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
Cancer Immunotherapy	**Hands-on analysis** to design personalized cancer vaccines and harness the patient's own immune system to fight cancer

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.

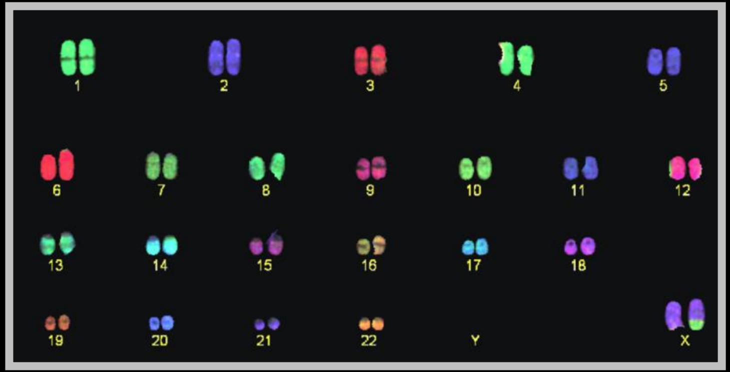
What is Cancer?

“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”

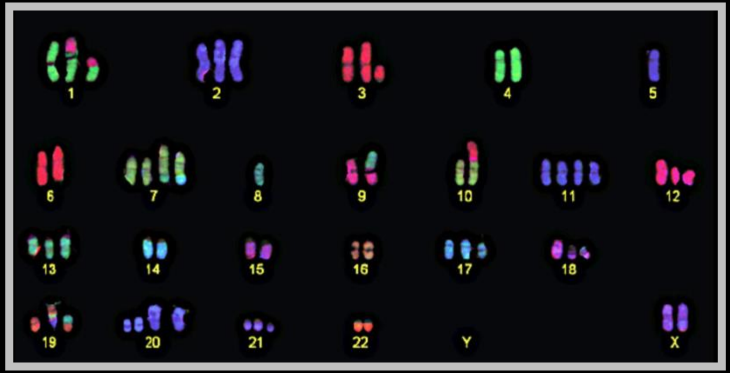
Source: https://www.cancer.gov

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.



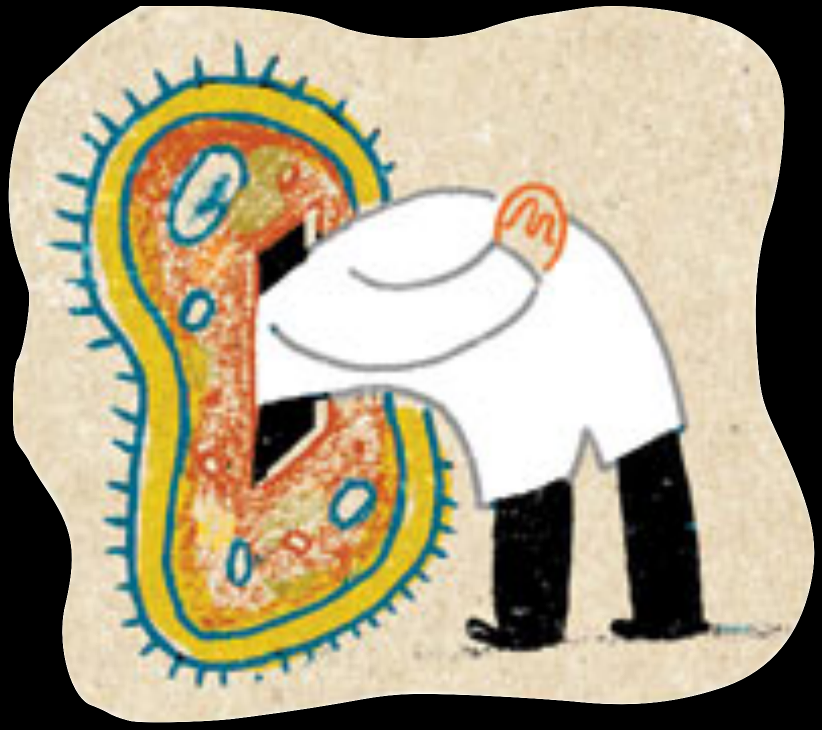
Healthy 46 chromosomes



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



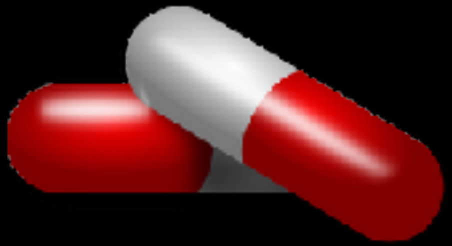
Finding Cancer Drivers

Motivation for adopting a genomics approach…

- Cancer is caused by mutations to specific genes

AGCT → AGAT

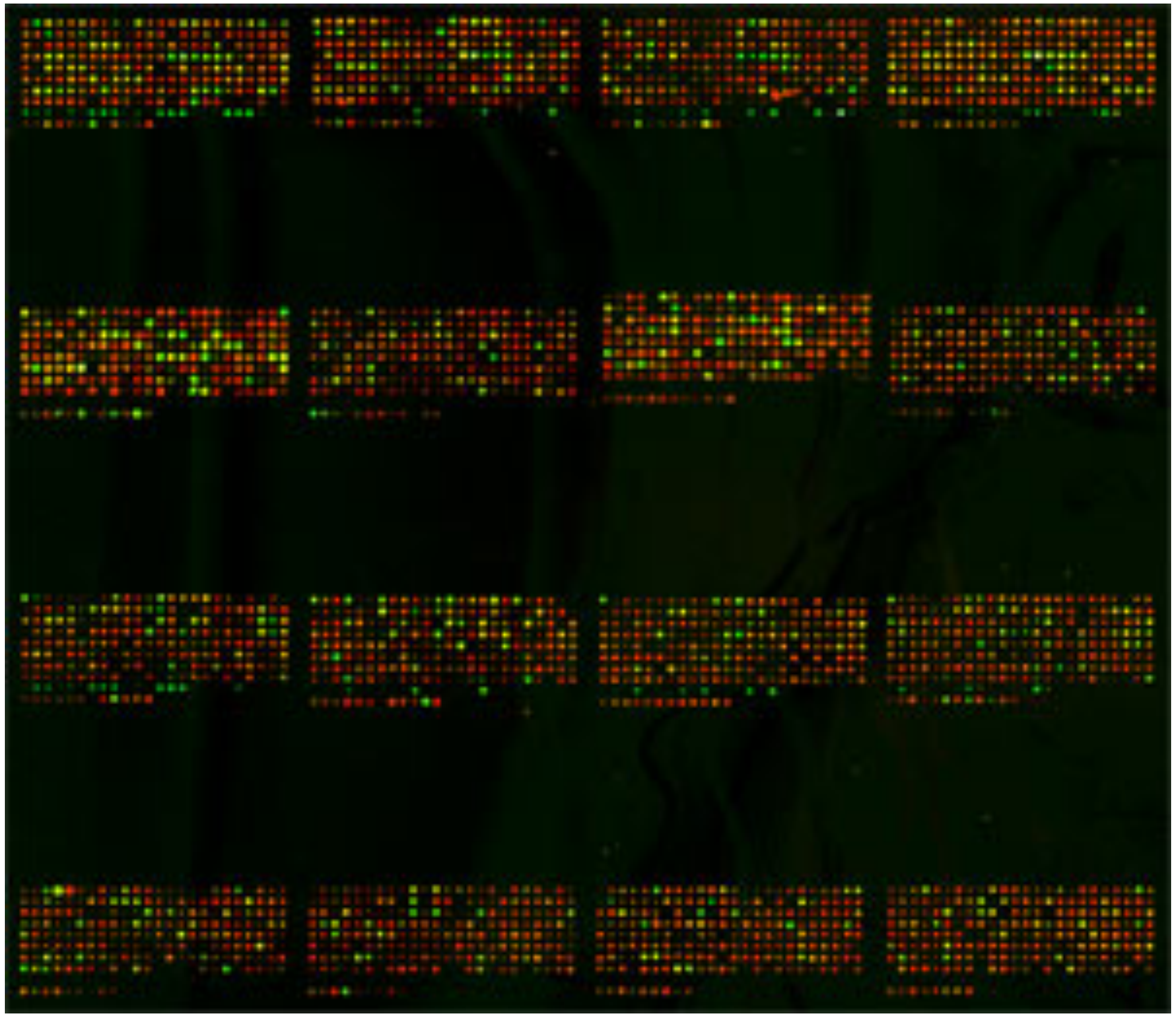
- Knowing which genes and proteins enables the development of **targeted treatments**



- <u>1st major Goal</u>:
  Define ALL cancer genes!

Use A Cancer Genomics Approach

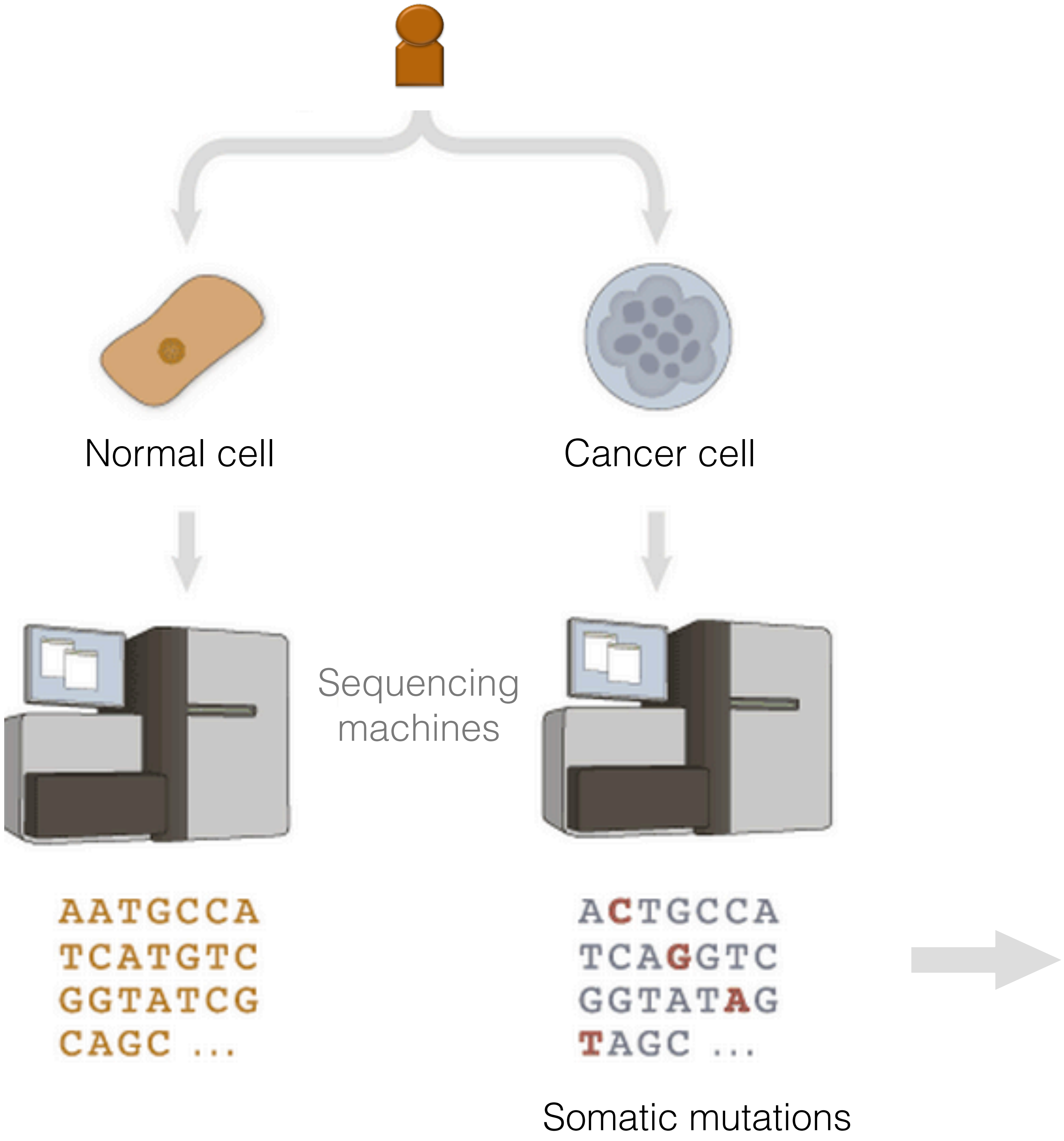
Arrays



Parallel Sequencing

```
ACTCAGCCCCAGCGGAGGTGAAGGACGTCCTTCCCCAGGAGCCGGTGAGA
AGCGCAGTCGGGGGCACGGGGATGAGCTCAGGGGCCTCTAGAAAGATGTA
GCTGGGACCTCGGGAAGCCCTGGCCTCCAGGTAGTCTCAGGAGAGCTACT
CAGGGTCGGGCTTGGGGAGAGGAGGAGCGGGGGTGAGGCCAGCAGCAGGG
GACTGGACCTGGGAAGGGCTGGGCAGCAGAGACGACCCGACCCGCTAGAA
GGTGGGGTGGGGAGAGCATGTGGACTAGGAGCTAAGCCACAGCAGGACCC
CCACGAGTTGTCACTGTCATTTATCGAGCACCTACTGGGTGTCCCCAGTG
TCCTCAGATCTCCATAACTGGGAAGCCAGGGGCAGCGACACGGTAGCTAG
CCGTCGATTGGAGAACTTTAAAATGAGGACTGAATTAGCTCATAAATGGA
AAACGGCGCTTAAATGTGAGGTTAGAGCTTAGAATGTGAAGGGAGAATGA
GGAATGCGAGACTGGGACTGAGATGGAACCGGCGGTGGGGAGGGGGAGGG
GGTGTGGAATTTGAACCCCGGGAGAGAAAGATGGAATTTTGGCTATGGAG
GCCGACCTGGGGATGGGGAAATAAGAGAAGACCAGGAGGGAGTTAAATAG
GGAATGGGTTGGGGGCGGCTTGGTAACTGTTTGTGCTGGGATTAGGCTGT
TGCAGATAATGGAGCAAGGCTTGGAAGGCTAACCTGGGGTGGGGCCGGGT
TGGGGTCGGGCTGGGGGCGGGAGGAGTCCTCACTGGCGGTTGATTGACAG
TTTCTCCTTCCCCAGACTGGCCAATCACAGGCAGGAAGATGAAGGTTCTG
TGGGCTGCCCCGACCCGCTAGAAGGTGGGGTGGGGAGAGCATGTGGACTA
GGAGCTAAGCCACAGCAGGACCCCCACGAGTTGTCACTGTCATTTATCGA
GCACCTACTGGGTGTCCCCAGTGTCCTCAGATCTCCATAACTGGGAAGCC
AGGGGCAGCGAC
```

Finding Cancer Associated Mutations

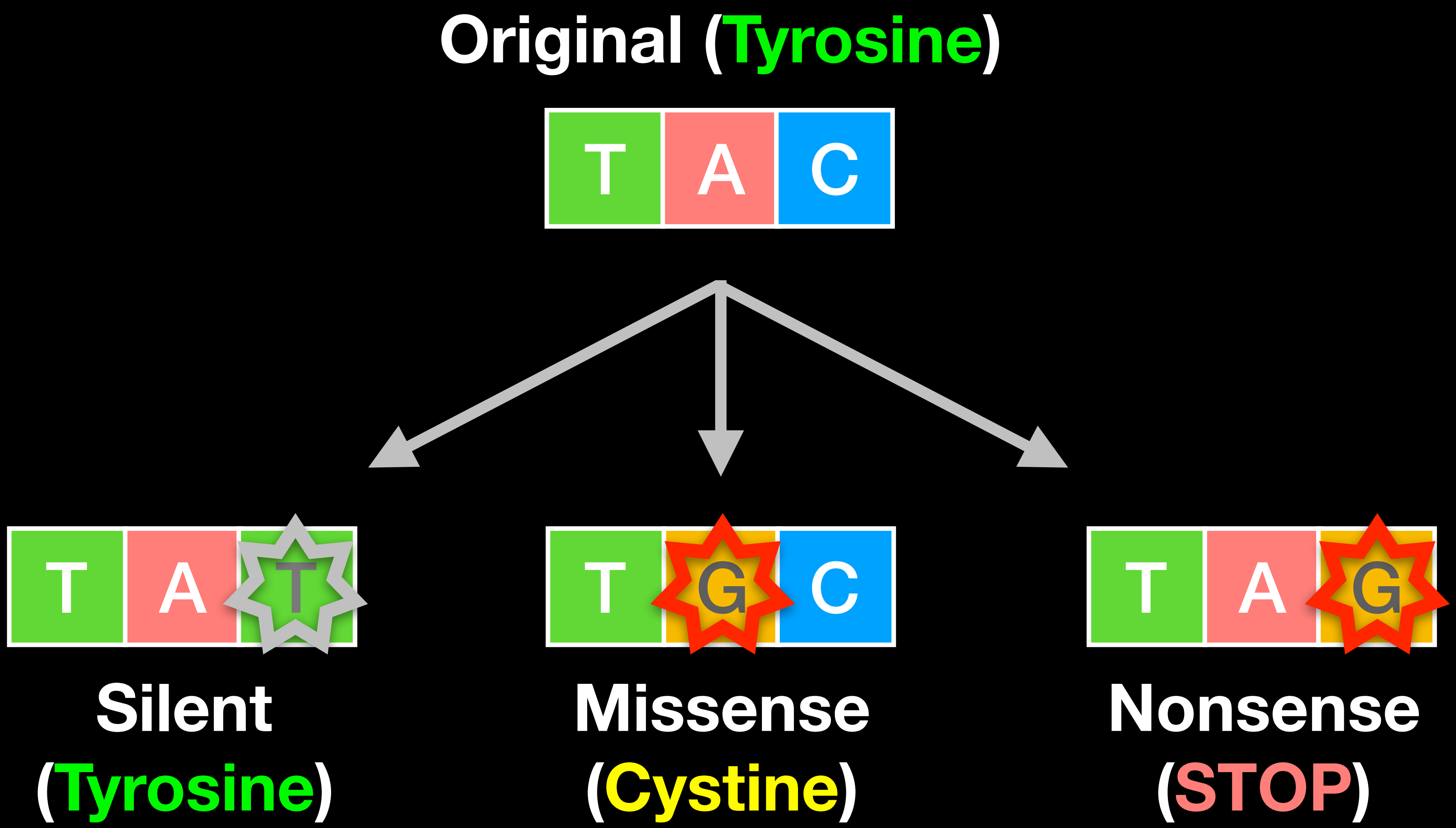


Identify all mutations specific to tumor cells

Filter out silent mutations

Mutations detected:

