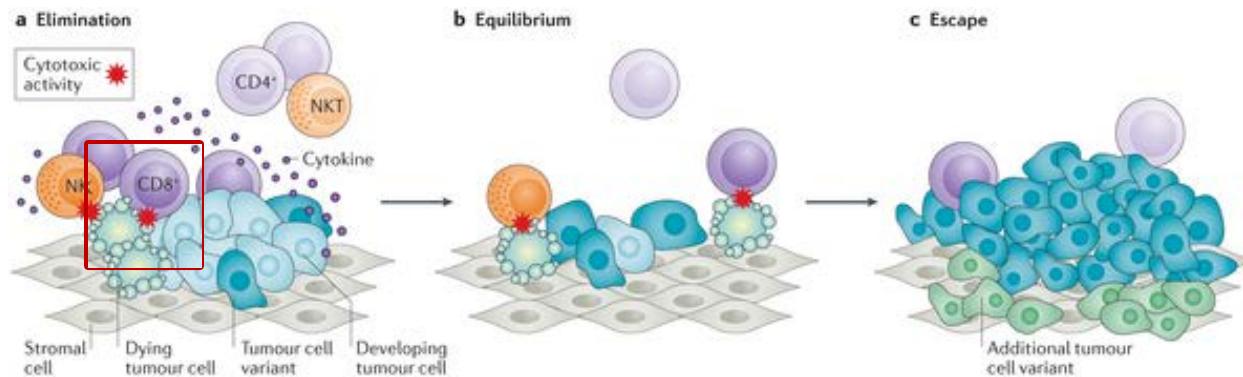
# Overview

---

- 
- Background Biology: The immune system can detect and eliminate mutated cancer cells
  - Background Immunoinformatics:
    - Mutations can be detected using patient sequencing data
    - Mutations that can be recognized by T cells can be predicted
  - Hands on: Design a personalized cancer vaccine

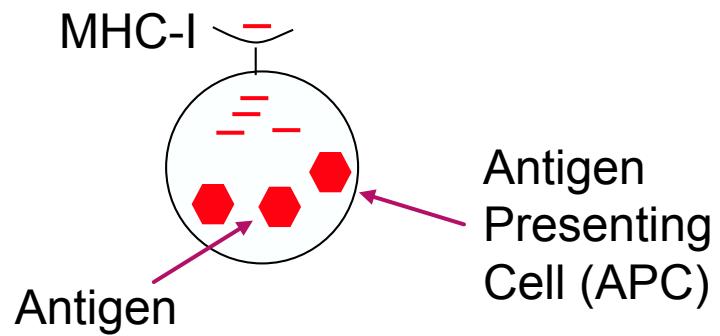
# Cancer immune surveillance and escape

- Mutations in cells occur frequently
- The immune system has the capacity to detect and eliminate such mutated cells, and will do so on a regular basis
- Only when mutated cells find ways to hide from- or suppress an attack by the immune system, they can grow and spread unhindered leading to clinically apparent tumors

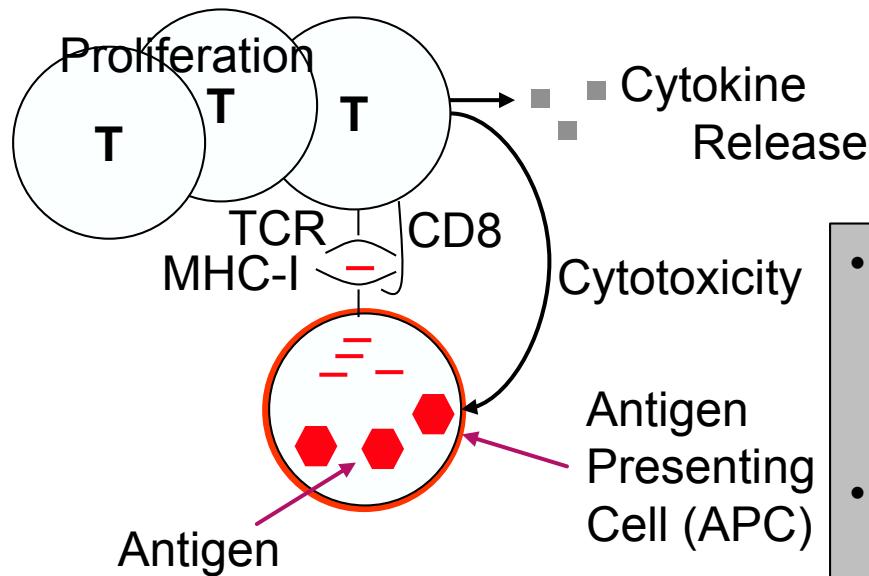


# CD8<sup>+</sup> T cell epitope recognition

---



# CD8<sup>+</sup> T cell epitope recognition

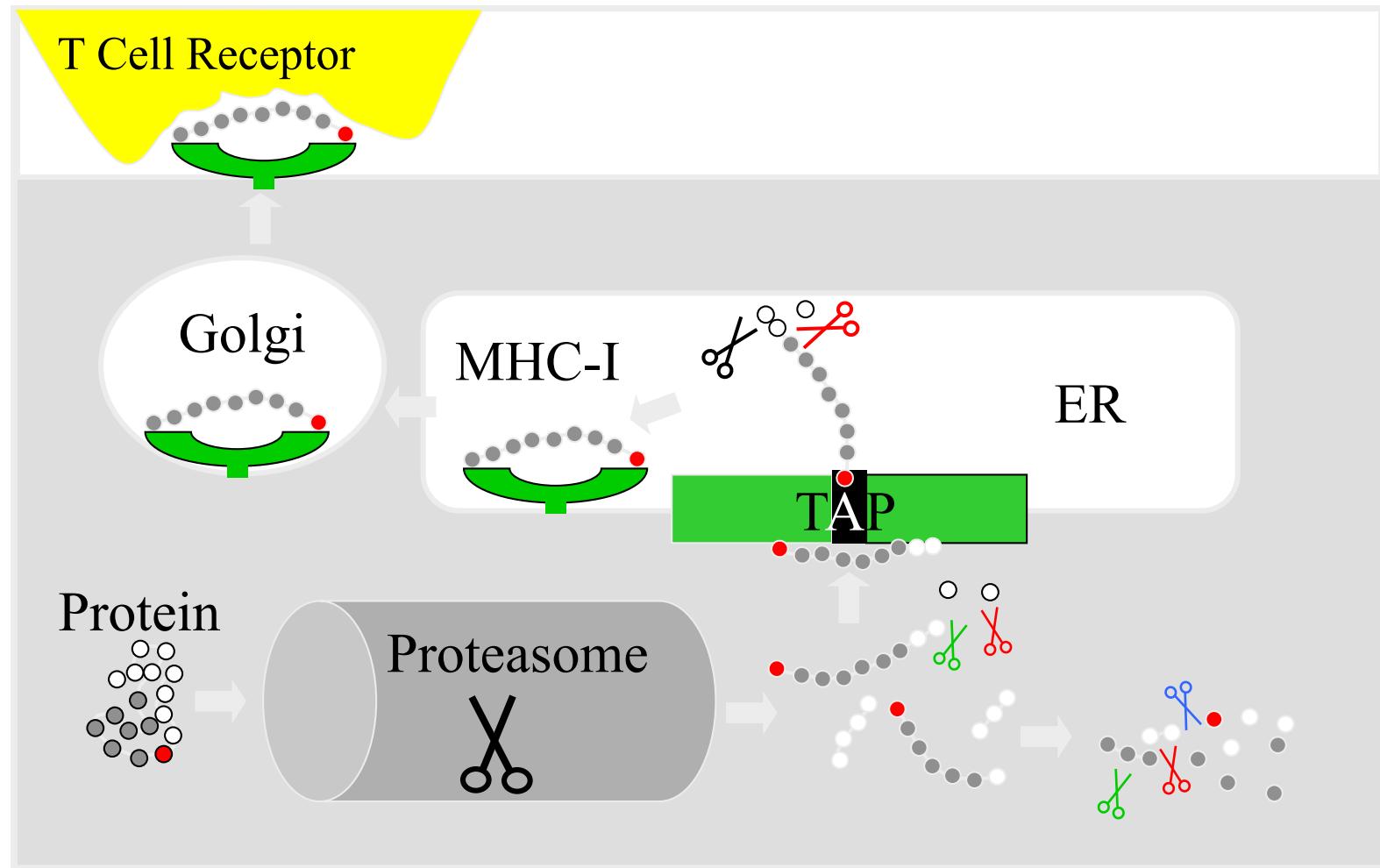


- What determines what peptides are presented by MHC molecules on the cell surface?
- How do T cells distinguish self- from non-self peptides?

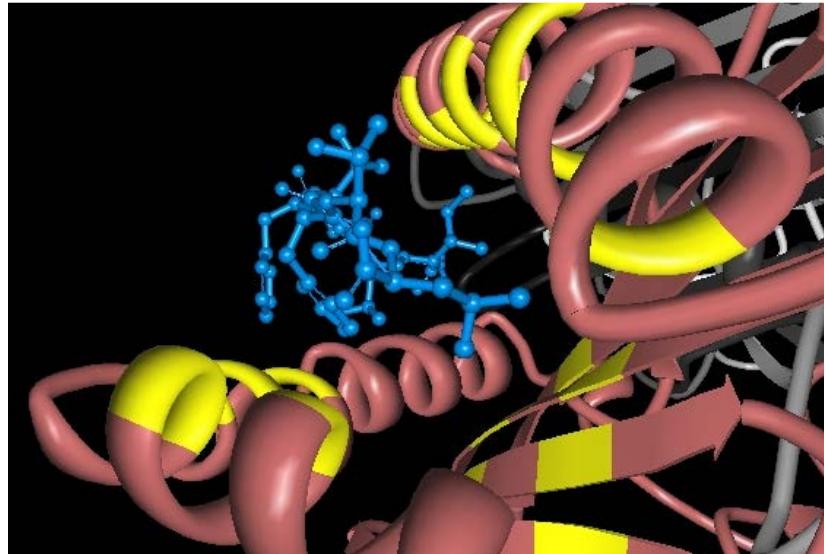
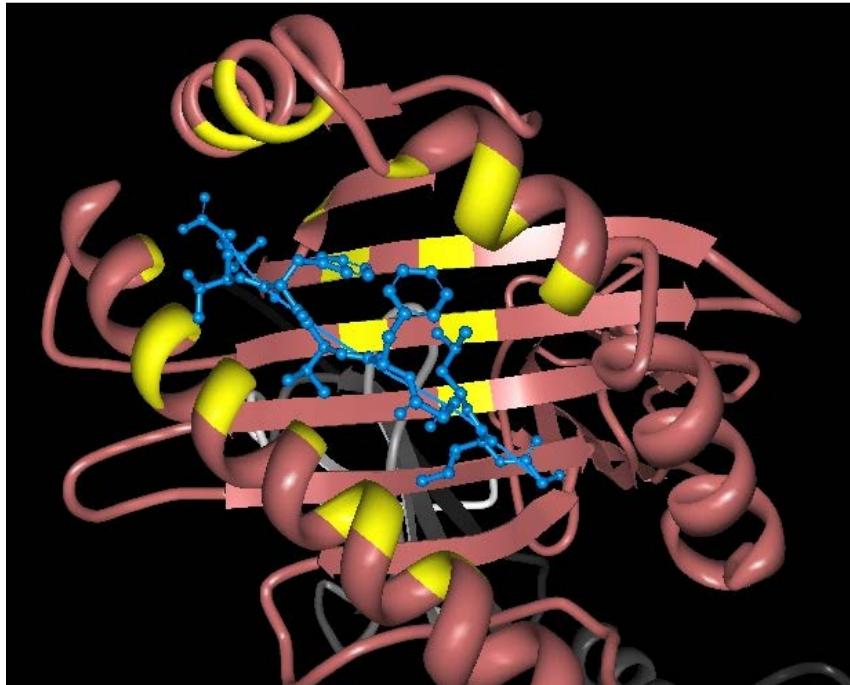
Antigen   
Epitope

The sequence of the antigen peptide is shown as a horizontal line of letters, with the epitope highlighted in red. A bracket below the sequence is labeled 'Epitope'.

