

**wellcome**  
connecting  
science

# connecting people with science

**Day 3: Driver Gene Identification and  
Oncoplot**

29th October 2025



# Driver Gene Identification and Oncoplots

**Nivedita Mukherjee**

Research Scholar (PhD)

niveditamukherjee997@gmail.com

Computational Cancer Genomics Lab

**Sabarinathan Radhakrishnan**, Assistant Professor

National Centre for Biological Sciences (NCBS - TIFR), Bangalore

*Materials partially borrowed from:*

**Patricia Basurto-Lozada**, PhD Candidate LIIGH-UNAM

**Nyasha Chambwe**, Assistant Professor, Institute of Molecular Medicine, Feinstein Institutes for Medical Research

**Federico Abascal**, PhD Wellcome Sanger Institute



# Outline

- What is a “cancer driver” gene/mutation?
- Clinical relevance of cancer drivers
- Identification of cancer drivers

dNdScv

Oncoplots

What is a “cancer driver” gene/mutation?

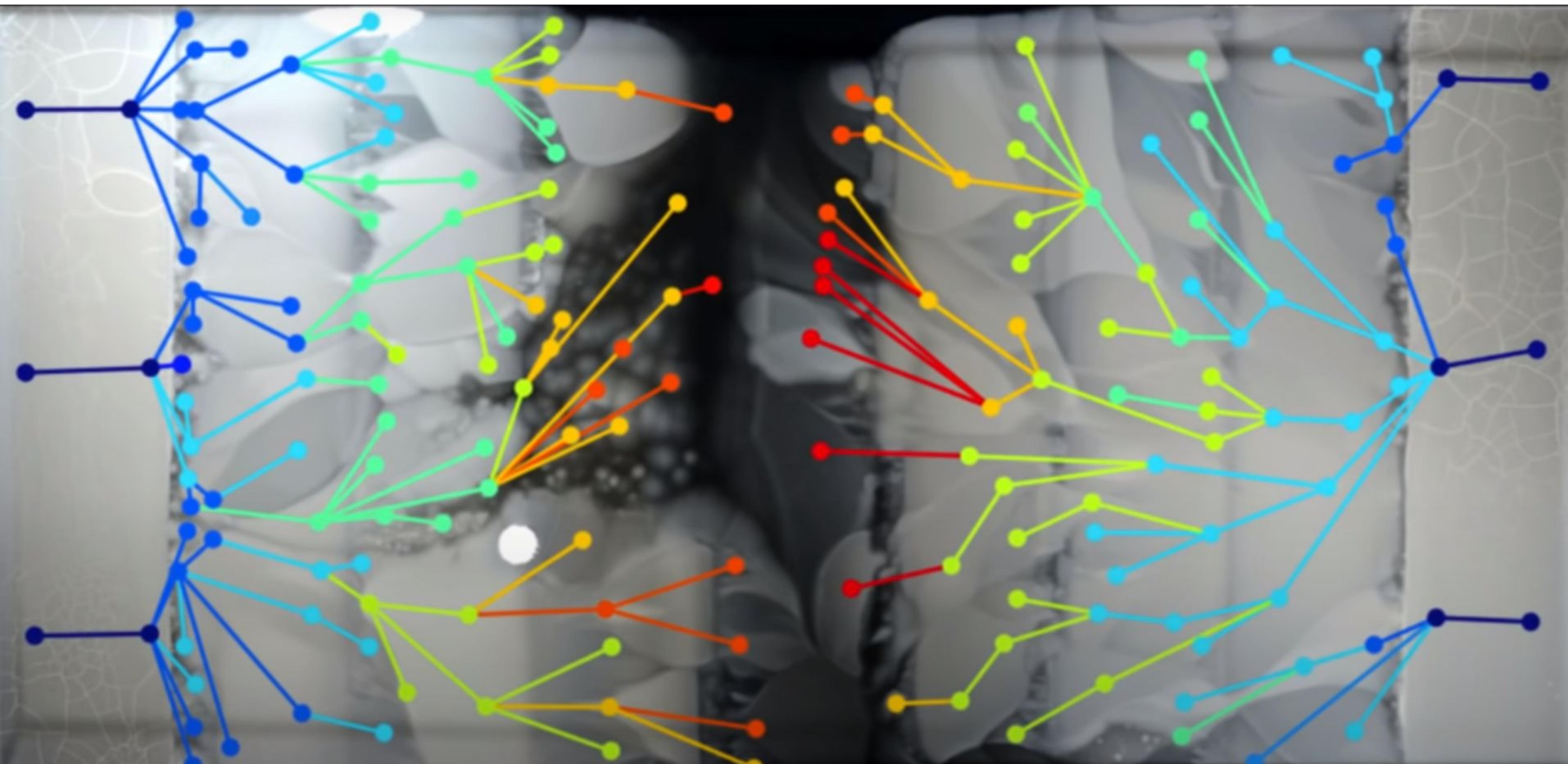
*“Nothing in biology makes sense except in the light of evolution.”*

– Theodosius G. Dobzhansky

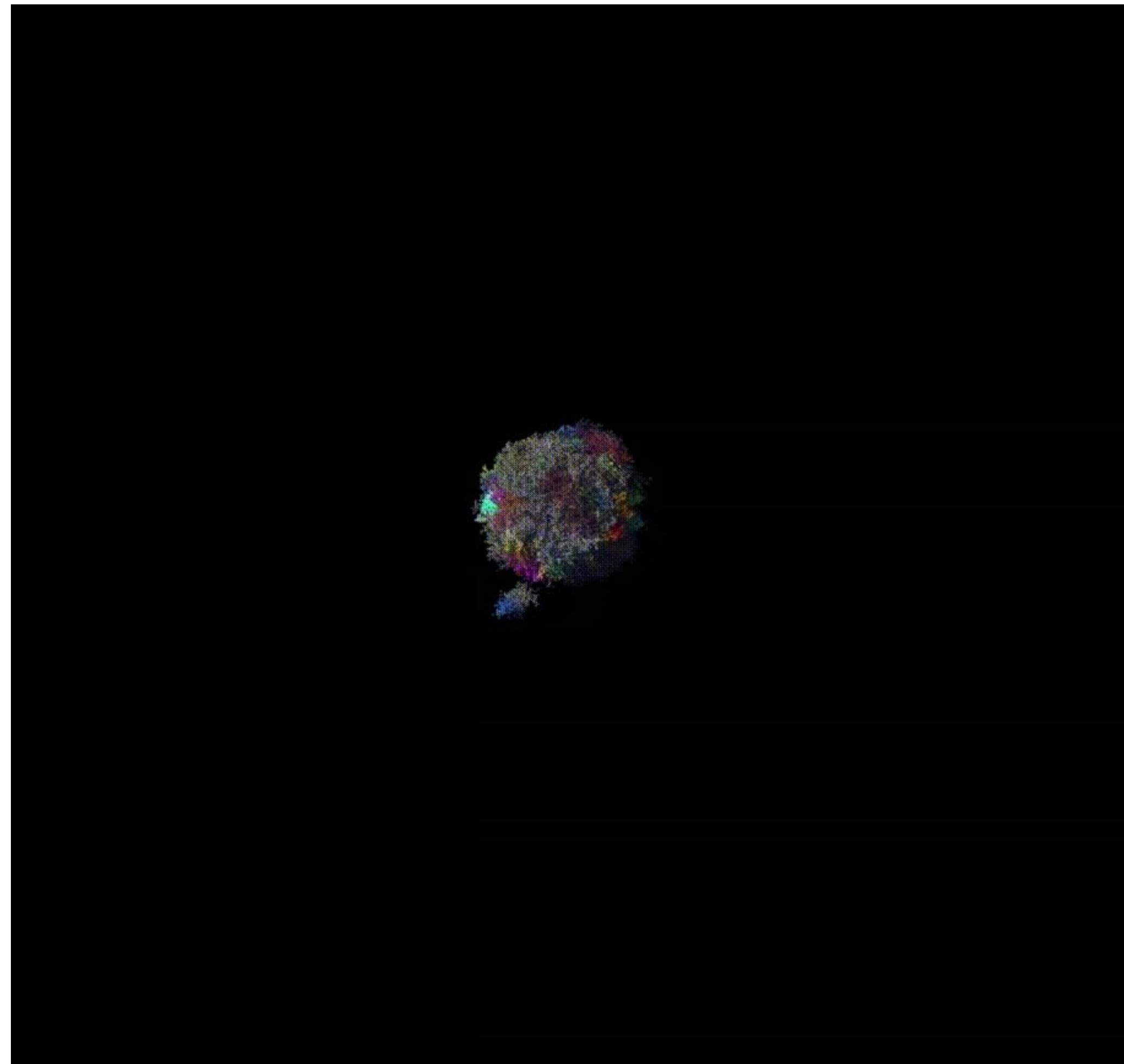
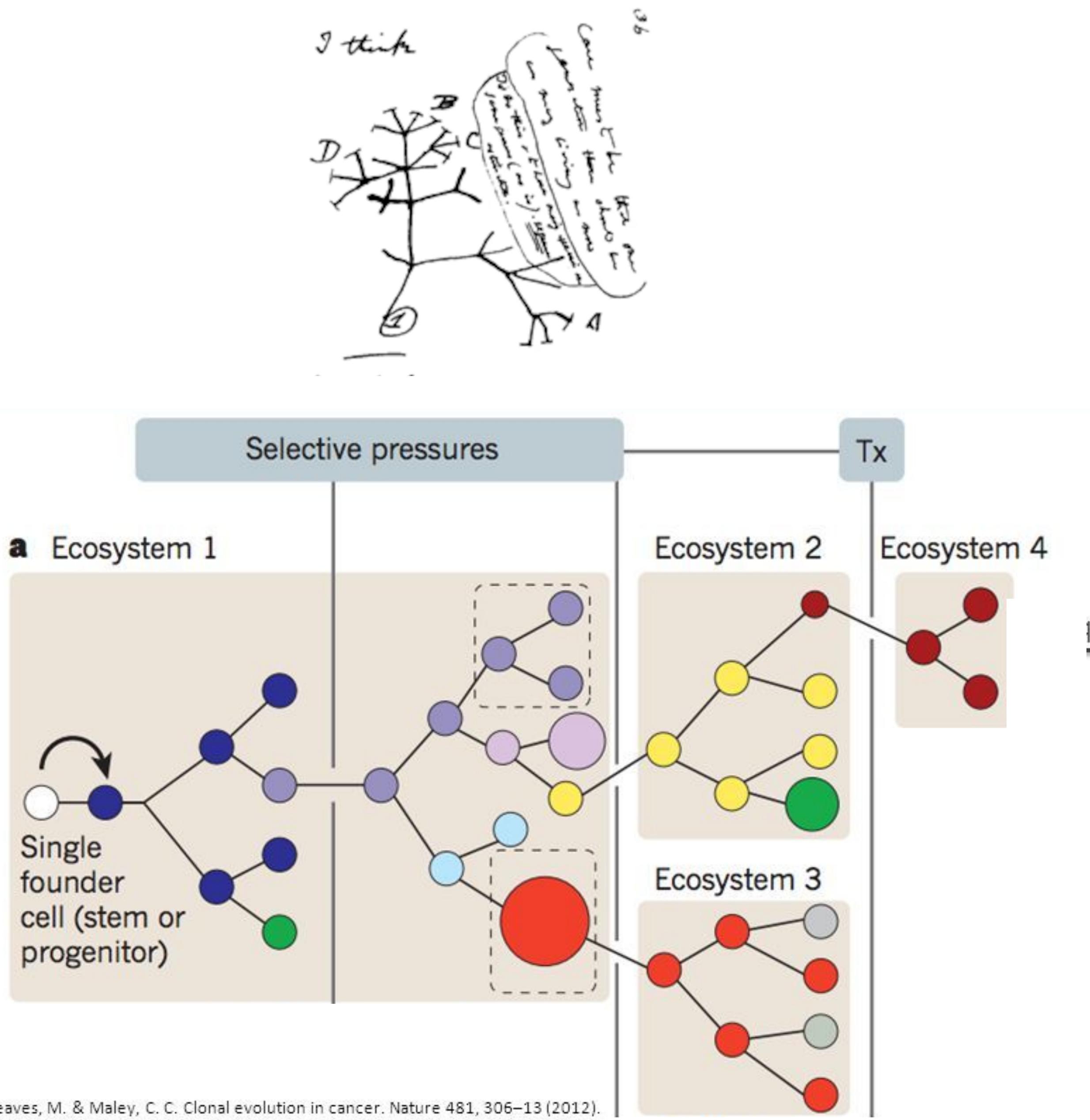
# The Evolution of Bacteria on a “Mega-Plate” Petri Dish



# The Evolution of Bacteria on a “Mega-Plate” Petri Dish



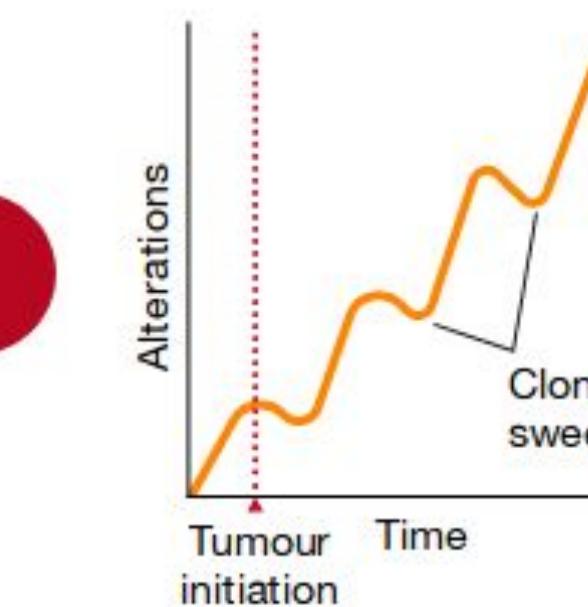
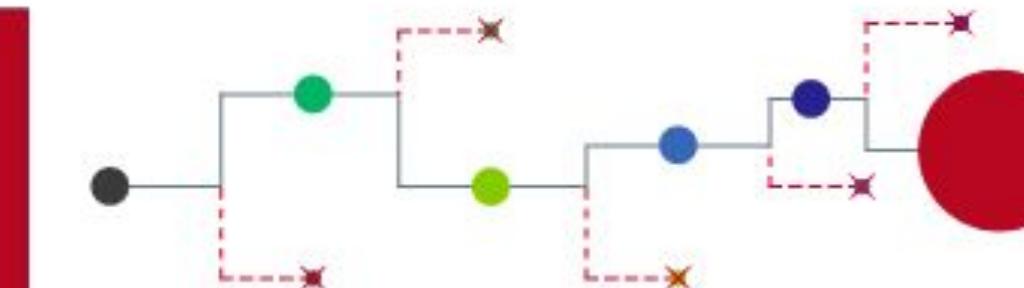
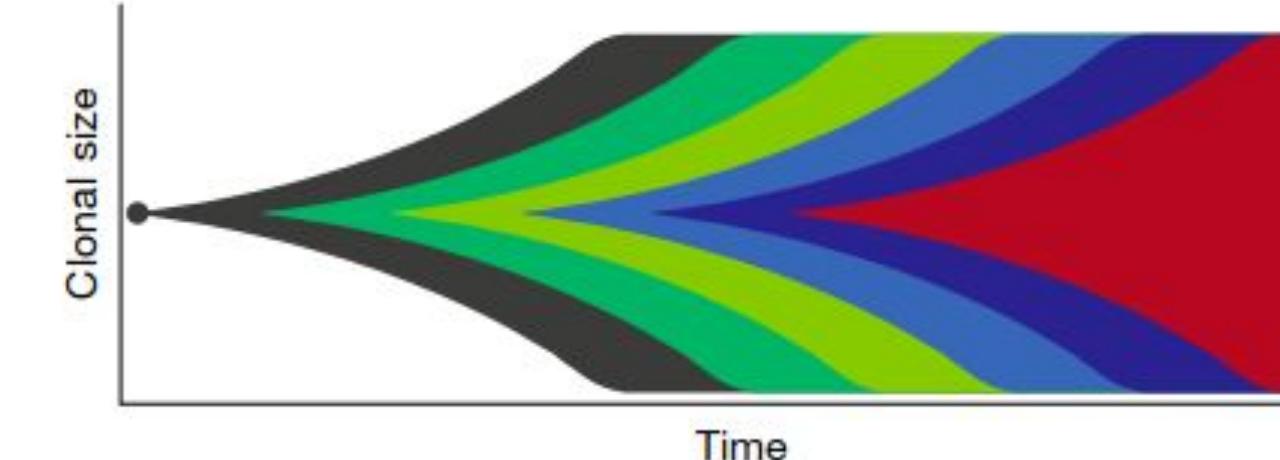
# Cancer: A microcosm of Darwinian evolution



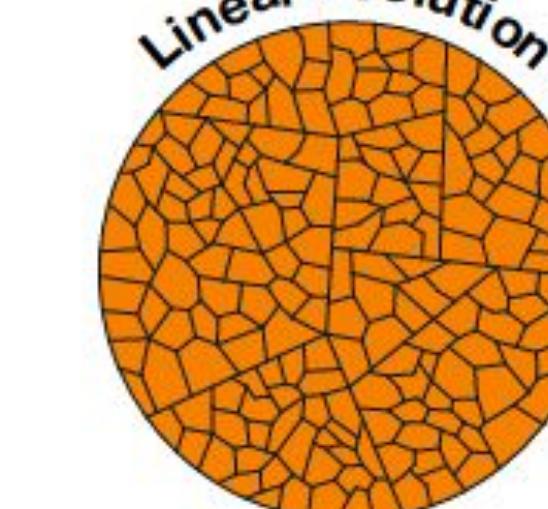
DOI:10.1038/nature14971  
<https://youtu.be/lpytolxRu0o>

# Models of tumor evolution

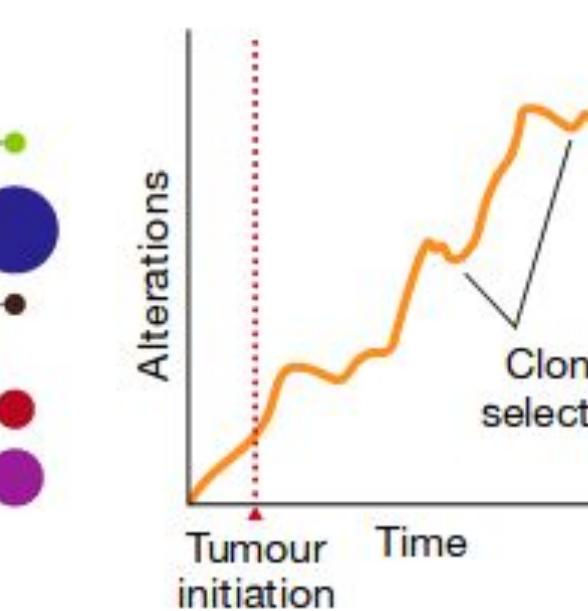
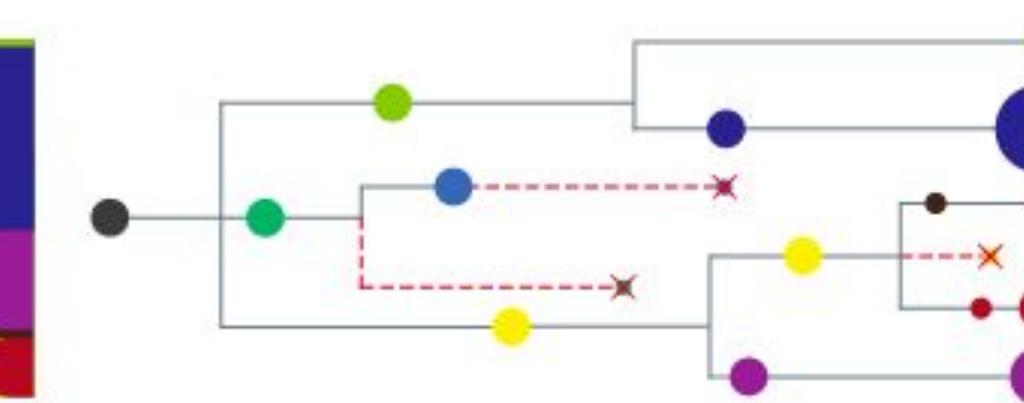
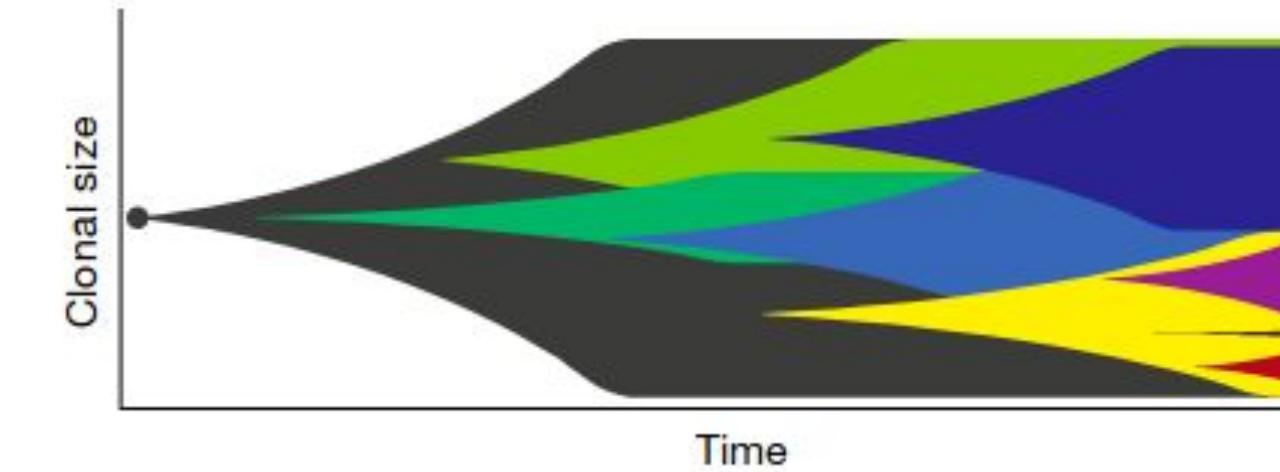
**A Linear evolution**