Point mutations



Original (Tyrosine)

T A C

T A T

Silent (Tyrosine)

T G C

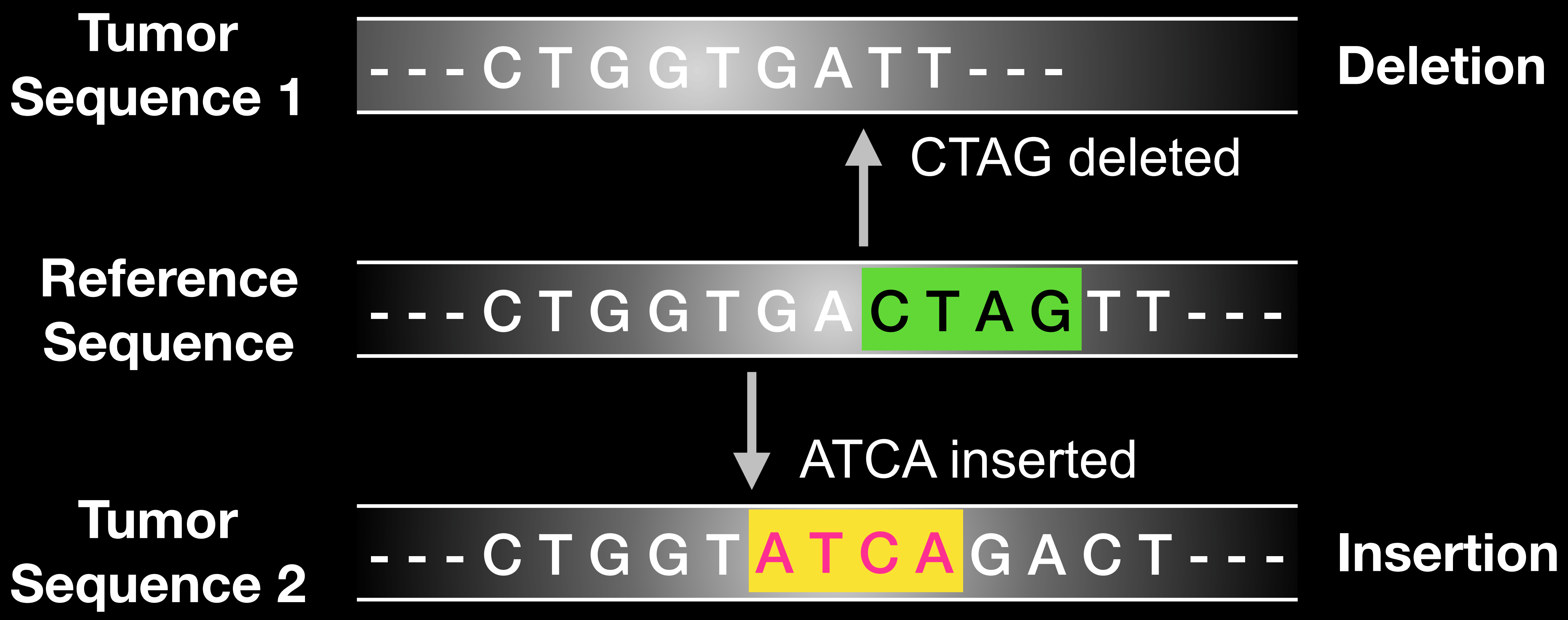
Missense (Cystine)

T A G

Nonsense (STOP)

Mutations detected:

Indels



Tumor Sequence 1

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

Deletion

CTAG deleted

Reference Sequence

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

ATCA inserted

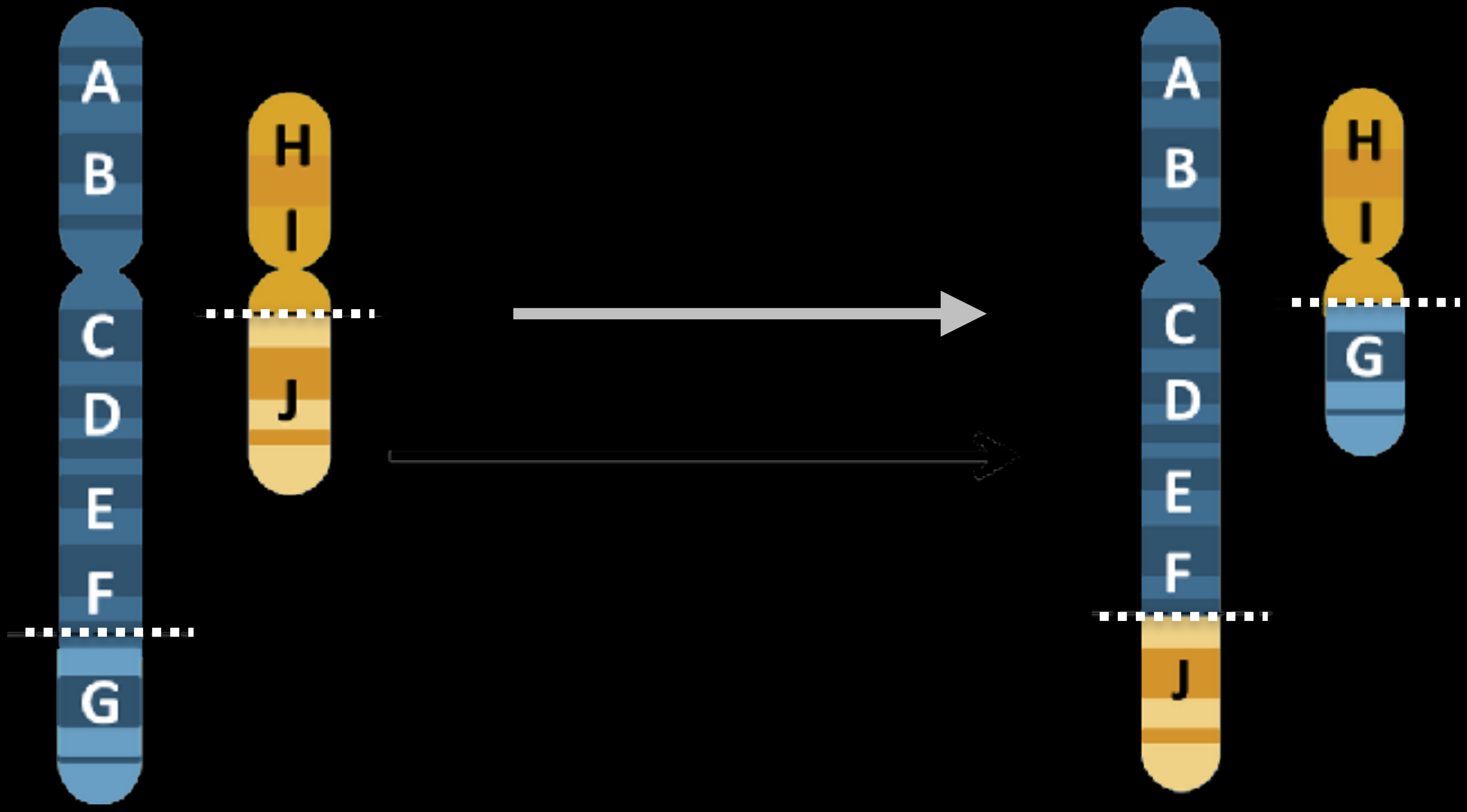
Tumor Sequence 2

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

Insertion

Mutations detected:

Translocations



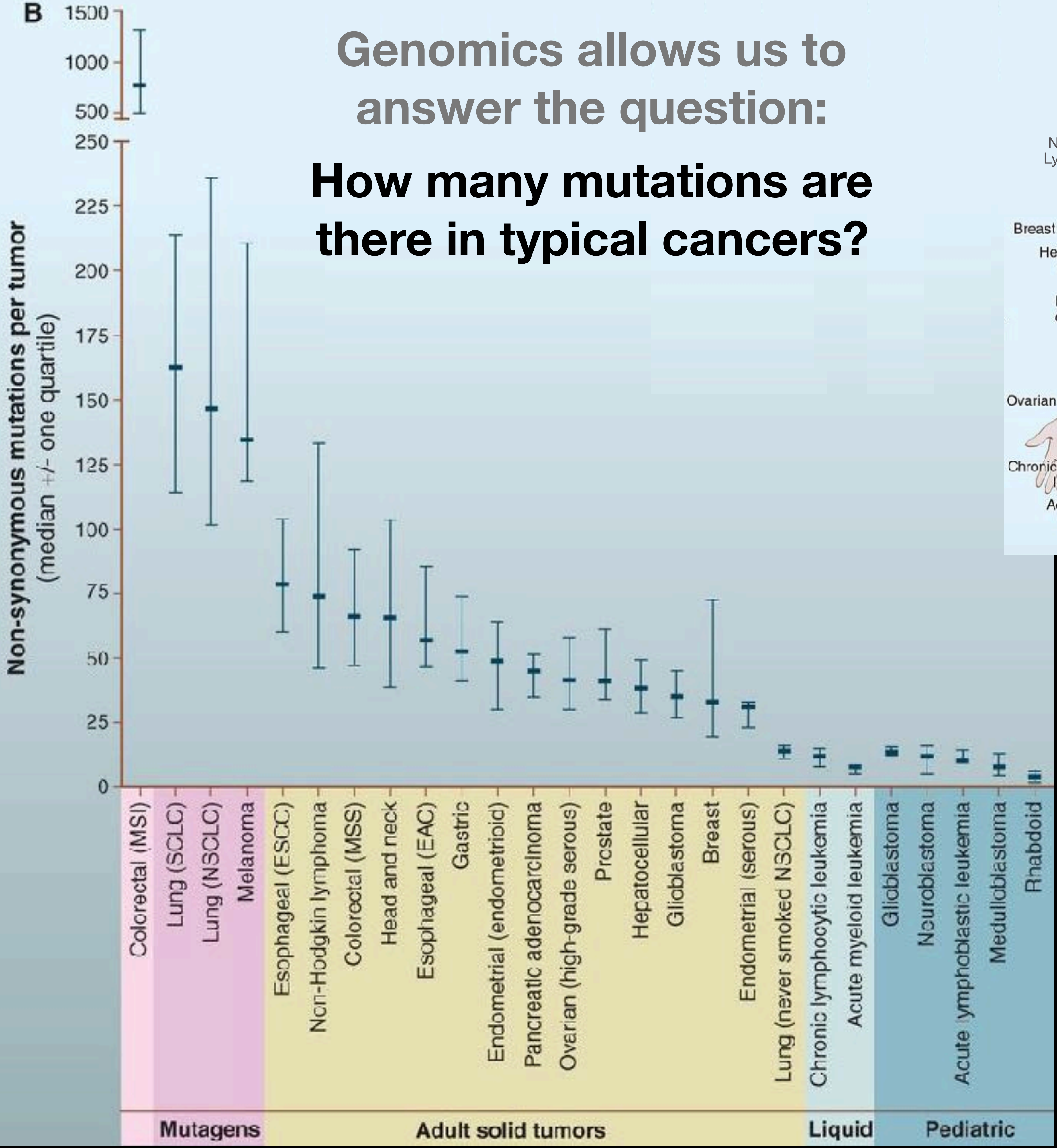
What can go wrong in cancer genomes?

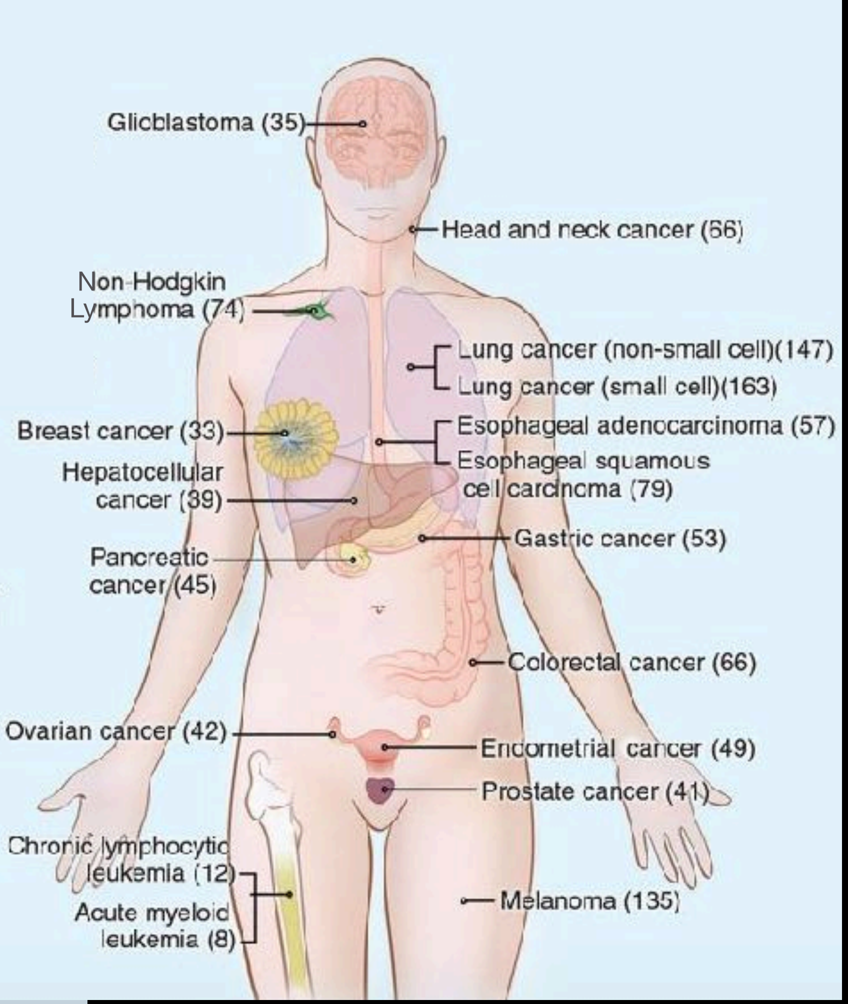
Type of change	**Some common technology to study changes**
DNA mutations	WGS, WXS
DNA structural variations	WGS
Copy number variation (CNV)	CGH array, SNP array, WGS
DNA methylation	Methylation array, RRBS, WGBS
mRNA expression changes	mRNA expression array, RNA-seq
miRNA expression changes	miRNA expression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

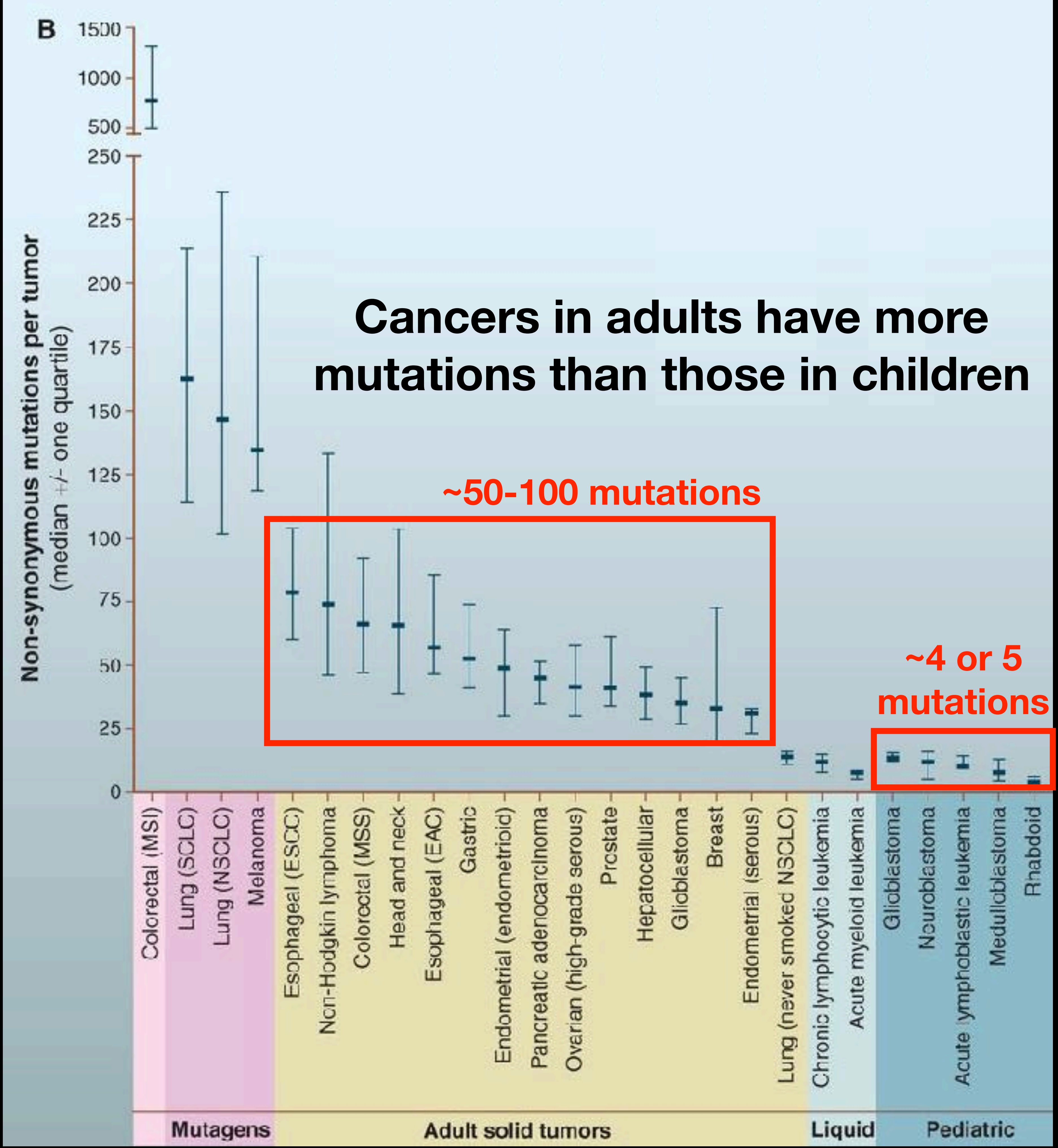
How many mutations are there in typical cancers?