# MHC I - Antigen processing and presentation pathway



# MHC:peptide binding mode

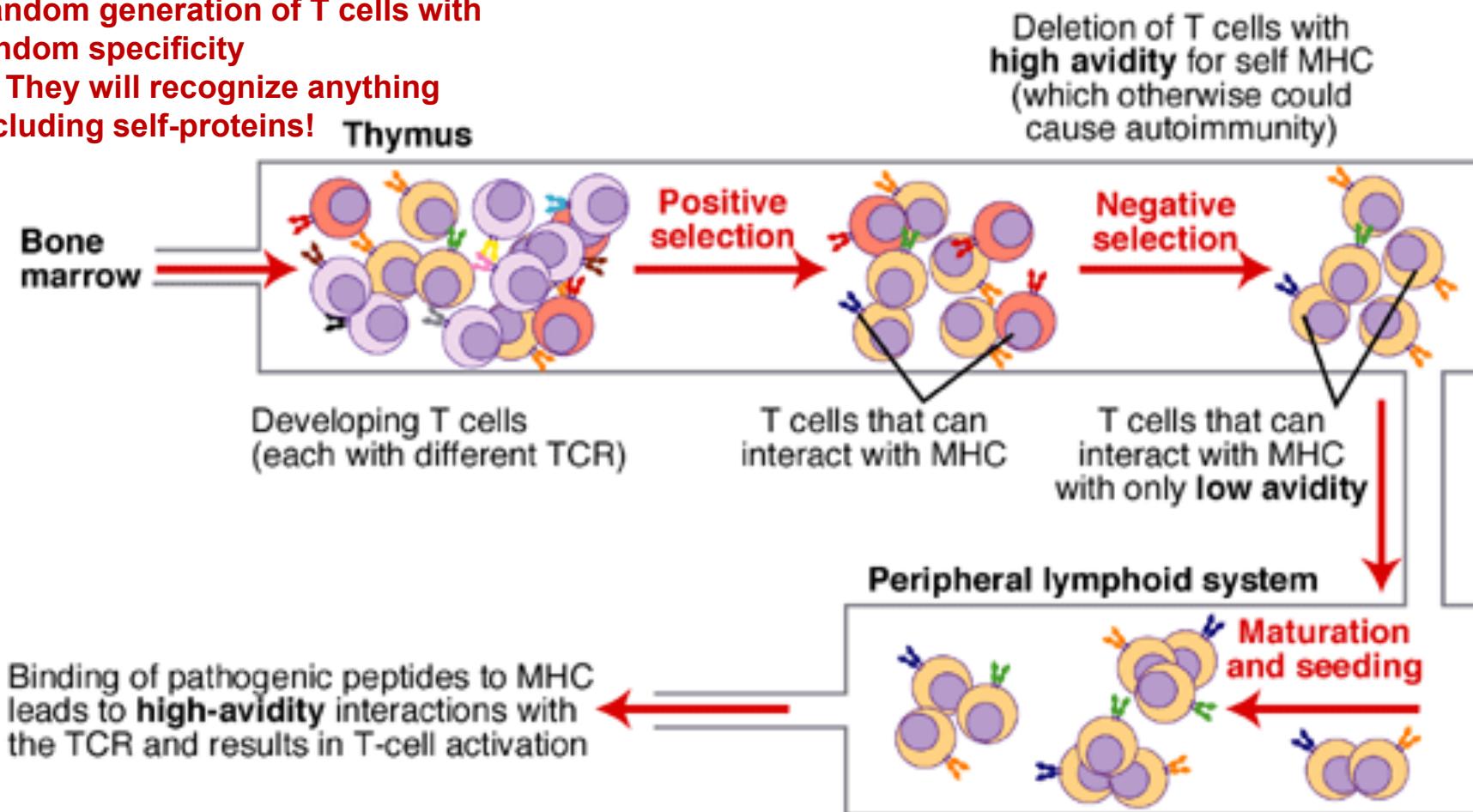


- Each human expresses up to 6 different MHC-I molecules in every cell
- >3000 MHC variants are known
- Distinct binding specificities → individual epitope repertoire

# Self –reactive T cells are deleted during maturation

Random generation of T cells with random specificity

→ They will recognize anything including self-proteins! Thymus

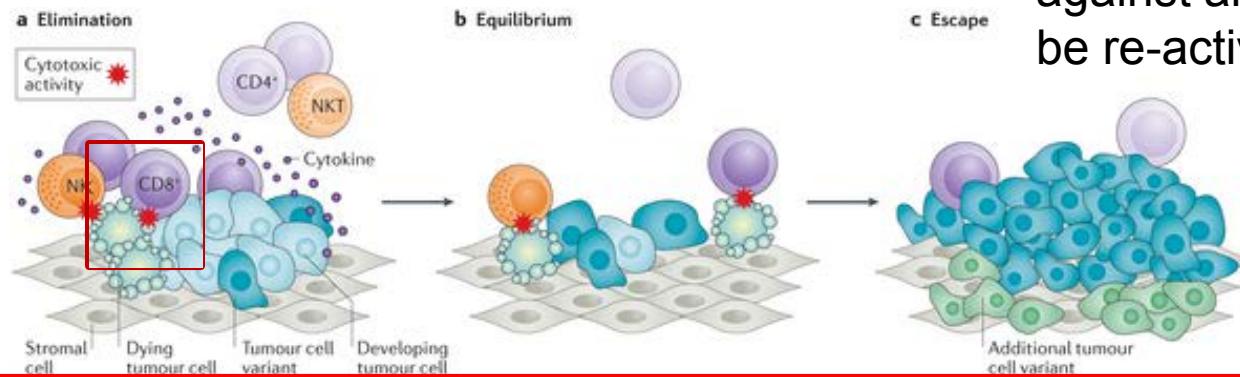


The repertoire of T cells is shaped by both positive and negative selection

Expert Reviews in Molecular Medicine © 1999 Cambridge University Press

# Cancer immune surveillance and escape

- Mutations in cells occur frequently
- The immune system has the capacity to detect and eliminate such mutated cells, and will do so on a regular basis
- Only when mutated cells find ways to hide from- or suppress an attack by the immune system, they can grow and spread unhindered leading to clinically apparent tumors



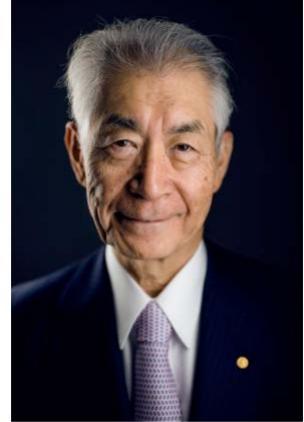
Can the immune response against an 'escaped' cancer be re-activated?

# Nobel Prize 2018 in Medicine

- Identification of the molecules PD-1 and CTLA-4 that function as 'T cell brakes' (immune checkpoints)
- Blockade of PD-1 and CTLA-4 results in activation of T cells which has "*fundamentally changed the outcome for certain groups of patients with advanced cancer*"
- "*Similar to other cancer therapies, adverse side effects are seen, which can be serious and even life threatening. They are caused by an overactive immune response leading to autoimmune reactions [...]”*



© Nobel Media AB. Photo: A.  
Mahmoud  
James P. Allison



© Nobel Media AB. Photo: A.  
Mahmoud  
Tasaku Honjo

*“for their discovery of cancer therapy by inhibition of negative immune regulation”*

# Rationale for Personalized 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.

## Newsroom

[Home](#) / [Newsroom](#) / [Releases](#) / Is the Next Big Step in Cancer Therapy Personalized Vaccines?

## Press Releases

UC San Diego Health In The News

News Features

Trending Topics

Podcast

## Is the Next Big Step in Cancer Therapy Personalized Vaccines?

*UC San Diego Health and La Jolla Institute launch clinical trial harnessing an individual's immune system in a syringe*

October 12, 2018 | Yadira Galindo



Ezra Cohen, MD, UC San Diego Health physician scientist, administered the first-of-its-kind personalized cancer vaccine to Tamara Strauss, while Aaron Miller, MD, PhD, UC San Diego Health physician scientist, Tamara's mother, Iris Strauss, and Stephen Schoenberger, PhD, La Jolla Institute for Allergy and Immunology professor of immunology, look on.

After sequencing Tamara's tumor and normal tissue, the team identified mutations expressed solely by cancer cells in her body. Schoenberger and LJI's Bjoern Peters, PhD, developed a novel algorithm to select mutations that are recognized by the immune system.

This algorithm was deployed to recognize the

neoantigens that generated the strongest T cell response from Tamara's tissue samples. These neoantigens were then presented to Tamara's own T cells and cultured over a two-week period using 50 milliliters of her blood to develop a personalized vaccine.

Several trials for personalized cancer vaccines are currently ongoing

# Personalized Cancer Immunotherapy



Ezra Cohen, MD, UC San Diego Health physician scientist, administered the first-of-its-kind personalized cancer vaccine to Tamara Strauss, while Aaron Miller, MD, PhD, UC San Diego Health physician scientist, Tamara's mother, Iris Strauss, and Stephen Schoenberger, PhD, La Jolla Institute for Allergy and Immunology professor of immunology, look on.

After sequencing Tamara's tumor and normal tissue, the team identified mutations expressed solely by cancer cells in her body.

Schoenberger and LJI's Bjoern Peters, PhD, developed a novel algorithm to

select mutations that are recognized by the immune system.

This algorithm was deployed to recognize the

neoantigens that generated the strongest T cell response from Tamara's tissue samples. These neoantigens were then presented to Tamara's own T cells and cultured over a two-week period using 50 milliliters of her blood to develop a personalized vaccine.