**Linear evolution**



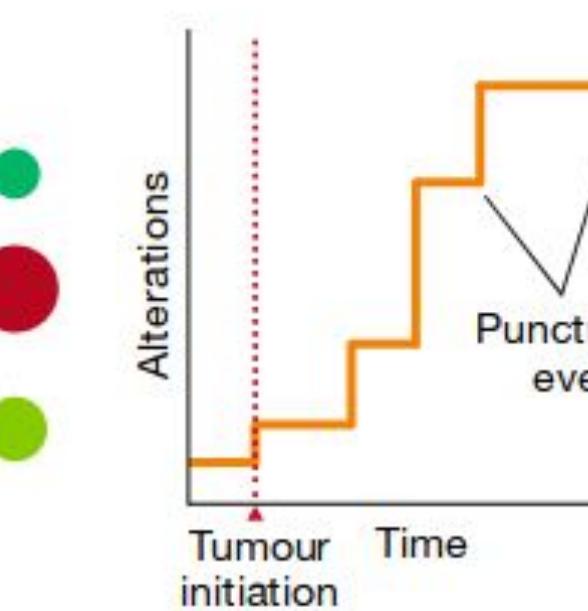
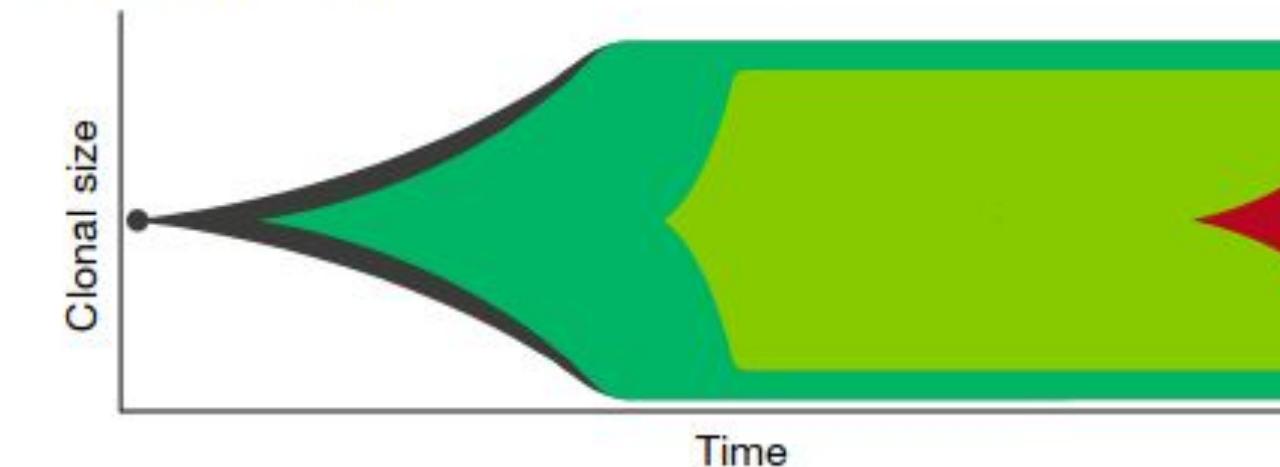
**B Branched evolution**



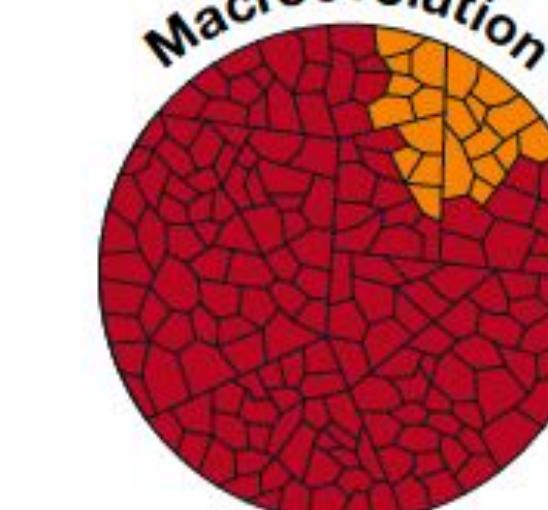
**Branched evolution**



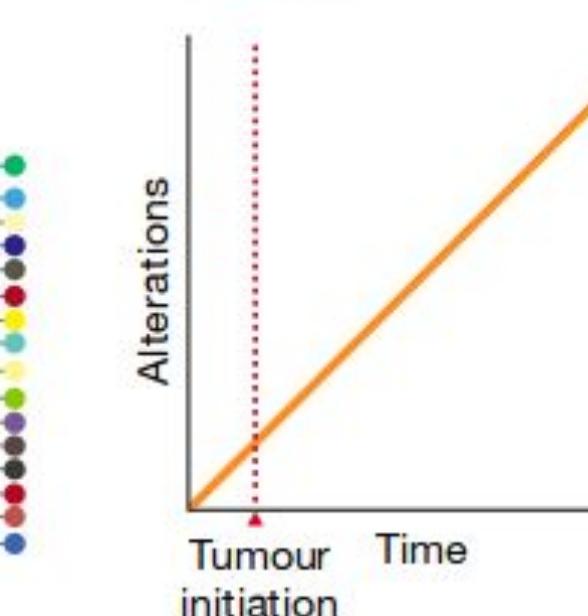
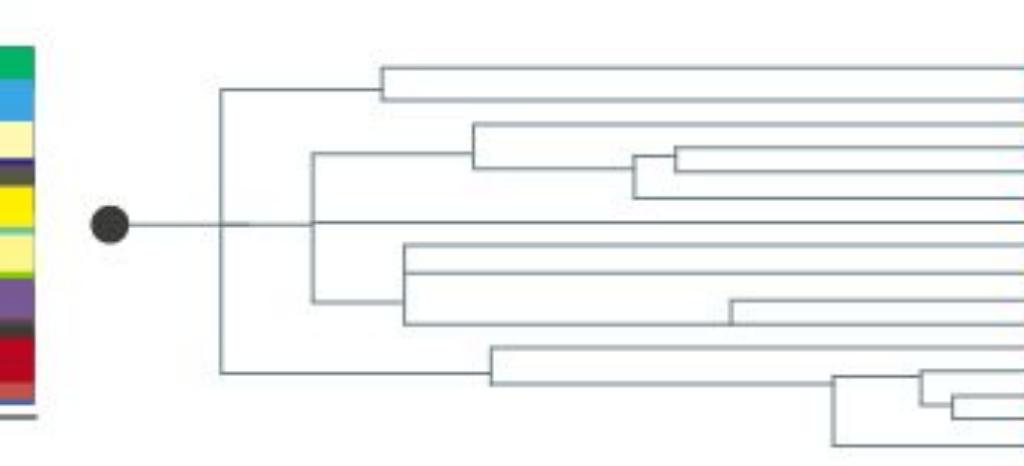
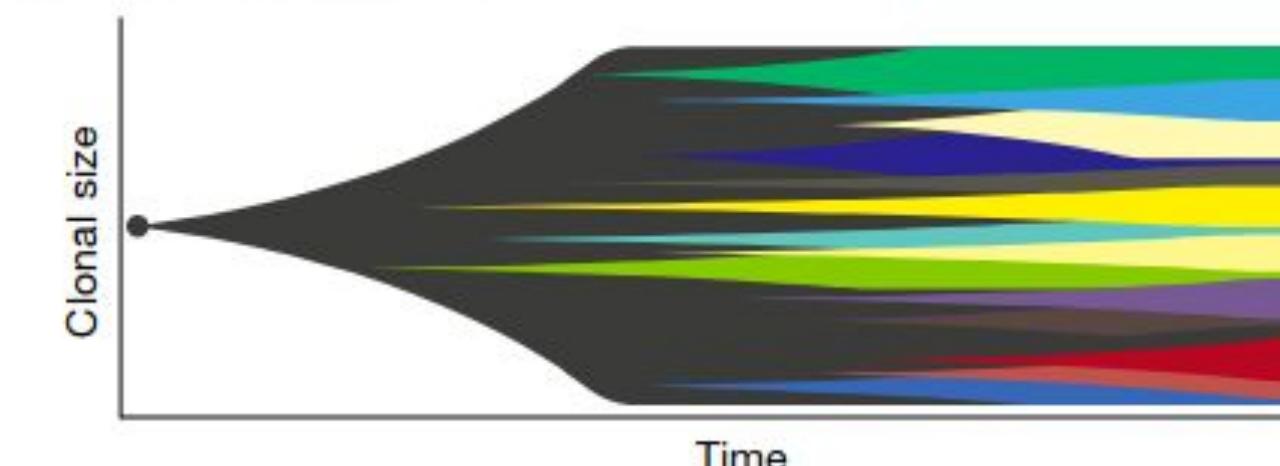
**C Macroevolution**



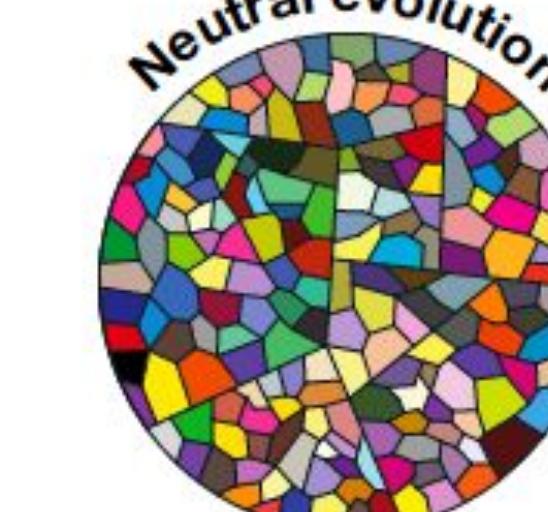
**Macroevolution**



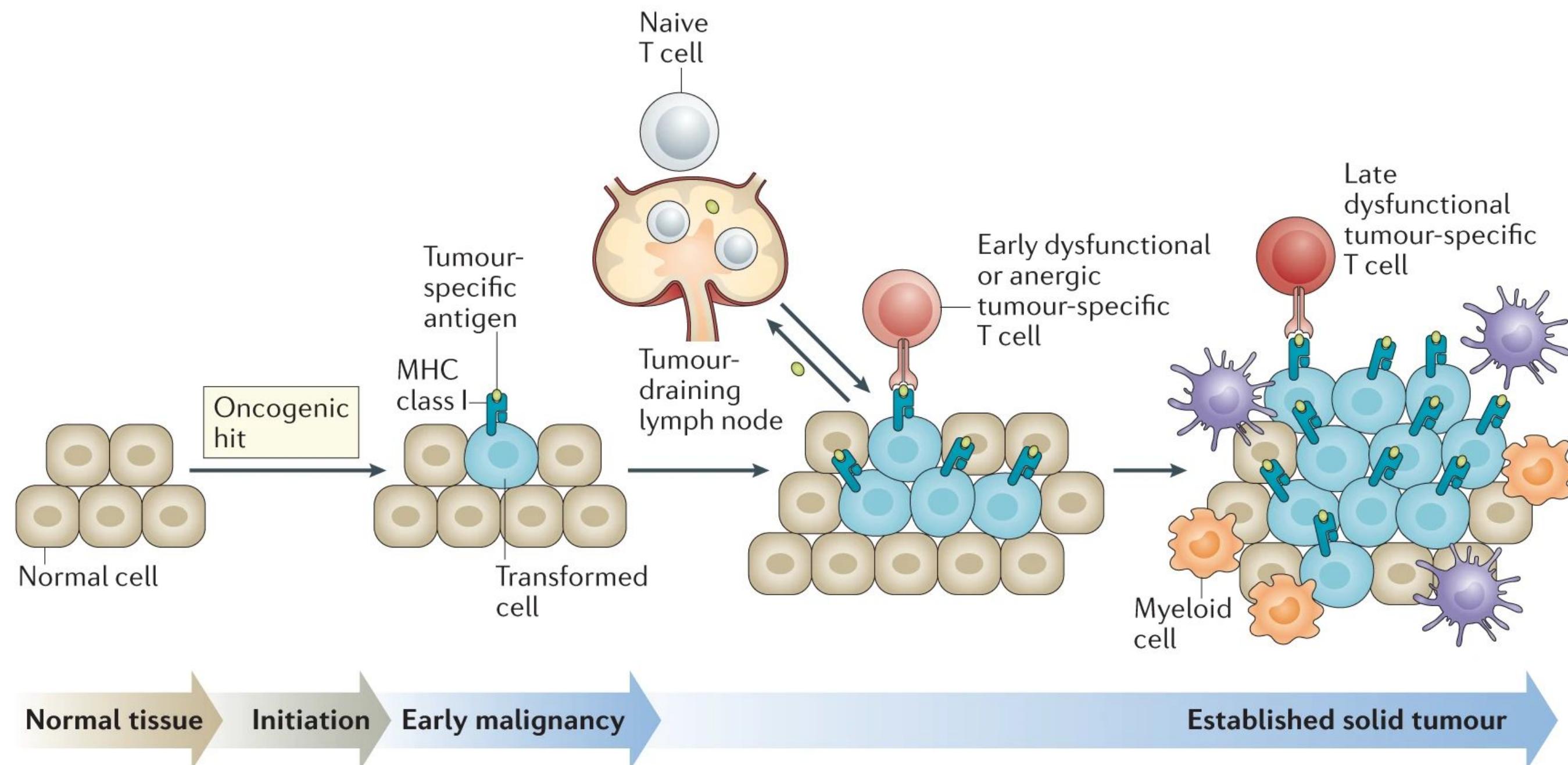
**D Neutral evolution**



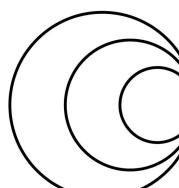
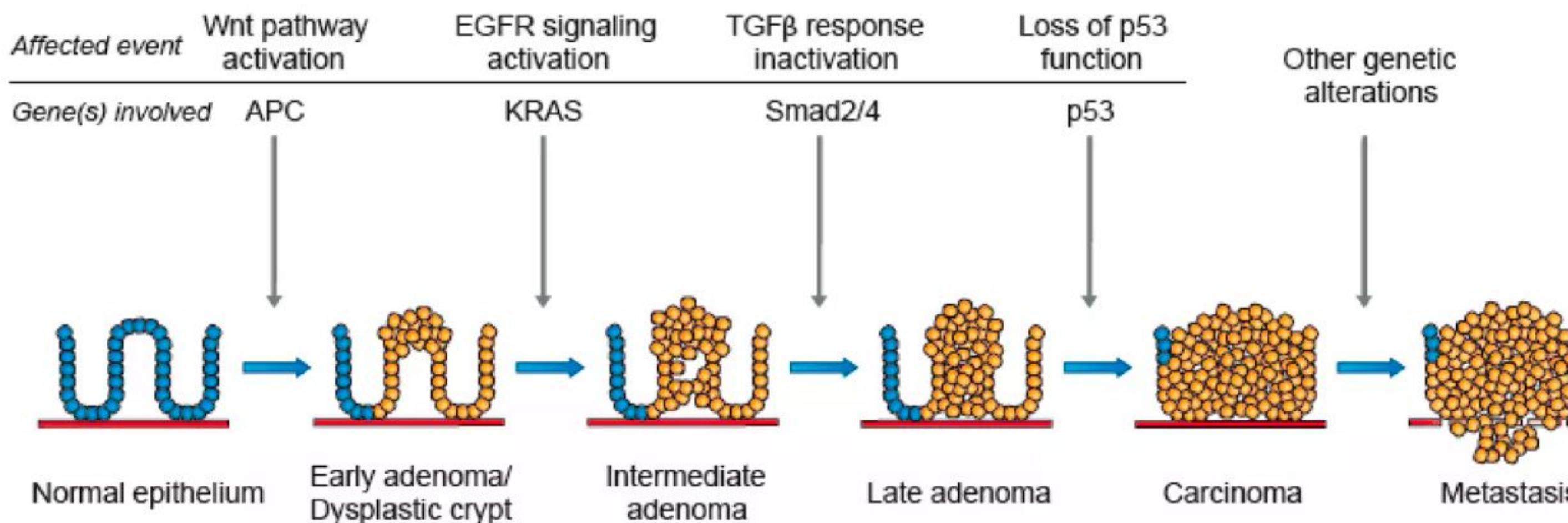
**Neutral evolution**



# Tumor initiation, progression and evolution



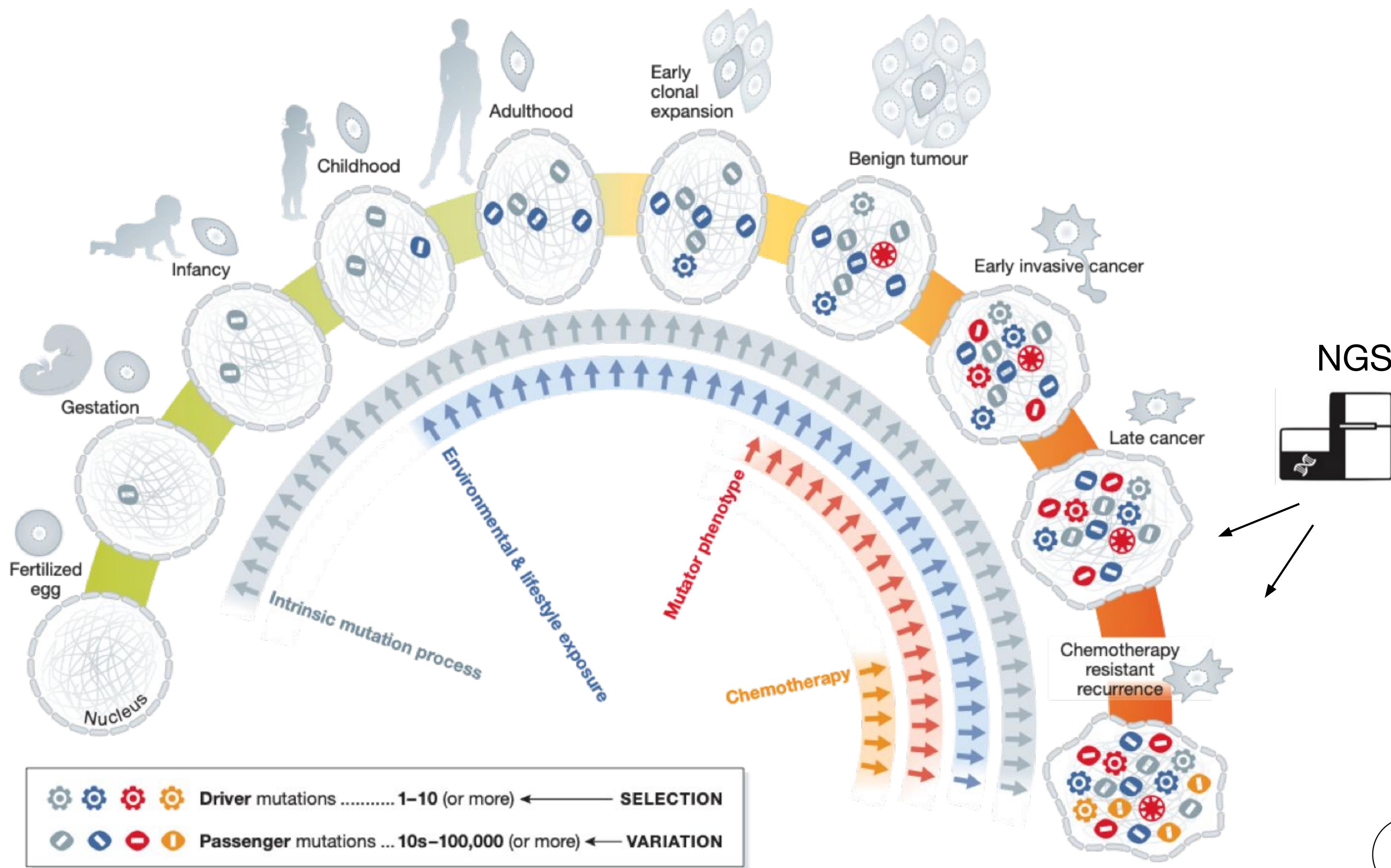
Philip and Schietinger,  
*Nature Reviews Immunology* 22,  
209–223 (2022)



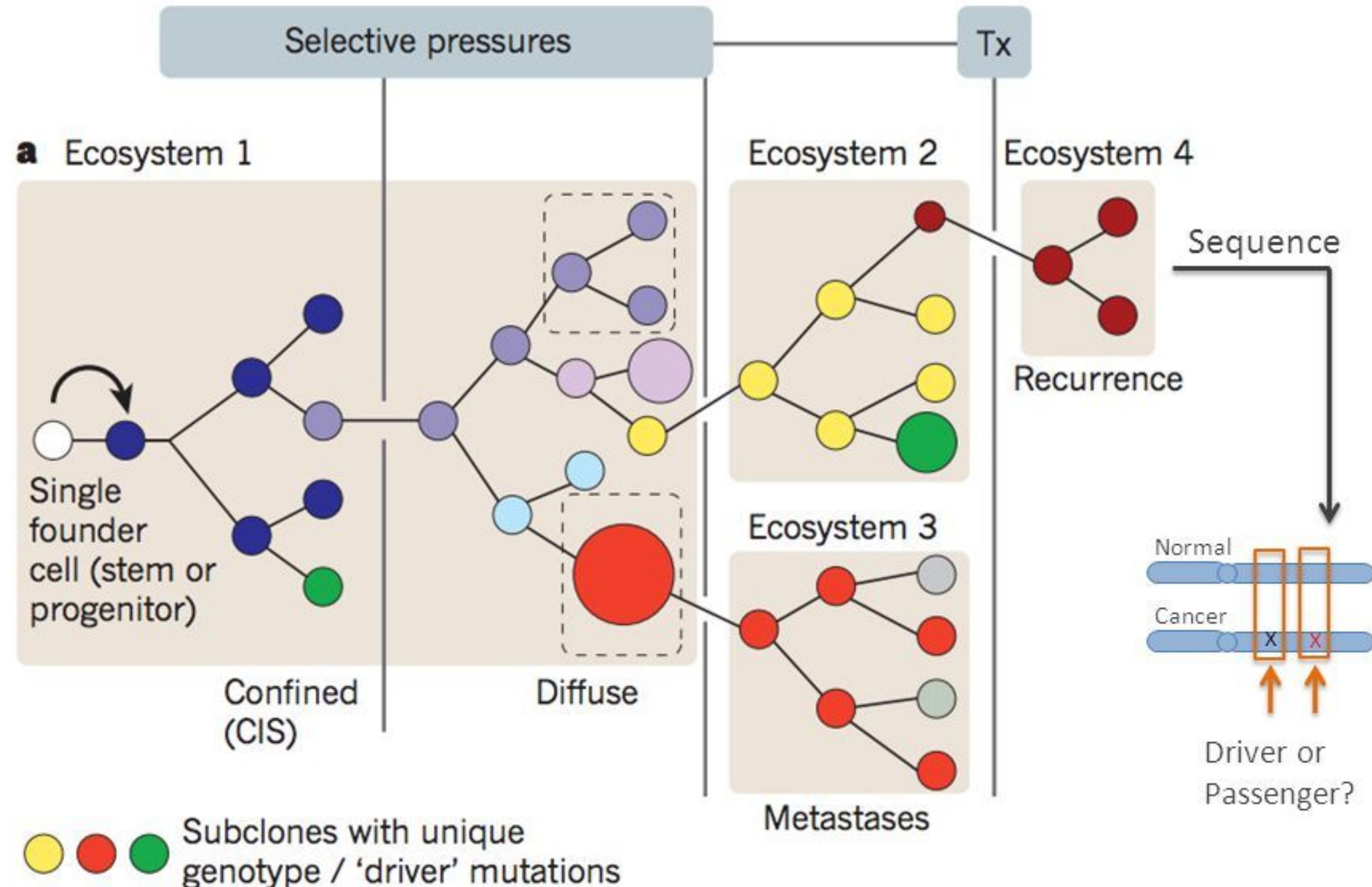
wellcome  
connecting  
science



# Somatic mutations accumulate over the lifetime of an individual



# Drivers and Passengers



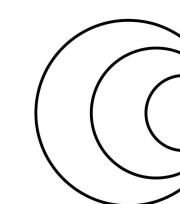
**Drivers** evolve under natural selection