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

Vogelstein et al. Science (2013)



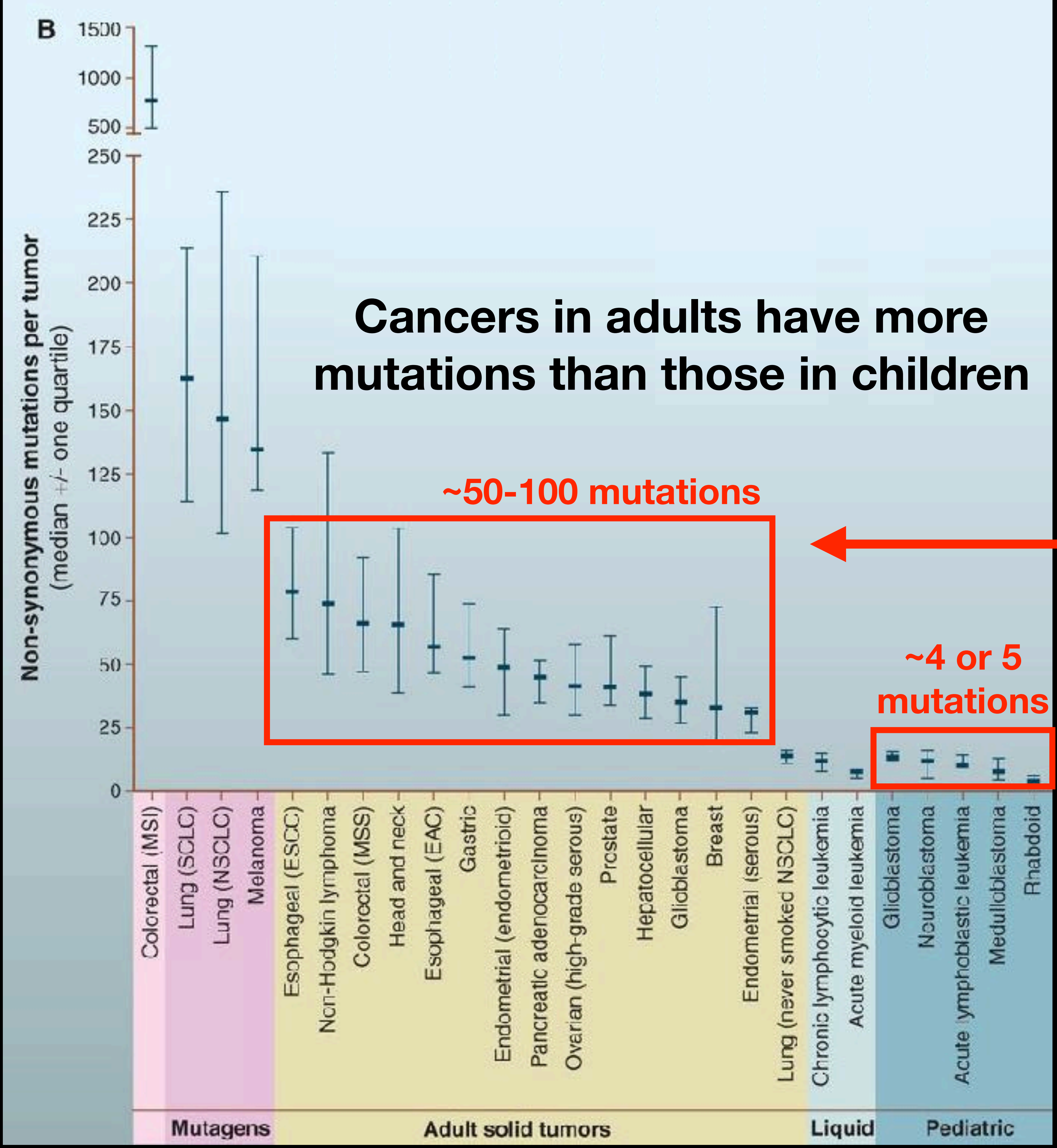
Cancers in adults have more mutations than those in children

~50-100 mutations

~4 or 5 mutations

Vogelstein et al. Science (2013)

Cancers in adults have more mutations than those in children

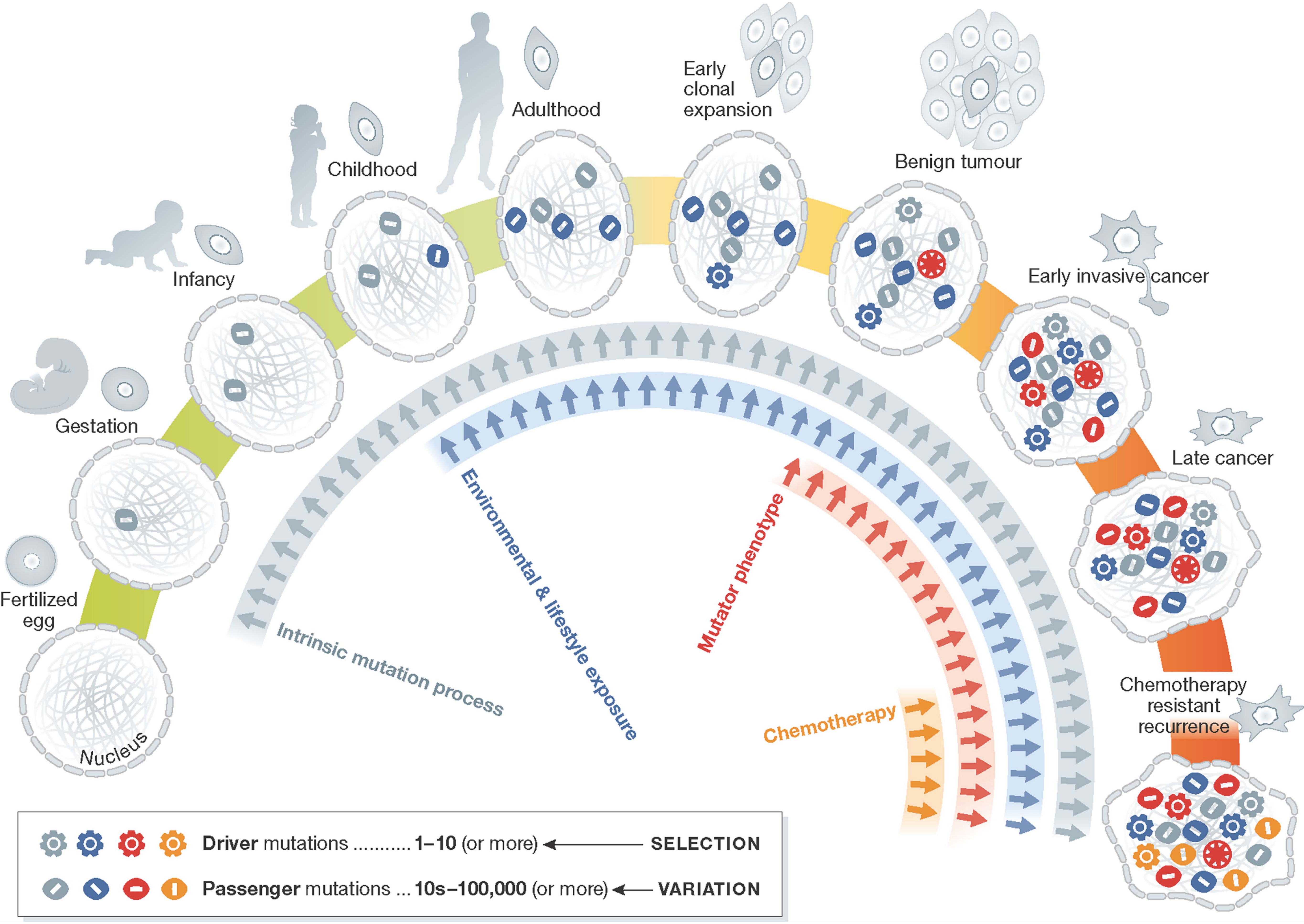


~50-100 mutations

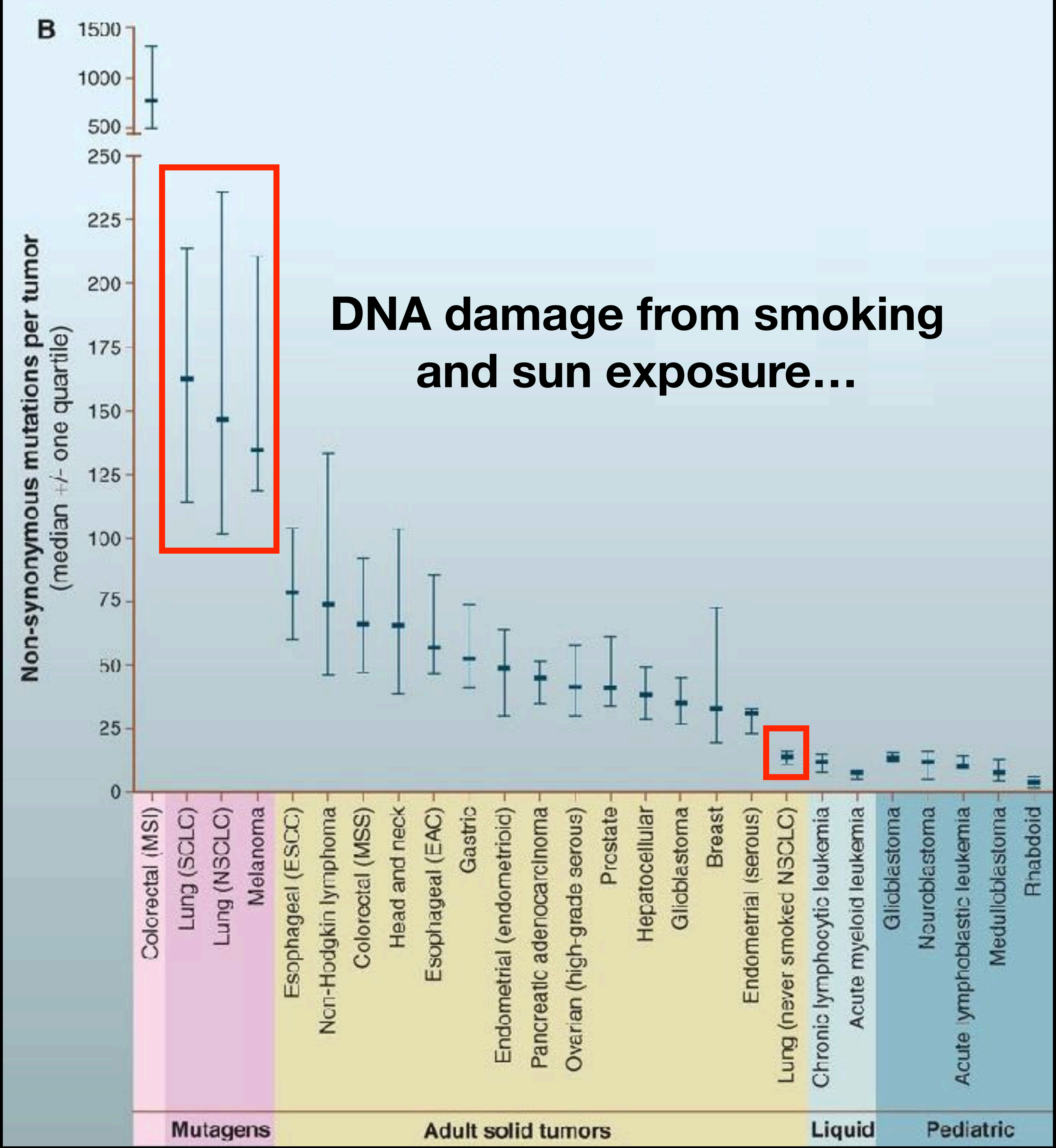
~4 or 5 mutations

Most of these mutations are likely "passenger" mutations

Vogelstein et al. Science (2013)



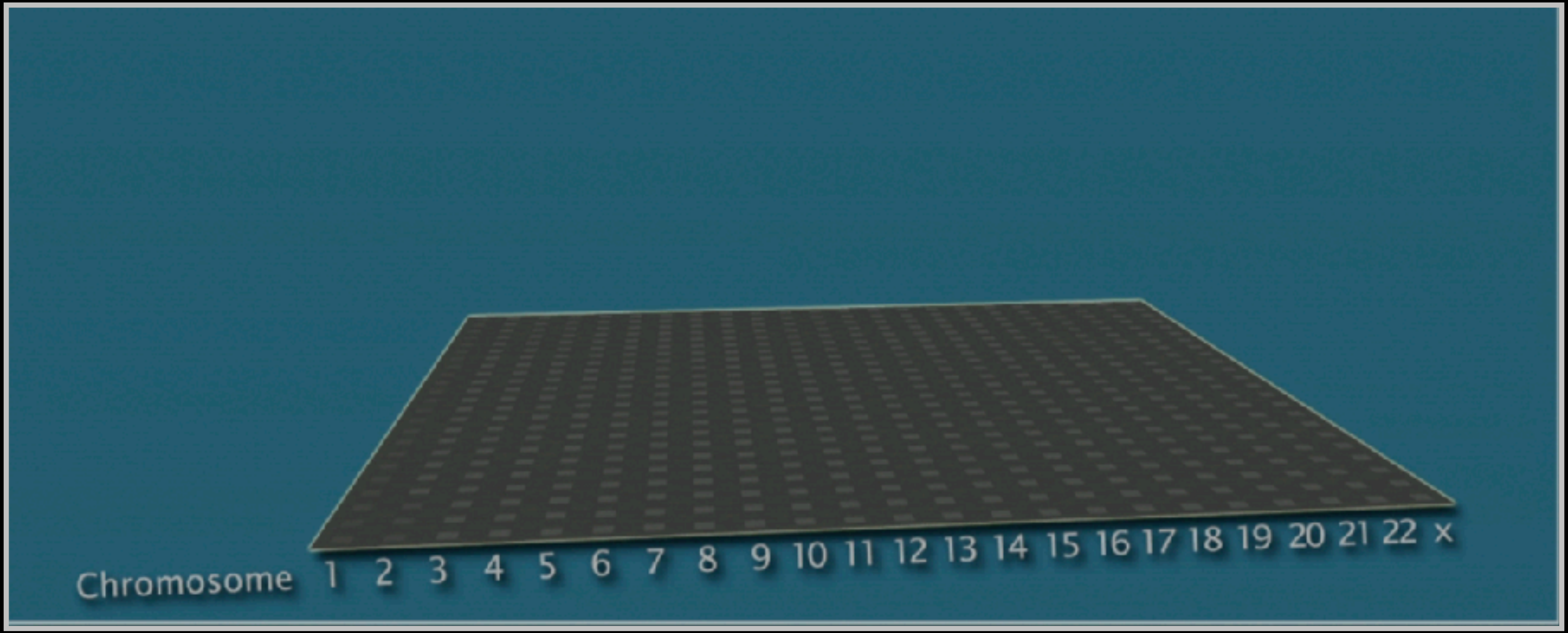
Mike Stratton. EMBO Molecular Medicine (2013)



DNA damage from smoking and sun exposure…

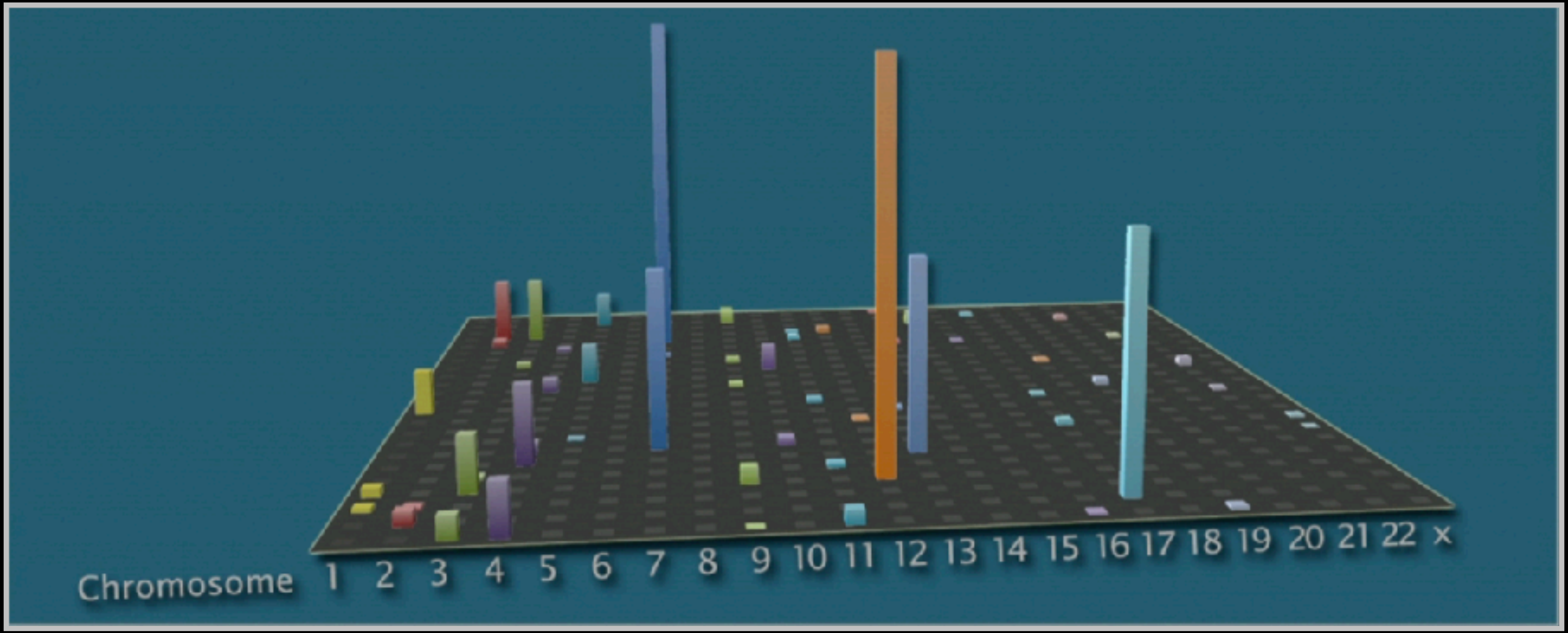
Vogelstein et al. Science (2013)