# Personalized Cancer Immunotherapy



Ezra Cohen, MD, UC San Diego Health physician scientist, administered the first-of-its-kind personalized cancer vaccine to Tamara Strauss, while Aaron Miller, MD, PhD, UC San Diego Health physician scientist, Tamara's mother, Iris Strauss, and Stephen Schoenberger, PhD, La Jolla Institute for Allergy and Immunology professor of immunology, look on.

After sequencing Tamara's tumor and normal tissue, the team identified mutations expressed solely by cancer cells in her body.

Schoenberger and LJI's Bjoern Peters, PhD, developed a novel algorithm to

select mutations that are recognized by the immune system.

This algorithm was deployed to recognize the

neoantigens that generated the strongest T cell response from Tamara's tissue samples. These neoantigens were then presented to Tamara's own T cells and cultured over a two-week period using 50 milliliters of her blood to develop a personalized vaccine.

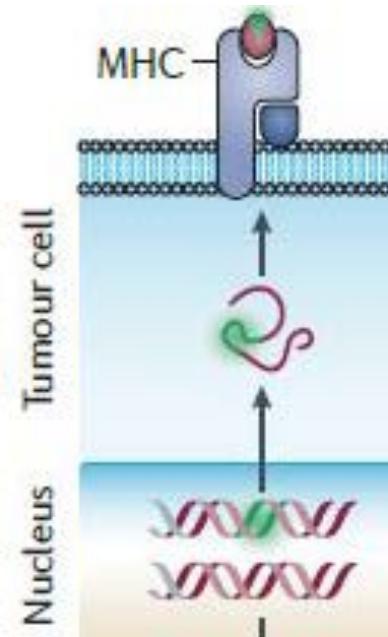
# Overview

---

- Background Biology: The immune system can detect and eliminate mutated cancer cells
- Background Immunoinformatics:
  - Mutations can be detected using patient sequencing data
  - Mutations that can be recognized by T cells can be predicted
- Hands on: Design a personalized cancer vaccine

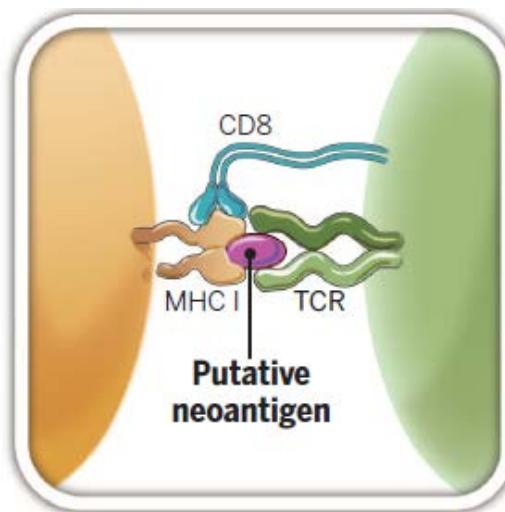
# Neoepitopes (Neoantigens)

- 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

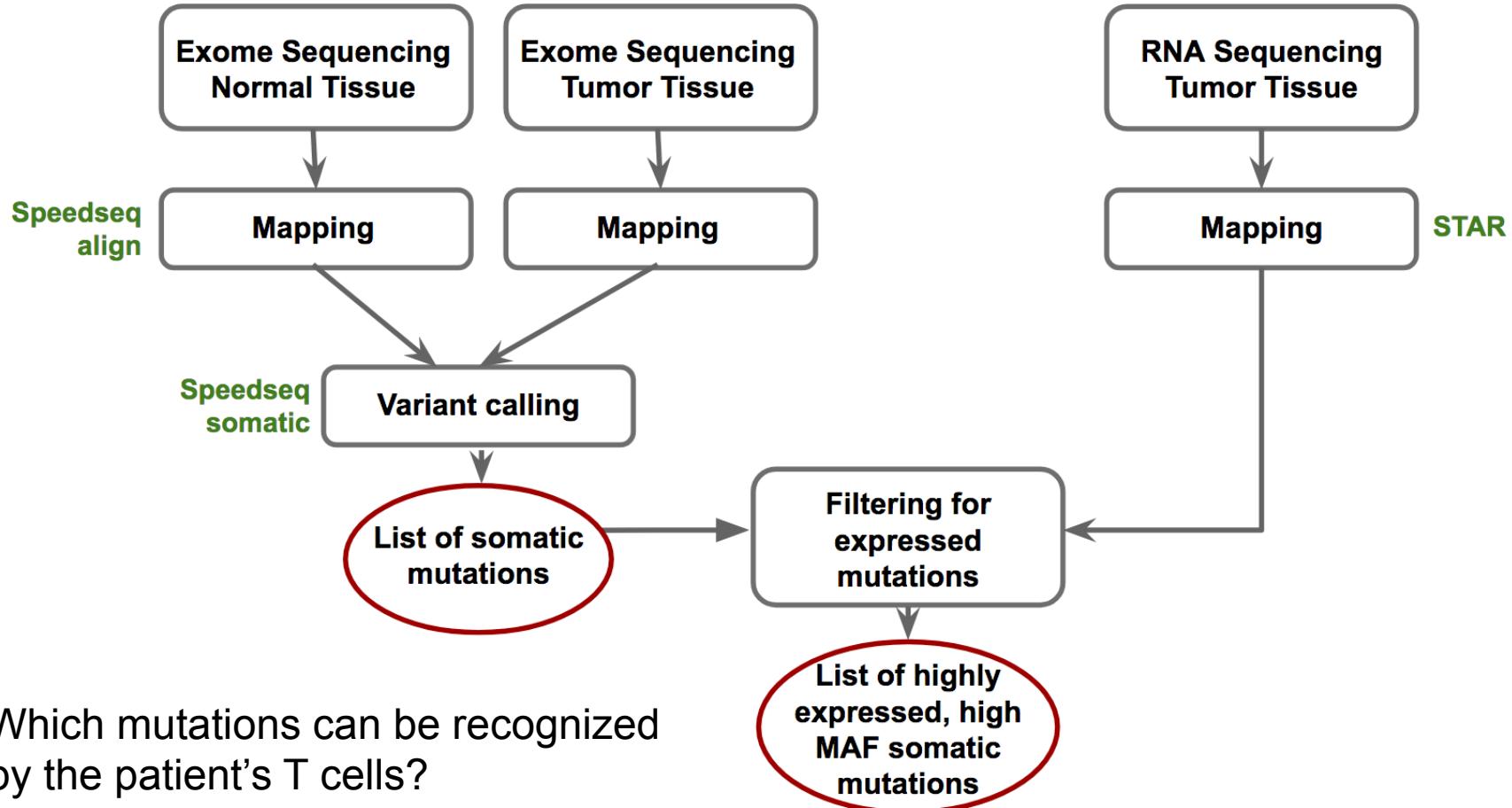


**Neoepitopes** can be  
recognized by tumor-infiltrating  
lymphocytes (**TILs**)

**Neoepitopes** are highly  
tumor-specific!



# 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

# Hands On Part 2.1

---

- Identify mutated sites in the provided p53 protein sequences
- Identify sequence regions that contain all 9-mer peptides that are only found in the tumor (that contain a mutation)

# Input data from patient:

---

## >P53\_HUMAN Cellular tumor antigen p53 - Healthy Tissue

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPSQAMDDLMLSPDDIEQWFTEDPGP  
DEAPRMPEAAPPVAPAPAAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIIYKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENLRKKGEPHHELP  
PGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  
GSRAHSSHLSKGQSTSRRHKKLMFKTEGPDS

## >P53\_HUMAN Cellular tumor antigen p53 - Tumor Tissue

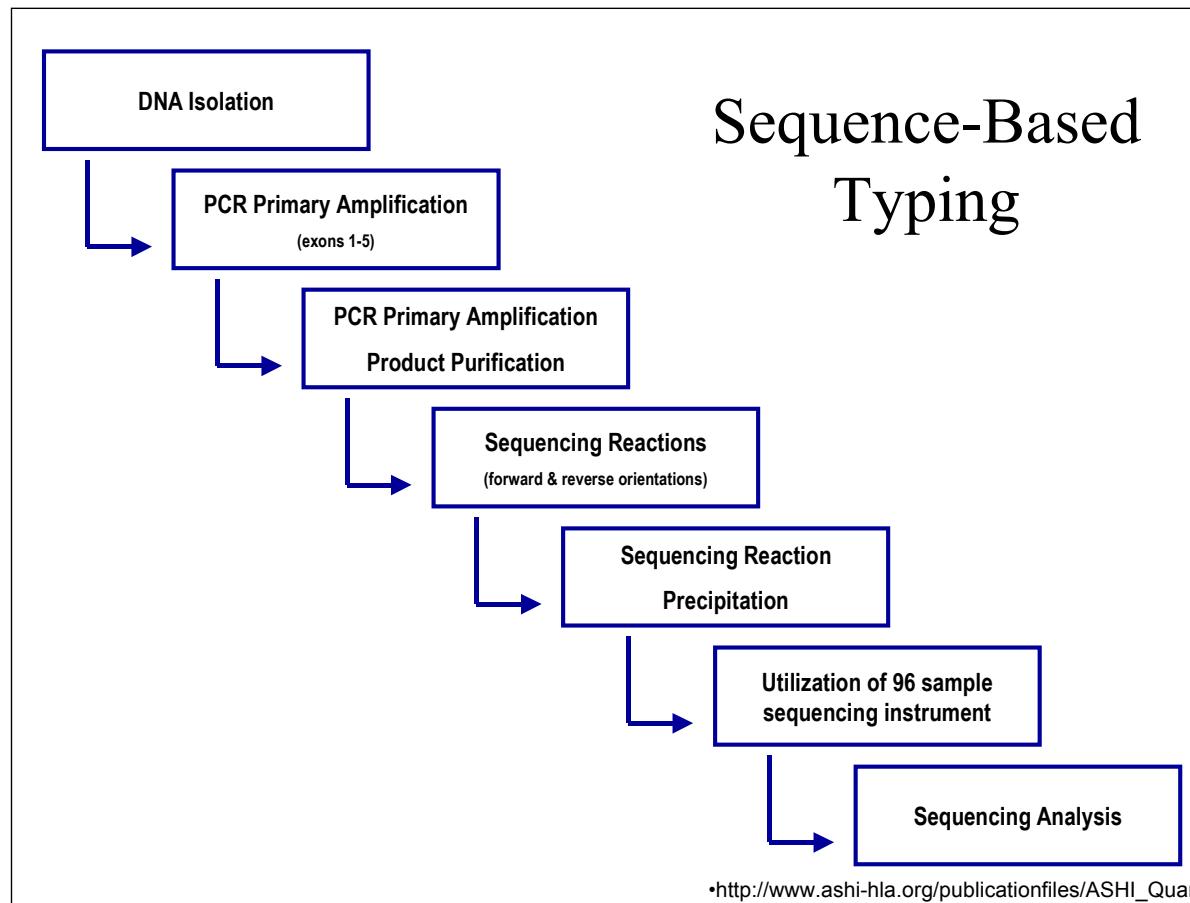
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPSQAMLDLMLSPDDIEQWFTEDPGP  
DEAPWMPEAAPPVAPAPAAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIIYKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFVHSVVVPYEPPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILTIITLEV

# Overview

---

- Background Biology: The immune system can detect and eliminate mutated cancer cells
- Background Immunoinformatics:
  - Mutations can be detected using patient sequencing data
  - Mutations that can be recognized by T cells can be predicted
- Hands on: Design a personalized cancer vaccine

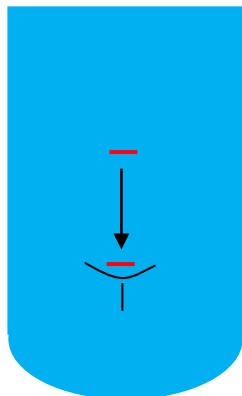
# HLA Typing: Targeted sequencing of HLA locus



HLA genotype of patient can be determined in the hospital/lab or can also be determined with immunoinformatics tools using the sequencing data

# Measuring and predicting MHC:peptide binding

## Experimental Basis: MHC Binding Assay



List of peptides with allele specific binding affinity

Sequence	IC <sub>50</sub>
QIVTMFEAL	3.6
LKGPDIFYKG	308
NFCNLTSAF	50,000
AQSQCRTFR	38,000
CTYAGPFGM	143
CFGNTAVAK	50,000
...	

$\log(\text{IC}_{50}) \sim \text{Binding free Energy}$

low IC<sub>50</sub> → high affinity

## 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 V 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 ...

**Impossible to measure all peptides**

→ Predict binding peptides using machine learning

Find function  $F_i$  in  $F_1, F_2, F_3, \dots$   
 $F_i(\text{Sequence}) \approx \text{Affinity}$

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

# 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)

 No file chosen

Choose sequence format

## Choose a Prediction Method

Prediction Method

 [Help on prediction method selections](#)

## Specify what to make binding predictions for

MHC source species

Show only frequently occurring alleles:  [?](#)  
Select MHC allele(s)[Select HLA allele reference set:](#)  [?](#) [?](#)

## Specify Output

Sort peptides by

Show

Output format

Email address (optional)

 [?](#)

## 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)

 No file chosen

Choose sequence format

## Choose a Prediction Method

Prediction Method

## Specify what to make binding predictions for

MHC source species

Show only frequently occurring alleles:  [?](#)

Select MHC allele(s)

 [?](#)[Select HLA allele reference set:](#)  [?](#)

## Specify Output

Sort peptides by

Show

Output format

Email address (optional)

 [?](#)

Prediction Method Version	2013-02-22 [ <a href="#">Older versions</a> ]
<b>Specify Sequence(s)</b>	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. <a href="#">(Browse for sequences in NCBI)</a>	<pre>&gt;Region 1 SPLPSQAMLDLMLSPDD &gt;Region 2 DPGPDEAPWMPEAAPPV</pre>
Or select file containing sequence(s)	<input type="button" value="Choose File"/> No file chosen
Choose sequence format	<input type="button" value="auto detect format"/>
<b>Choose a Prediction Method</b>	
Prediction Method	<input type="button" value="IEDB recommended"/> <input type="button" value="Help on prediction method selections"/>
<b>Specify what to make binding predictions for</b>	
MHC source species	<input type="button" value="human"/>
Show only frequently occurring alleles:	<input checked="" type="checkbox"/> <a href="#">?</a>
Select MHC allele(s)	<input type="button" value="Allele"/> <input type="button" value="Length"/> <input type="button" value="Upload allele file"/> <a href="#">?</a>
Select HLA allele reference set:	<input type="button" value=""/> <a href="#">?</a>
Sort peptides by	<input type="button" value="HLA-A*01:01"/> <input type="button" value="HLA-A*02:01"/> <input type="button" value="HLA-A*02:06"/> <input type="button" value="HLA-A*03:01"/> <input type="button" value="HLA-A*11:01"/> <input type="button" value="HLA-A*23:01"/> <input type="button" value="HLA-A*24:02"/> <input type="button" value="HLA-A*25:01"/> <input type="button" value="HLA-A*26:01"/> <input type="button" value="HLA-A*29:02"/> <input type="button" value="HLA-A*30:01"/> <input type="button" value="HLA-A*30:02"/> <input type="button" value="HLA-A*31:01"/> <input type="button" value="HLA-A*32:01"/>
Show	<input type="button" value=""/>
Output format	<input type="button" value=""/>
Email address (optional)	<input type="text"/> <a href="#">?</a>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

Prediction Method Version	2013-02-22 [ <a href="#">Older versions</a> ]
<b>Specify Sequence(s)</b>	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. <a href="#">(Browse for sequences in NCBI)</a>	<pre>&gt;Region 1 SPLPSQAMLDLMLSPDD &gt;Region 2 DPGPDEAPWMPEAAPPV</pre>
Or select file containing sequence(s)	<input type="button" value="Choose File"/> No file chosen
Choose sequence format	<input type="button" value="auto detect format"/>
<b>Choose a Prediction Method</b>	
Prediction Method	<input type="button" value="IEDB recommended"/> <a href="#">Help on prediction method selections</a>
<b>Specify what to make binding predictions for</b>	
MHC source species	<input type="button" value="human"/>
Show only frequently occurring alleles: <input checked="" type="checkbox"/> <a href="#">?</a>	<input type="button" value="Allele"/> <input type="button" value="Length"/>
Select MHC allele(s)	<input type="button" value="HLA-A*02:01"/> <a href="#">?</a>
Select HLA allele reference set: <input type="checkbox"/> <a href="#">?</a>	<input type="button" value="Upload allele file"/> <a href="#">?</a>
Sort peptides by	<input type="button" value="Percentile Rank"/>
Show	<input type="button" value="All predictions"/>
Output format	<input type="button" value="XHTML table"/> <a href="#">?</a>
Email address (optional)	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

Prediction Method Version	2013-02-22 [ <a href="#">Older versions</a> ]									
<b>Specify Sequence(s)</b>										
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. <a href="#">(Browse for sequences in NCBI)</a>	<pre>&gt;Region 1 SPLPSQAMLDLMLSPDD &gt;Region 2 DPGPDEAPWMPEAAPPV</pre>									
Or select file containing sequence(s)	<input type="button" value="Choose File"/> No file chosen									
Choose sequence format	<input type="button" value="auto detect format"/>									
<b>Choose a Prediction Method</b>										
Prediction Method	<input type="button" value="IEDB recommended"/> <input type="button" value="Help on prediction method selections"/>									
<b>Specify what to make binding predictions for</b>										
MHC source species	<input type="button" value="human"/>									
Show only frequently occurring alleles: <input checked="" type="checkbox"/> <a href="#">?</a> Select MHC allele(s) <a href="#">Select HLA allele reference set:</a> <input type="checkbox"/> <a href="#">?</a>	<table style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 30%;">Allele</th> <th style="width: 30%;">Length</th> <th style="width: 40%;"></th> </tr> <tr> <td>HLA-A*02:01</td> <td>9</td> <td><input type="button" value=""/></td> </tr> <tr> <td><input type="button" value=""/></td> <td><input type="button" value=""/></td> <td><a href="#">Upload allele file</a> <a href="#">?</a></td> </tr> </table>	Allele	Length		HLA-A*02:01	9	<input type="button" value=""/>	<input type="button" value=""/>	<input type="button" value=""/>	<a href="#">Upload allele file</a> <a href="#">?</a>
Allele	Length									
HLA-A*02:01	9	<input type="button" value=""/>								
<input type="button" value=""/>	<input type="button" value=""/>	<a href="#">Upload allele file</a> <a href="#">?</a>								
<b>Specify Output</b>										
Sort peptides by	<input type="button" value="Percentile Rank"/>									
Show	<input type="button" value="All predictions"/>									
Output format	<input type="button" value="XHTML table"/>									
Email address (optional)	<input type="text"/> <a href="#">?</a>									
<input type="button" value="Submit"/> <input type="button" value="Reset"/>										

# MHC-I Binding Predictions

 Loading... please wait.

© 2005-2017 | [IEDB Home](#)

Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services

# 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	MLDMLMLSPD	Consensus (ann/comblib_sidney2008/smm)	2.9
HLA-A*02:01	1	7	15	9	AMLDMLMLSP	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

---

- Background Biology: The immune system can detect and eliminate mutated cancer cells
-  Background Immunoinformatics: Mutations that can be recognized by T cells can be predicted
- Hands on: Design a personalized cancer vaccine

## Hands On Part 2.2

---

- 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 to select peptides that can be presented to T cells in this patient
- Step 3: Determine if the identified peptides are specific for the tumor
- Final question: Which peptide would you choose?

Bonus problem: Start with DNA sequencing data

# HLA binding Prediction on IEDB

## Peptides:

>D41L  
SPLPSQAMLD**L**MLSPDD  
>R65W  
DPGPDEAP**W**MPEAAPPV  
>R213V  
YLDDRNTF**V**HSVVVVPYE  
>D259V  
**V**ILTIITLE**V**