**Passengers** evolve under genetic drift



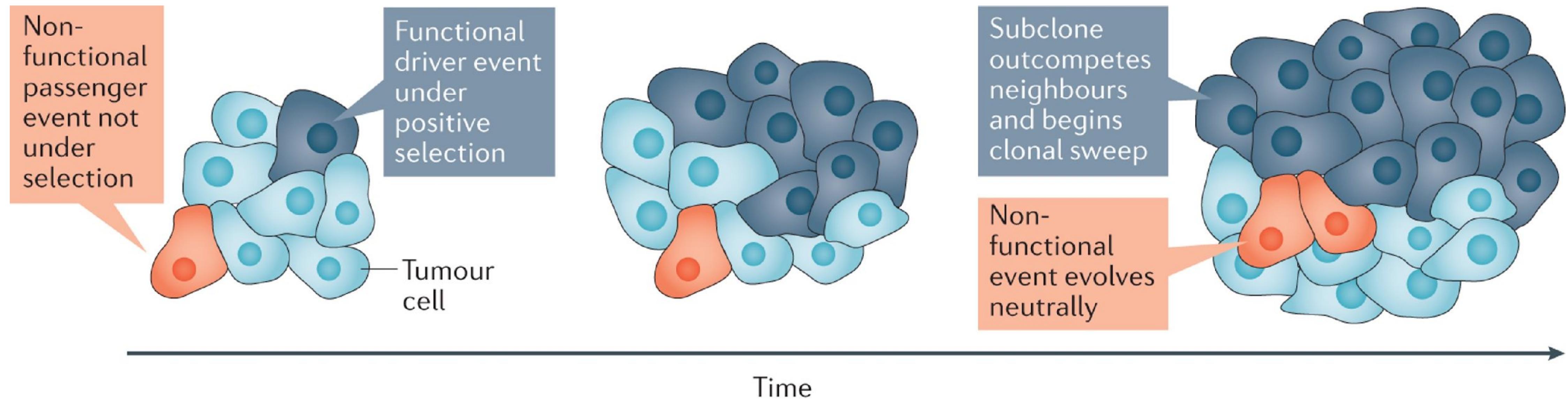
Alamy

Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* 481, 306–13 (2012).



wellcome  
connecting  
science

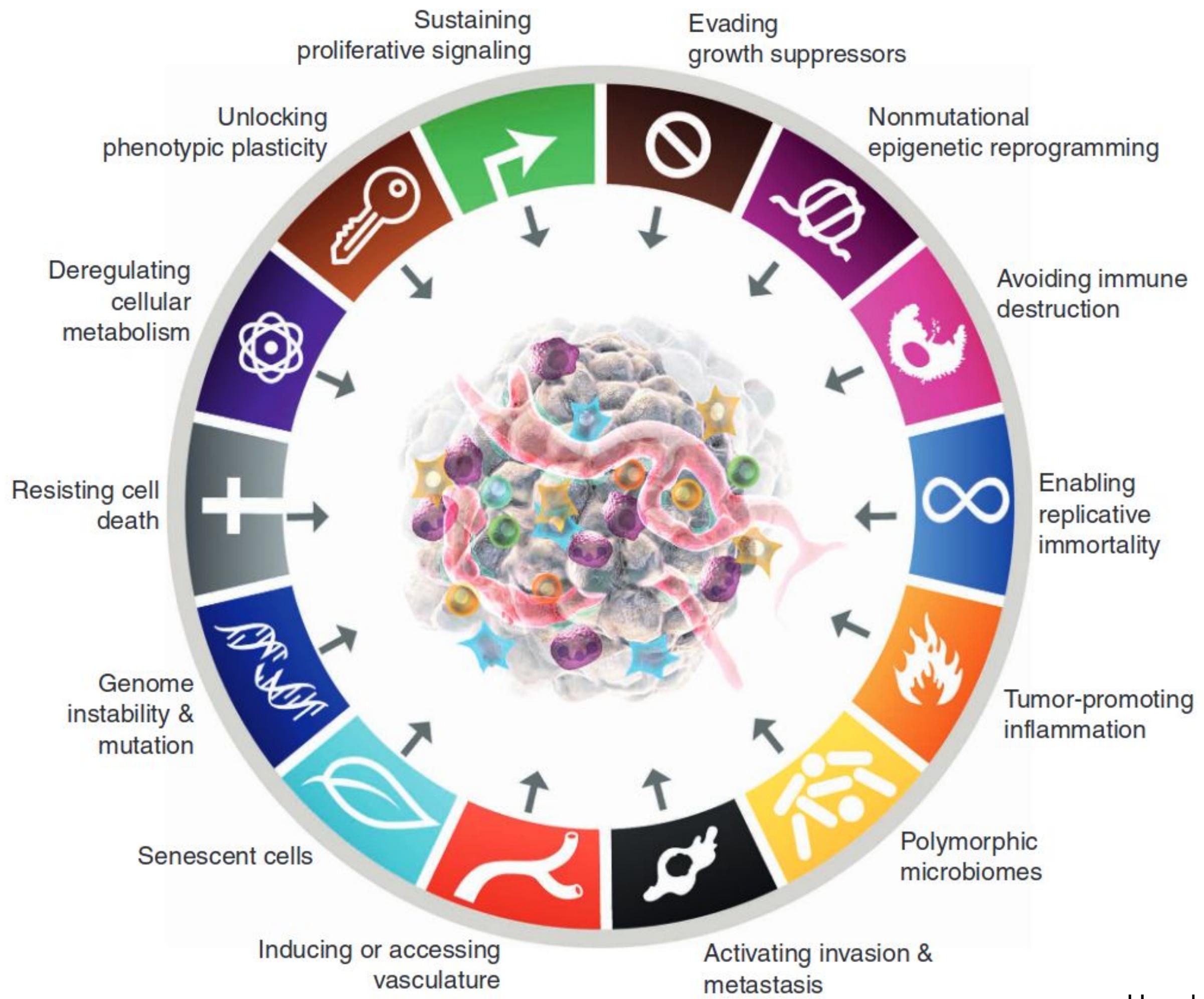
# Functional and non-functional intra-tumor heterogeneity in tumor evolution



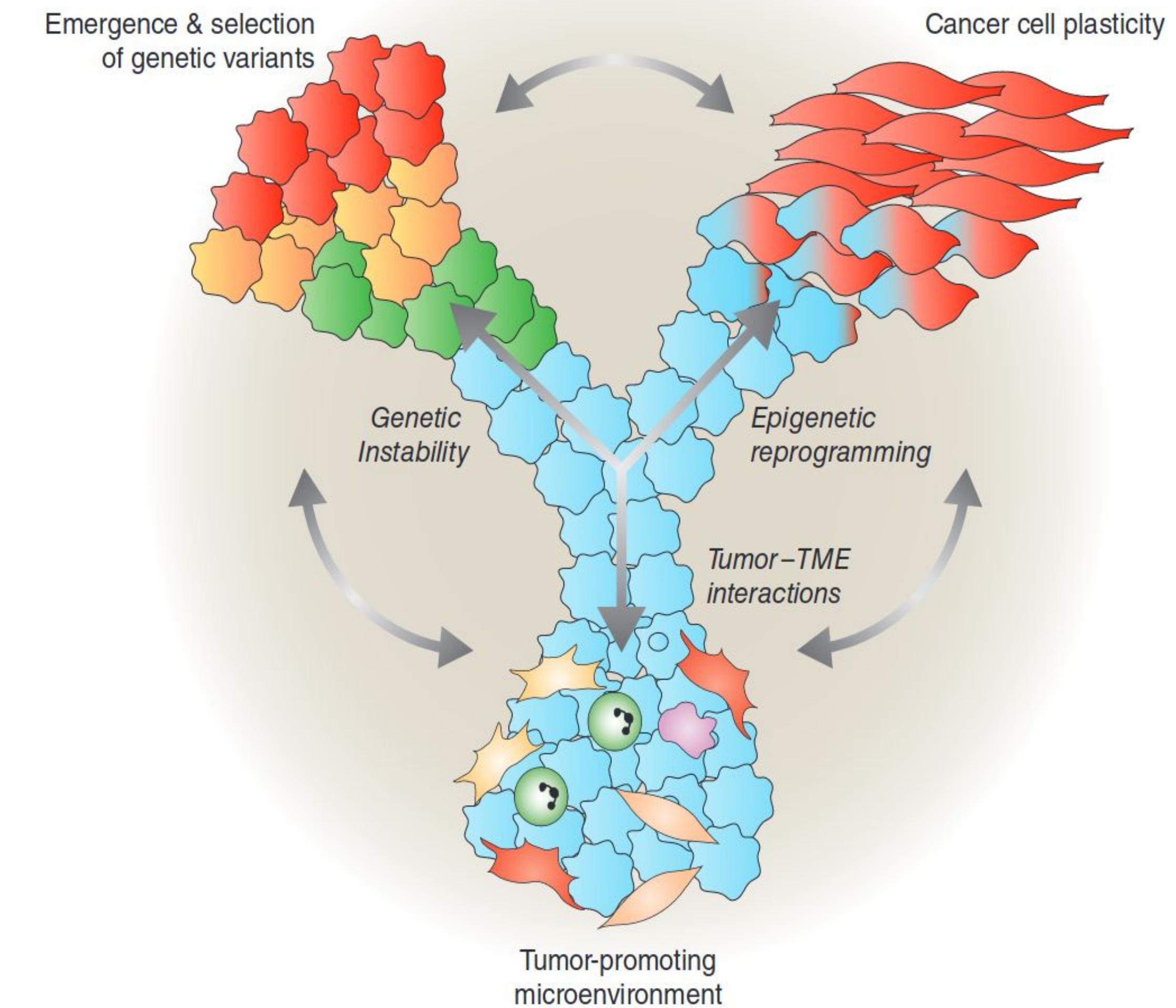
1. The increased rate of phenotypic variation in cancers compared with normal tissues means that new subclones arise and compete.
2. A minority contain a **driver event**, such as a genetic mutation or copy number alteration, that grants a **selective advantage**.
3. These subclones may grow at a faster rate than their neighbors and outcompete them in a '**selective sweep**'.

# Drivers and Passengers

Drivers are causal alterations that enable the hallmarks of cancer



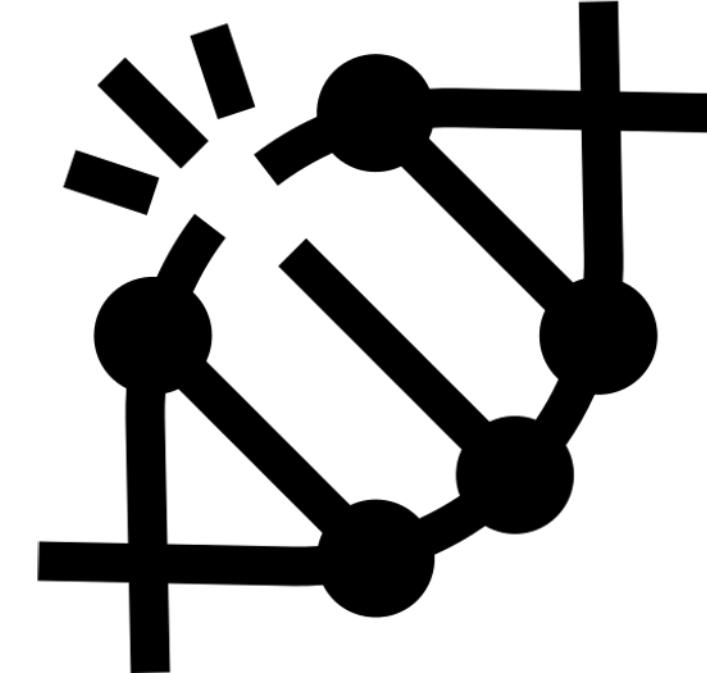
Hanahan D. Cancer  
Discov. 2022;12(1):31-46



Ciriello et al

# Driver mutations

- Mutations in cancer can be classified into:
  - Driver mutations
  - Passenger mutations
- **Driver mutations**
  - Provide a selective advantage to the cell.
  - Promote cancer development
- Passenger mutations
  - Neutral mutations.



## Driver genes

Genes that harbor driver mutations are called “**cancer driver genes**”.

# A genetic basis of cancer: The first identified driver genes

RB1: a tumor suppressor gene

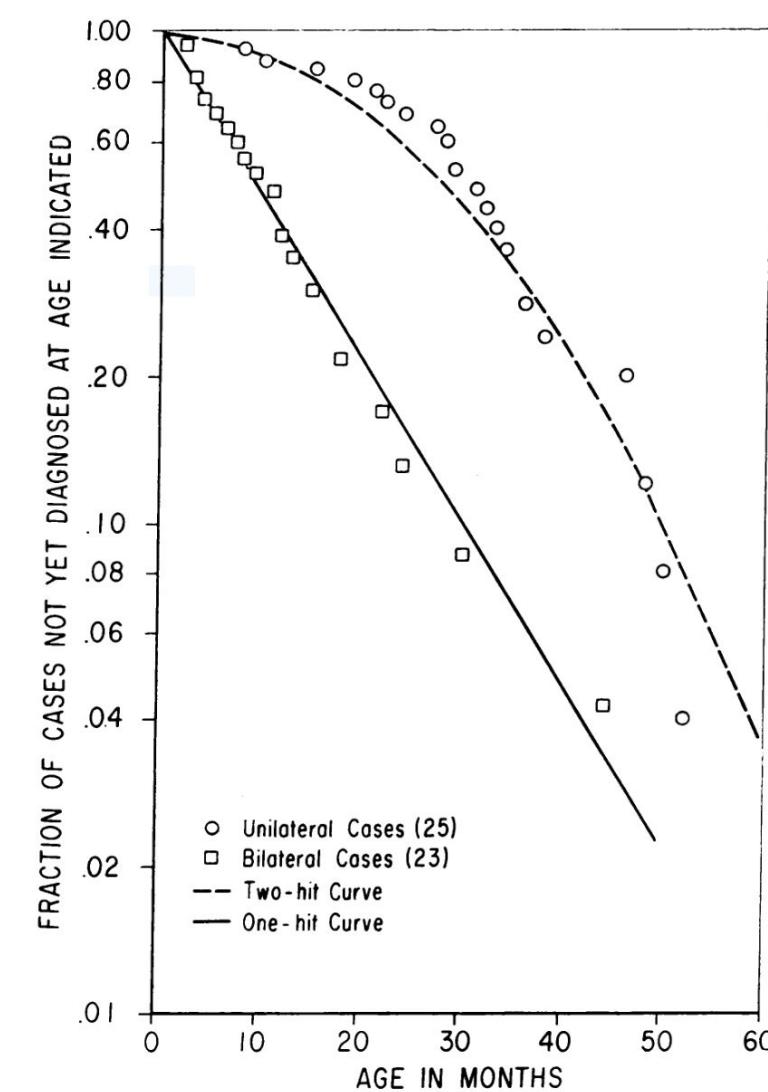
SRC: a proto-oncogene

## Mutation and Cancer: Statistical Study of Retinoblastoma

ALFRED G. KNUDSON, JR.

Graduate School of Biomedical Sciences and M. D. Anderson Hospital and Tumor Institute,  
The University of Texas at Houston, Houston, Texas 77025

Communicated by James V. Neel, February 8, 1971

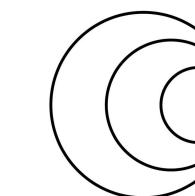
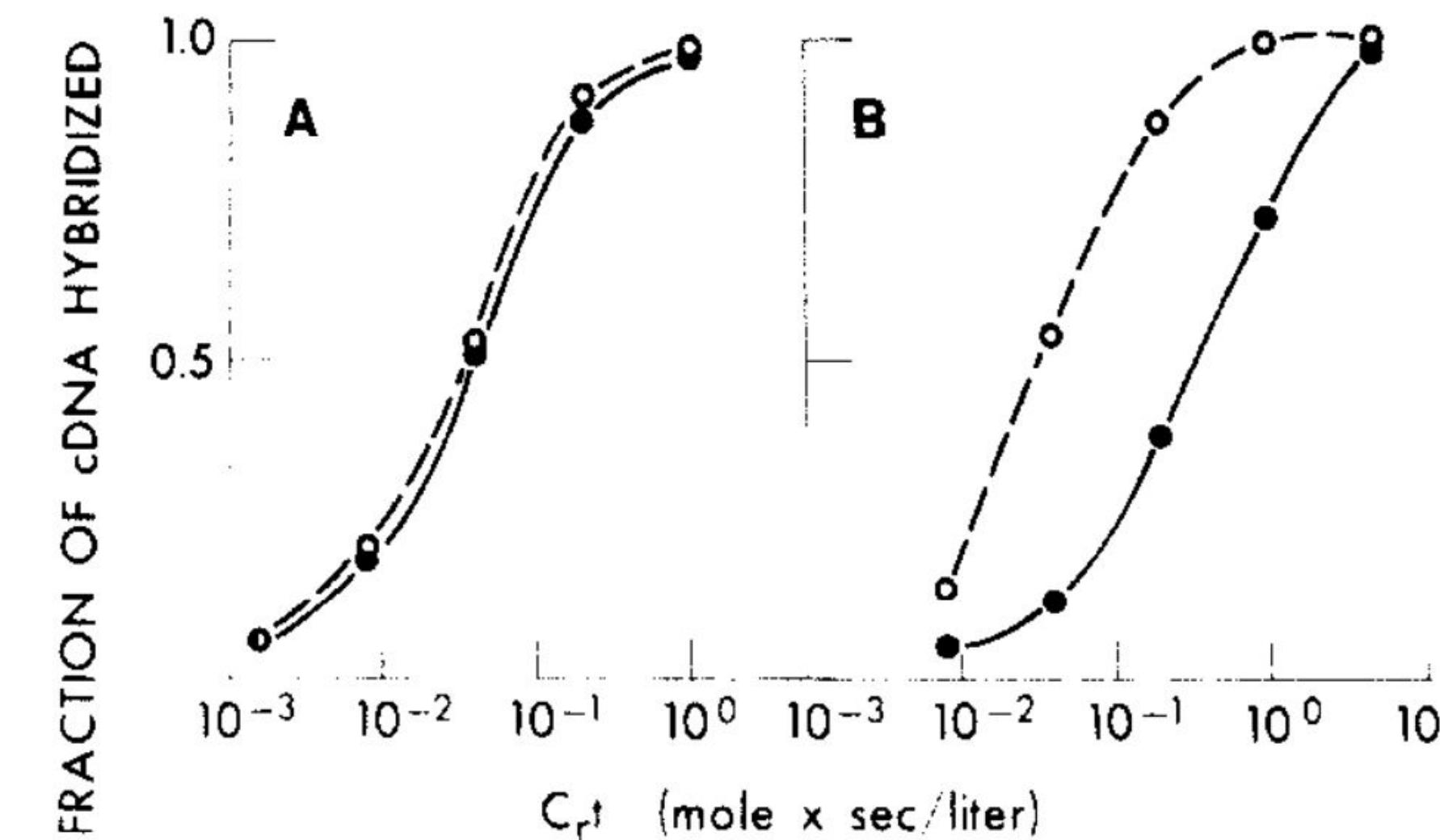


## Detection and Enumeration of Transformation-Defective Strains of Avian Sarcoma Virus with Molecular Hybridization

DOMINIQUE STEHELIN,<sup>1</sup> DONALD J. FUJITA, THOMAS PADGETT,  
HAROLD E. VARMUS, AND J. MICHAEL BISHOP<sup>2</sup>

Department of Microbiology, University of California, San Francisco, California 94143