Genomic approaches can identify the genes most commonly mutated in cancer



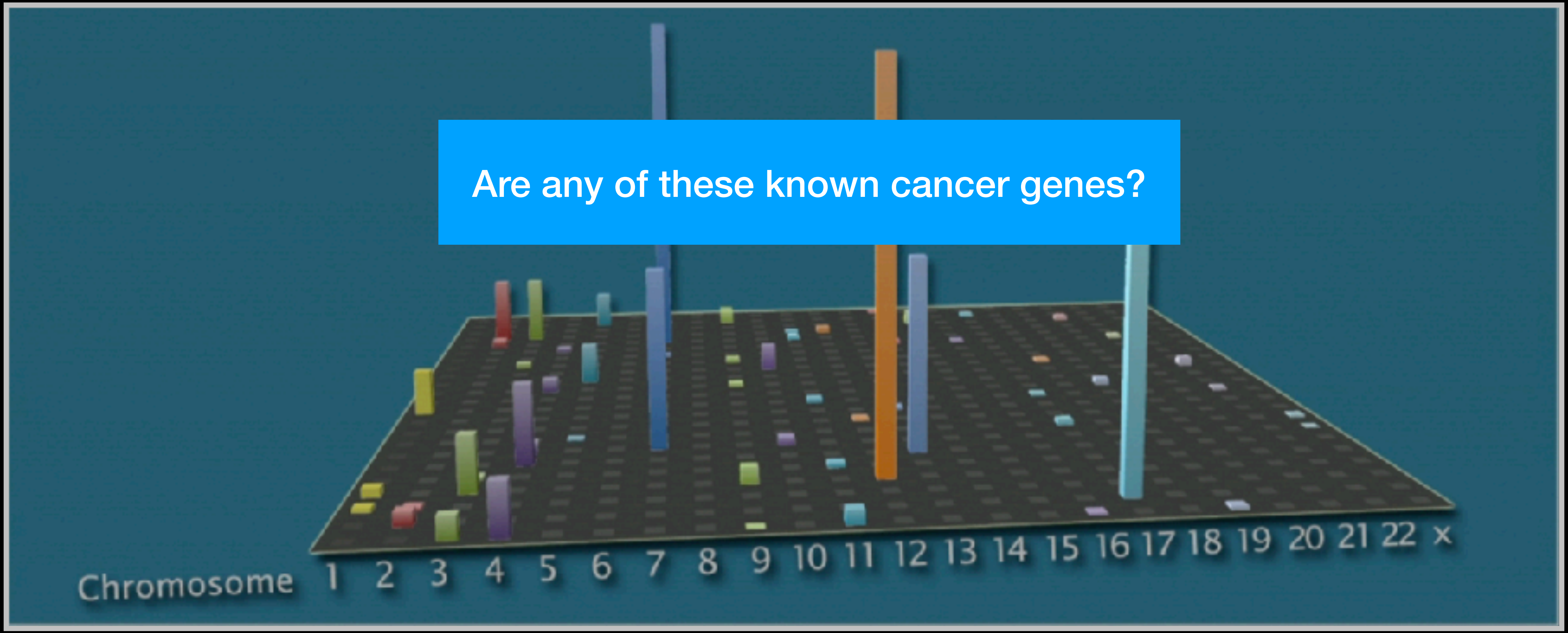
Arrange all genes in a matrix, ordered by chromosomes

Identifying genes most commonly mutated in cancer



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

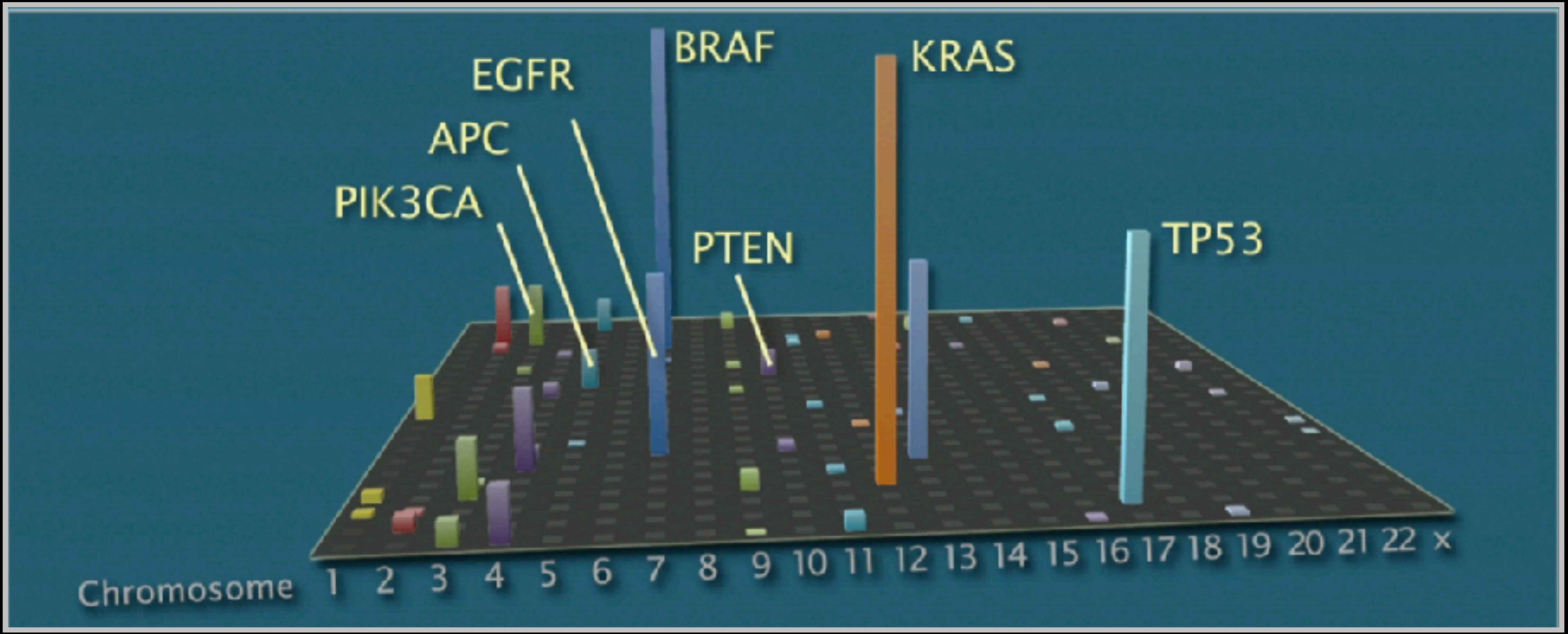
Identifying genes most commonly mutated in cancer



Are any of these known cancer genes?

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

Identifying genes most commonly mutated in cancer



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

Three Main Types of Cancer Genes:

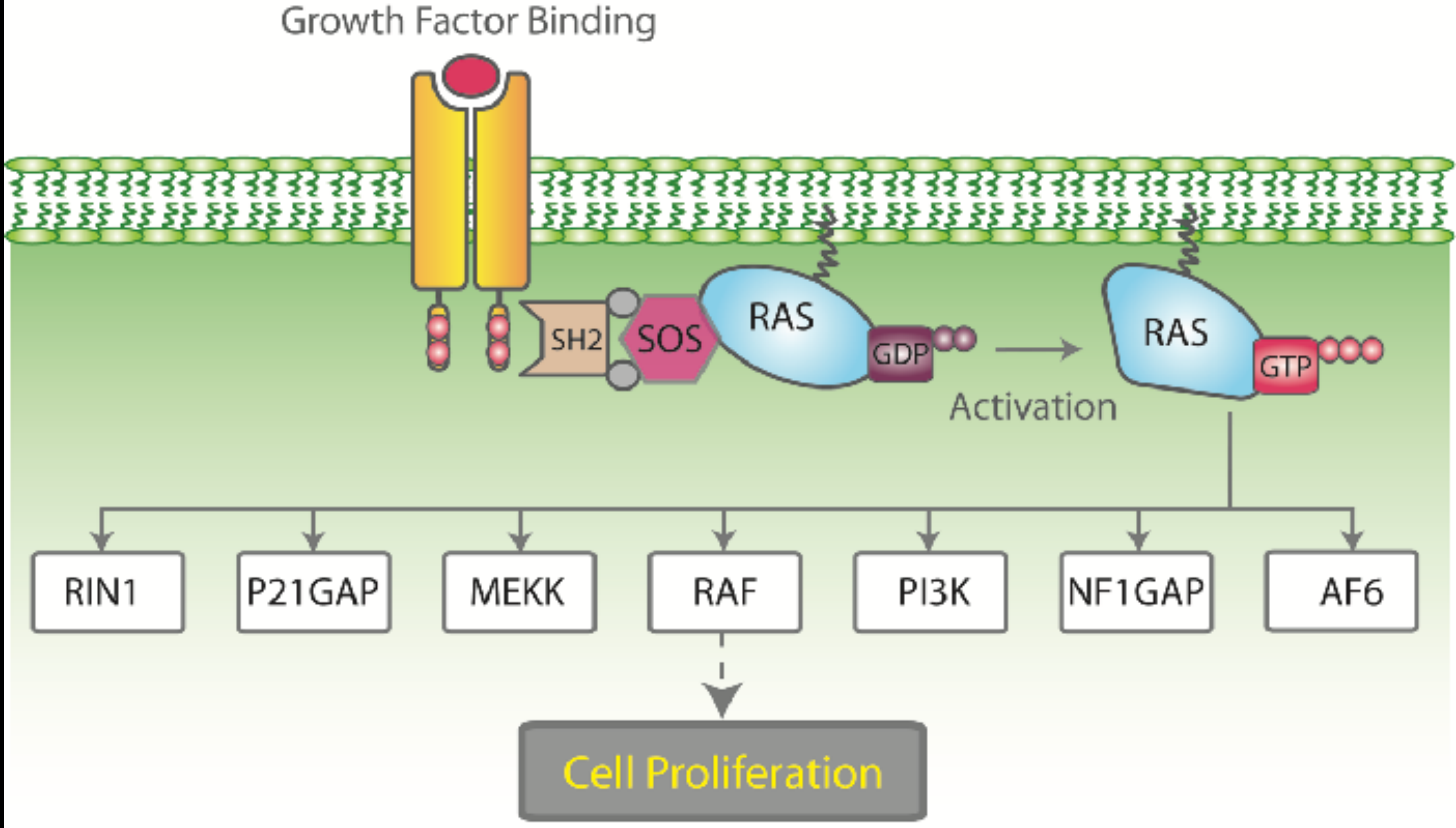
- Oncogenes, such as **Ras**, normally function to accelerate cell division and growth. They can be mutated to act like stuck gas pedals.

- 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.

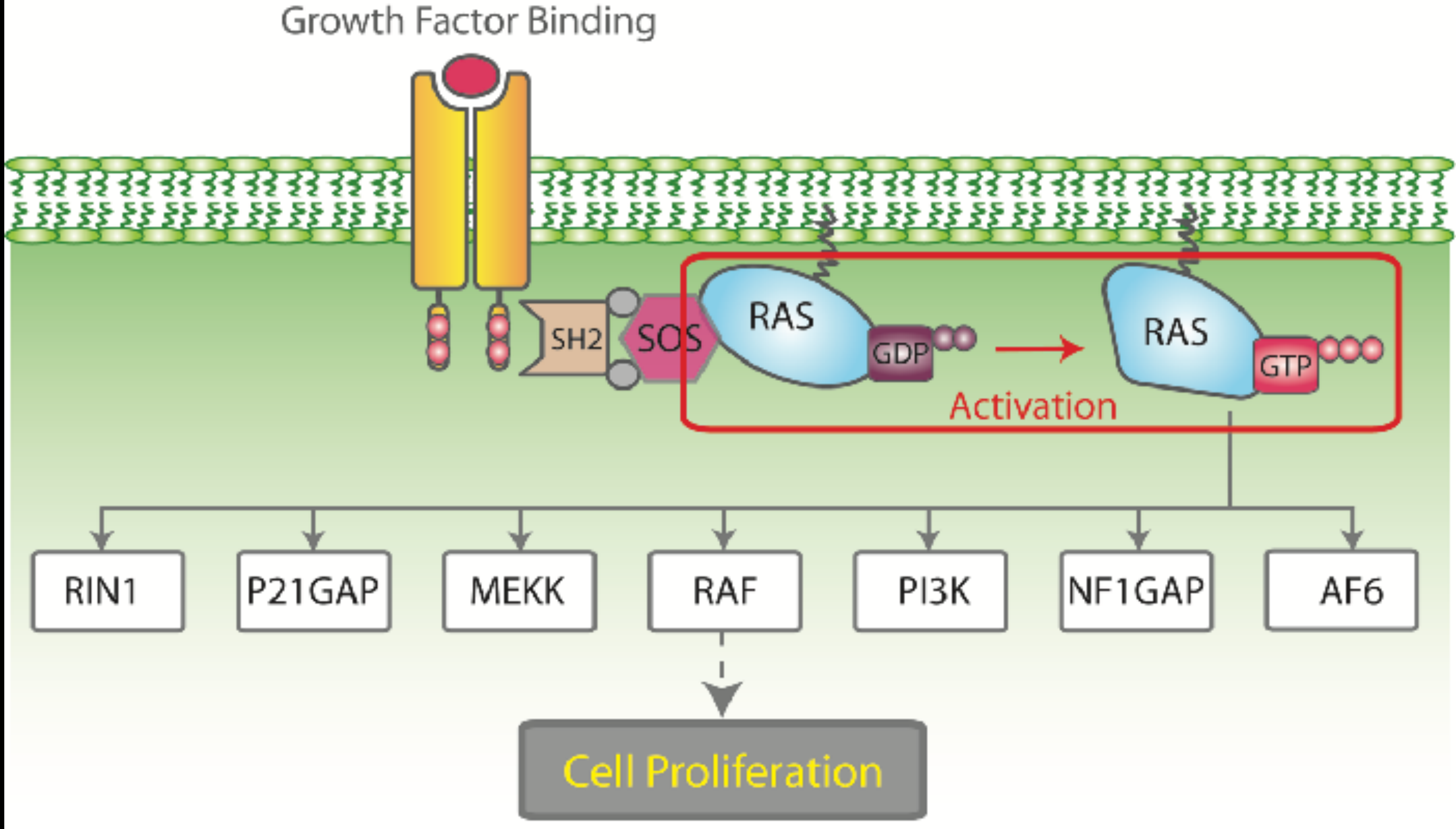
Cell growth and survival genes

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

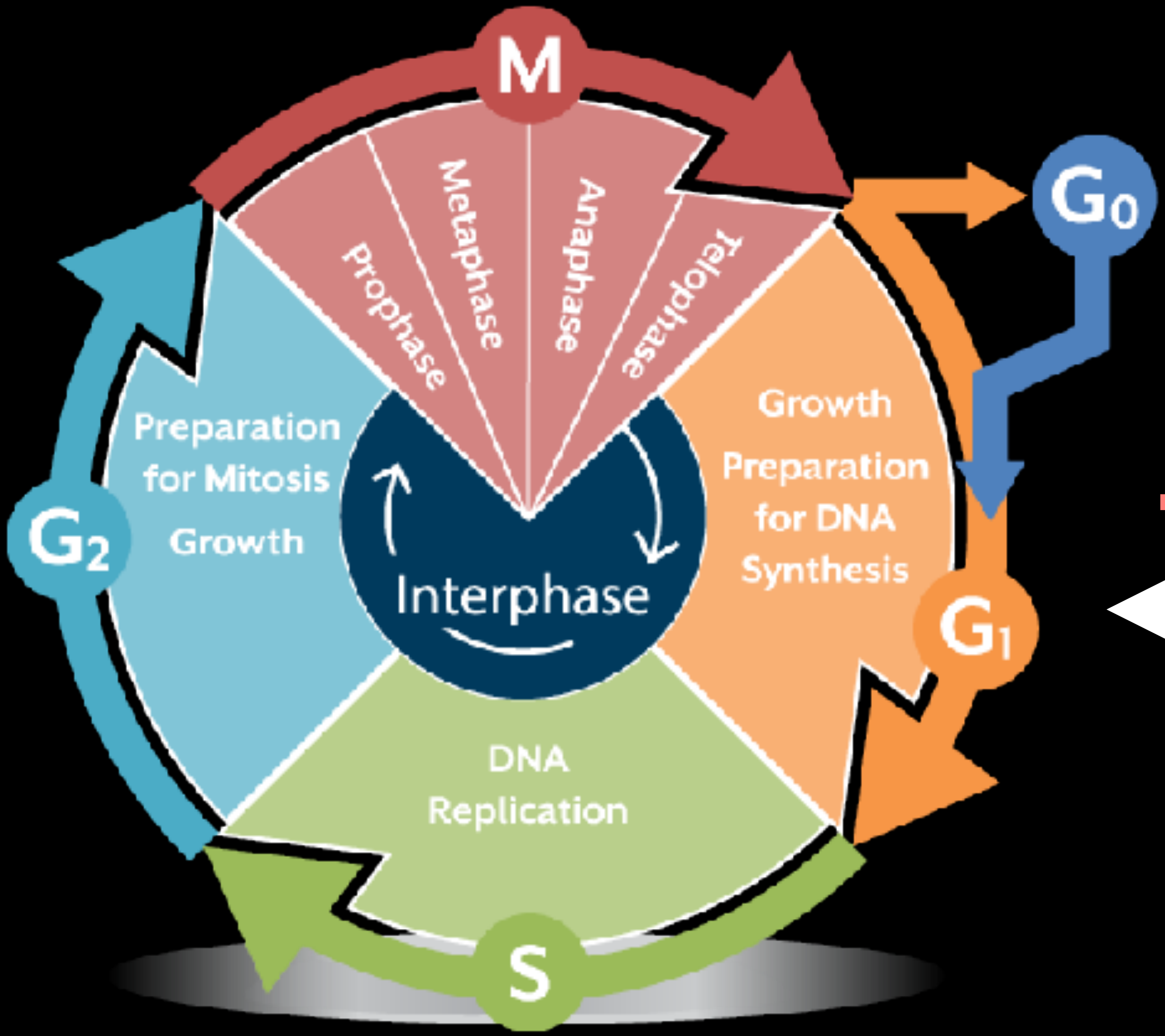
Some stimulate
the cell cycle

Cyclin D1

CDK4

+

Oncogenes



Some inhibit
the cell cycle

P53

RB

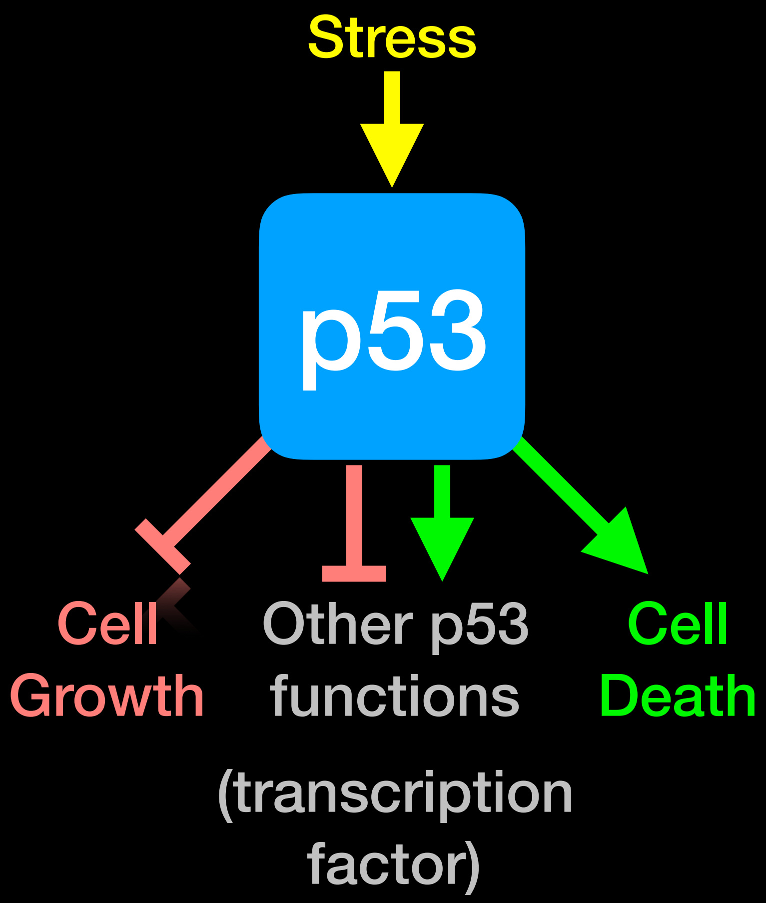
-

Suppressor genes

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.