## **HLA Alleles:**

HLA-A\*02:01  
HLA-A\*68:01  
HLA-B\*07:02  
HLA-B\*35:01

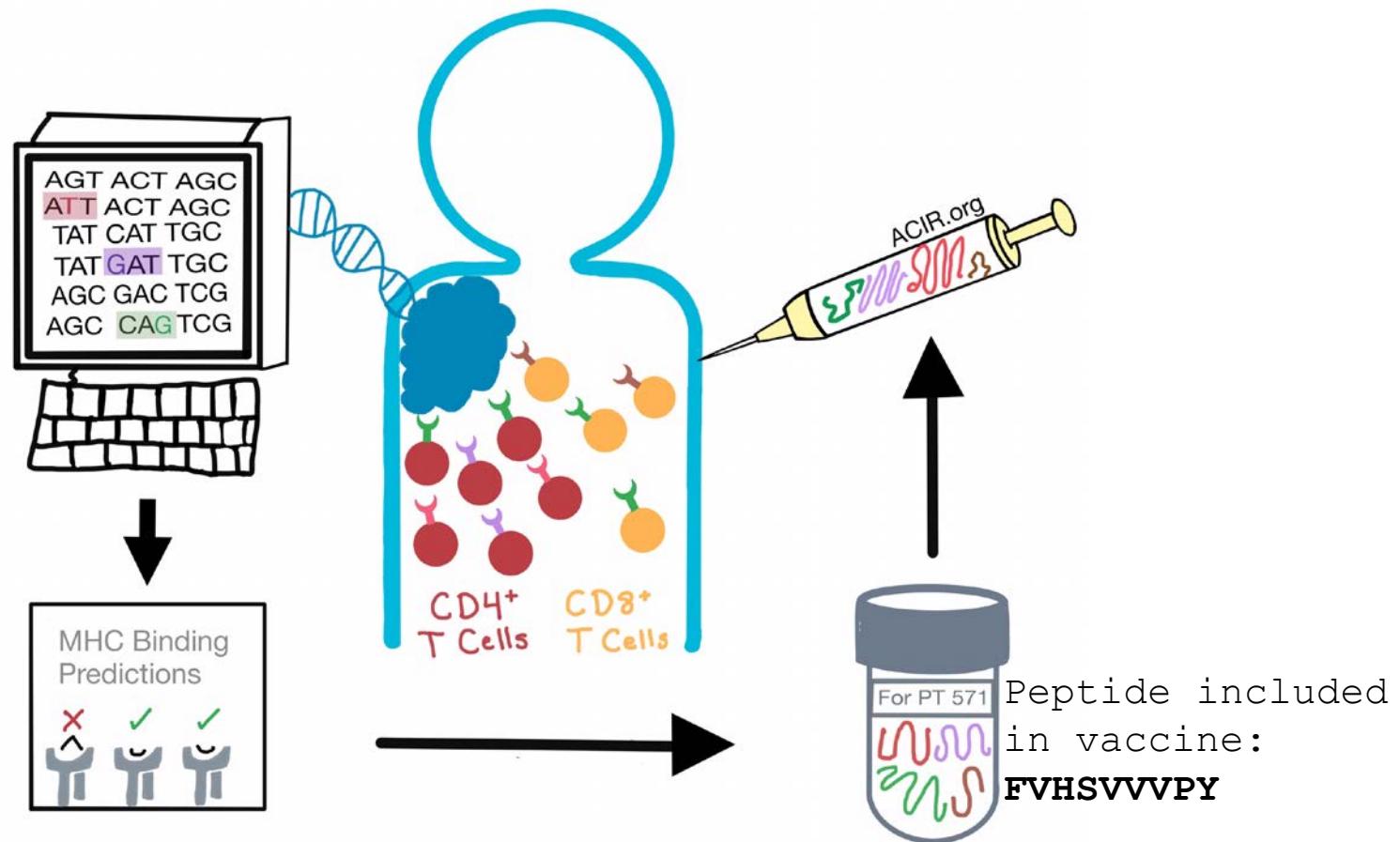
## Length:

9

# MHC-I Binding Predictions

Prediction Method Version	2013-02-22 <a href="#">[Older versions]</a>															
<b>Specify Sequence(s)</b>																
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. <a href="#">(Browse for sequences in NCBI)</a>	<pre>&gt;D41L SPLPSQAMLDLMLSPDD &gt;R65W DPGPDEAPWMPEAAPPV &gt;R213V YLDDRNTFVHSVVVPYE &gt;D259V ILTIITLEV</pre>															
Or select file containing sequence(s)	<input type="file"/> Choose File No file chosen															
Choose sequence format	<input type="button" value="auto detect format"/>															
<b>Choose a Prediction Method</b>																
Prediction Method	<input type="radio"/> IEDB recommended 2.19 <a href="#">Help on prediction method selections</a>															
Show all the method versions:	<input type="checkbox"/>															
<b>Specify what to make binding predictions for</b>																
MHC source species	<input type="button" value="human"/>															
Show only frequently occurring alleles:	<input checked="" type="checkbox"/>															
Select MHC allele(s)	<input type="checkbox"/>															
Select HLA allele reference set:	<input type="checkbox"/>															
	<table border="1"> <thead> <tr> <th>Allele</th> <th>Length</th> <th>Remove</th> </tr> </thead> <tbody> <tr> <td>HLA-A*02:01</td> <td>9</td> <td><input type="checkbox"/></td> </tr> <tr> <td>HLA-A*68:01</td> <td>9</td> <td><input type="checkbox"/></td> </tr> <tr> <td>HLA-B*07:02</td> <td>9</td> <td><input type="checkbox"/></td> </tr> <tr> <td>HLA-B*35:01</td> <td>9</td> <td><input type="checkbox"/></td> </tr> </tbody> </table> <input type="button"/> <input type="button"/> <a href="#">Upload allele file</a> ?	Allele	Length	Remove	HLA-A*02:01	9	<input type="checkbox"/>	HLA-A*68:01	9	<input type="checkbox"/>	HLA-B*07:02	9	<input type="checkbox"/>	HLA-B*35:01	9	<input type="checkbox"/>
Allele	Length	Remove														
HLA-A*02:01	9	<input type="checkbox"/>														
HLA-A*68:01	9	<input type="checkbox"/>														
HLA-B*07:02	9	<input type="checkbox"/>														
HLA-B*35:01	9	<input type="checkbox"/>														
<b>Specify Output</b>																
Sort peptides by	<input type="button" value="Percentile Rank"/>															
Show	<input type="button" value="All predictions"/>															
Output format	<input type="button" value="XHTML table"/>															
Email address (optional)	<input type="text"/>															
<input type="button" value="Submit"/>																

# Personalized Vaccines for Cancer Immunotherapy



# Contact & Feedback

---

Bjoern Peters

[bpeters@lji.org](mailto:bpeters@lji.org)

Zeynep Kosaloglu Yalcin

[zeynep@lji.org](mailto:zeynep@lji.org)

**Feedback link** < <https://goo.gl/forms/jfrqJHVq0SzCF4JC3> >

# Bonus Slides (For Reference)

# Workflow:

- **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 to select peptides that can be **presented to T cells** in this patient
- **Step 3:** Determine if the identified peptides are specific for the tumor (i.e. are they found anywhere else?)
- **Final question:** Which peptide would you choose?

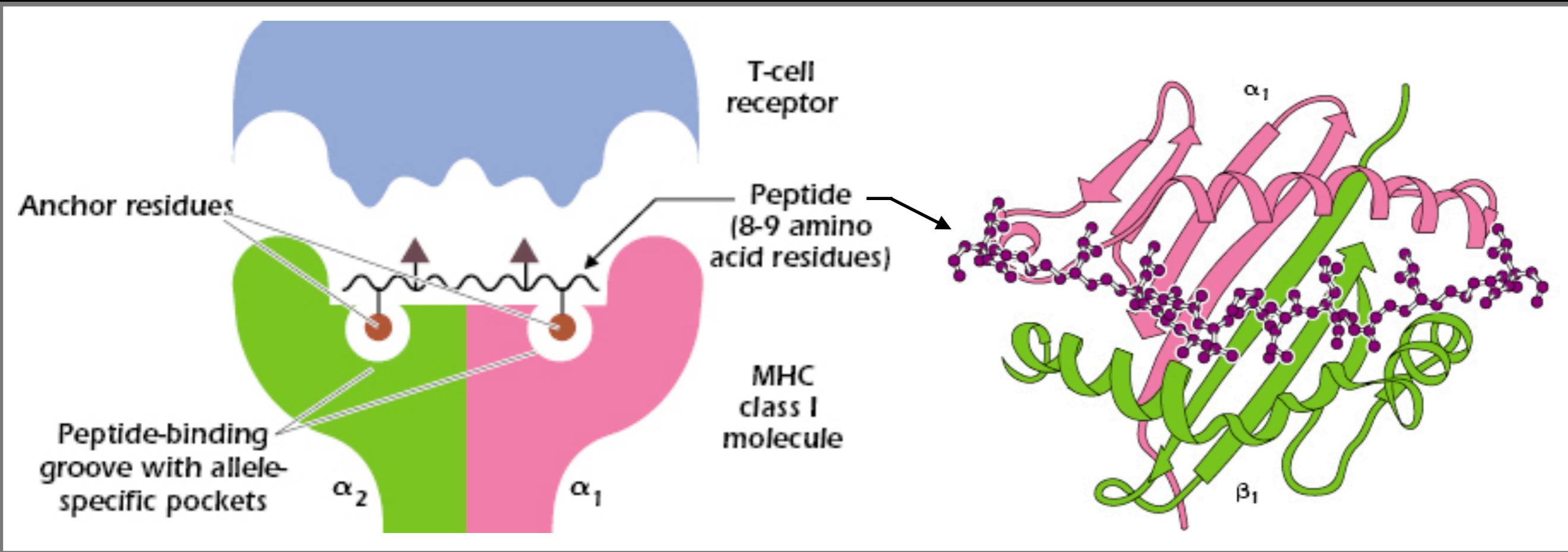
# Workflow:

- **Step 1:** Identify sequence regions that contain all 9-mer peptides that are only found in the tumor
  - ▶ What are the tumor specific amino-acids?
  - ▶ What peptides include these amino acids?
- **Step 2:** Run HLA binding prediction to identify 9-mer peptides in the sequence regions unique to the tumor to select peptides that can be **presented to T cells** in this patient
- **Step 3:** Determine if the identified peptides are specific for the tumor (i.e. are they found anywhere else?)
  - **Which peptide would you choose?**

# Workflow:

- **Step 1:** Identify sequence regions that contain all 9-mer peptides that are only found in the tumor
  - ▶ What are the tumor specific amino-acids?
  - ▶ What peptides include these amino acids?
- **Step 2:** Run HLA binding prediction to identify 9-mer peptides in the sequence regions unique to the tumor to select peptides that can be **presented to T cells** in this patient
- **Step 3:** Determine if the identified peptides are specific for the tumor (i.e. are they found anywhere else?)
  - **Which peptide would you choose?**

# Depictions of the peptide bound MHC and T-cell receptor



## Note:

- Anchor residues in the peptide bind to the allele-specific pockets of the MHC molecule.
- Certain MHC molecules (alleles) preferentially bind peptides with specific anchor residues in the 8- or 9-amino-acid peptide sequence.
- We want our tumor specific residues to be within 8 to 9-mer sequences bound by a patient HLA alleles!