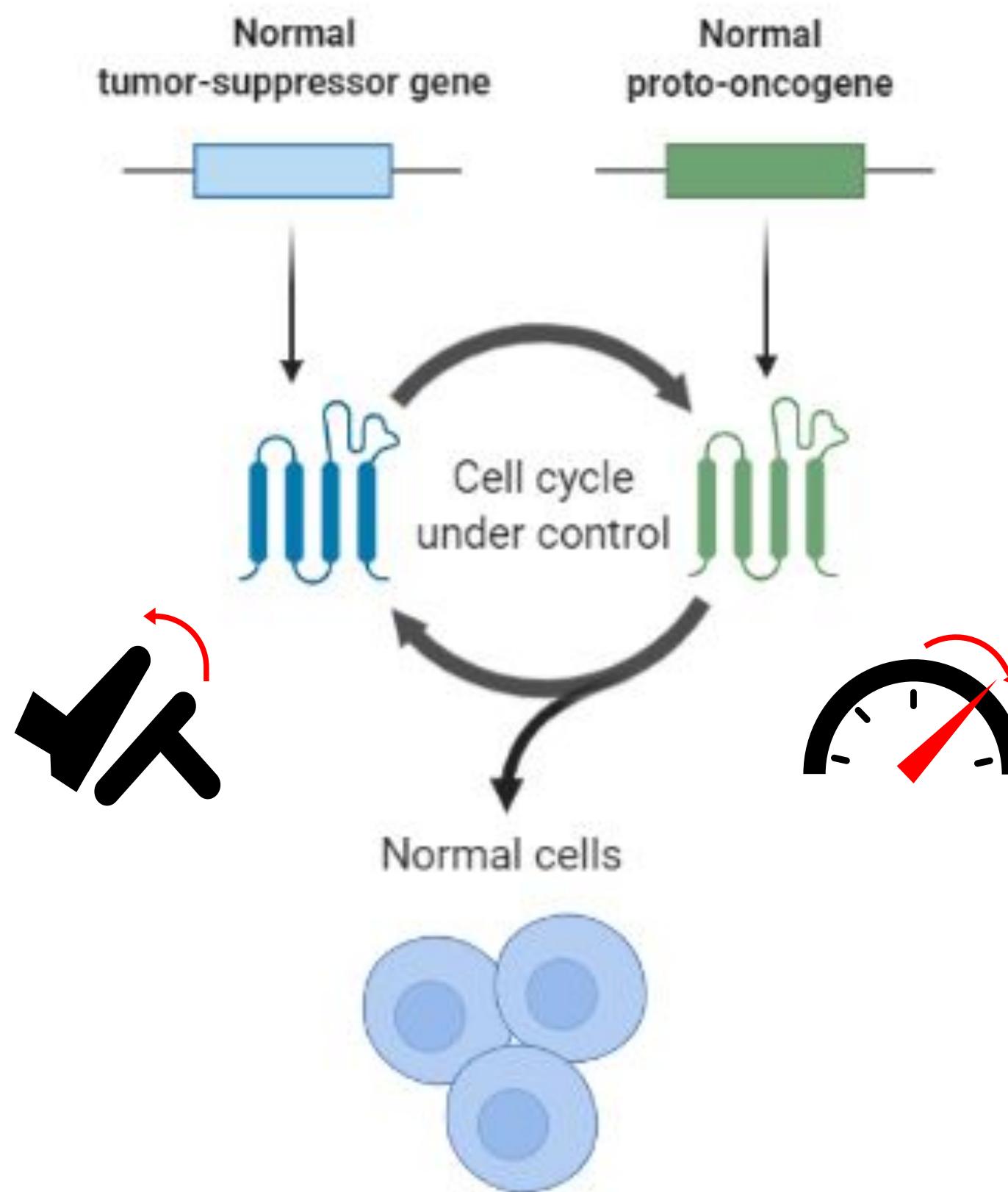
Accepted September 27, 1976



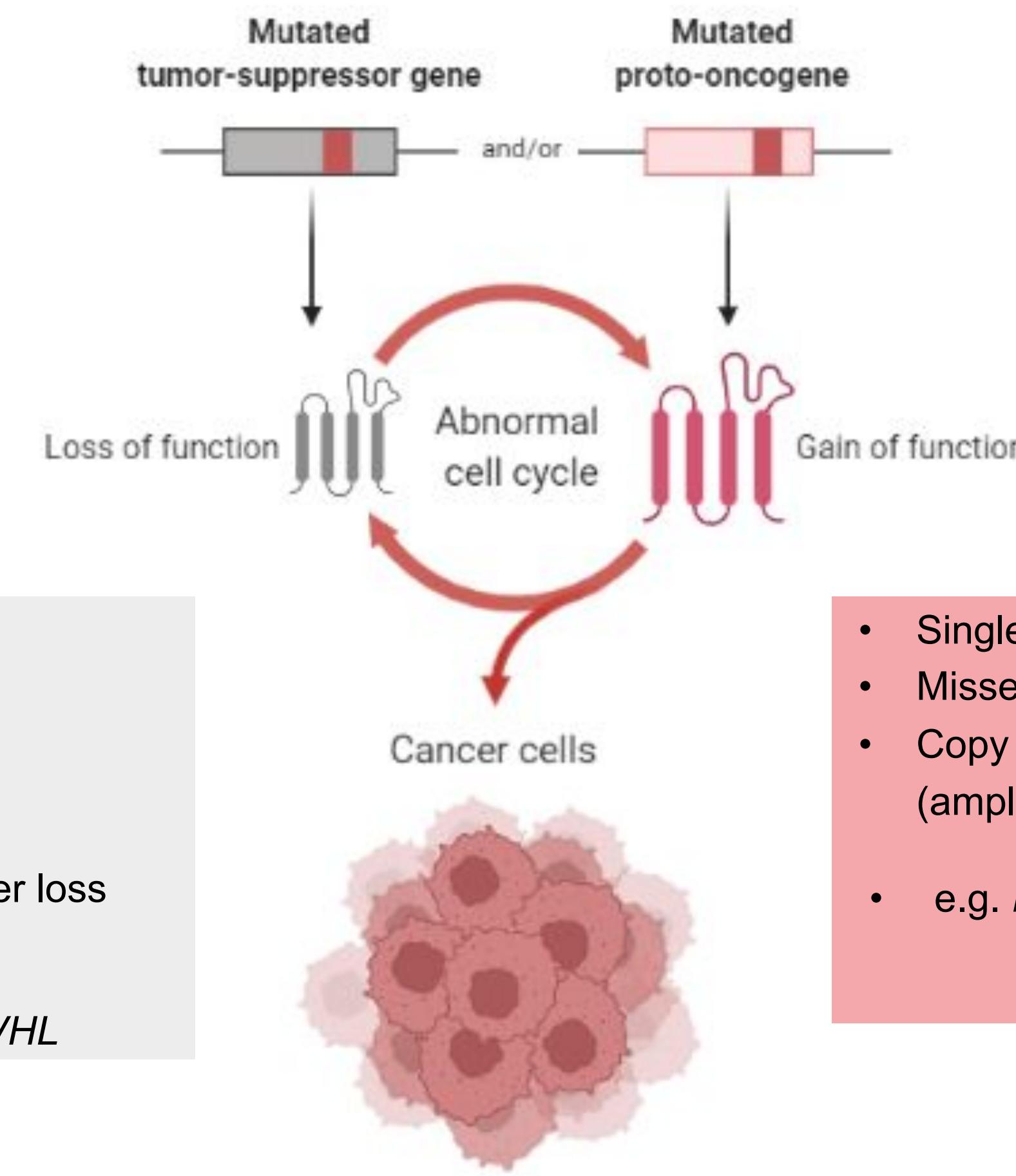
wellcome  
connecting  
science

# Oncogenes and Tumor Suppressor Genes

## Normal Cell Division



## Malignant Cell Division



- Double hit
- Missense, splice site, nonsense mutations
- Copy number loss (Deletions)
- e.g. *TP53*, *VHL*

- Single hit
- Missense mutations
- Copy number gain (amplifications)
- e.g. *KRAS*, *MYC*

# Clinical relevance of cancer drivers

# Why identify drivers?

- Cancer monitoring (e.g. ctDNA)
- Precision medicine in cancer treatment

Bottleneck: Requires identification of actionable genetic alterations

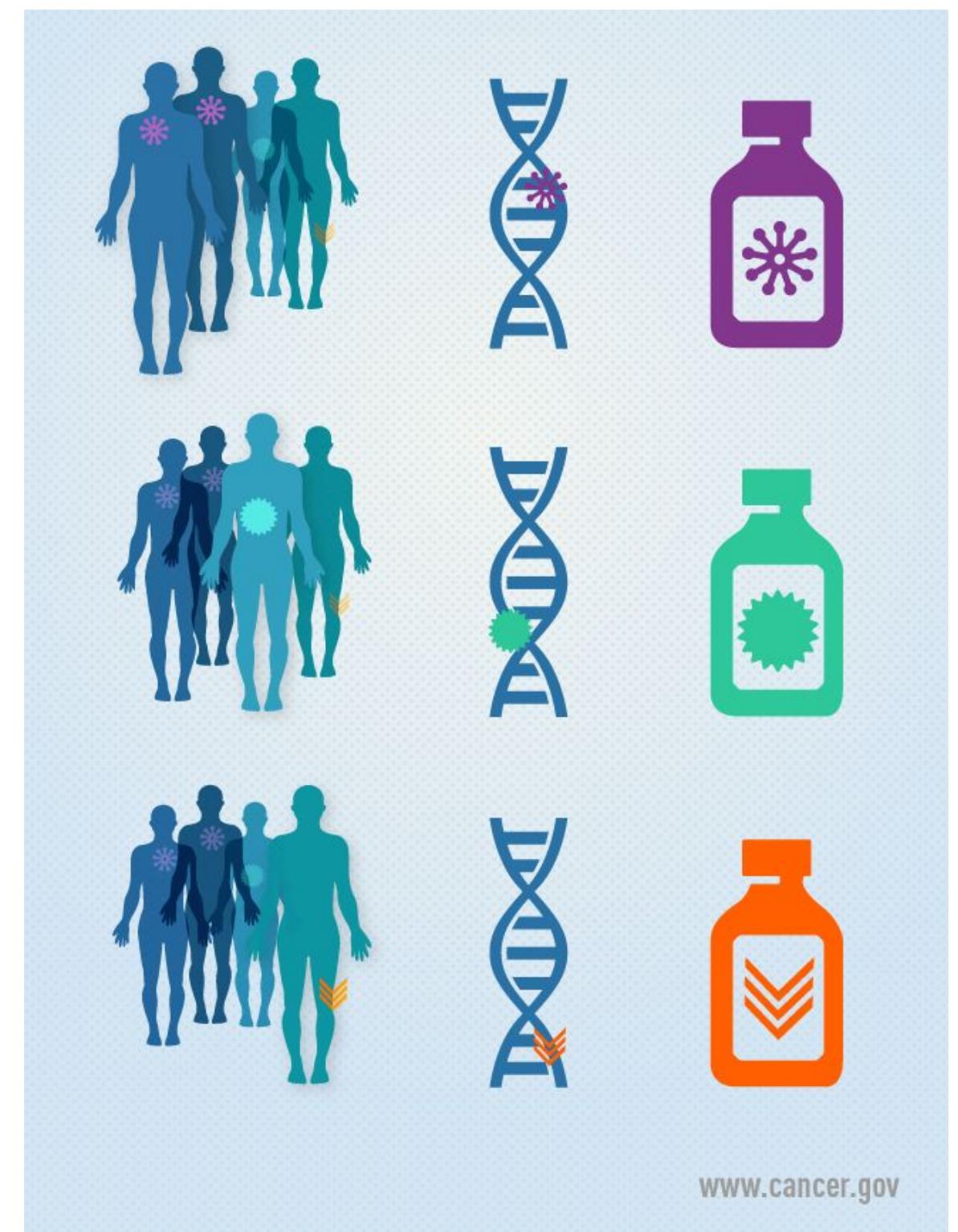
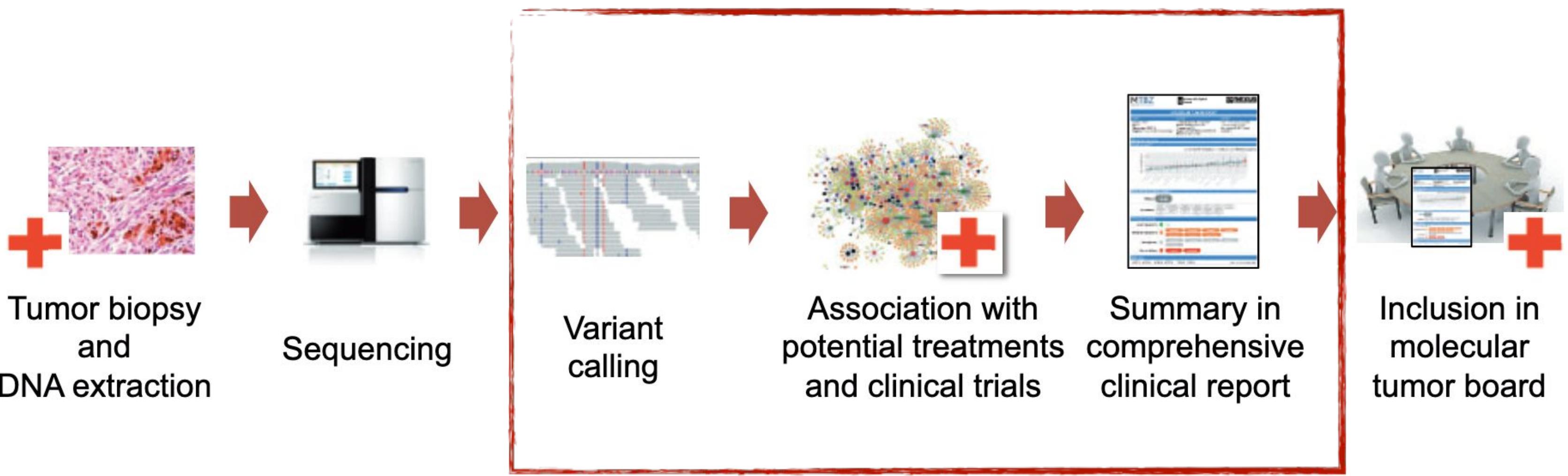
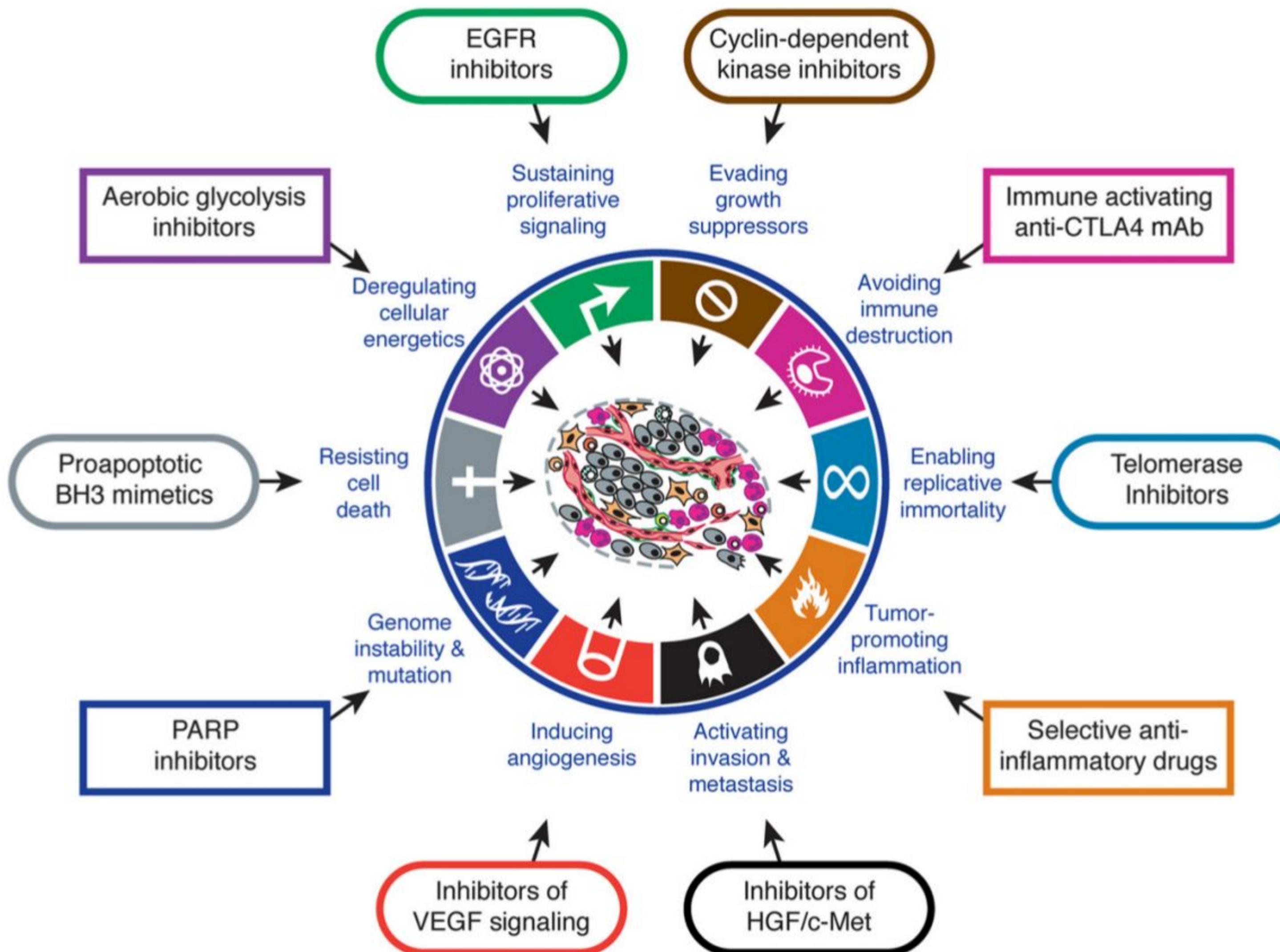


Image adapted from <https://www.cancer.gov/>

# Therapeutic targeting of the hallmarks of cancer



# Reporting summary

## MUTATION SUMMARY

Mutational burden: 833 SNVs (292 non-synonymous), 17 CNVs (affecting 2028 genes)

High mutational burden indicates potential benefit of anti-CTLA4 therapy

HLA Type: A\*03:02, A\*03:02, B\*07:01, B\*08:01, C\*07:02, C\*07:02

Somatic mutations in melanoma-typical genes (based on mycancergenome.org and customer requests):

BRAF: p.Val600Glu    CDKN2a/b: Copy number loss x1

Melanoma-typical genes without somatic mutations (based on mycancergenome.org and customer requests):

CTNNB1, CDK4/6, EGFR, GNA11, GNAQ, KIT, MEK1, MET, NF1, NRAS, PIK3CA, PTEN

## CANCER TYPE-SPECIFIC THERAPIES

Gene	Treatment	Mutation	Variant Frequency / Copy number	Pathway	Confidence	References
BRAF	Dabrafenib, Vemurafenib	c.1799T>A p.Val600Glu	38.1%	MAPK signaling	I.A*	2,3,4,5

## NON CANCER TYPE-SPECIFIC THERAPIES

Gene	Treatment	Mutation	Variant Frequency / Copy number	Pathway	Confidence	References
BRCA2	Olaparib	Copy number loss	x1 (hemizygous deletion)	DNA repair	I.C	7,12,13,14, 26, 27
SRC	Dasatinib, Bosutinib, Ponatinib	Copy number gain	x3	Growth signal transduction	I.B	18,19, 20,25

## INVESTIGATIONAL THERAPIES

Gene	Treatment	Mutation	Variant Frequency / Copy number	Pathway	Confidence	References
CDKN2A	PD0332991 (Palbociclib)	Copy number loss	x1 (heterozygous deletion)	Cell cycle	I.B*	6,7,8, 9,10,11

## THERAPIES WITH POTENTIAL LACK OF BENEFIT

Therapy	Responsible Mutation	Description	References
Sunitinib, Sorafenib	FLT3 Copy loss	Inhibitor treatment not advisable for copy loss. In addition, absence of cKIT mutation	21,22,23
Regorafenib	FRK Copy loss	Inhibitor treatment not advisable for copy loss.	24

## COMPREHENSIVE LIST OF MUTATIONS

See Appendix for comprehensive overview and details of all 833 SNVs and 17 CNVs.

# Association of genomic alterations with potential treatments and clinical trials

**DGIdb**

The previous version of DGIdb can be found at [old.dgide.org](http://old.dgide.org) until June 1st, 2024.

[About](#) [Publications](#) [Types/Directionalities](#) [Interaction Score](#) [FAQ](#) [Known Data Clients](#) [Contact](#) [Current Contributors](#) [Acknowledgements](#)

## About

### Mining the Druggable Genome for Personalized Medicine

Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Freshour S\*, Kiwala S\*, Cotto KC\*, Coffman AC, McMichael JF, Song J, Griffith M, Griffith OL, Wagner AH. Nucleic Acids Research. 2020 Nov 25; doi: <https://doi.org/10.1093/nar/gkaa1084>. PMID: 33237278

In the era of clinical sequencing and personalized medicine, investigators are frequently presented with lists of mutated or otherwise altered genes implicated in disease for a specific patient or cohort. Numerous resources exist to help form hypotheses about how such genomic events might be targeted therapeutically. However, utilizing these resources typically involves tedious manual review of literature, clinical trial records, and knowledgebases. Few currently exist which collect and curate these resources and provide a simple interface for searching lists of genes against the existing compendia of known or potential drug-gene interactions. The drug-gene interaction database (DGIdb) attempts to address this challenge. Using a combination of expert curation and text-mining, drug-gene interactions have been mined from DrugBank, PharmGKB, ChEMBL, Drug Target Commons, and others. Genes have also been categorized as potentially druggable according to membership in selected pathways, molecular functions and gene families from the

**Oncokb™** Levels of Evidence Actionable Genes Oncology Therapies Cancer Genes API / License About News FAQ 🔍

## Welcome to Oncokb™

MSK's Precision Oncology Knowledge Base

An FDA-Recognized Human Genetic Variant Database\*

858

Genes

7729

Alterations

137

Cancer Types

138

Drugs

Search Gene / Alteration / Cancer Type / Drug / Genomic Variant

Therapeutic Levels

Diagnostic Levels

Prognostic Levels

FDA Levels

① Level 1  
FDA-approved drugs  
53 Genes

② Level 2  
Standard care  
26 Genes

③ Level 3  
Clinical evidence  
33 Genes

④ Level 4  
Biological evidence  
27 Genes

Level R1/R2  
Resistance  
11 Genes

**CANCER GENOME INTERPRETER**

HOME ANALYSIS ABOUT FAQ CONTACT Login

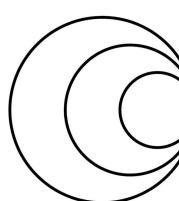
[Cancer Biomarkers](#) [Validated Oncogenic Mutations](#) [Cancer Genes](#) [Cancer Bioactivities](#)

**Cancer Biomarkers database** Last update: 2022/10/17

The Cancer Biomarkers database is curated and maintained by [several clinical and scientific experts](#) in the field of precision oncology supported by the European Union's Horizon 2020 funded [project](#). This database is currently being integrated with knowledge databases of other institutions in a [collaborative effort](#) of the [Global Alliance for Genomics and Health](#). The contribution of the community is encouraged and proposals of edition or comments about the information contained in this database can be given by contacting us [here](#) or by using the feedback icon located at the left of each entry of the table. The database follows the data model originally described by [Dienstmann et al.](#) This table provides a summary of the content of the database that can be interactively browsed. Additional information, including the genomic coordinates of the variants, can be accessed via the download feature. This database is licensed under a [Creative Commons Public Domain Dedication \(CC0 1.0 Universal\)](#). When referring to this database, please cite: Cancer Genome Interpreter Annotates The Biological And Clinical Relevance Of Tumor Alterations; doi: <https://doi.org/10.1101/140475>.