Stress

p53

Cell Growth

Other p53 functions

Cell Death

(transcription factor)

Do it Yourself!

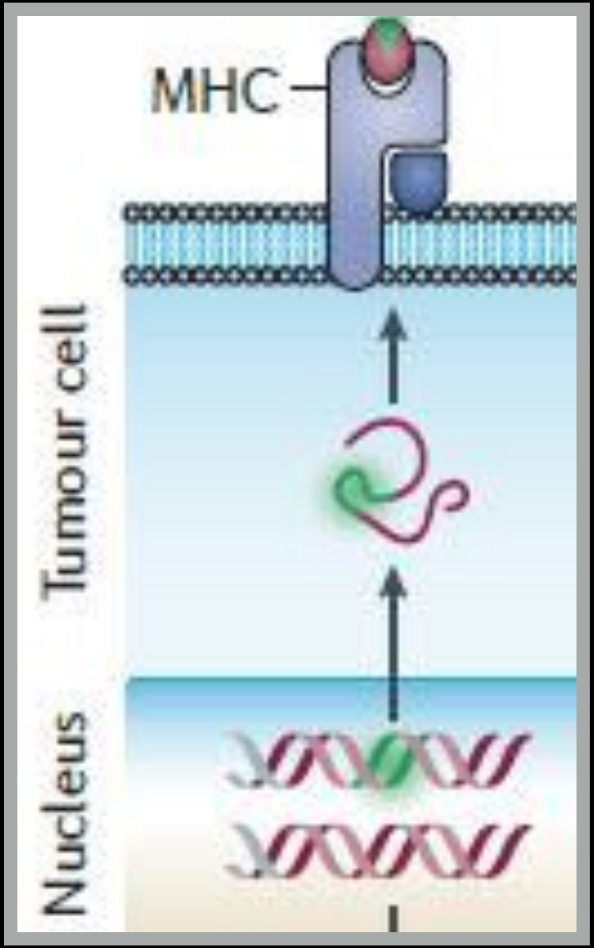
Hands-on time!

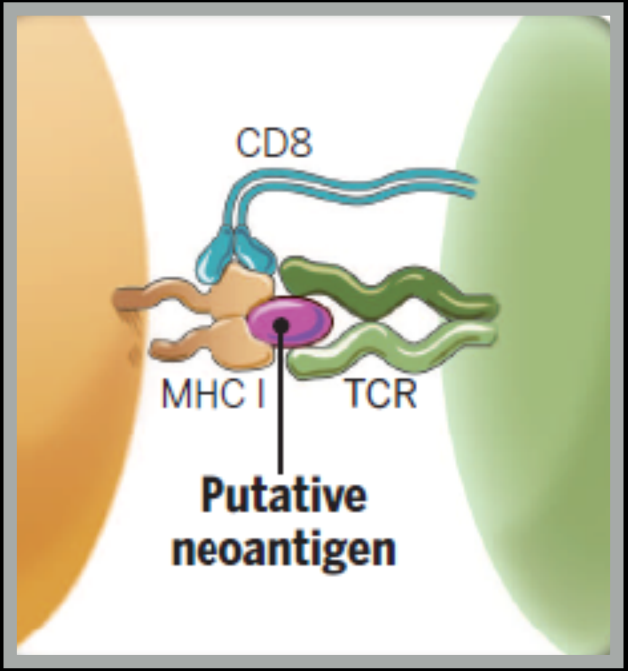
https://bioboot.github.io/bggn213_S18/lectures/#18

Part 1 Only Please

Cancer Immunotherapy

- 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)

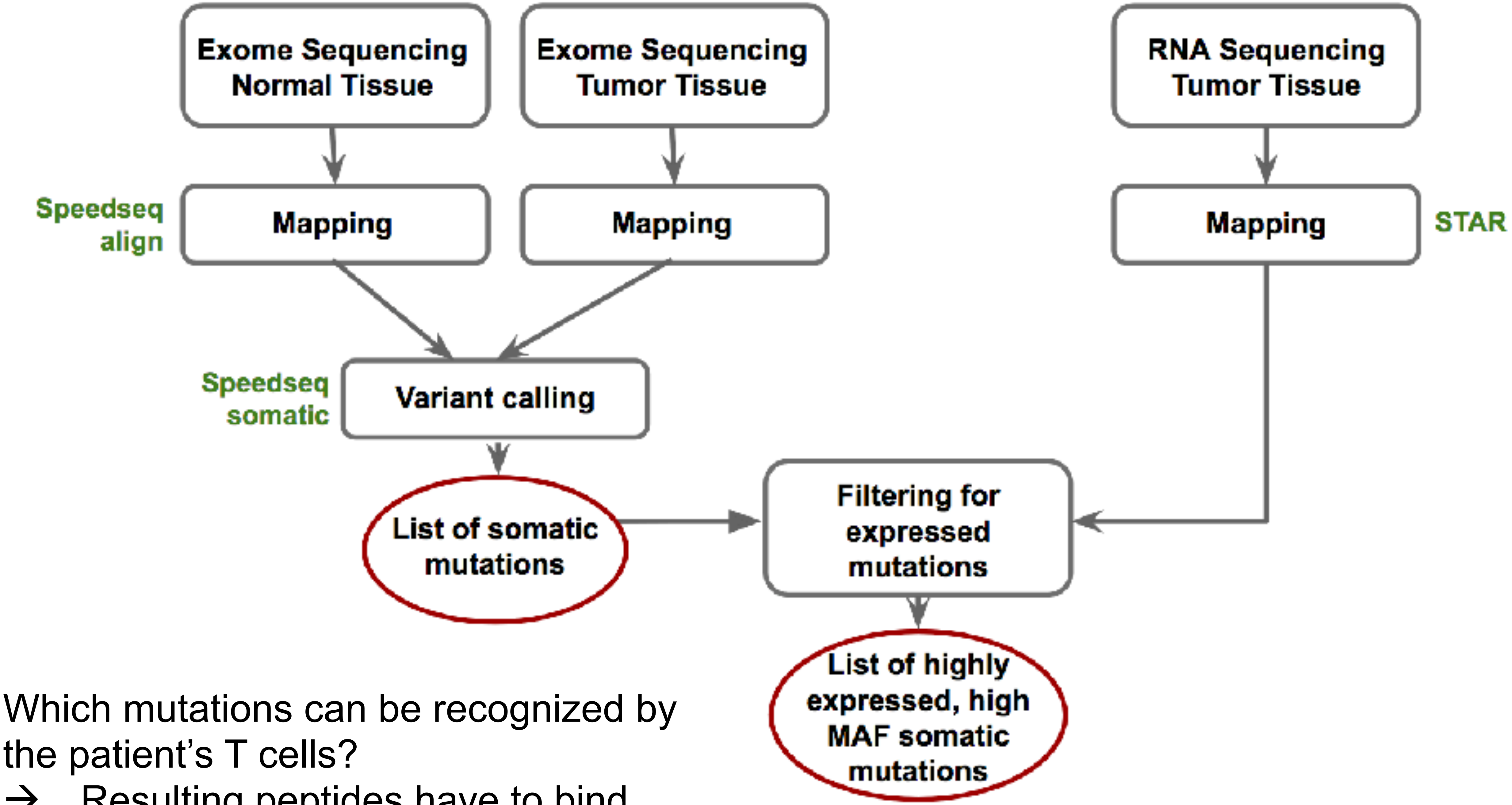
- **Neoepitopes** are highly tumor-specific!

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

- **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 the 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.

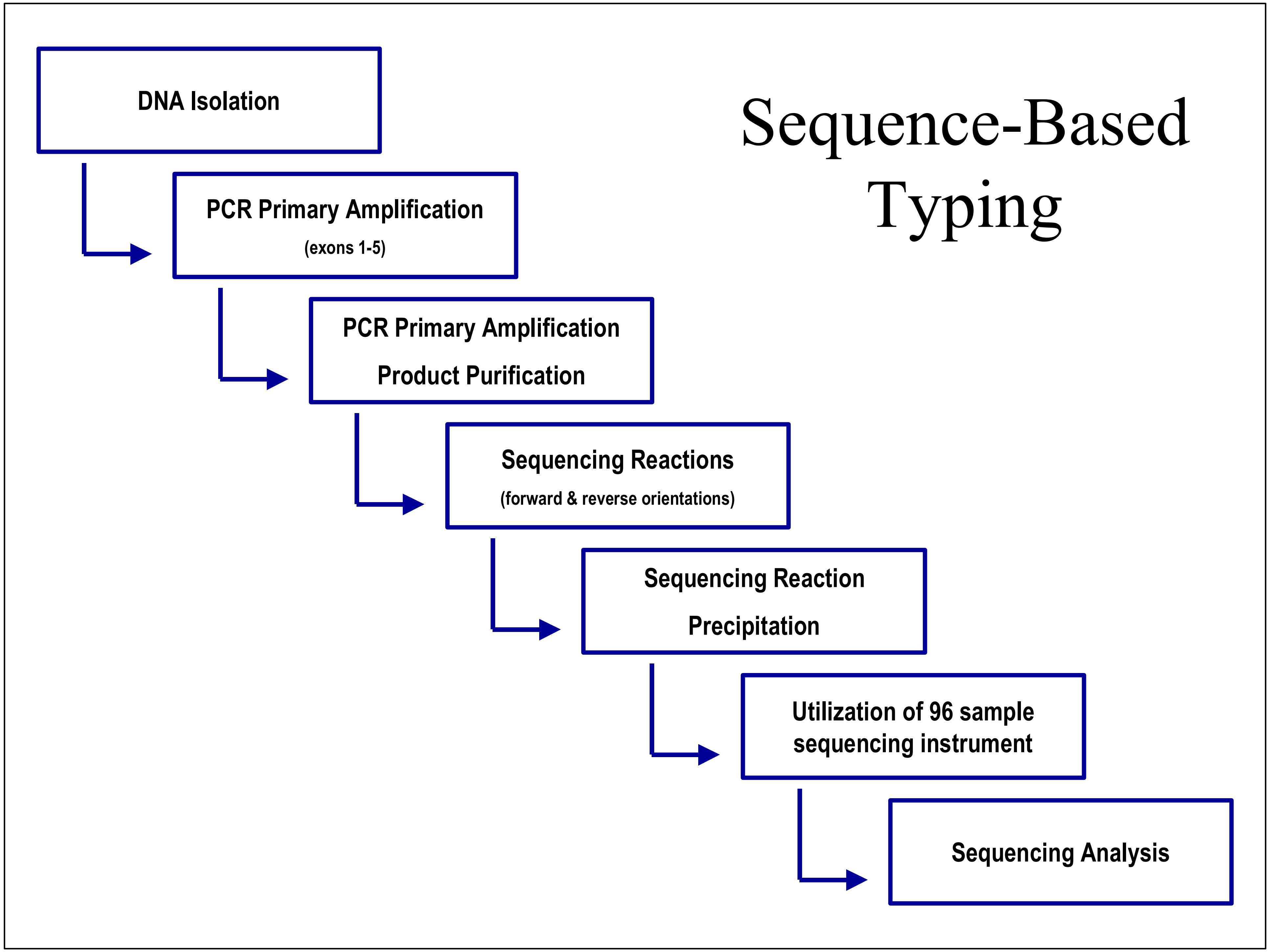
Q. How can such a vaccine be designed?

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



Sequence-Based Typing

DNA Isolation

PCR Primary Amplification
(exons 1-5)

PCR Primary Amplification
Product Purification

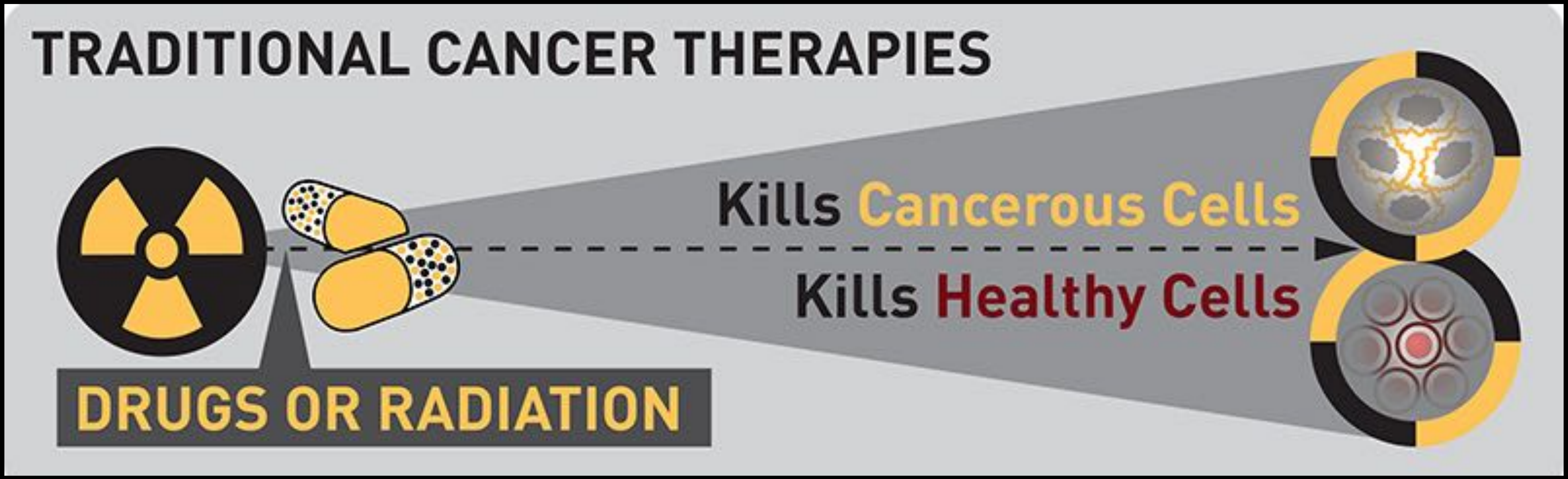
Sequencing Reactions
(forward & reverse orientations)

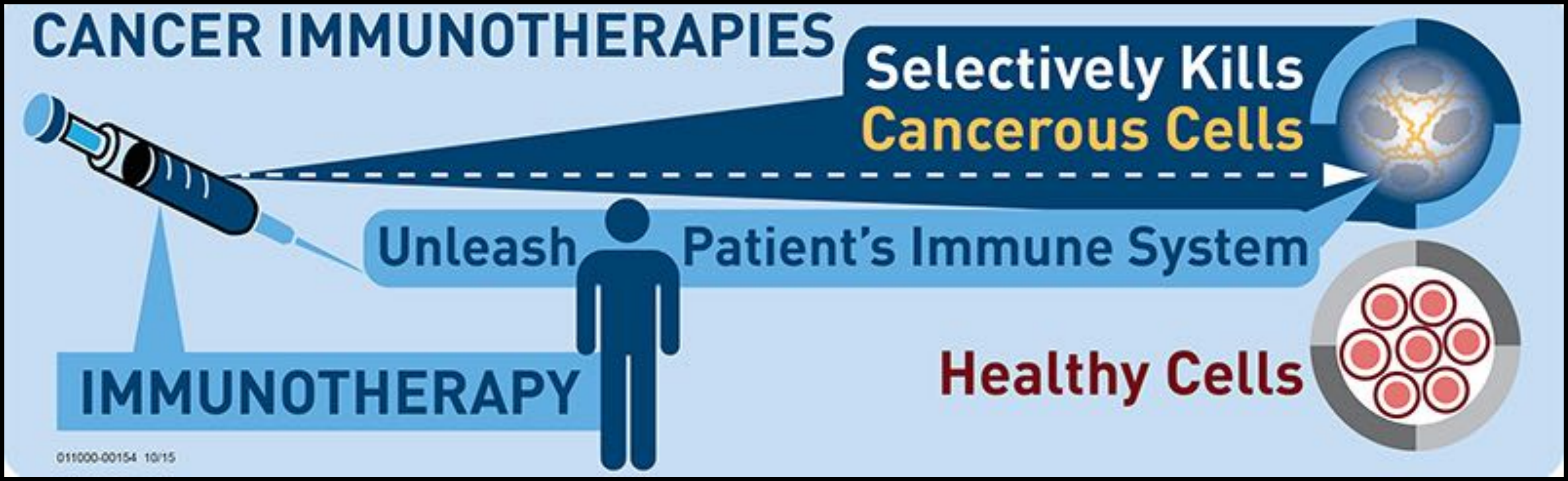
Sequencing Reaction
Precipitation

Utilization of 96 sample
sequencing instrument

Sequencing Analysis

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


TRADITIONAL CANCER THERAPIES
Kills Cancerous Cells
Kills Healthy Cells
DRUGS OR RADIATION