# Input data from patient:

---

## >P53\_HUMAN Cellular tumor antigen p53 - Healthy Tissue

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPSQAMDDLMLSPDDIEQWFTEDPGP  
DEAPRMPEAAPPVAPAPAAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIIYKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENLRKKGEPHHELP  
PGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  
GSRAHSSHLSKGQSTSRRHKKLMFKTEGPDS

## >P53\_HUMAN Cellular tumor antigen p53 - Tumor Tissue

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPSQAMLDLMLSPDDIEQWFTEDPGP  
DEAPWMPEAAPPVAPAPAAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIIYKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFVHSVVVPYEPPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILTIITLEV

## Tumor Specific Site

\*  
|. | S | P | L | P | S | Q | A | M | L | D | L | M | L | S | P | D | D | K | L | P | Q | . | . |

## Tumor Specific Site

\*  
|. | S P L P S Q A M | L | D | L | M | L | S | P | D | D | K | L | P | Q | . .  
| L | D | L | M | L | S | P | D | D |

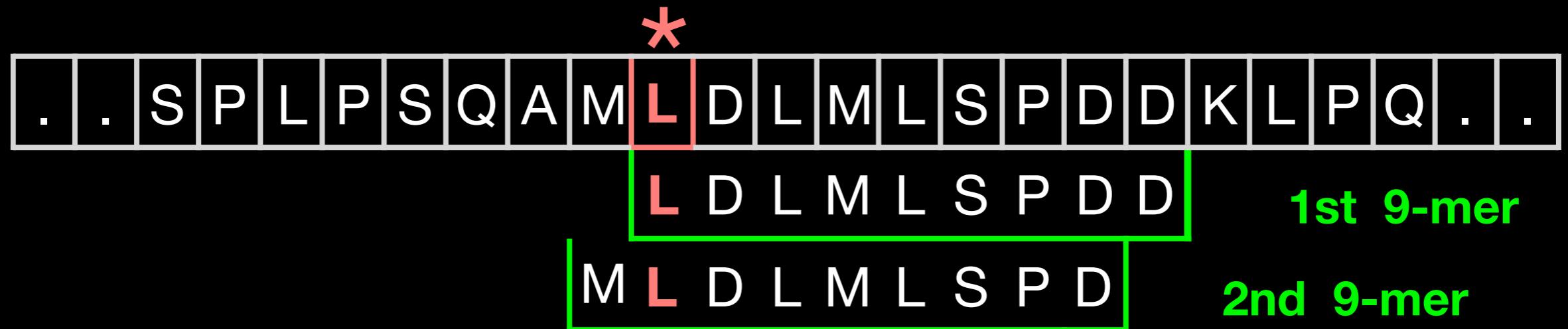
**1st 9-mer**

## Tumor Specific Site

\*  
|. | S P L P S Q A M | L | D | L | M | L | S | P | D | D | K | L | P | Q | . .

L D L M L S P D D      **1st 9-mer**

M L D L M L S P D      **2nd 9-mer**



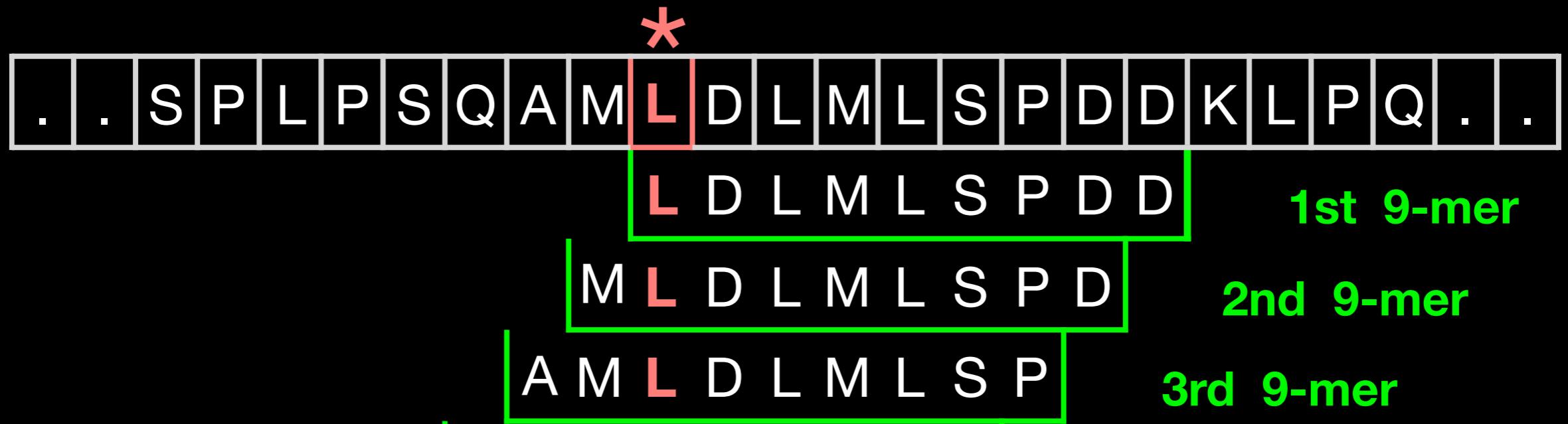
## Tumor Specific Site

.\*.SPLPSQAM**L**|D|L|M|L|S|P|D|D|K|L|P|Q|.\*.

**L D L M L S P D D**      **1st 9-mer**

**M L D L M L S P D**      **2nd 9-mer**

**A M L D L M L S P**      **3rd 9-mer**



## Tumor Specific Site

..SPLPSQAMLDLMLSPDDKLPQ..  
\*  
L D L M L S P D D **1st 9-mer**  
M L D L M L S P D **2nd 9-mer**  
A M L D L M L S P **3rd 9-mer**  
Q A M L D L M L S **4th 9-mer**

# Tumor Specific Site

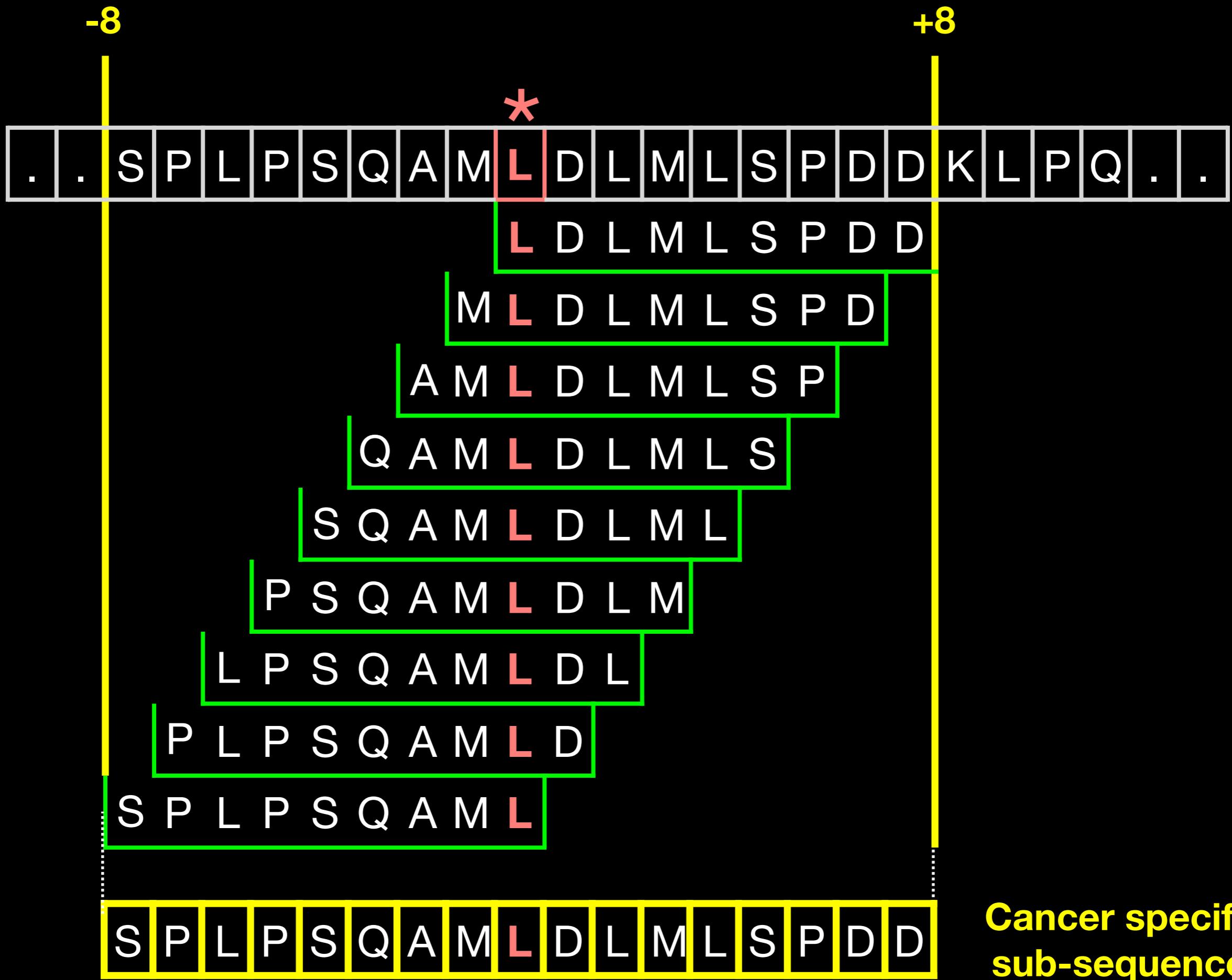
The diagram illustrates the extraction of 9-mers from a protein sequence. The sequence is shown as a row of boxes, each containing a single amino acid. A red asterisk (\*) marks the start of the sequence. Below the sequence, green boxes highlight 9-mers starting at each position, with the last one being incomplete.

- 1st 9-mer:** L D L M L S P D D
- 2nd 9-mer:** M L D L M L S P D
- 3rd 9-mer:** A M L D L M L S P
- 4th 9-mer:** Q A M L D L M L S
- 5th 9-mer:** S Q A M L D L M L
- 6th 9-mer:** P S Q A M L D L M
- 7th 9-mer:** L P S Q A M L D L
- 8th 9-mer:** P L P S Q A M L D
- 9th 9-mer:** S P L P S Q A M L

-8 +8

..SPLPSQAMLDLMLSPDDKLQP..