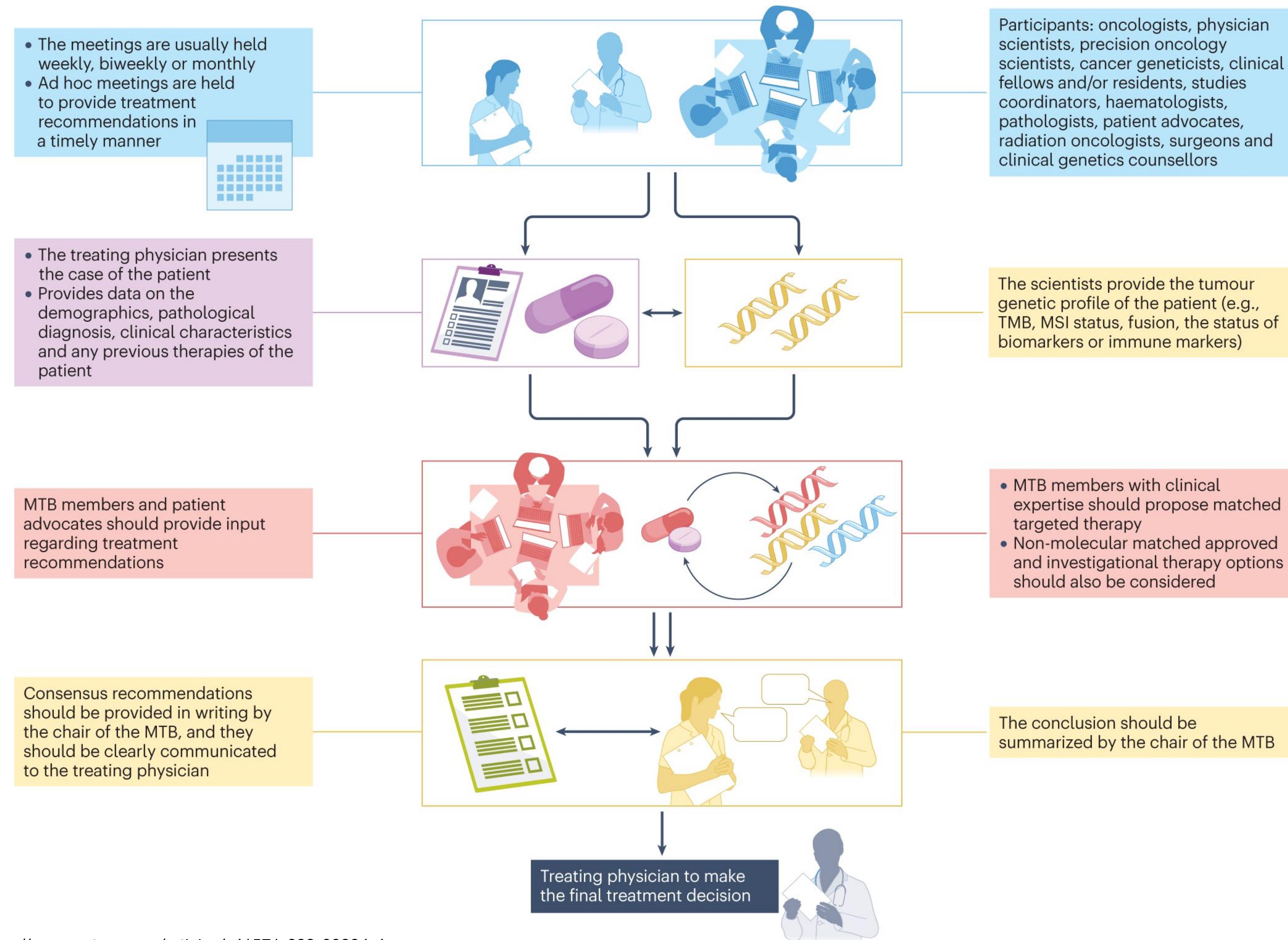
Download Feedback

Biomarker	Drug	Effect	Evidence
ABL1 (E255V,Y253H,F359V)	Ponatinib (BCR-ABL inhibitor 3rd gen&PA)	Responsive	NCCN guidelines
ABL1 (F359V,F359C,F359I,Y)	Dasatinib (BCR-ABL inhibitor 2nd gen)	Responsive	NCCN guidelines
ABL1 (I242T,M244V,K247R)	Imatinib (BCR-ABL inhibitor 1st gen&KIT)	Resistant	European Leukemi
ABL1 (T315A,F317L,F317V,I)	Bosutinib (BCR-ABL inhibitor 3rd gen)	Responsive	NCCN guidelines
ABL1 (T315A,F317L,F317V,I)	Nilotinib (BCR-ABL inhibitor 2nd gen)	Responsive	NCCN guidelines
ABL1 (T315I)	Asciminib (Kinase inhibitor)	Responsive	FDA guidelines
ABL1 (T315I)	Nilotinib (BCR-ABL inhibitor 2nd gen)	Resistant	NCCN guidelines
ABL1 (T315I)	Bosutinib (BCR-ABL inhibitor 3rd gen)	Resistant	NCCN guidelines
ABL1 (T315I)	Axitinib (VEGFR inhibitor)	Responsive	Pre-clinical

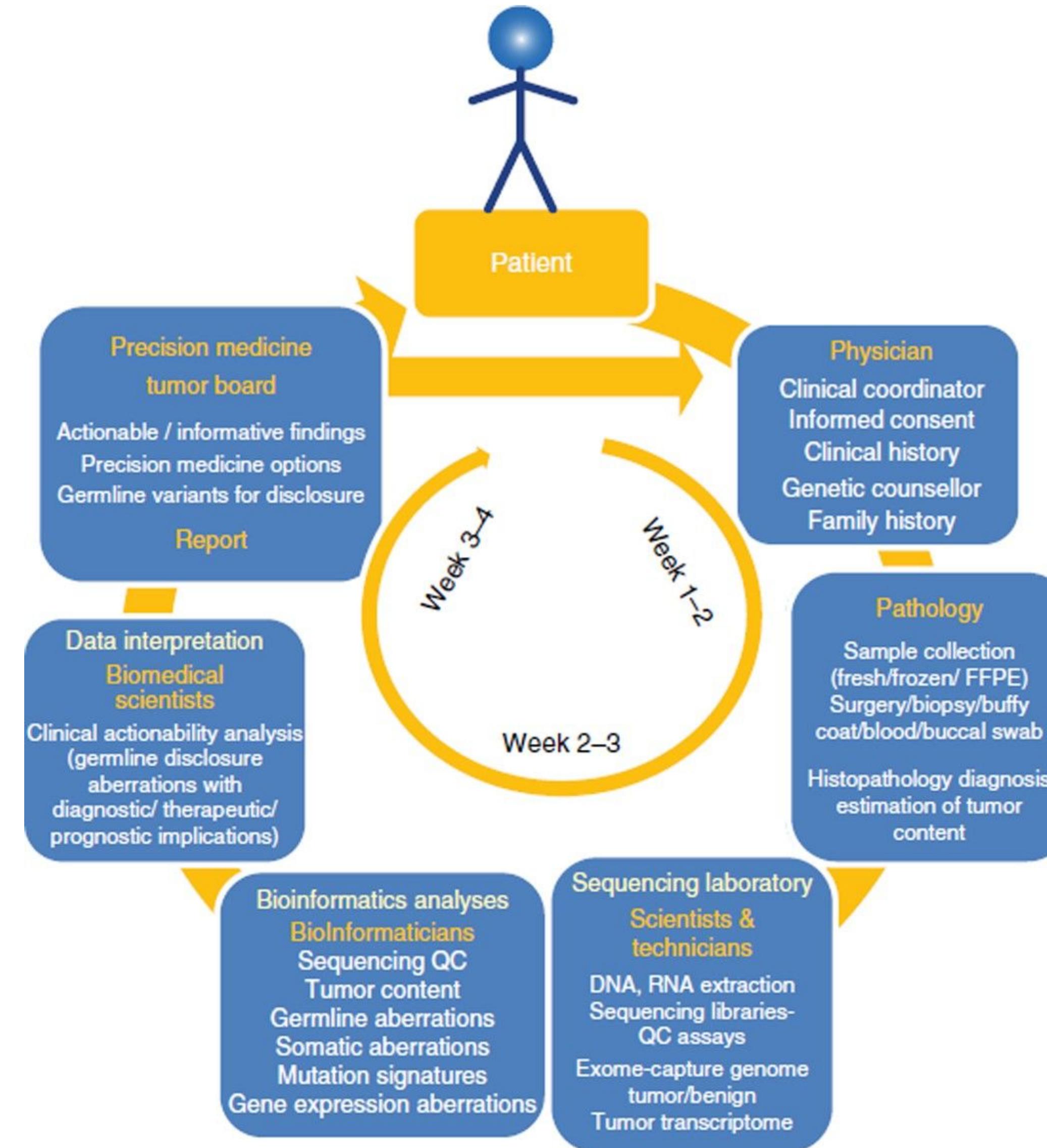


wellcome  
connecting  
science

# Molecular tumour board (MTB)

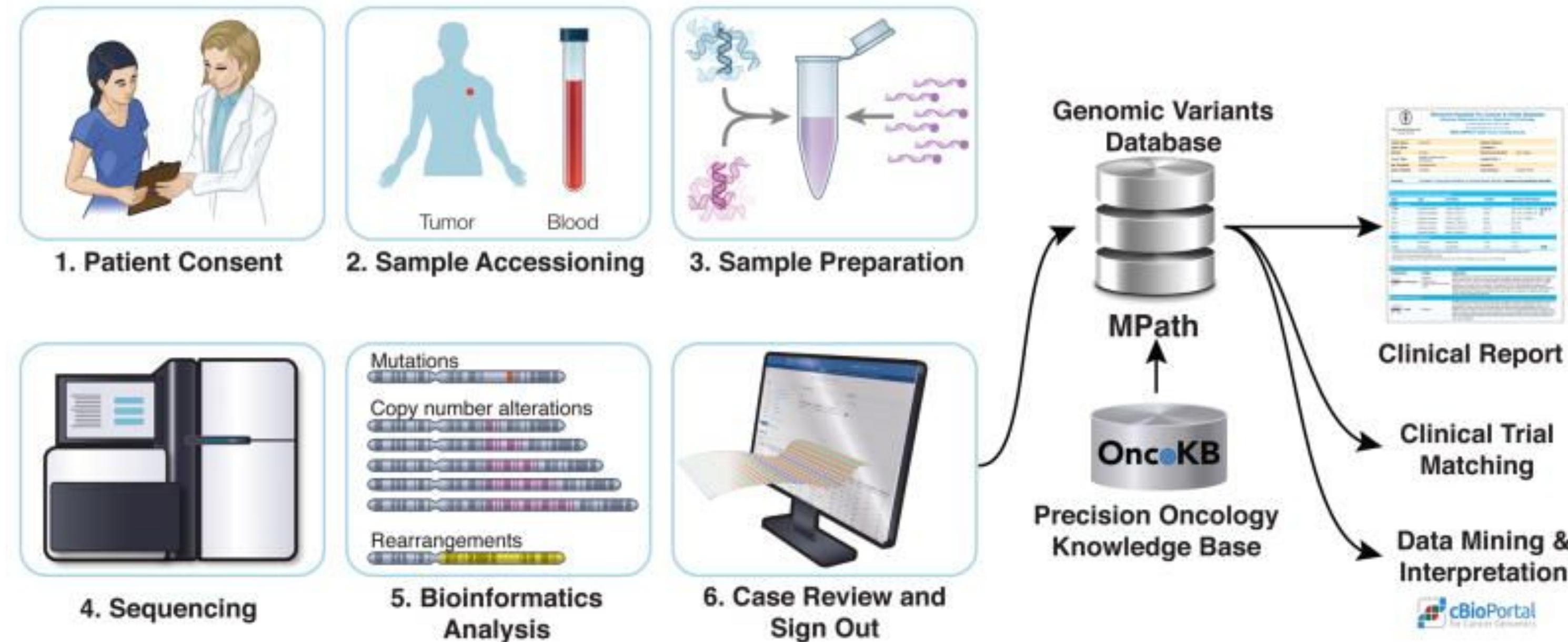


# Workflow of integrative clinical sequencing for precision oncology



University of Michigan (Ann Arbor, MI,  
USA)

# MSK-IMPACT: Integrated Mutation Profiling of Actionable Cancer Targets

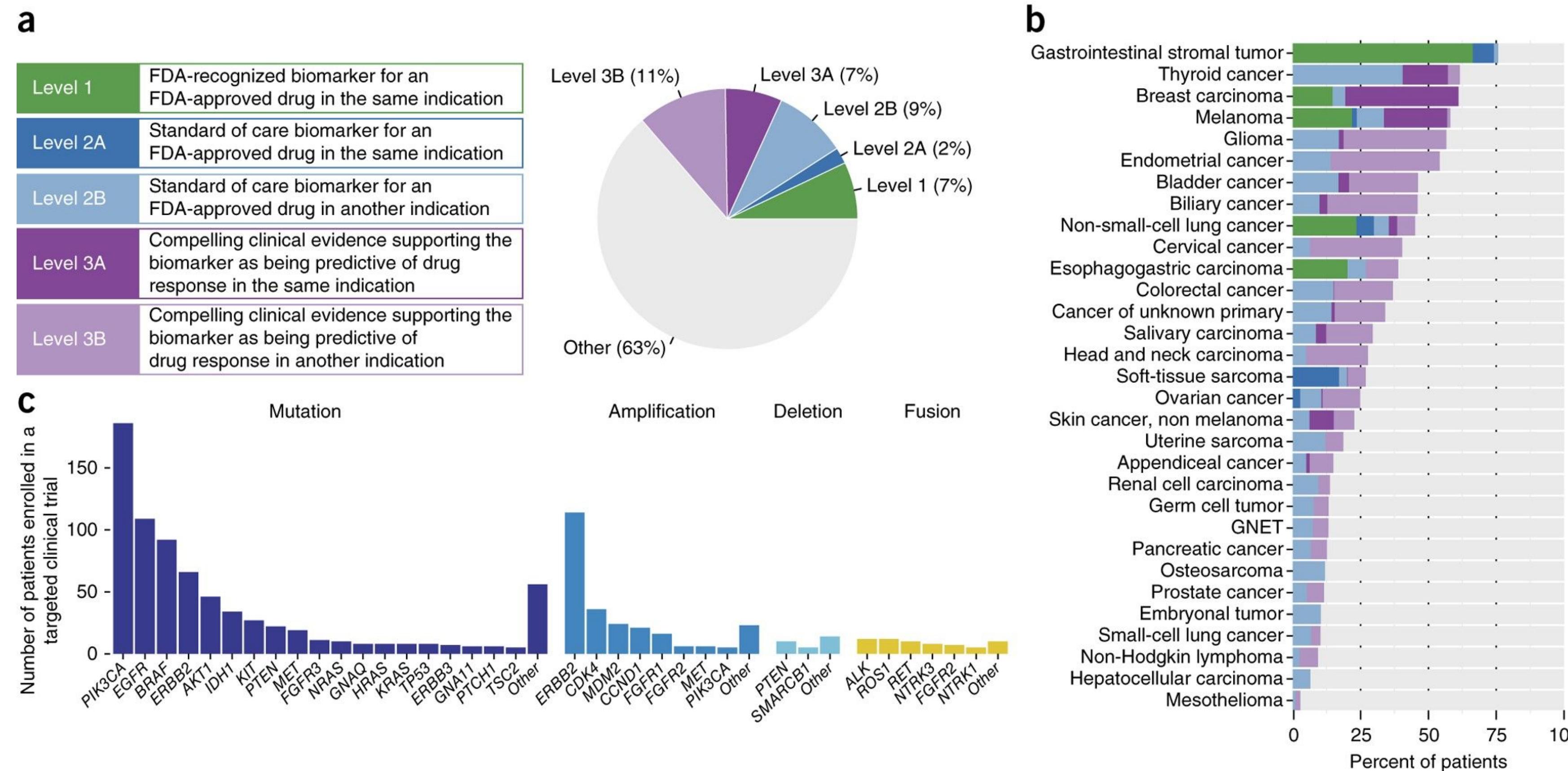


January 2014 and May 2016: **12,670 advanced tumours from 11,369 patients**

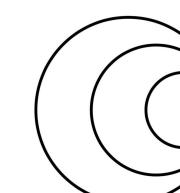
Gene panel covering: ~400 cancer related genes

median turnaround time of <21 days

# Clinical actionability of somatic alterations revealed by MSK-IMPACT



36.7% of patients (n=3,792) harbored at least one actionable alteration

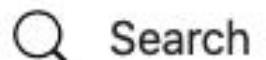


wellcome  
connecting  
science

# Targeted panels to whole-genome sequencing

nature medicine

[View all journals](#)



[Search](#)

[Log in](#)

[Explore content](#) ▾

[About the journal](#) ▾

[Publish with us](#) ▾

[Sign up for alerts](#)

[RSS feed](#)

[nature](#) > [nature medicine](#) > [analyses](#) > [article](#)

Analysis | [Open access](#) | Published: 11 January 2024

## Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme

[Alona Sosinsky](#), [John Ambrose](#), [William Cross](#), [Clare Turnbull](#), [Shirley Henderson](#), [Louise Jones](#),  
[Angela Hamblin](#), [Prabhu Arumugam](#), [Georgia Chan](#), [Daniel Chubb](#), [Boris Noyvert](#), [Jonathan Mitchell](#), [Susan Walker](#), [Katy Bowman](#), [Dorota Pasko](#), [Marianna Buongermino Pereira](#), [Nadezda Volkova](#), [Antonio Rueda-Martin](#), [Daniel Perez-Gil](#), [Javier Lopez](#), [John Pullinger](#), [Afshan Siddiq](#), [Tala Zainy](#), [Tasnim Choudhury](#), ... [Nirupa Murugaesu](#) [+ Show authors](#)

[Download PDF](#)



### Associated content

[The grand challenge of moving cancer whole-genome sequencing into the clinic](#)

Dilara Akhounova & Mark A. Rubin

[Nature Medicine](#) | [News & Views](#) | 11 Jan 2024

[Sections](#)

[Figures](#)

[References](#)

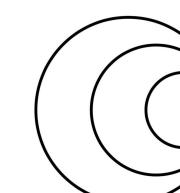
[Abstract](#)

[Main](#)

[Results](#)

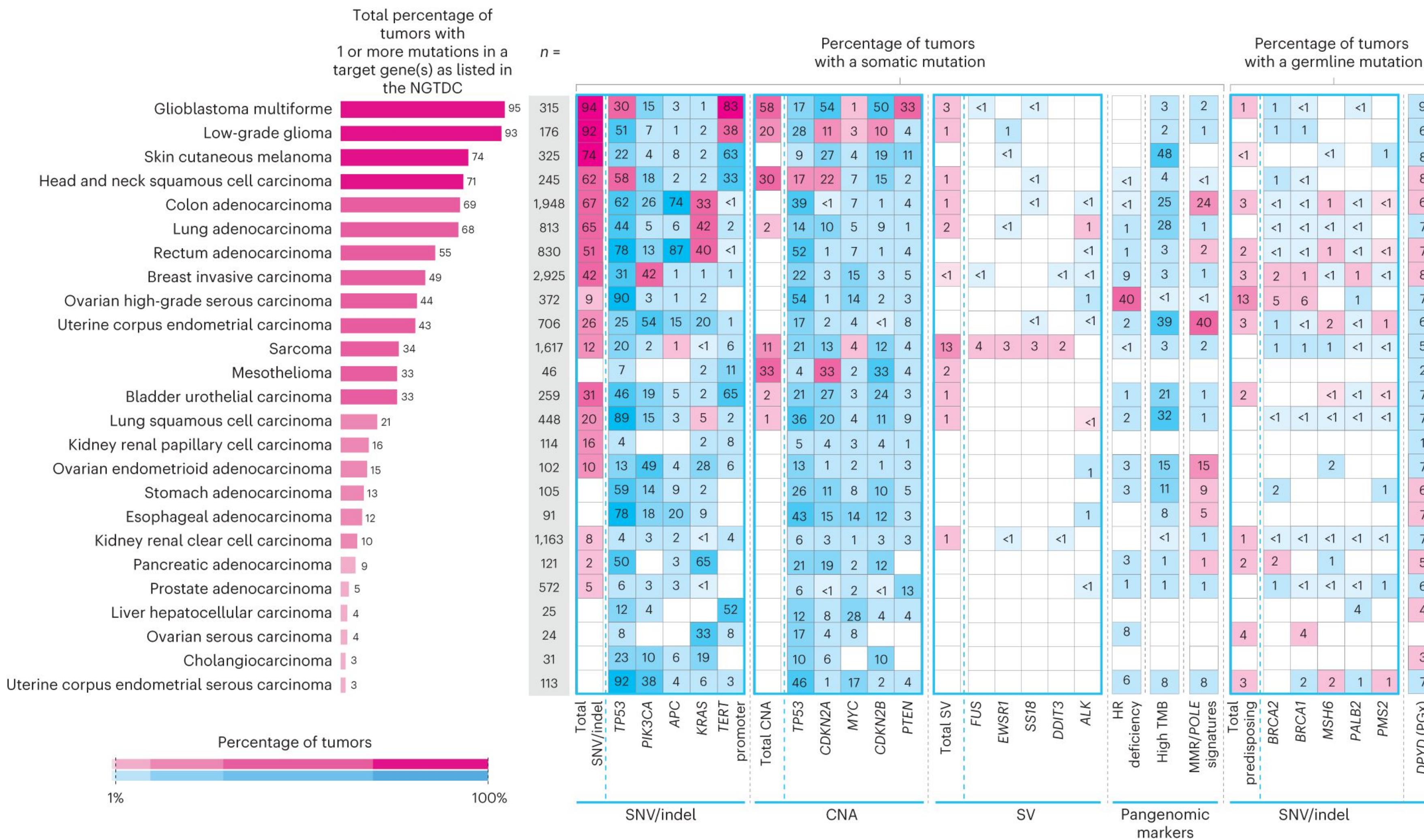
[Nature Medicine](#) **30**, 279–289 (2024) | [Cite this article](#)

**54k** Accesses | **7** Citations | **1590** Altmetric | [Metrics](#)



wellcome  
connecting  
science

# Prevalence of different types of mutations identified using WGS in genes indicated for testing in the NGTDC



“We note that the clinical actionability of these mutations will be dependent on the individual case and clinical circumstances, such as the stage of the tumor and associated comorbidities of the participant. This highlights the need for clinical interpretation and discussion where clinically appropriate within a GTAB.”

NGTDC: National Genomic Test Directory for Cancer

GTAB: Genomic Tumour Advisory Board

# Cancer Gene panel testing in India

**strand**

Diagnostics Genomic Wellness More ▾ | [✉](#) [📞](#) [📍](#)

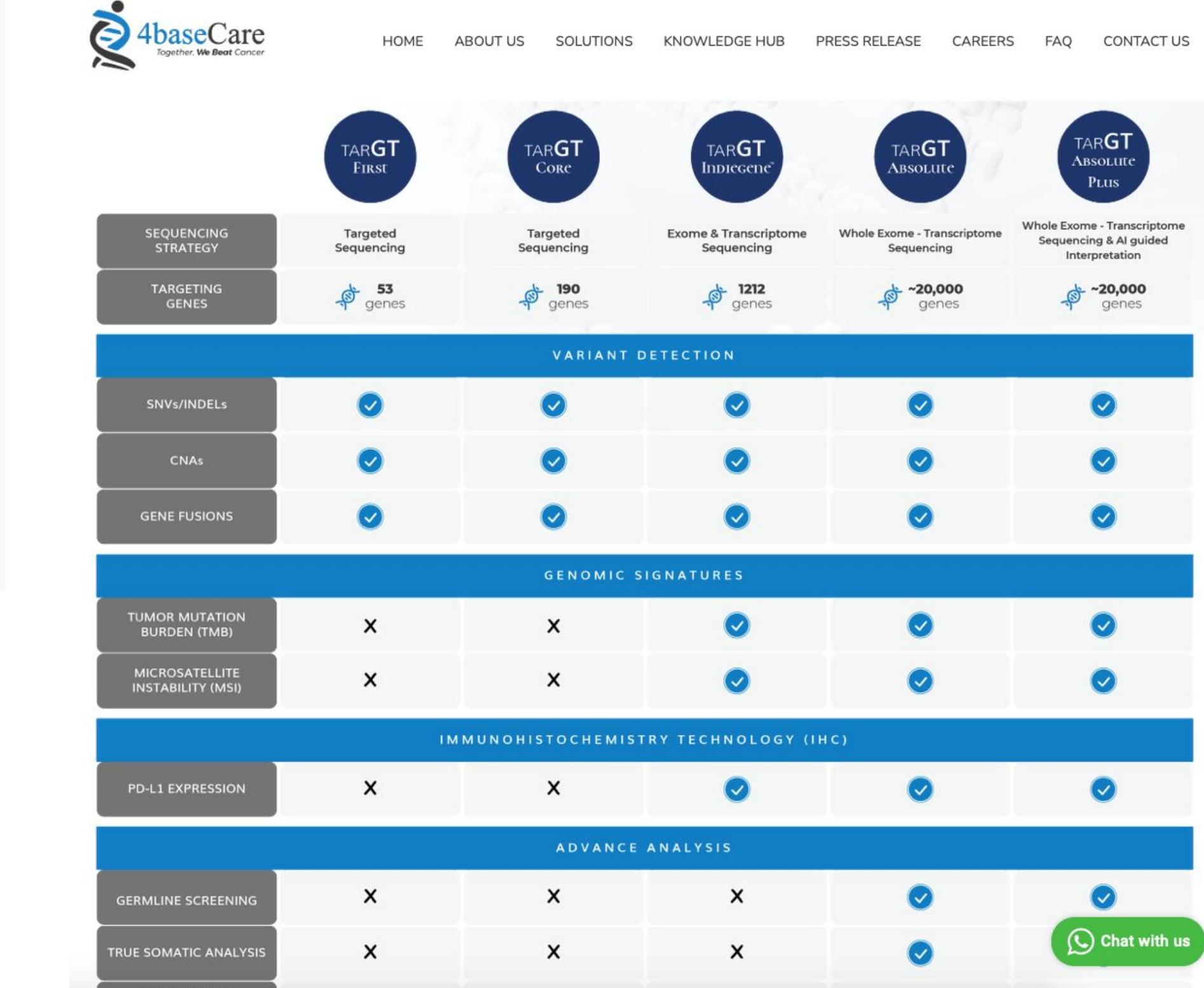
## Explore our Test Catalog

[Search by disorder](#) [Search by test/gene](#)

🔍

Test Name	No.of genes	Select
Germline Panel – Prostrate Cancer	12 ↗	+ <span style="color: #0056b3;">+/-</span>
Germline Panel – Colorectal Cancer	14 ↗	+ <span style="color: #0056b3;">+/-</span>
Germline Panel – Endocrine Cancer	10 ↗	+ <span style="color: #0056b3;">+/-</span>
Germline Panel – Gynaecological Cancer	12 ↗	+ <span style="color: #0056b3;">+/-</span>
Germline Panel – Nervous System Cancer	16 ↗	+ <span style="color: #0056b3;">+/-</span>
Germline Panel – Pancreatic Cancer	14 ↗	+ <span style="color: #0056b3;">+/-</span>

<https://strandls.com/testcatalog/>



The chart compares five cancer gene panel tests from 4baseCare based on various sequencing strategies, variant detection, genomic signatures, IHC technology, and advanced analysis.