CANCER IMMUNOTHERAPIES
Selectively Kills
Cancerous Cells
Unleash Patient's Immune System
Healthy Cells
IMMUNOTHERAPY
011000-00154 10/15

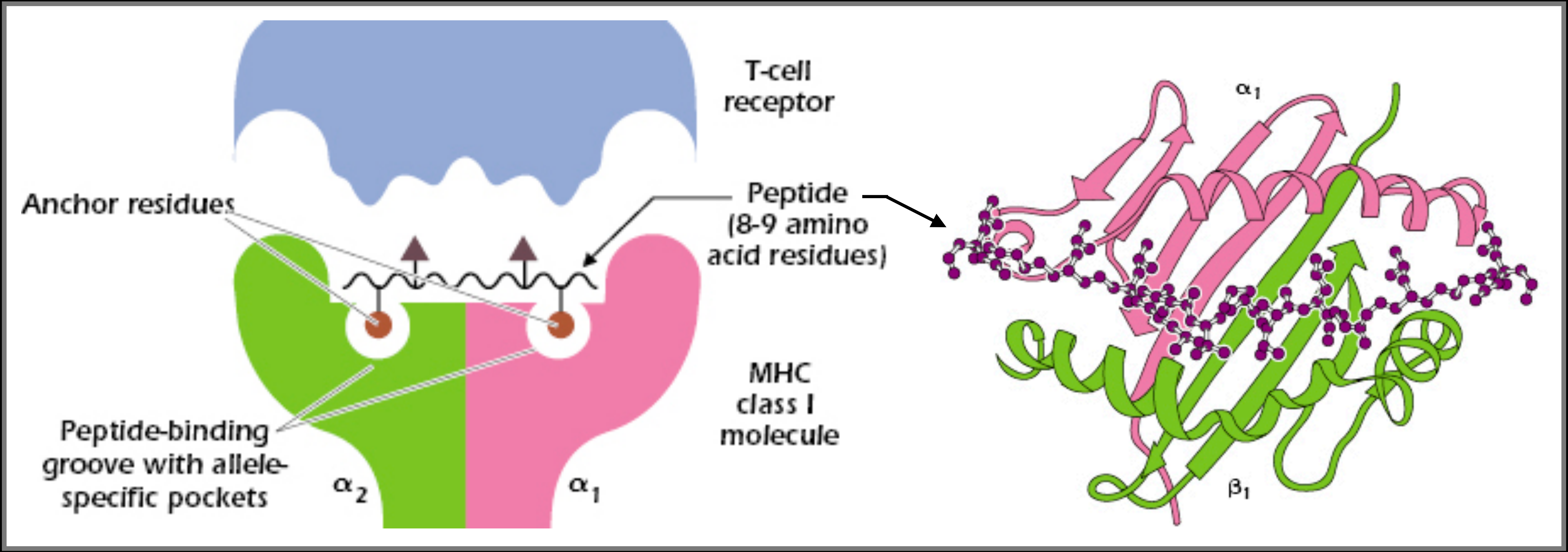
Do it Yourself!

Hands-on time!

https://bioboot.github.io/bggn213_S18/lectures/#18

Part 2: Designing a personalized cancer vaccine

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!

Reference: https://oncohemakey.com/how-t-cells-recognize-antigen-the-role-of-the-major-histocompatibility-complex/

Bonus Slides

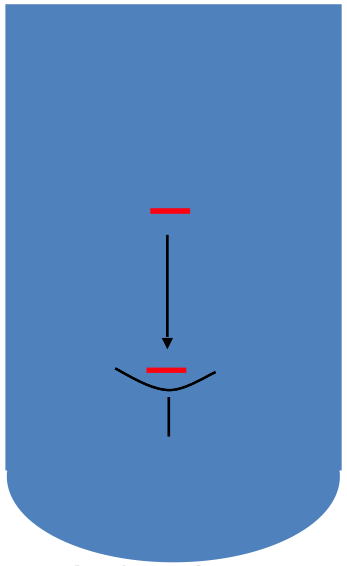
(For Reference)

Slide from: **Bjoern Peters** (LIAI)

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}) ~ Binding free Energy

low IC_{50} → high affinity

Impossible to measure all peptides

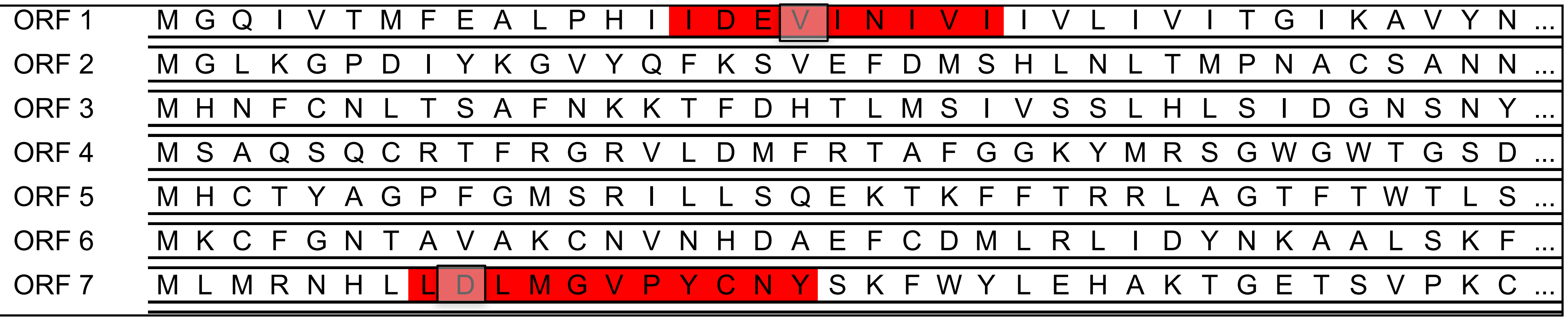
→ Predict binding peptides using machine learning

Find function F_i in F_1, F_2, F_3, ...

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 ...

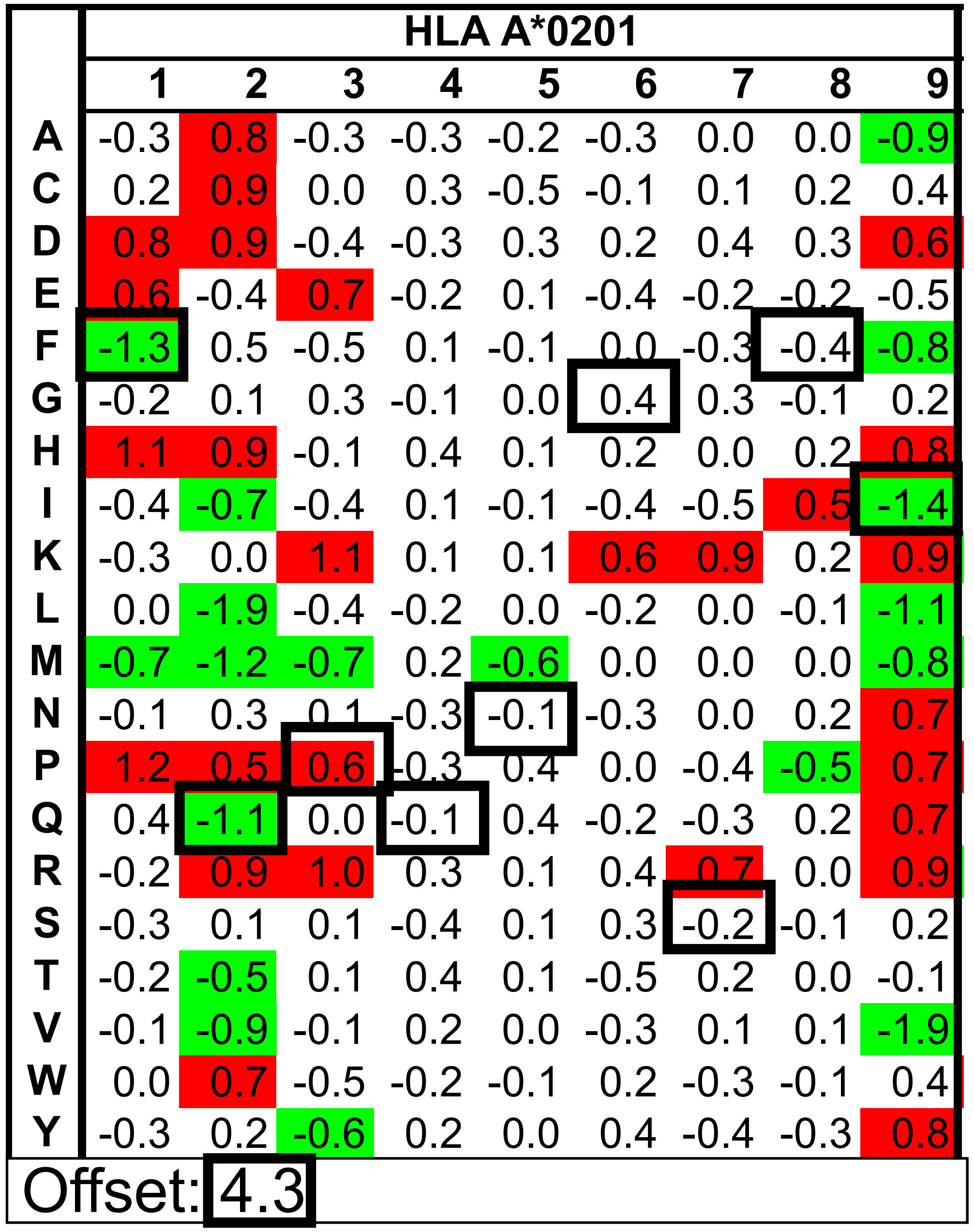
Slide from: **Bjoern Peters** (LIAI)

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
...	