\*  
L D L M L S P D D  
M L D L M L S P D  
A M L D L M L S P  
Q A M L D L M L S  
S Q A M L D L M L  
P S Q A M L D L M  
L P S Q A M L D L  
P L P S Q A M L D  
S P L P S Q A M L



# Mutated sites

---

## >P53\_HUMAN Cellular tumor antigen p53 - Healthy Tissue

MEEPQSDPSVEPPLSQETFSDLWKL~~L~~PENNVLSP~~L~~PSQAM~~D~~DLMLSPDDIEQWFTEDPGP  
DEAP~~R~~MPEAAPPVAPAPAA~~P~~TPAAPAPAPS~~W~~PLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMF~~C~~QLAKTCPVQLWVDSTPPPGTRVRAMA~~I~~YKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNT~~F~~RHSVVVPYE~~P~~EVGS~~D~~CTTIHYNYMCNS  
SCMGGMNRRPIL~~T~~IITLE~~D~~SSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP  
PGSTKR~~A~~L~~P~~NNTSSSPQP~~K~~KPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  
GSRAHSSH~~L~~KS~~K~~G~~Q~~TSRHKKLMFKTEGP~~D~~SD

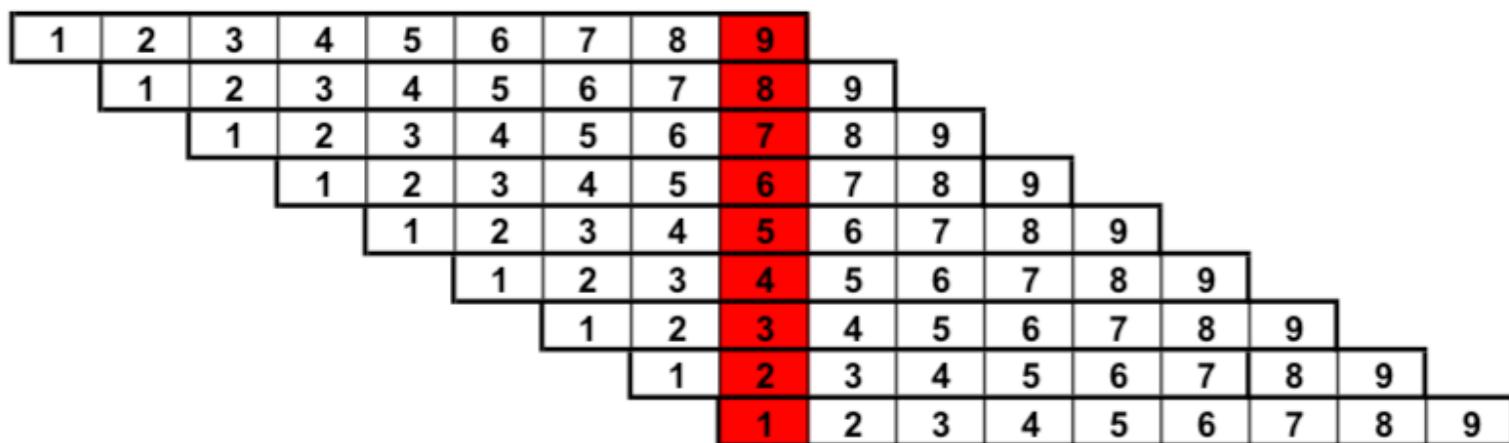
## >P53\_HUMAN Cellular tumor antigen p53 - Tumor Tissue

MEEPQSDPSVEPPLSQETFSDLWKL~~L~~PENNVLSP~~L~~PSQAM~~L~~DLMLSPDDIEQWFTEDPGP  
DEAP~~W~~MPEAAPPVAPAPAA~~P~~TPAAPAPAPS~~W~~PLSSSVPSQKTYQGSYGFRLGFLHSGTAK  
SVTCTYSPALNKMF~~C~~QLAKTCPVQLWVDSTPPPGTRVRAMA~~I~~YKQSQHMTEVVRRCPHHE  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNT~~F~~VHSVVVPYE~~P~~EVGS~~D~~CTTIHYNYMCNS  
SCMGGMNRRPIL~~T~~IITLE~~V~~

How long do the sequence regions that contain all 9-mer peptides with the mutation need to be?

## Mutation

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 X +1 +2 +3 +4 +5 +6 +7 +8 +9 +10



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

→ 17-mer peptide will cover all 9-mers that contain the mutation

# Mutated regions

## >P53\_HUMAN Cellular tumor antigen p53 - Healthy Tissue

MEEPQSDPSVEPLSQETFSDLWKL[PENNVL]SPLPSQAM**D**DLMSPDDIEQWFTEDPGP  
DEAP**R**MPEAAPPVAPAPAAPTPAAPAPAPS[WPLSSSVPSQKTYQGSYGFRLGFLHSGTAK]  
SVTCTYS[PALNKMF]CQLAKTCPQLWVDSTPPPGTRVRAMA[IYKQSQHMTEVVRRCPH]E  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTF**R**HSV[VVPYE]PPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILT[IITLE]DSSGNLLGRNSFEVRVCACPGRDRRTEENLRKKGEPHHELP  
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  
GSRAHSSHLSKKGQSTS[RHKKLMFKTEGP]DSD

## >P53\_HUMAN Cellular tumor antigen p53 - Tumor Tissue

MEEPQSDPSVEPLSQETFSDLWKL[PENNVL]SPLPSQAM**I**DLMSPDDIEQWFTEDPGP  
**DEAPW**MPEAAPPVAPAPAAPTPAAPAPAPS[WPLSSSVPSQKTYQGSYGFRLGFLHSGTAK]  
SVTCTYS[PALNKMF]CQLAKTCPQLWVDSTPPPGTRVRAMA[IYKQSQHMTEVVRRCPH]E  
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTF**V**HSV[VVPYE]PPEVGSDCTTIHYNYMCNS  
SCMGGMNRRPILT[IITLE]**V**

### >D41L

SPLPSQAM**L**DLSPDD

### >R65W

DPGPDEAP**W**MPEAAPPV

### >R213V

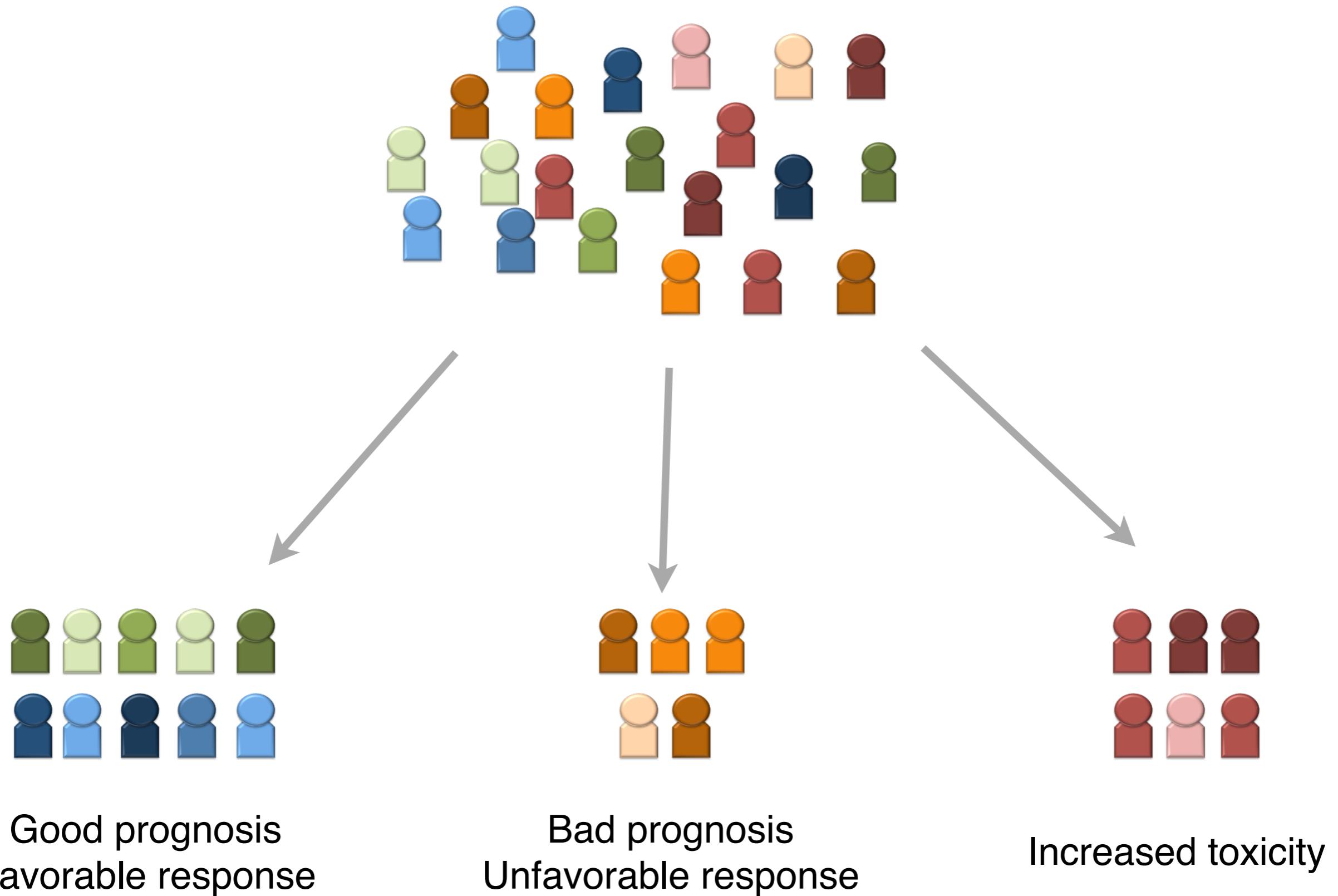
YLDDRNTF**V**HSV[VVPYE]

### >D259V

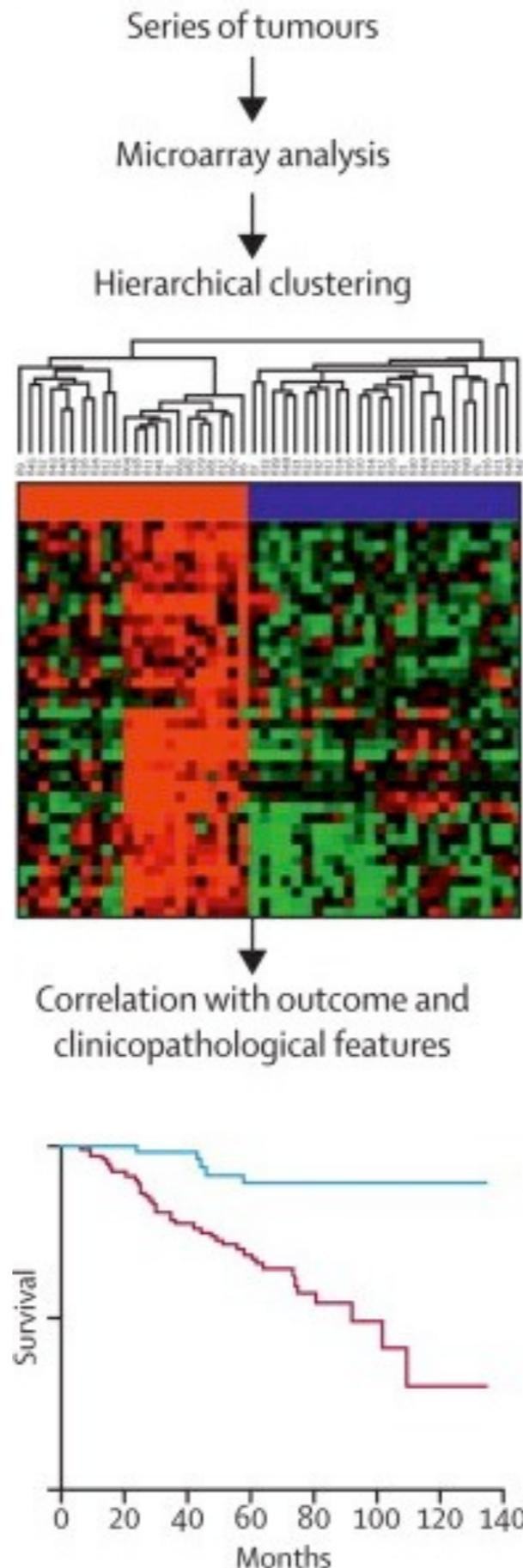
ILT[IITLE]**V**

Genetic and genomic approaches  
can identify a cancers molecular  
signature to usefully stratify  
tumors for treatment