SEQUENCING STRATEGY					
TARGETING GENES	Targeted Sequencing	Targeted Sequencing	Exome & Transcriptome Sequencing	Whole Exome - Transcriptome Sequencing	Whole Exome - Transcriptome Sequencing & AI guided Interpretation
53 genes	53 genes	190 genes	1212 genes	~20,000 genes	~20,000 genes
VARIANT DETECTION					
SNVs/INDELs	✓	✓	✓	✓	✓
CNAs	✓	✓	✓	✓	✓
GENE FUSIONS	✓	✓	✓	✓	✓
GENOMIC SIGNATURES					
TUMOR MUTATION BURDEN (TMB)	✗	✗	✓	✓	✓
MICROSATELLITE INSTABILITY (MSI)	✗	✗	✓	✓	✓
IMMUNOHISTOCHEMISTRY TECHNOLOGY (IHC)					
PD-L1 EXPRESSION	✗	✗	✓	✓	✓
ADVANCE ANALYSIS					
GERMLINE SCREENING	✗	✗	✗	✓	✓
TRUE SOMATIC ANALYSIS	✗	✗	✗	✓	✓
AU/ML BASED					

<https://4basecare.com/solutions/genomics/somatic-tests/comparison-of-our-tests/>

# Strand - Germline panel

Original Article | Published: 25 February 2016

## Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: implications of embracing a multi-gene panel in molecular diagnosis in India

[Ashraf U Mannan](#), [Jaya Singh](#), [Ravikiran Lakshmikeshava](#), [Nishita Thota](#), [Suhasini Singh](#), [T S Sowmya](#), [Avshesh Mishra](#), [Aditi Sinha](#), [Shivani Deshwal](#), [Megha R Soni](#), [Anbukayalvizhi Chandrasekar](#), [Bhargavi Ramesh](#), [Bharat Ramamurthy](#), [Shila Padhi](#), [Payal Manek](#), [Ravi Ramalingam](#), [Suman Kapoor](#), [Mithua Ghosh](#), [Satish Sankaran](#), [Arunabha Ghosh](#), [Vamsi Veeramachaneni](#), [Preveen Ramamoorthy](#), [Ramesh Hariharan](#) & [Kalyanasundaram Subramanian](#)

[Journal of Human Genetics](#) **61**, 515–522 (2016) | [Cite this article](#)

3194 Accesses | 45 Citations | 136 Altmetric | [Metrics](#)

- Sequenced 141 unrelated patients and families with BOC using the TruSight Cancer panel, which includes 13 genes strongly associated with risk of inherited BOC
- Detected pathogenic mutations in 51 (36.2%) cases, out of which 19 were novel mutations. When we considered familial breast cancer cases only, the detection rate increased to 52%.

# Targeted molecular profiling of solid tumours-Indian tertiary cancer centre experience

Research | Published: 20 March 2023

Volume 149, pages 7413–7425, (2023) [Cite this article](#)

Mamta Gurav, Sridhar Epari, Prachi Gogte, Trupti Pai, Gauri Deshpande, Nupur Karnik, Omshree Shetty

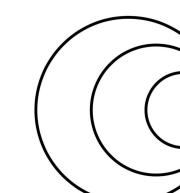
[✉](#) & Sangeeta Desai

Study included 1140 formalin Fixed paraffin embedded samples. NGS was performed using two targeted gene panels viz. Ampliseq Focus panel and Sophia Solid Tumor Plus Solution. Data was analyzed using Illumina's Local Run Manager and SOPHiA DDM software. Variant interpretation and annotations were done as per AMP/ACMG guidelines.

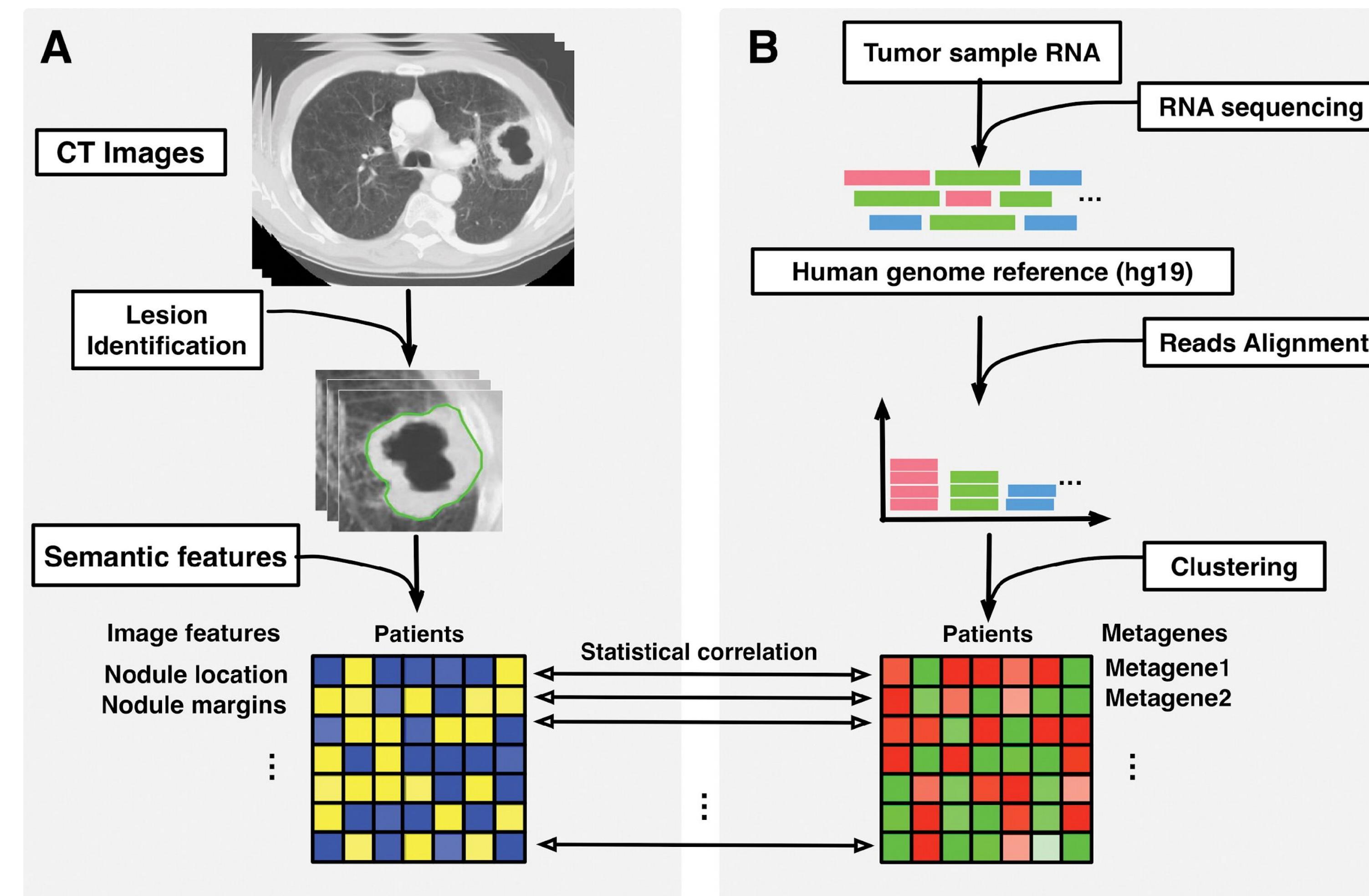
## Results

Total 896 cases were subjected to NGS after excluding cases with suboptimal nucleic acid quality/quantity. DNA alterations were detected in 64.9% and RNA fusions in 6.9% cases.

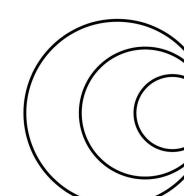
Among detected variants, 86.7% were clinically relevant aberrations. Mutation frequency among different solid tumours was 70.8%, 67.4%, 64.4% in non-small cell lung (NSCLC), lung squamous cell carcinomas and head neck tumours respectively. *EGFR*, *KRAS*, *BRAF*, *ALK* and *ROS1* were commonly altered in NSCLC. Gastrointestinal tumours showed mutations in 63.6% with predominant alterations in pancreatic (88.2%), GIST (87.5%), colorectal (78.7%), cholangiocarcinoma (52.9%), neuroendocrine (45.5%), gall bladder (36.7%) and gastric adenocarcinomas (16.7%). The key genes affected were *KRAS*, *NRAS*, *BRAF* and *PIK3CA*. NGS evaluation identified co-occurring alterations in 37.7% cases otherwise missed by conventional assays. Resistance mutations were detected in progressive lung tumours (39.5%) against *EGFR* TKIs and *ALK/ROS* inhibitors.



# Radiogenomics

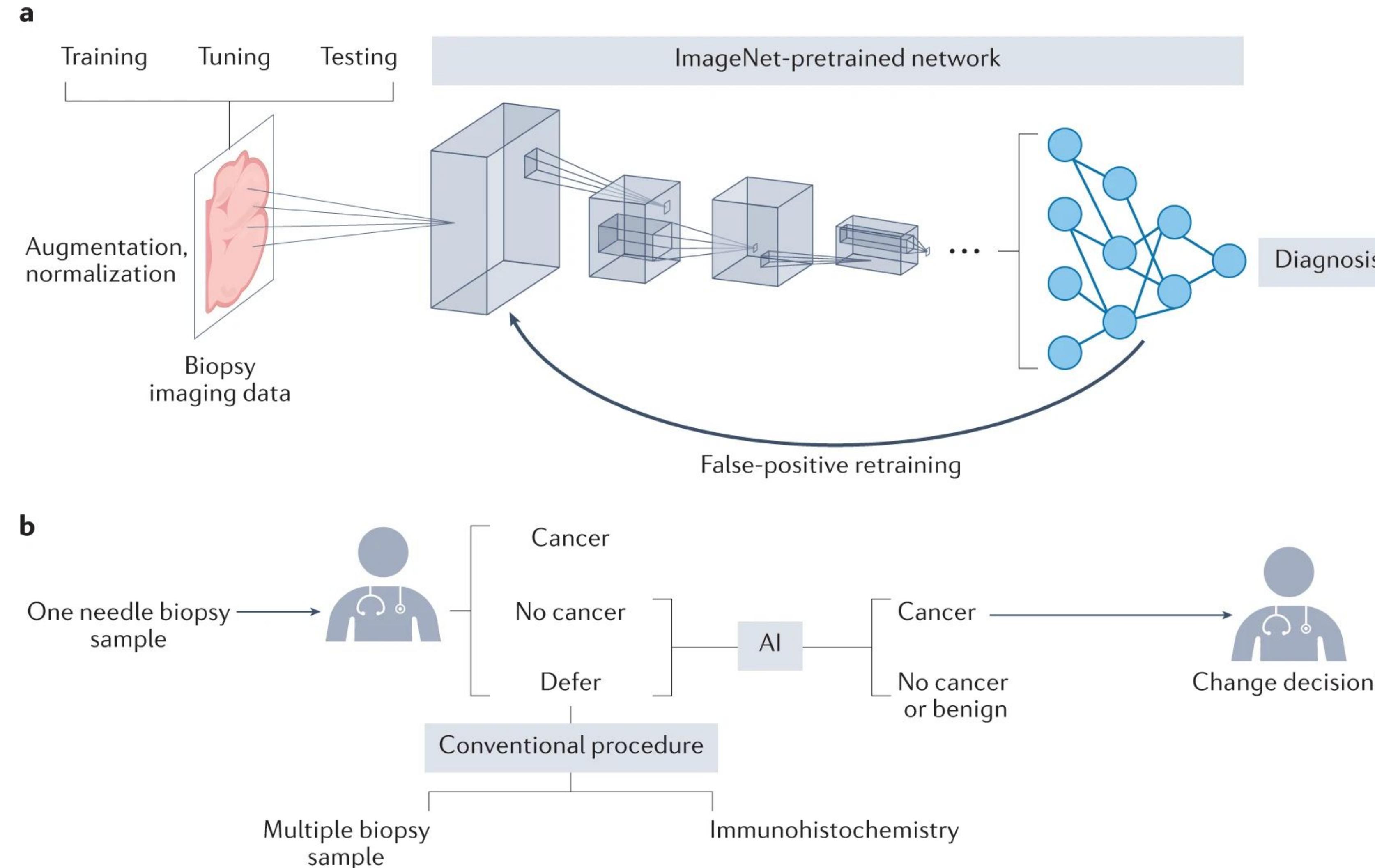


<https://pubs.rsna.org/doi/10.1148/radiol.2017161845>



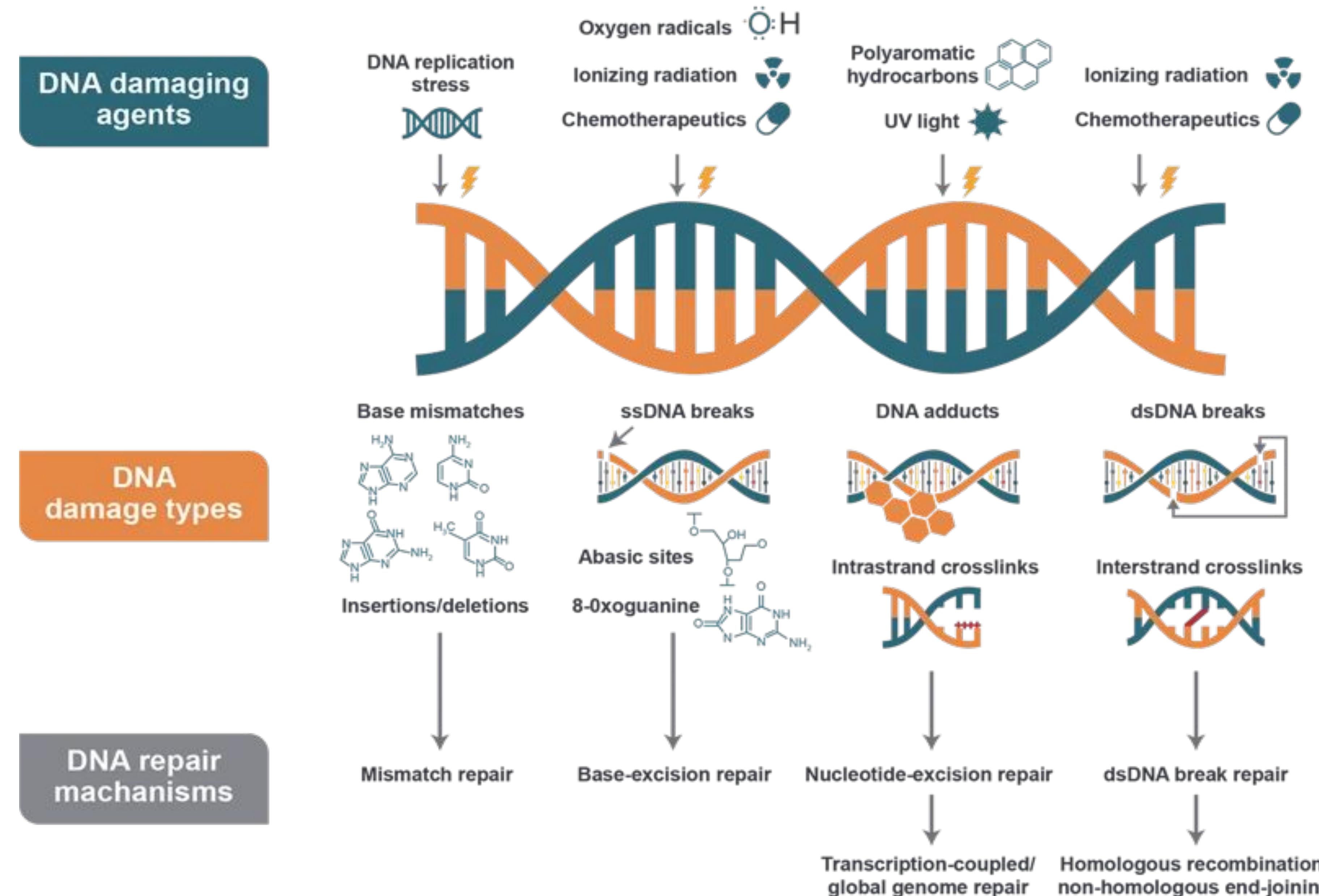
wellcome  
connecting  
science

# Data-driven artificial intelligence to support cancer diagnosis.

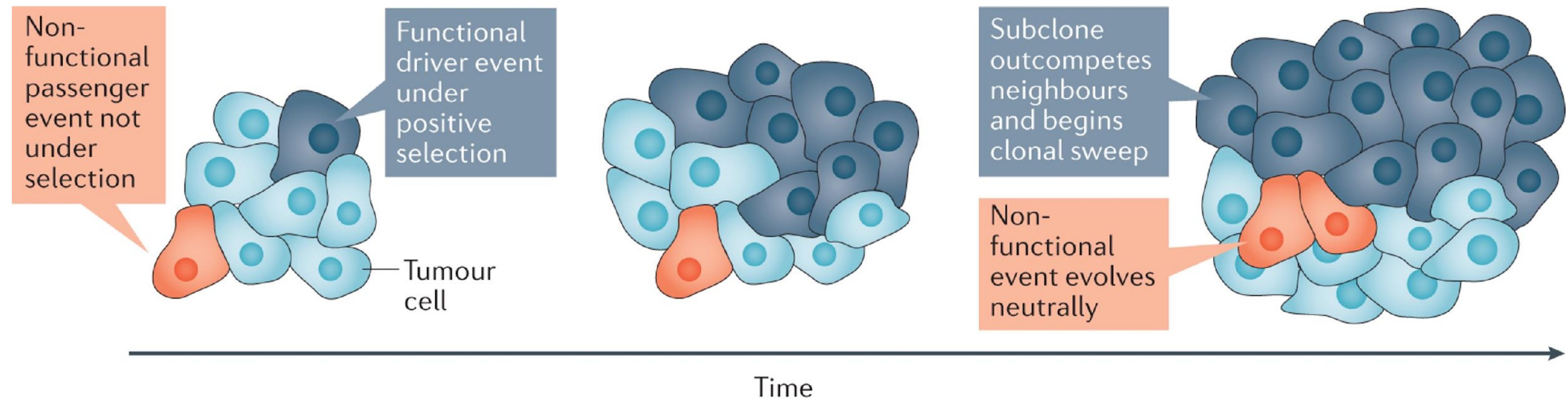


# Identification of cancer drivers

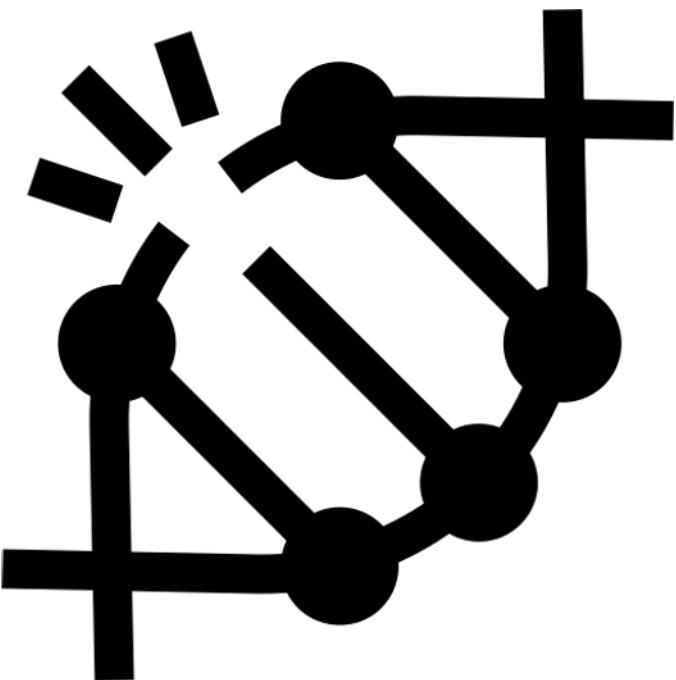
# Somatic mutations: Accumulation



# Clonal Evolution



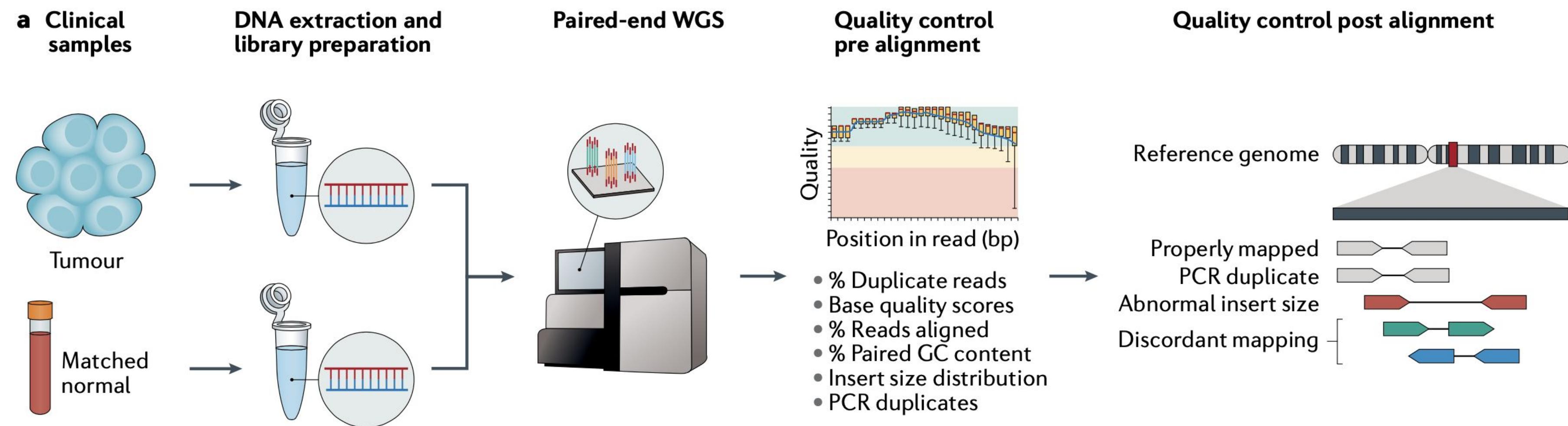
# Driver mutations



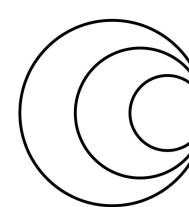
- Mutations in cancer can be classified into:
  - Driver mutations
  - Passenger mutations
- **Driver mutations**
  - Provide a selective advantage to the cell.
  - Promote cancer development
- Passenger mutations
  - Neutral mutations.