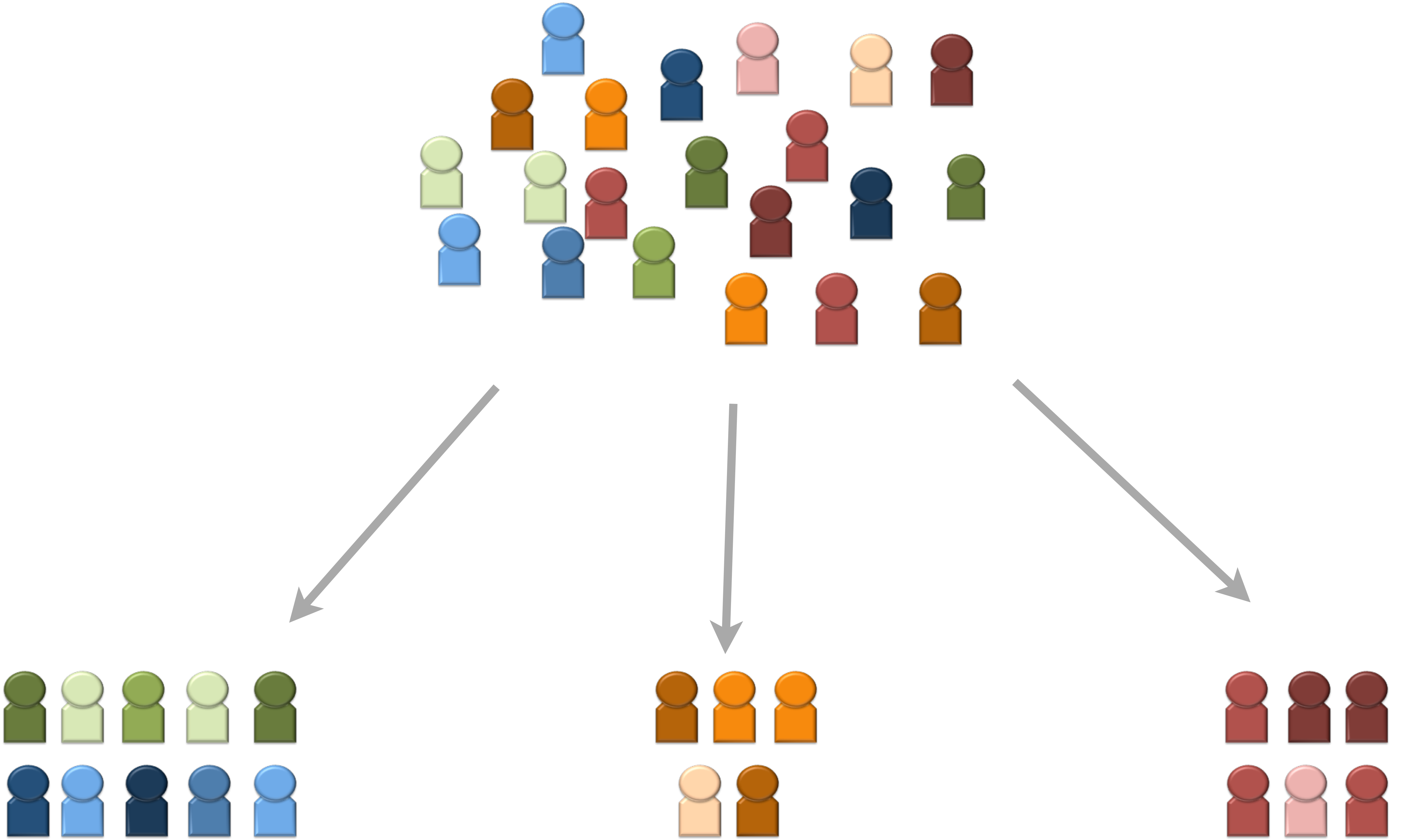
HLA A*0201

	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

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

Stratify tumors based on molecular patterns

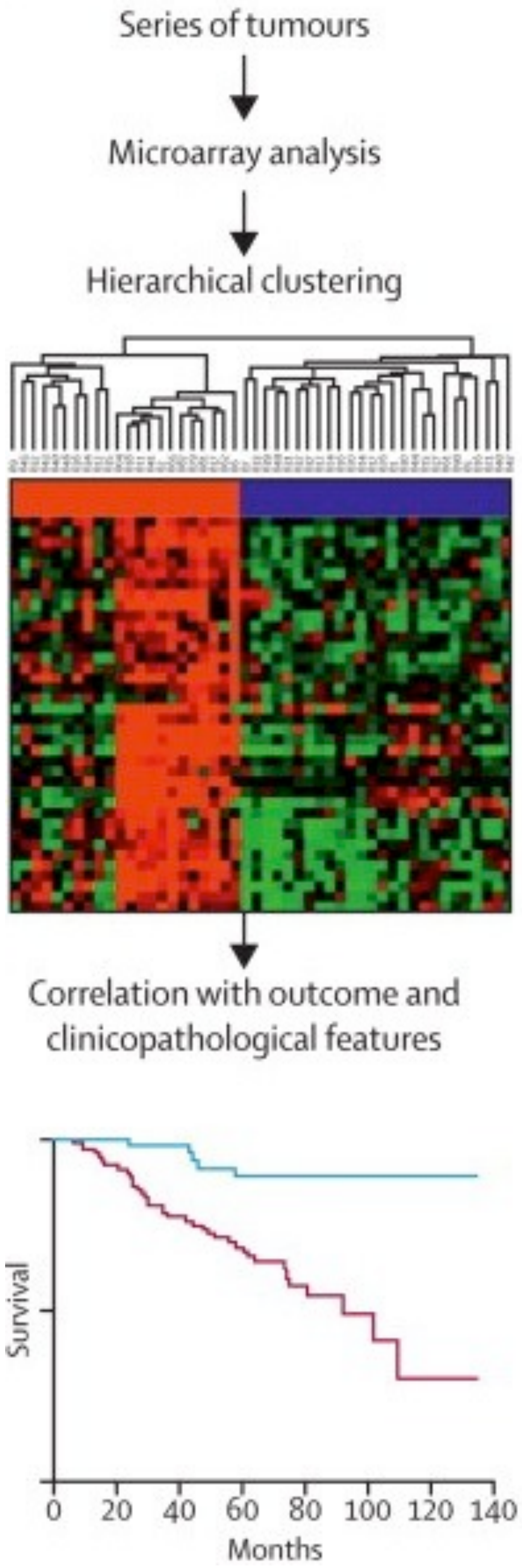


Good prognosis
Favorable response

Bad prognosis
Unfavorable response

Increased toxicity

Stratify tumors based on molecular patterns



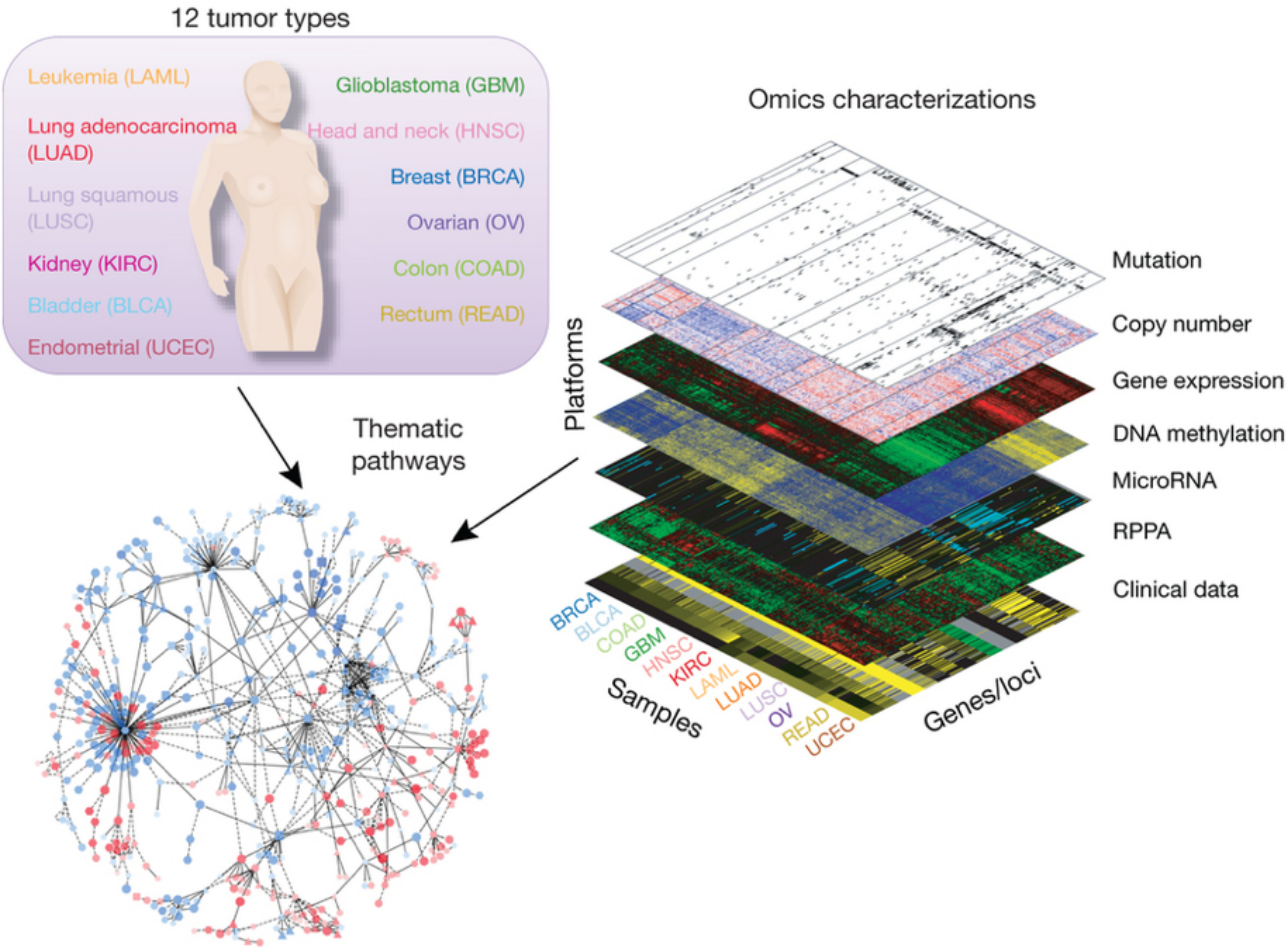
Series of tumours

Microarray analysis

Hierarchical clustering

Correlation with outcome and clinicopathological features

TCGA Pan-Cancer project

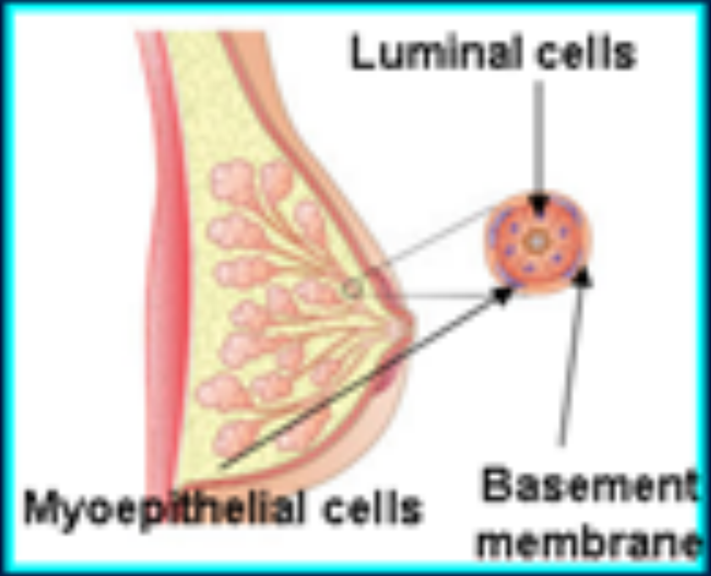


For example, breast cancer may be classified into various types based upon which proteins are expressed on the surface of the tumor cells. Breast tumors that express human epidermal growth factor 2 (HER2), estrogen receptor (ER), and progesterone receptor (PR), or are triple negative (do not express 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.

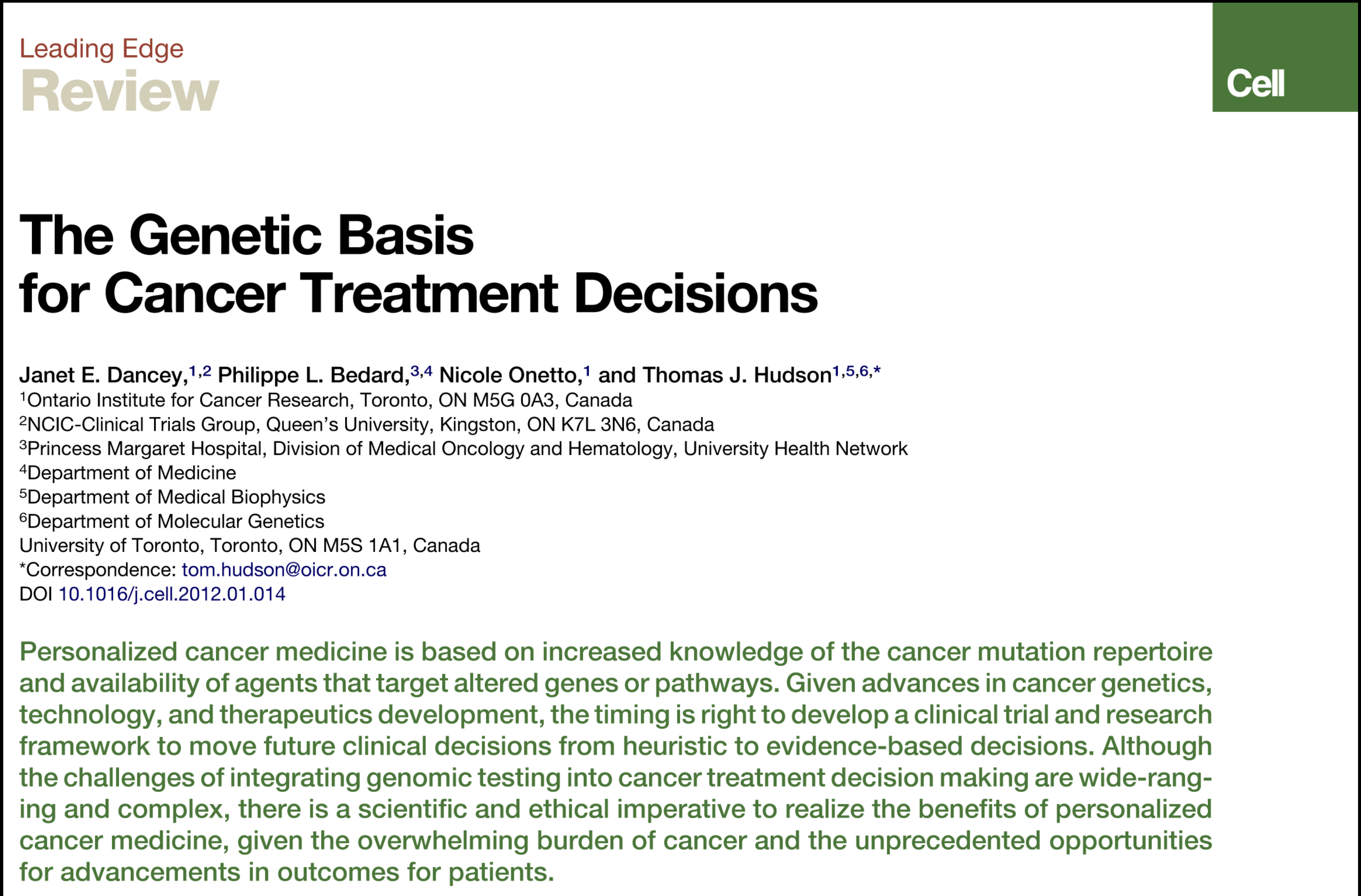
As more has been learned about the molecular signature of various cancer subtypes, therapies that are specifically targeted to those signatures have been developed. Conventional chemotherapy acts on all rapidly dividing cells and does not distinguish between cancer cells and normal cells.

Classification of Breast Cancer

- Infiltrating breast cancer
  - ER IHC
    - HER2 IHC
    - ER-
      - HER2 (+)
        - ER-, HER2+ classic
        - (trastuzumab, lapatinib)
      - HER2 (-)
        - CK5/6 EGFR IHC
          - ER-, HER2- CK5/6+ and/or EGFR+
            - Basal-like phenotype
          - ER-, HER2- CK5/6- and EGFR-
            - Non-basal, triple negative phenotype
          - (aggressive chemotherapy)
    - ER+
      - HER2 (+)
        - ER+ luminal B HER2 variant
      - HER2 (-)
        - ER+ luminal A
      - (tamoxifen and aromatase inhibitors)



Readings to find out more…



Leading Edge
Review

Cell

The Genetic Basis for Cancer Treatment Decisions

Janet E. Dancey,[1,2] Philippe L. Bedard,[3,4] Nicole Onetto,[1] and Thomas J. Hudson[1,5,6,*]
[1]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada
[2]NCIC-Clinical Trials Group, Queen's University, Kingston, ON K7L 3N6, Canada
[3]Princess Margaret Hospital, Division of Medical Oncology and Hematology, University Health Network
[4]Department of Medicine
[5]Department of Medical Biophysics
[6]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*

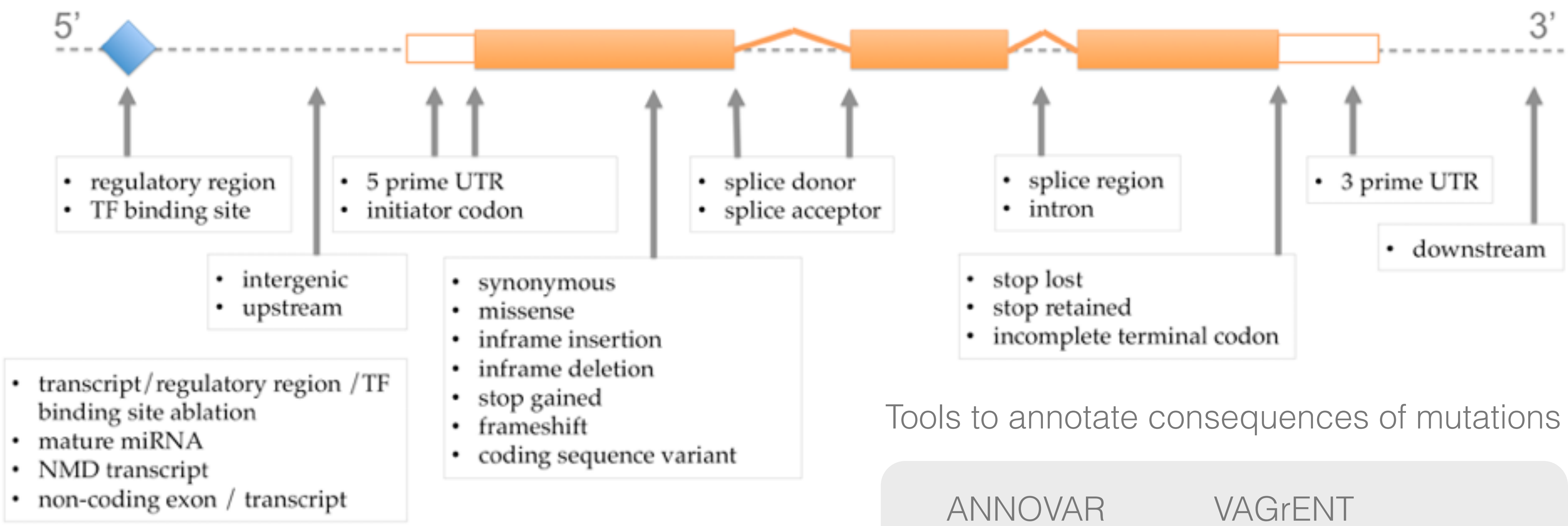
- Calling cancer's bluff with neoantigen vaccines
- Can genomics help detect early cancer and monitor treatment effectiveness?
- The increasing cost of cancer therapies

https://bioboot.github.io/bimm194_W18/readings/

1. Predict consequences of mutations

```
ACTGCCTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCCATGCTACTGCATCTCGGGTTAACTC
GACGTTTTTCATGCATGTGTGCACCCCAATATATATGCAACTTTTGTGCACCTCTGTCACGCGCGAGTTGGCA
CTGTCGCCCCTGTGTGCATGTGCACTGTCTCTCGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCGCTT
AACCCGTACTCCCATGCTACTGCATCTCGGGTTAACTCGACGTTTTGCATGCATGTGTGCACCCCAATATATA
TGCAACTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGTCTCTCGA
```

Map mutations into genome annotations to predict its possible effect



- regulatory region
- TF binding site

- intergenic
- upstream

- transcript/regulatory region /TF binding site ablation
- mature miRNA
- NMD transcript
- non-coding exon / transcript

- 5 prime UTR
- initiator codon

- synonymous
- missense
- inframe insertion
- inframe deletion
- stop gained
- frameshift
- coding sequence variant

- splice donor
- splice acceptor

- splice region
- intron

- stop lost
- stop retained
- incomplete terminal codon

- 3 prime UTR

- downstream

Tools to annotate consequences of mutations

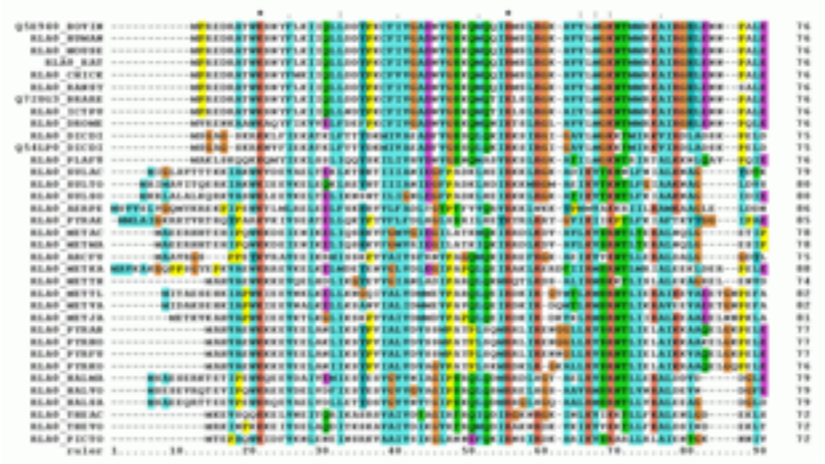
ANNOVAR
VAGrENT
Ensembl VEP
ASOoVIR
snpEff
annTools

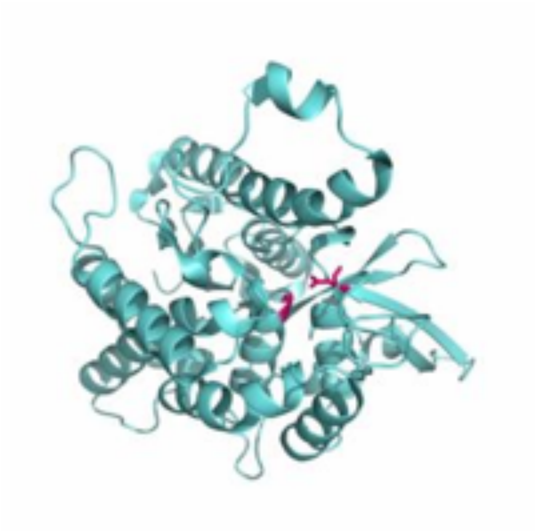
2. Assess the functional impact of nsSNVs

nsSNVs = non-synonymos Single Nucleotide Variant (missense)

ATC GAA GCA CGT
Met Glu Ala Gly

ATC GAC GCA CGT
Met Asp Ala Gly

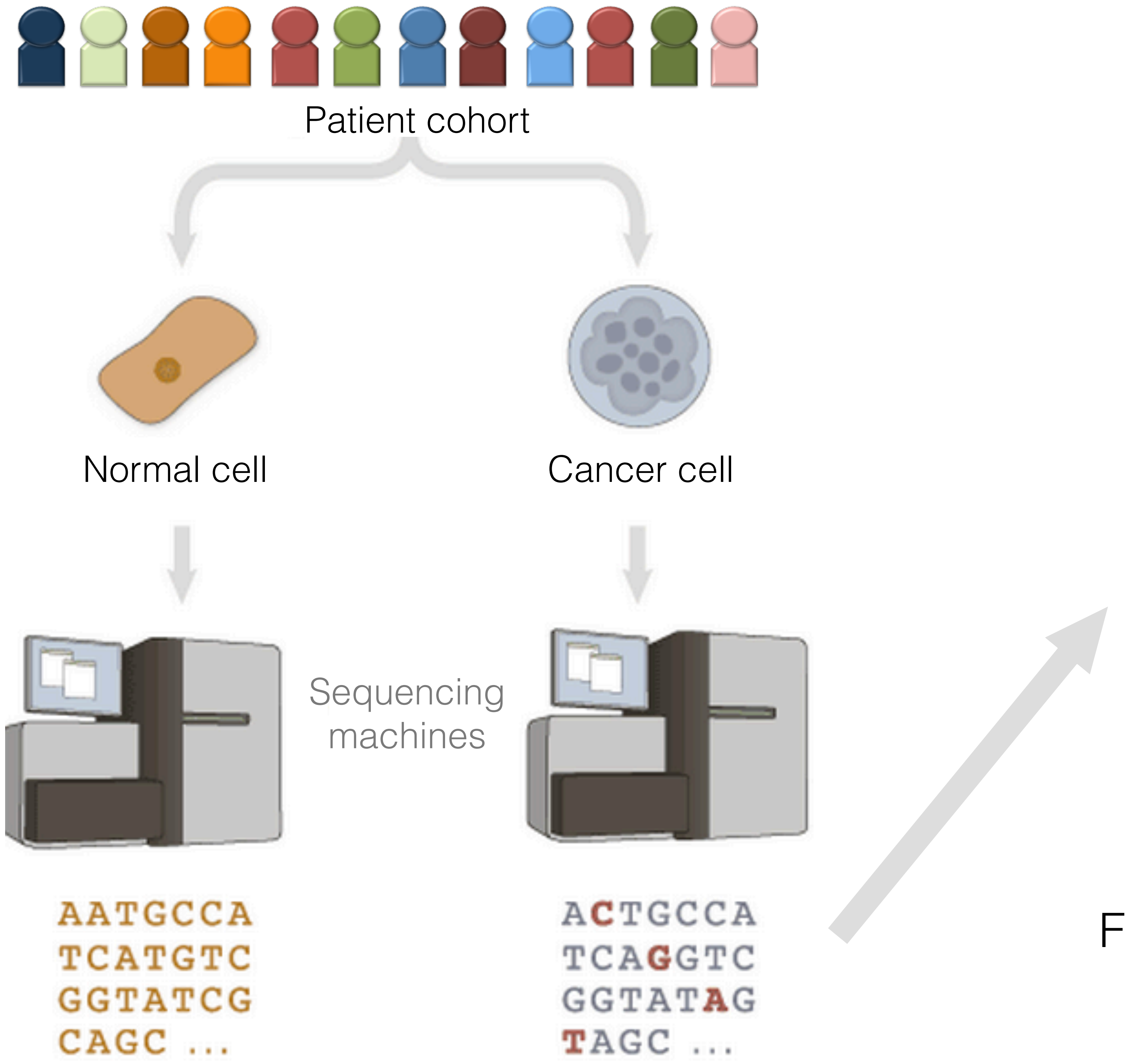




Computational methods to assess the functional impact of nsSNVs

- MutationTaster
- CanPredict
- SIFT
- LogRe
- PolyPhen2
- Condel
- MutationAssessor
- MutPred
- CHASM
- PMut
- SNPs&GO
- SNPeffect
- transFIC