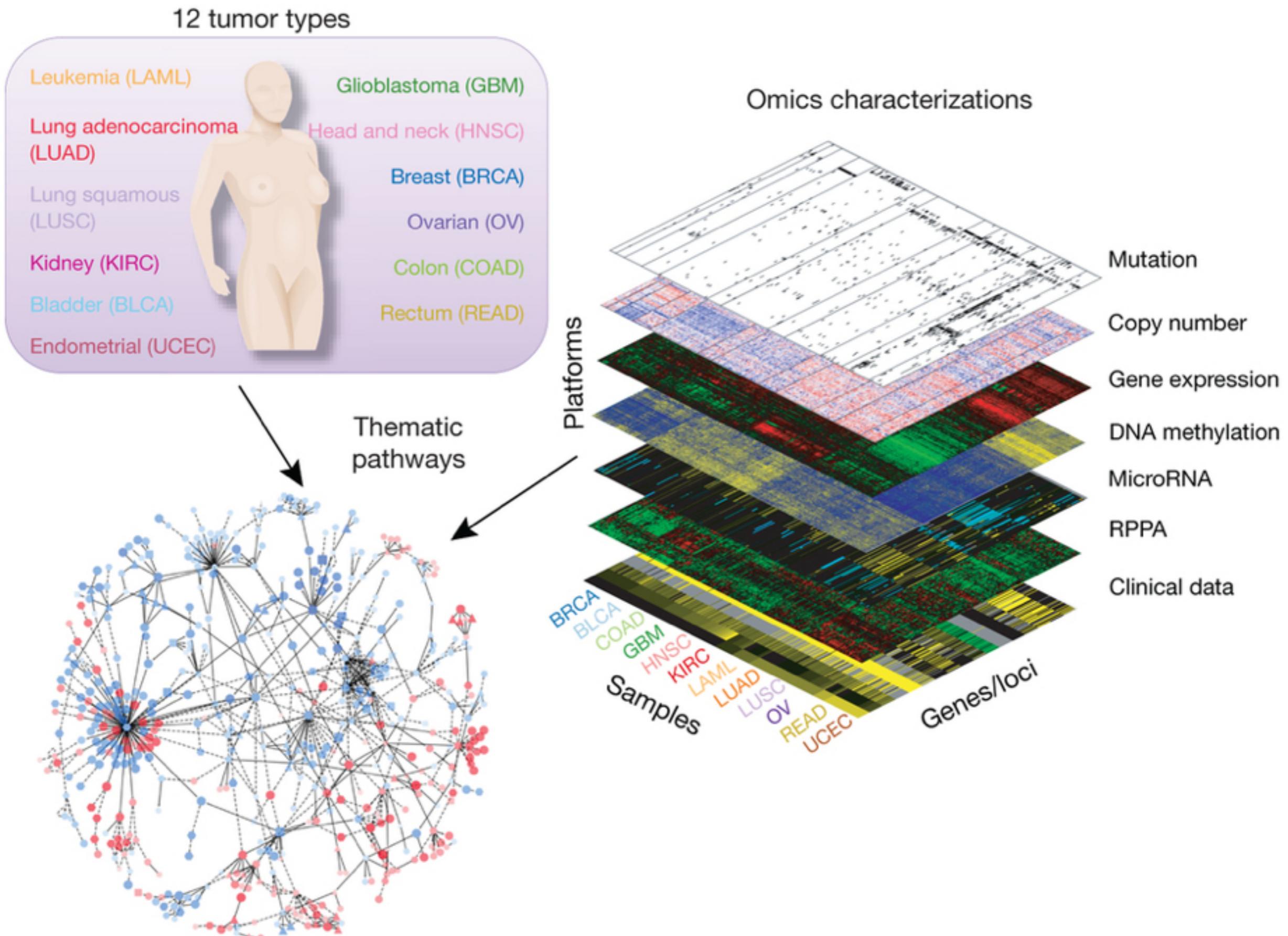
# Stratify tumors based on molecular patterns



# Stratify tumors based on molecular patterns



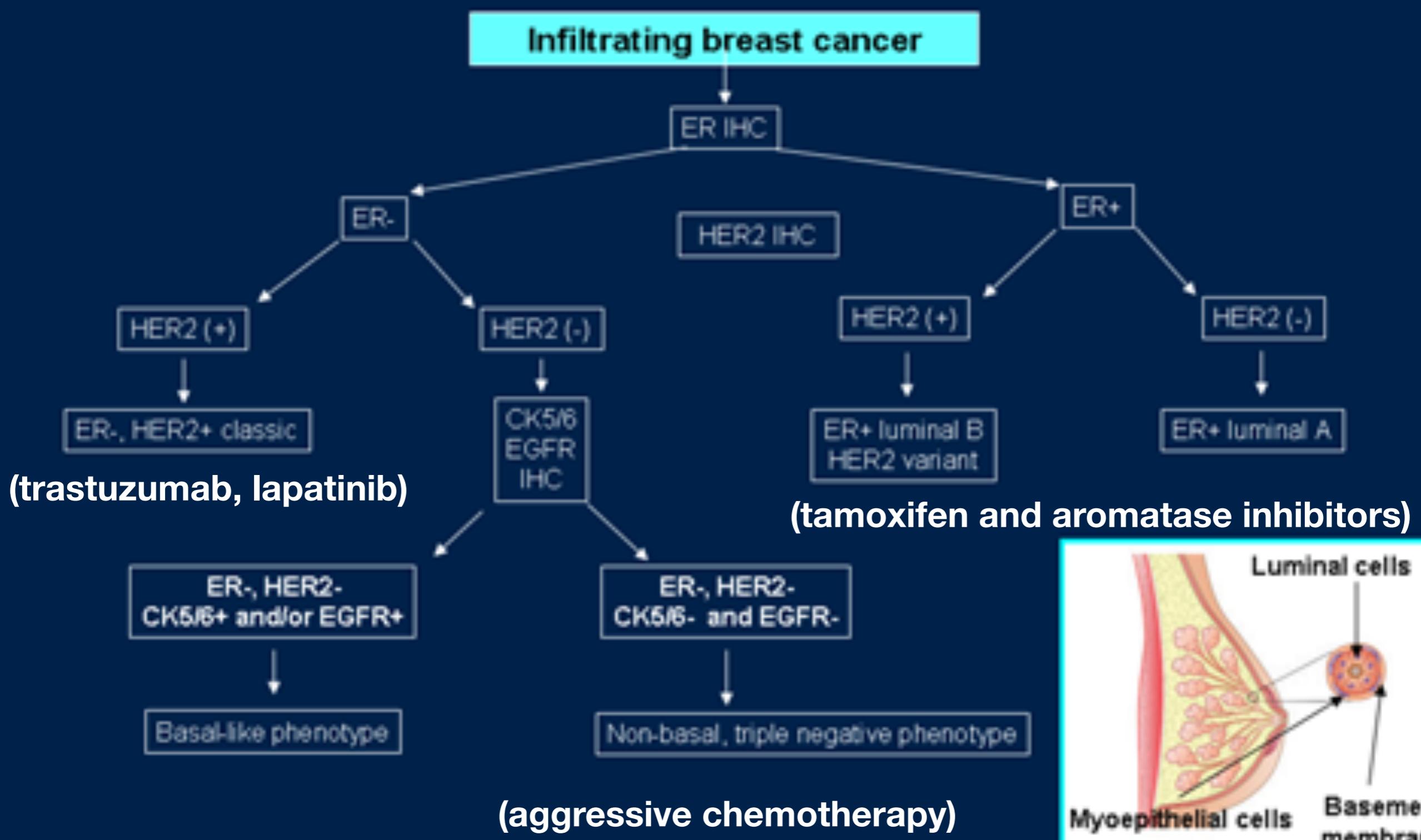
# TCGA Pan-Cancer project



For example, breast cancer may be classified into various types based upon which proteins are expressed on the surface of the tumor cells. Breast tumors that express human epidermal growth factor 2 (HER2), estrogen receptor (ER), and progesterone receptor (PR), or are triple negative (do not express HER2, ER, or PR) behave differently and have different prognoses. Tumors that are HER2 positive are treated with medications that bind to HER2 (e.g. trastuzumab, lapatinib) and inhibit its activity. ER and PR are hormone receptors, and ER/PR positive tumors are treated with antihormonal therapies (e.g. tamoxifen and aromatase inhibitors). Triple negative tumors have the poorest prognosis and are unlikely to respond to HER2-targeted therapies or antihormonal therapies. Such cancers are usually treated very aggressively with chemotherapy.

As more has been learned about the molecular signature of various cancer subtypes, therapies that are specifically targeted to those signatures have been developed. Conventional chemotherapy acts on all rapidly dividing cells and does not distinguish between cancer cells and normal cells.

# Classification of Breast Cancer



# *Readings to find out more...*

Leading Edge  
**Review**

Cell

## The Genetic Basis for Cancer Treatment Decisions

Janet E. Dancey,<sup>1,2</sup> Philippe L. Bedard,<sup>3,4</sup> Nicole Onetto,<sup>1</sup> and Thomas J. Hudson<sup>1,5,6,\*</sup>

<sup>1</sup>Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada

<sup>2</sup>NCIC-Clinical Trials Group, Queen's University, Kingston, ON K7L 3N6, Canada

<sup>3</sup>Princess Margaret Hospital, Division of Medical Oncology and Hematology, University Health Network

<sup>4</sup>Department of Medicine

<sup>5</sup>Department of Medical Biophysics

<sup>6</sup>Department of Molecular Genetics

University of Toronto, Toronto, ON M5S 1A1, Canada

\*Correspondence: tom.hudson@oicr.on.ca

DOI 10.1016/j.cell.2012.01.014

Personalized cancer medicine is based on increased knowledge of the cancer mutation repertoire and availability of agents that target altered genes or pathways. Given advances in cancer genetics, technology, and therapeutics development, the timing is right to develop a clinical trial and research framework to move future clinical decisions from heuristic to evidence-based decisions. Although the challenges of integrating genomic testing into cancer treatment decision making are wide-ranging and complex, there is a scientific and ethical imperative to realize the benefits of personalized cancer medicine, given the overwhelming burden of cancer and the unprecedented opportunities for advancements in outcomes for patients.

# Your Turn

Read and share your thoughts on the following class *Readings*

- Calling cancer's bluff with neoantigen vaccines
- Can genomics help detect early cancer and monitor treatment effectiveness?
- The increasing cost of cancer therapies

[https://bioboot.github.io/bimm194\\_W18/readings/](https://bioboot.github.io/bimm194_W18/readings/)