# Identification of genetic alterations in cancer genomes



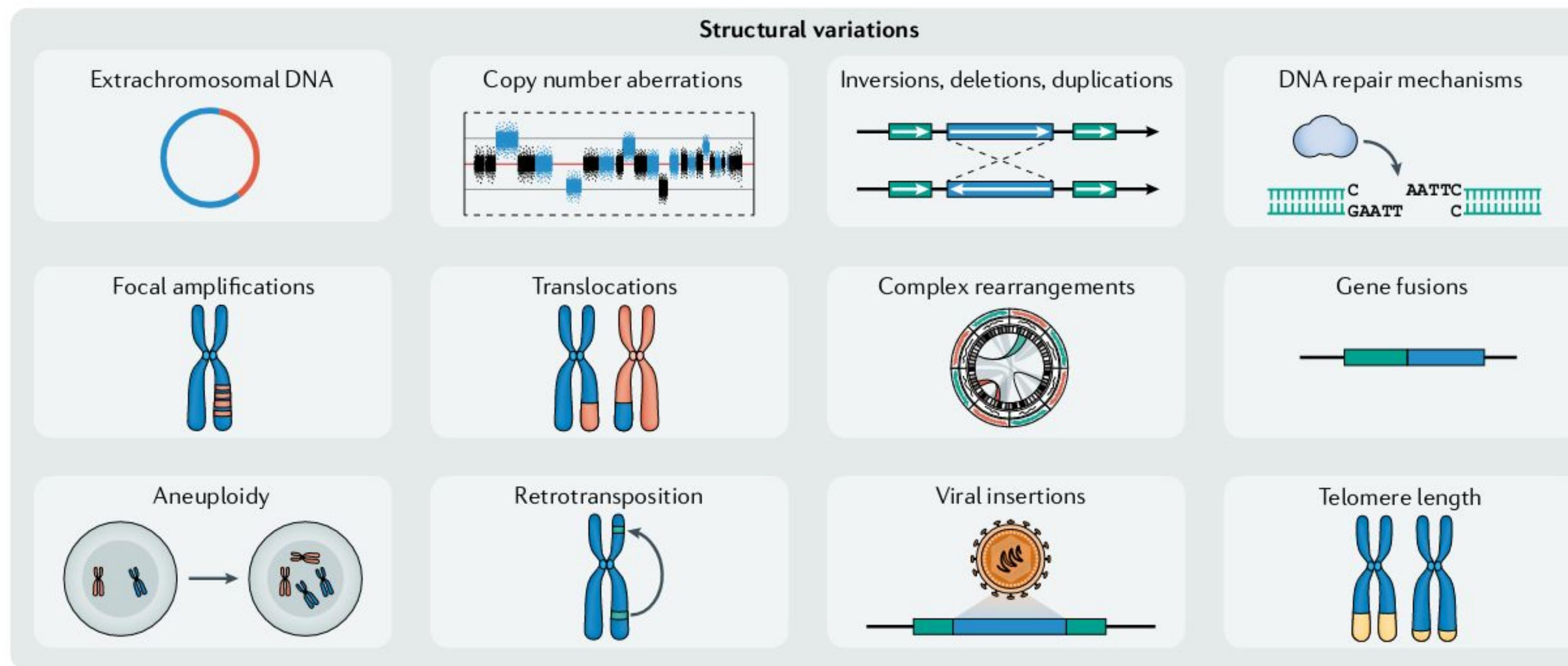
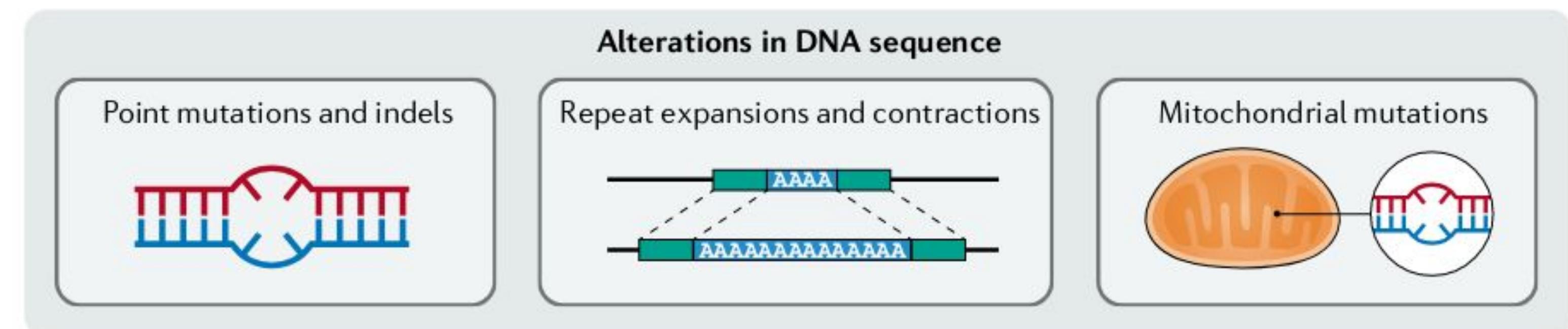
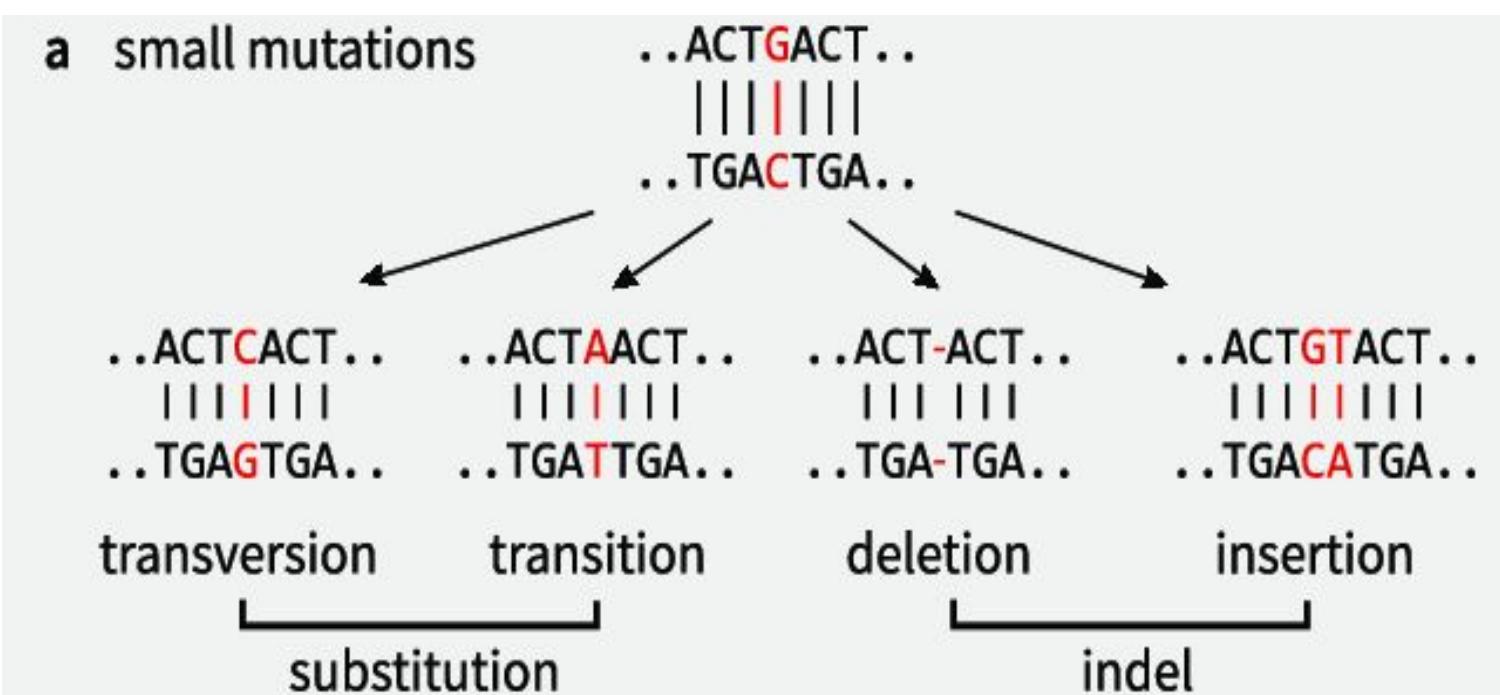
Nature Reviews Genetics 23, 298–314 (2022)



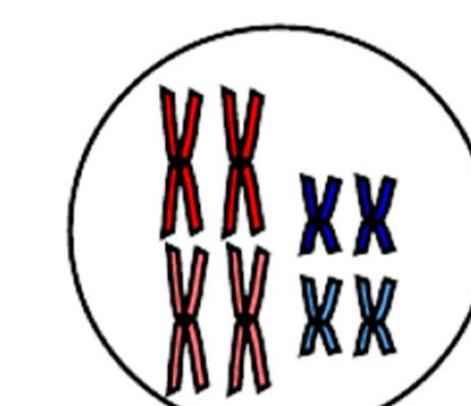
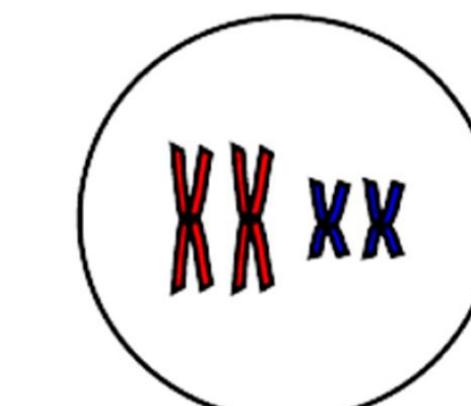
wellcome  
connecting  
science

# Identification of genetic alterations in cancer genomes

## Point mutations

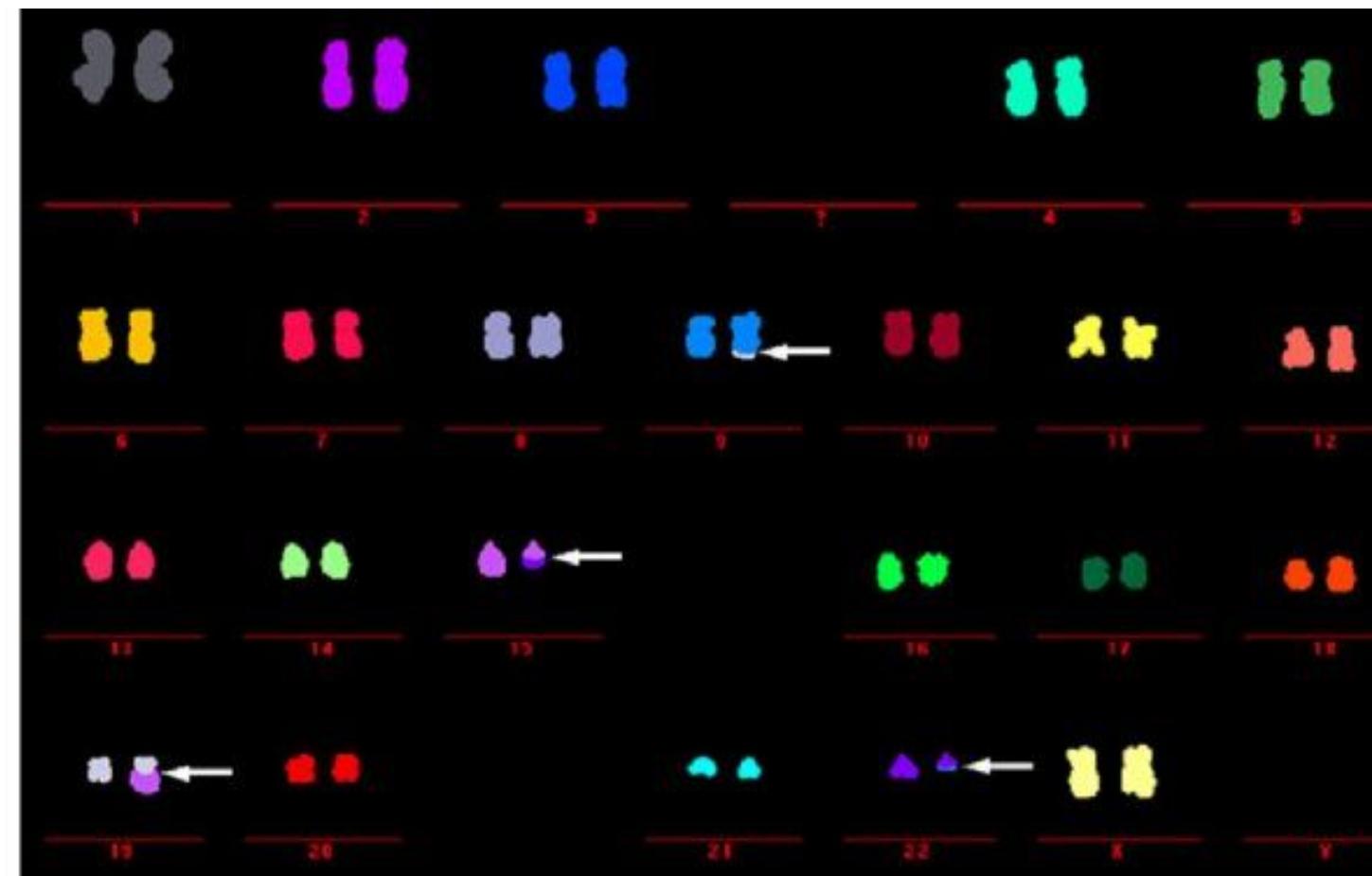


## Whole Genome Duplication



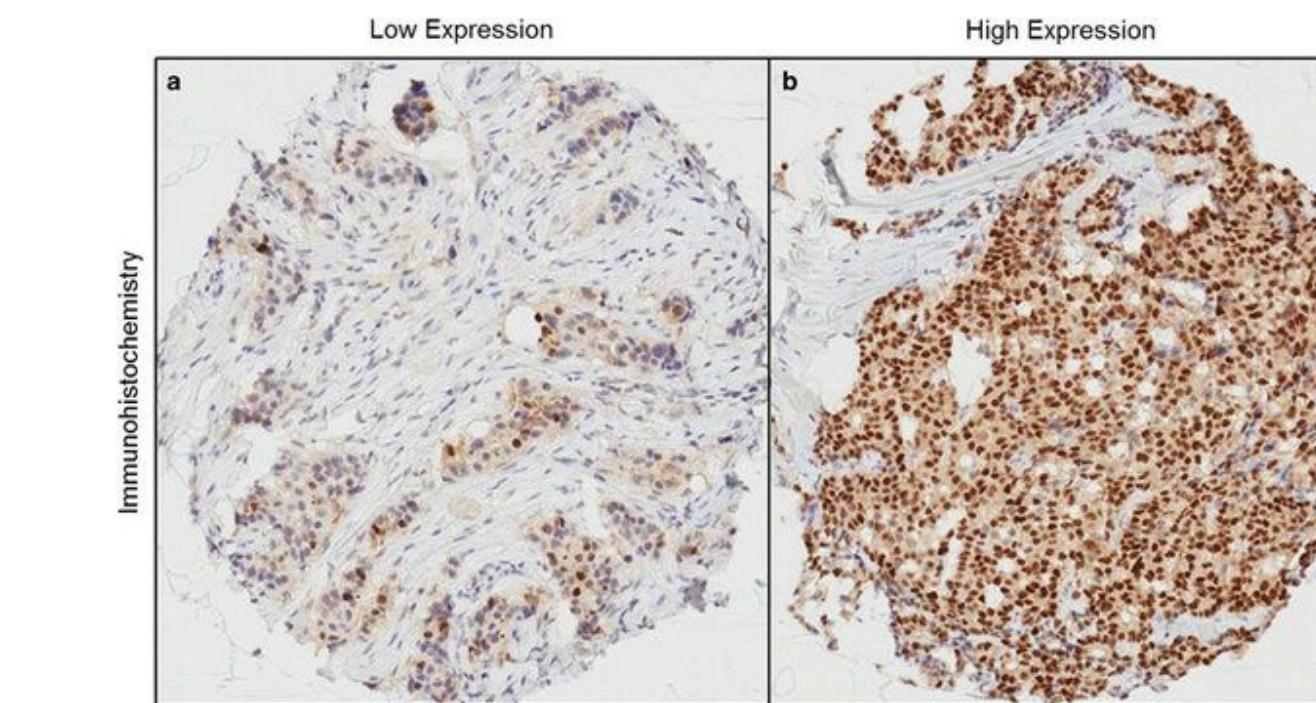
# Methods to identify genomic alterations in the cancer cells