3. Identify cancer drivers from somatic mutations



Patient cohort

Normal cell

Cancer cell

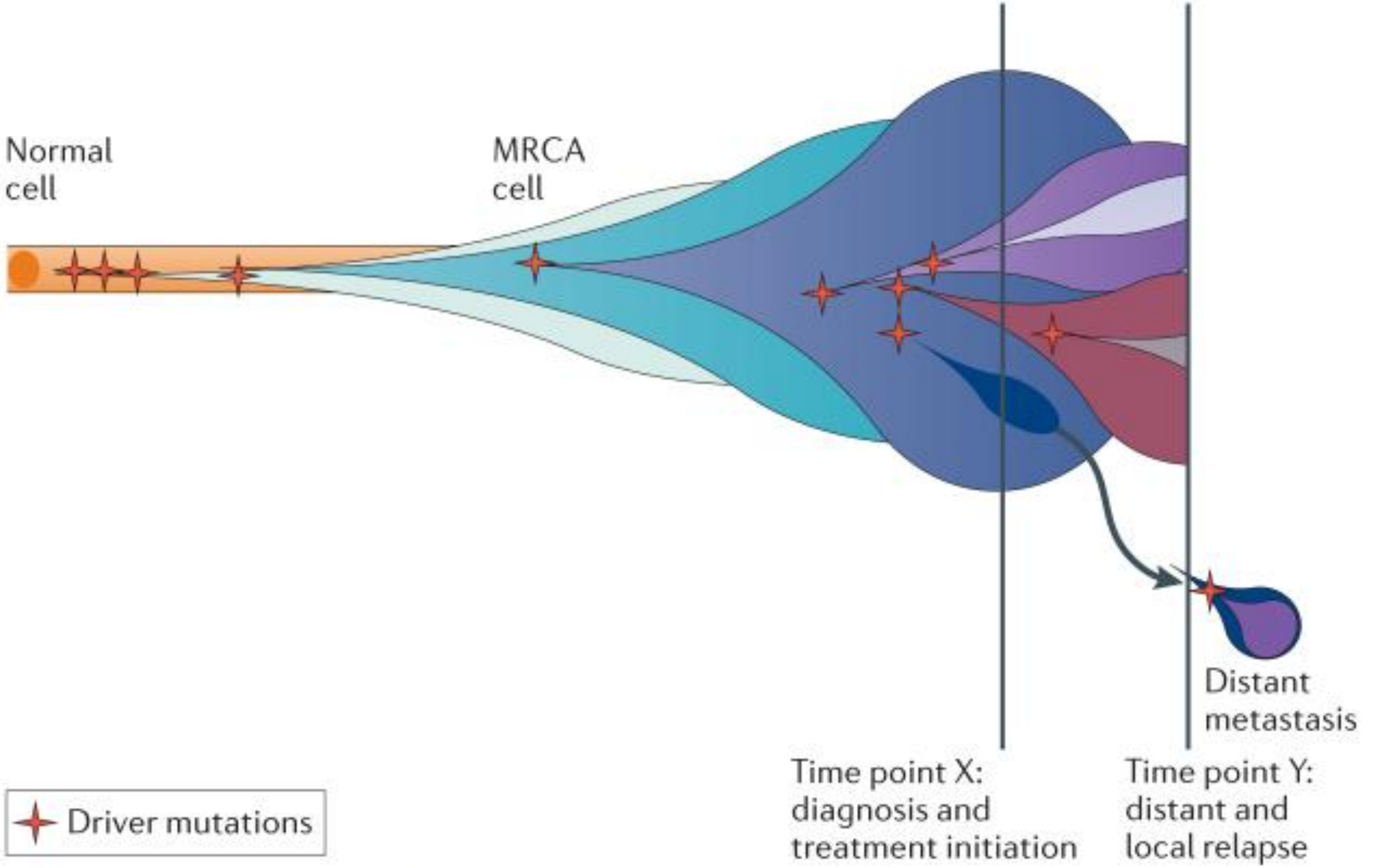
Sequencing machines

Somatic mutations

Which mutations are cancer drivers?

Find signals of selection across tumors

Cancer is an evolutionary process



Yates and Campbell et al, Nat Rev Genet 2012

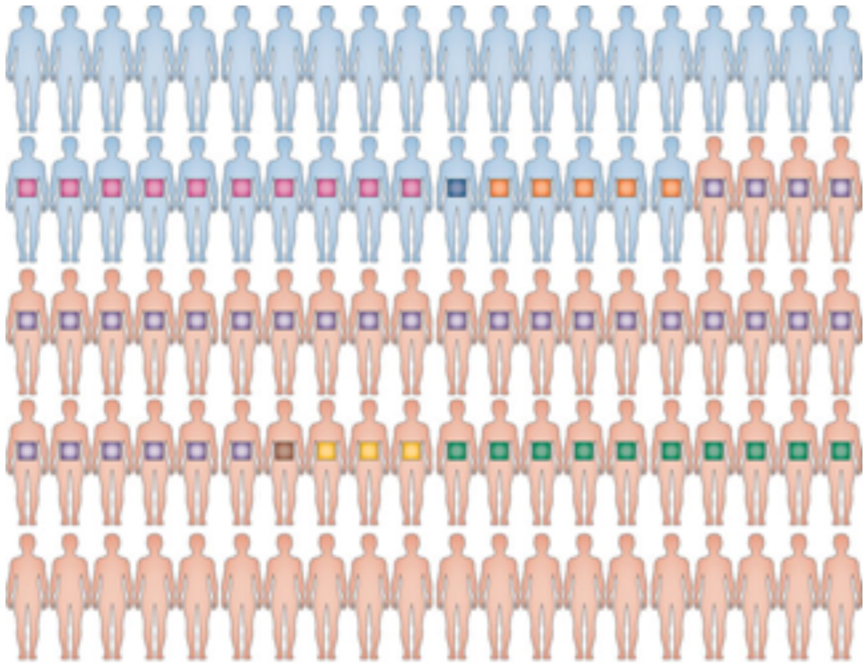
How to differentiate drivers from passengers?

ACTG**C**CTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCCATGCTACTGC
ATCTCGGGTTAACTCGACGTTTT**T**CATGCATGTGTGCACCCCAATATATATGCA**A**CTT
TTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGT
CTC**T**CGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCC
ATGCTACTGCATCTCGGGTTAACTCGACGTTTT**G**CATGCATGTGTGCACCCCAATATA
TATGCA**A**CTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCA
TGTGCACTGTCTC**T**CGAGTTTT**G**CATGCATGTGTGCACTGTGCACCTCTGTTACGTCT

How to differentiate drivers from passengers?

ACTG**C**CTACGTCTCACCGTCGACTTCAAATCGCTTAACCCGTACTCCCATGCTACTGC
ATCTCGGGTTAACTCGACGTTTT**T**CATGCATGTGTGCACCCCAATATATATGCA**A**CTT
TTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGT
CTC**T**CGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCC
ATGCTACTGCATCTCGGGTTAACTCGACGTTTT**G**CATGCATGTGTGCACCCCAATATA
TATGCA**A**CTTTTGTGCACCTCTGTCACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCA
TGTGCACTGTCTC**T**CGAGTTTT**G**CATGCATGTGTGCACTGTGCACCTCTGTTACGTCT

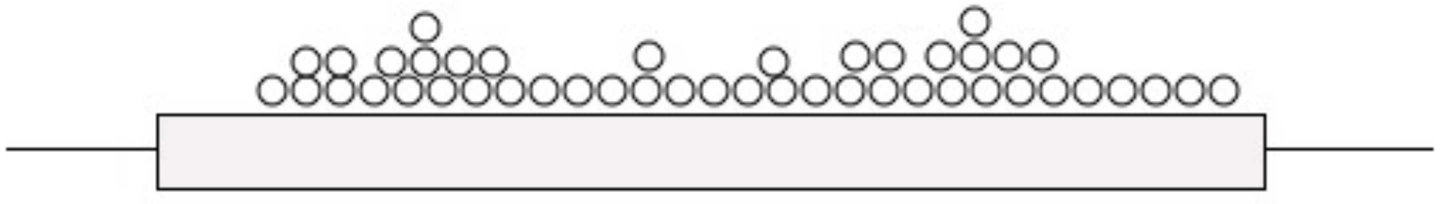
Find signals of positive selection across tumour re-sequenced genomes



Signals of positive selection

Recurrence

MuSiC-SMG / MutSigCV

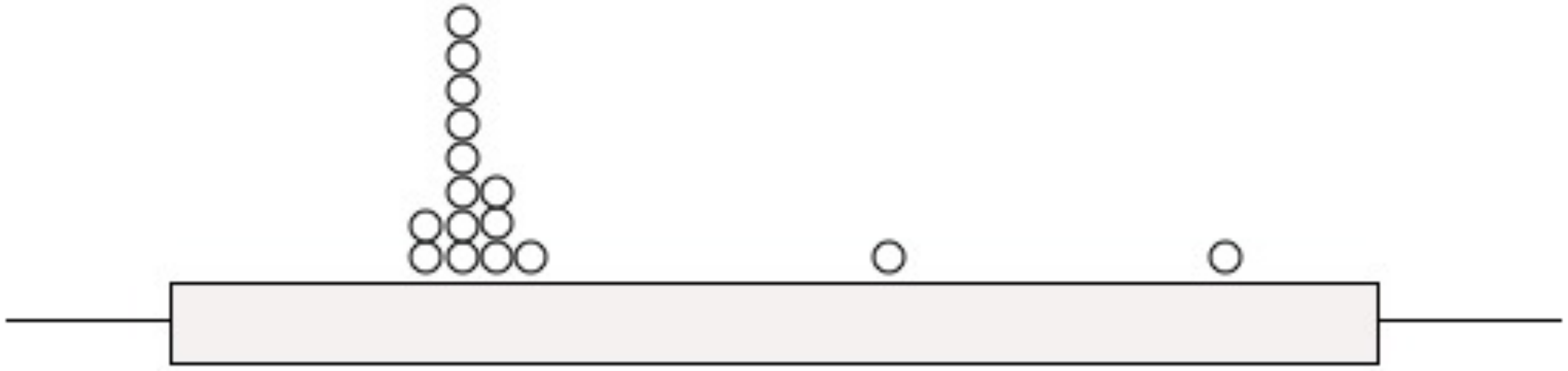


○ Mutation

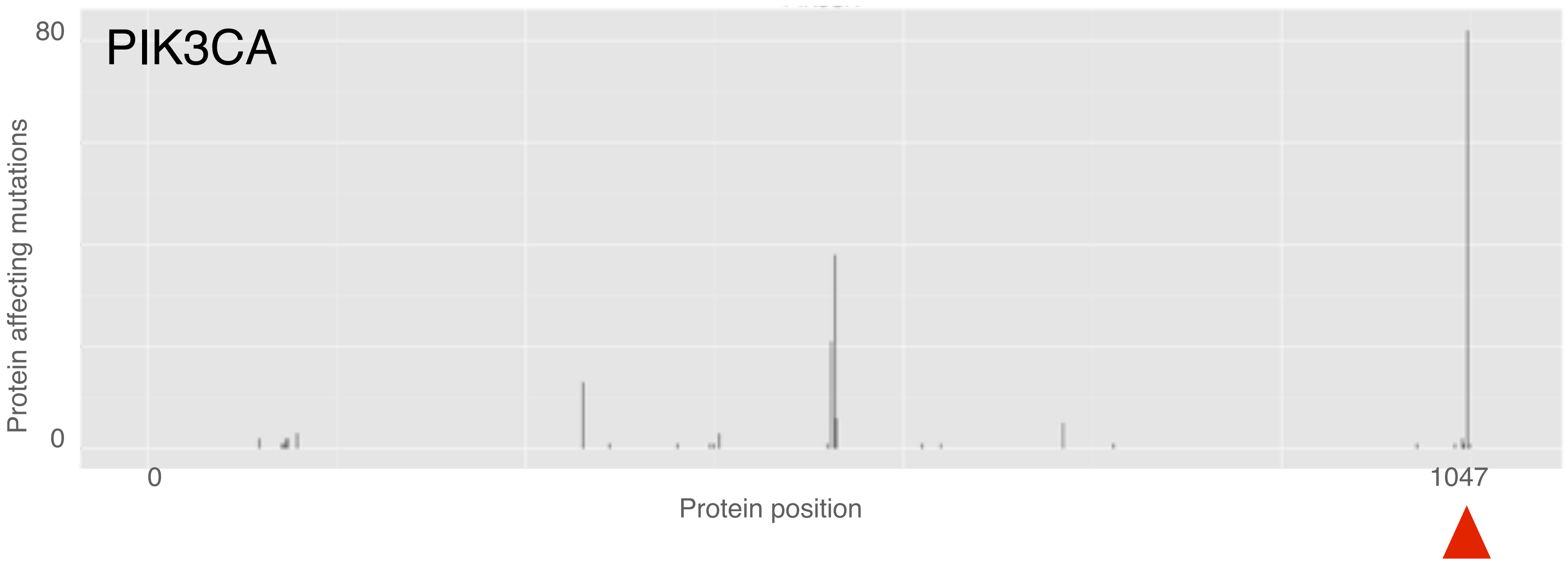
Identify genes mutated more frequently than background mutation rate

Mutation clustering

OncodriveCLUST



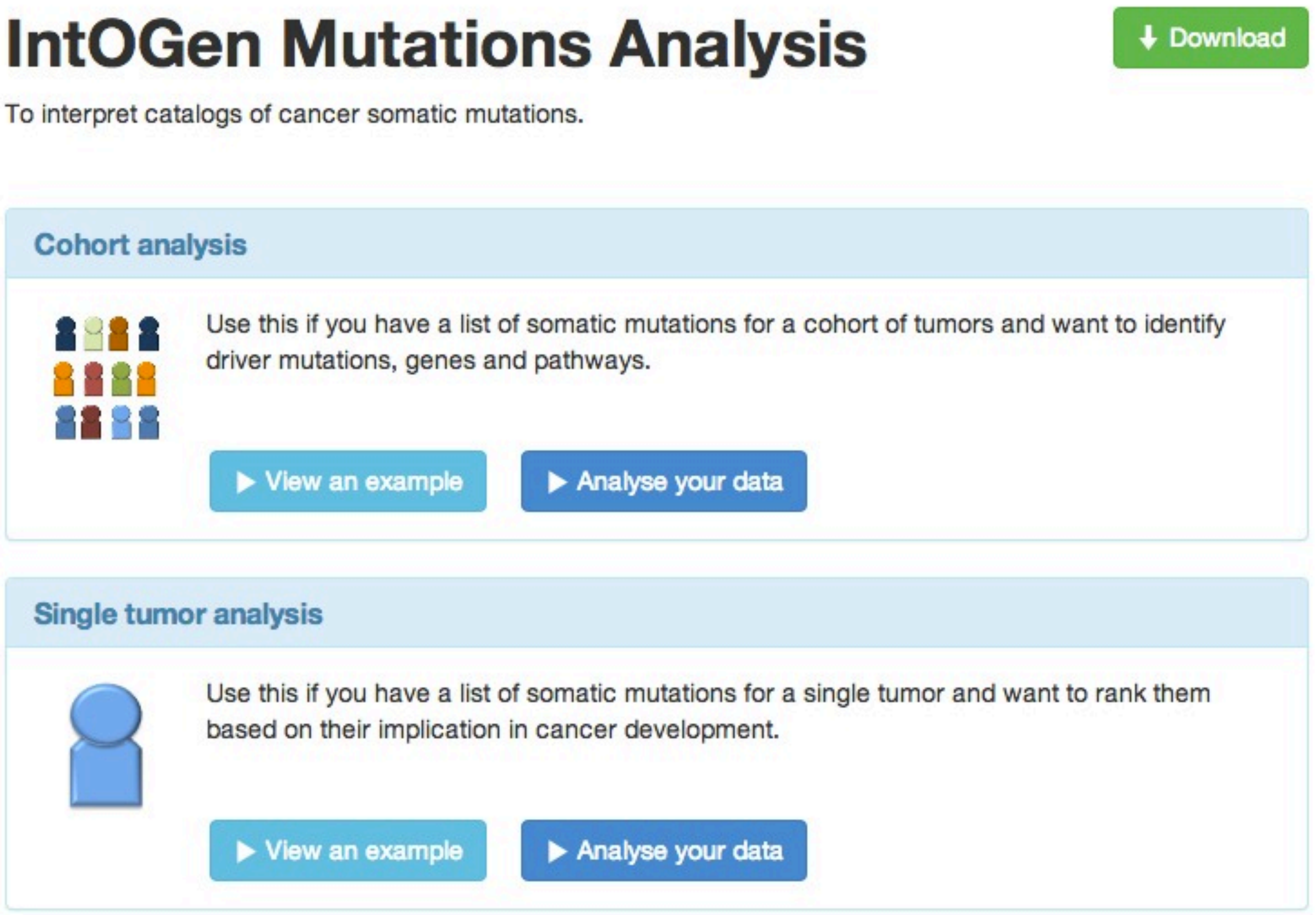
○ Mutation



H1047L

PIK3CA is recurrently mutated in the same residue in breast tumours

http://www.intogen.org/mutations/analysis



IntOGen Mutations Analysis

To interpret catalogs of cancer somatic mutations.

Cohort analysis

Use this if you have a list of somatic mutations for a cohort of tumors and want to identify driver mutations, genes and pathways.

Single tumor analysis

Use this if you have a list of somatic mutations for a single tumor and want to rank them based on their implication in cancer development.

Gonzalez-Perez et al, Nature Methods 2013