## Karyotyping



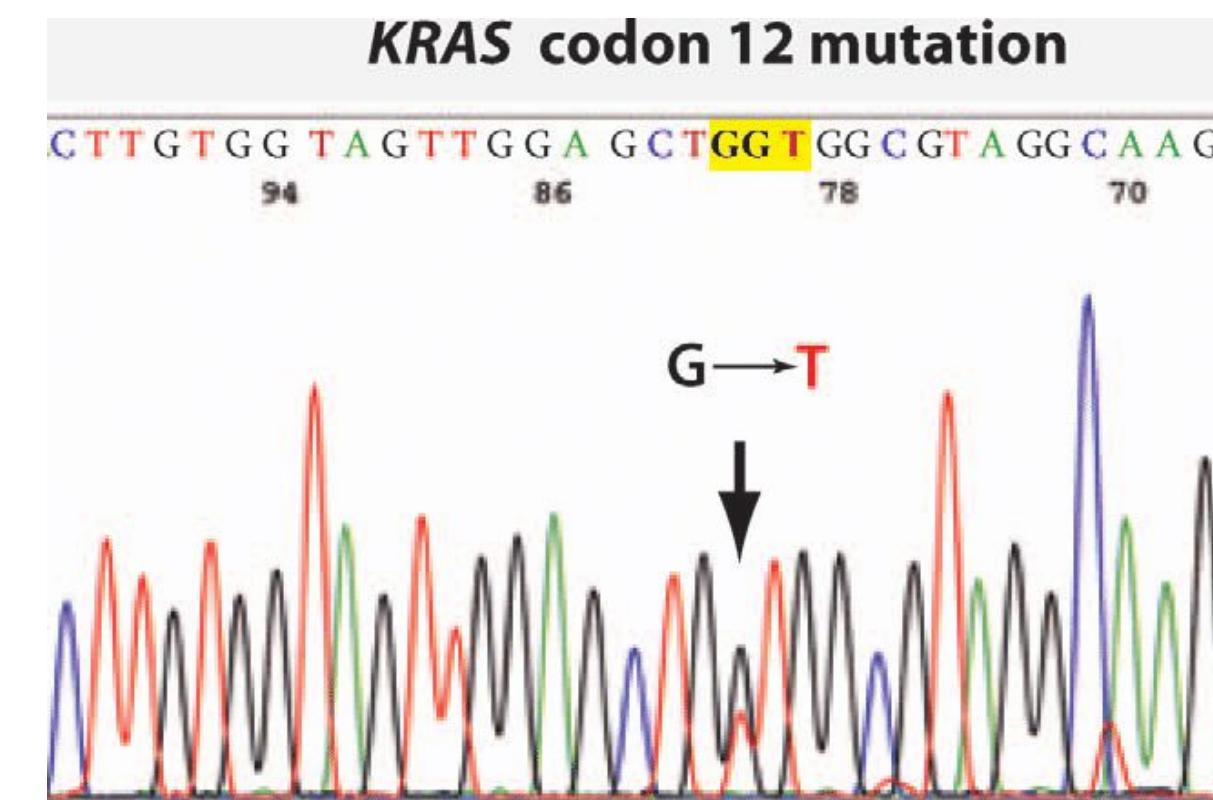
doi:10.1007/s12185-011-0769-z

## Immunohistochemistry (IHC)

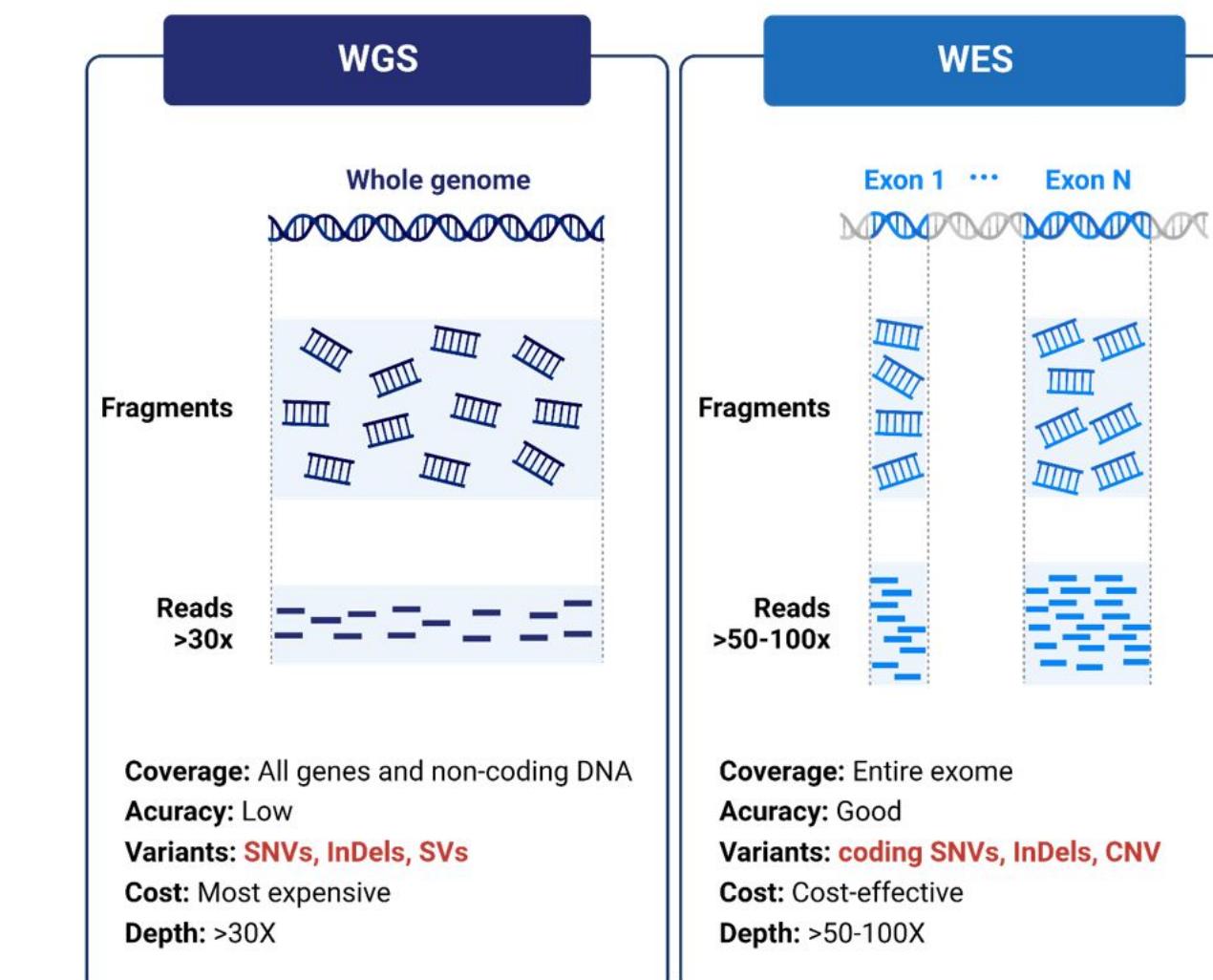


doi: 10.1038/labinvest.2016.73

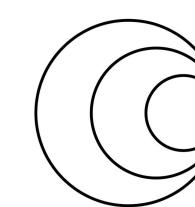
## Sanger sequencing



## NGS: high-throughput sequencing



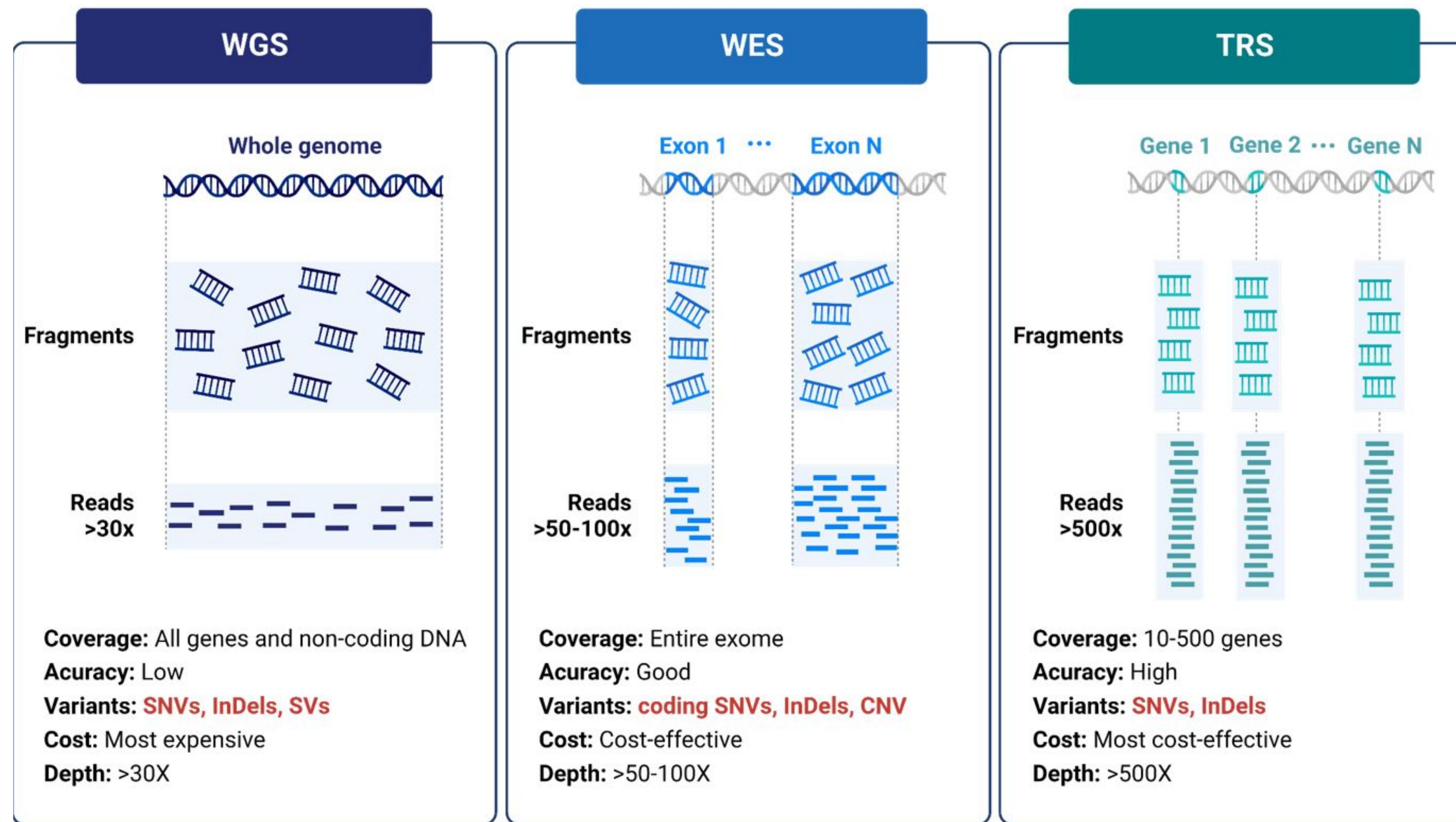
<https://www.novogene.com/>



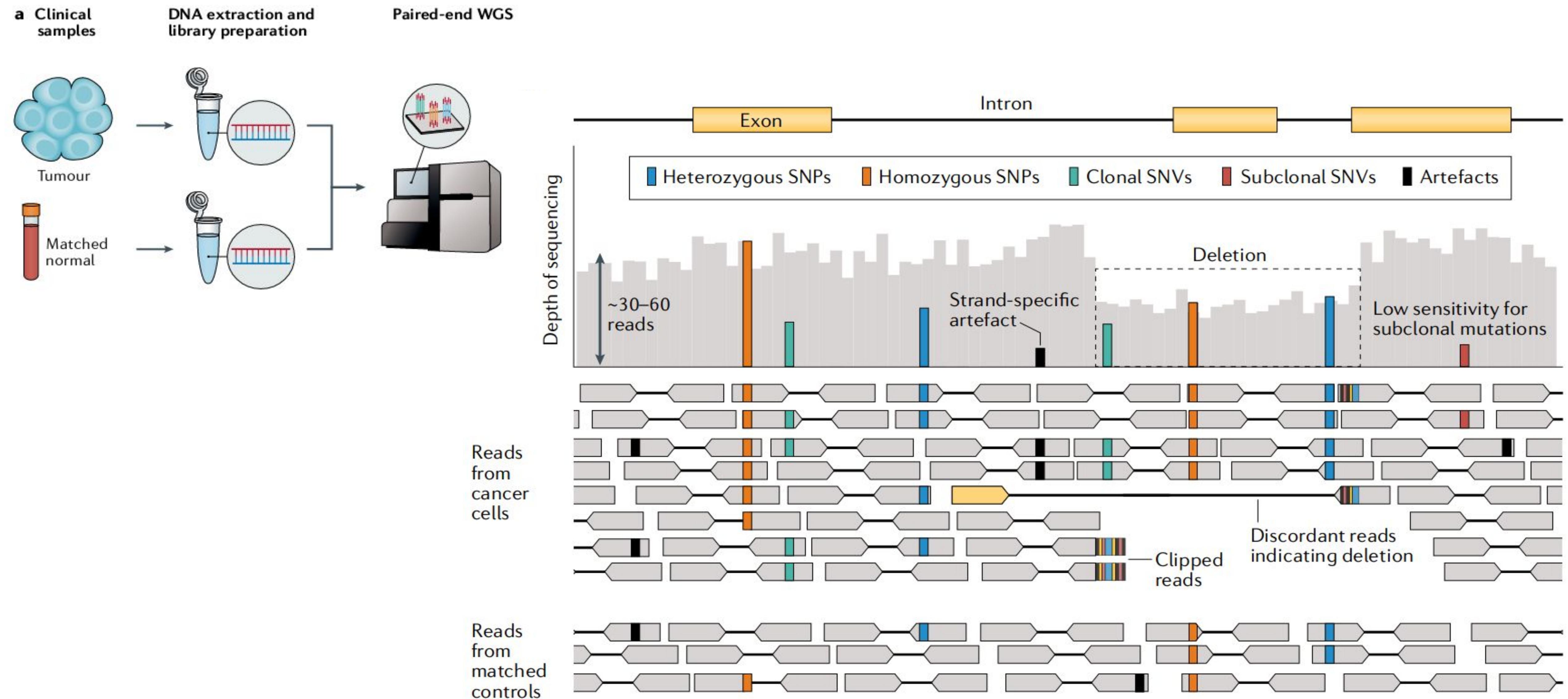
wellcome  
connecting  
science

# Methods to identify genomic alterations in the cancer cells

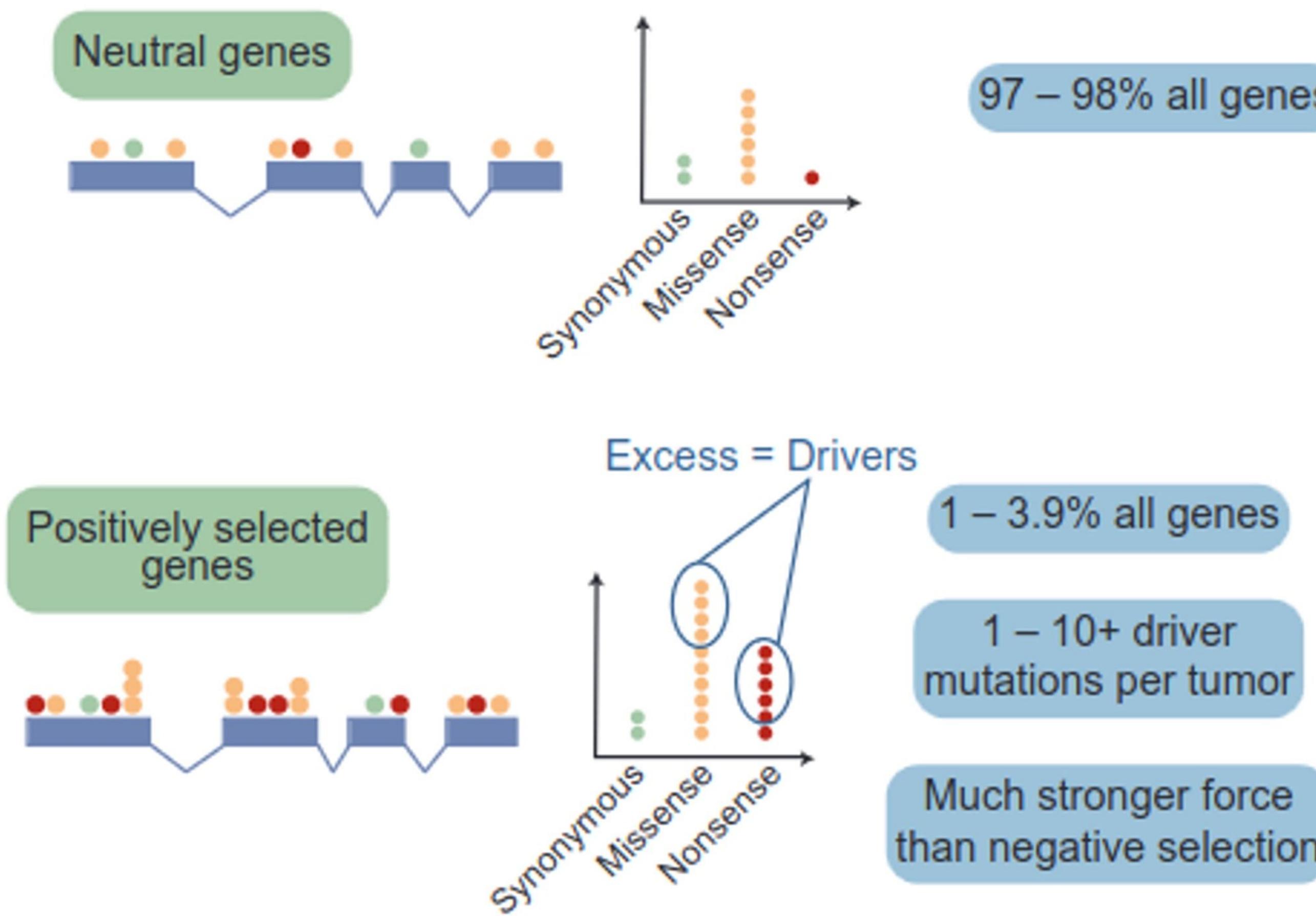
## NGS: high-throughput sequencing



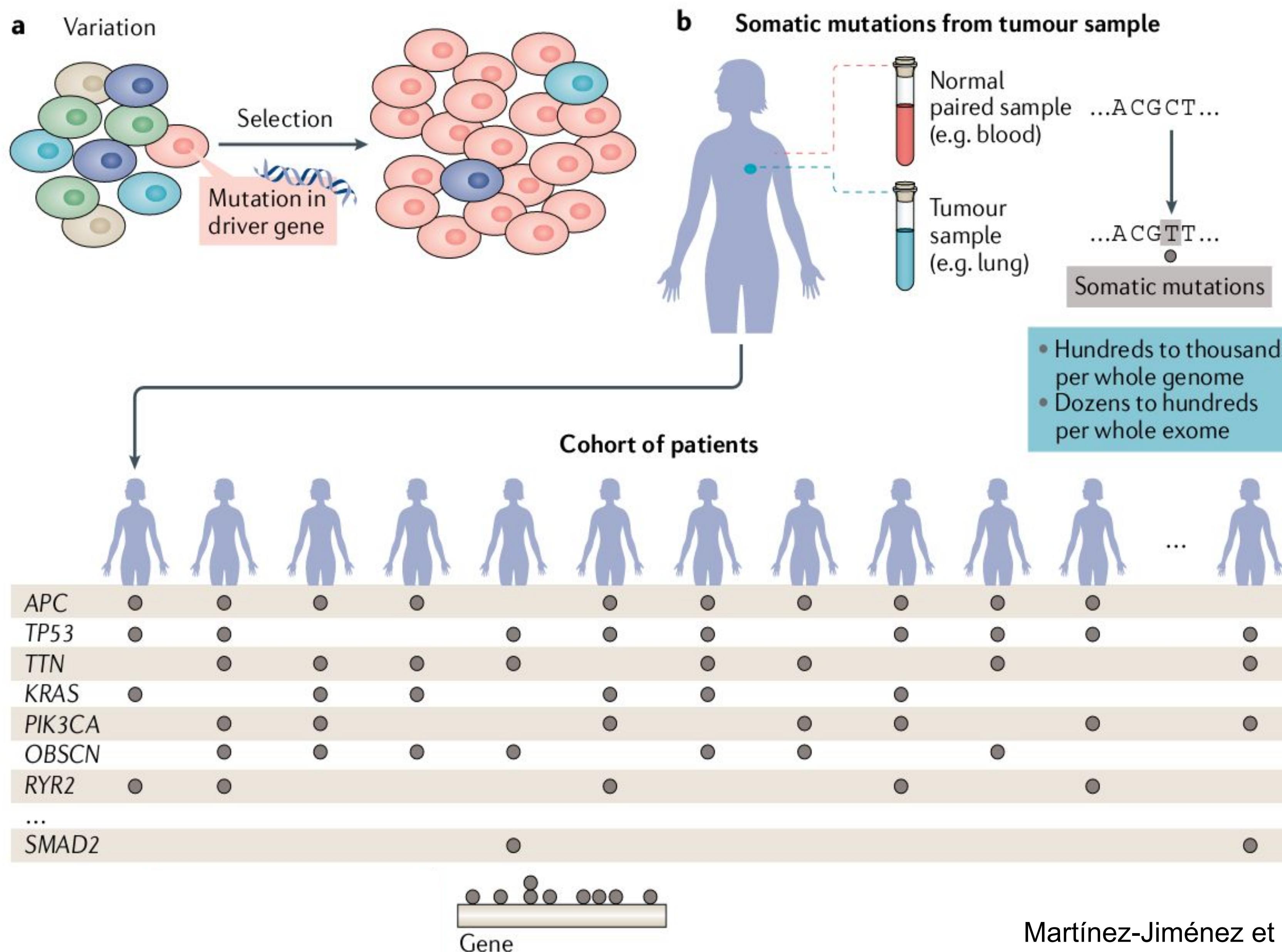
# Somatic and Germline variant identification through NGS



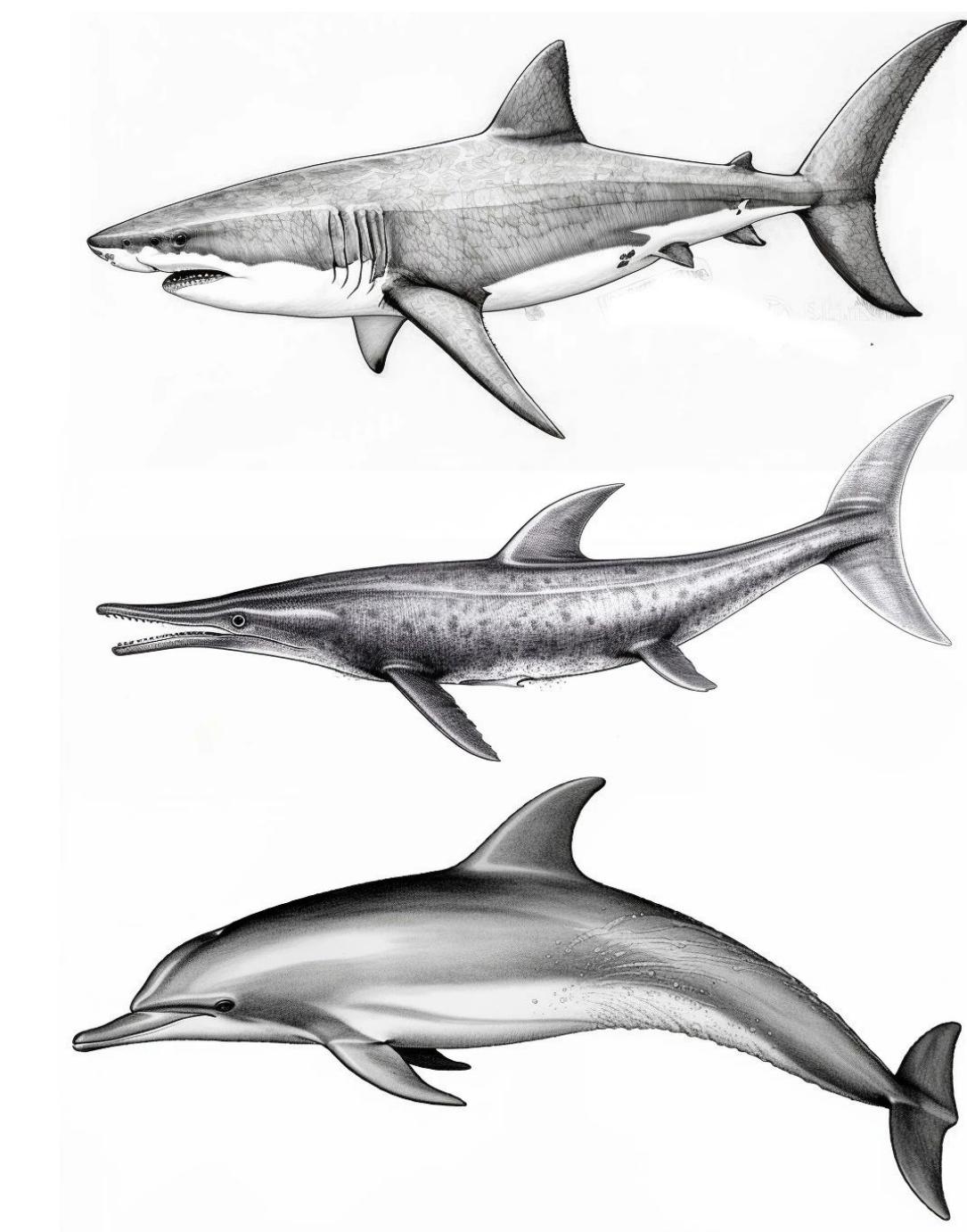
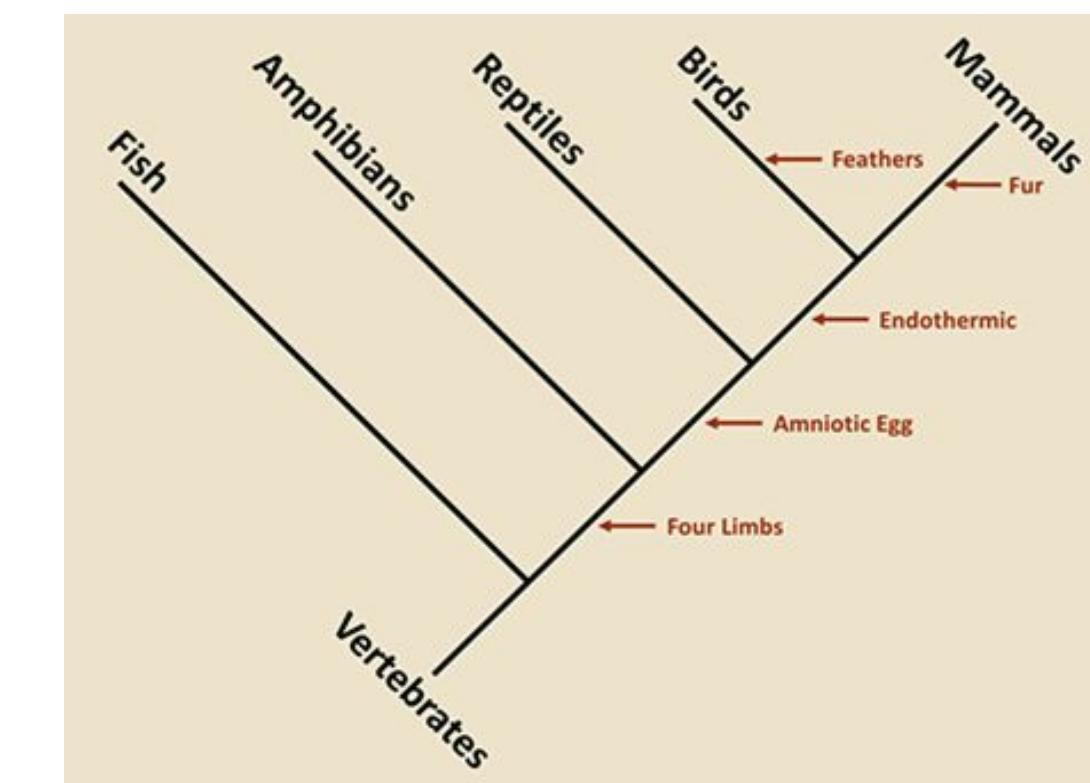
# Driver genes are under positive selection in tumors



# Identification of cancer driver genes through cohort-based analysis



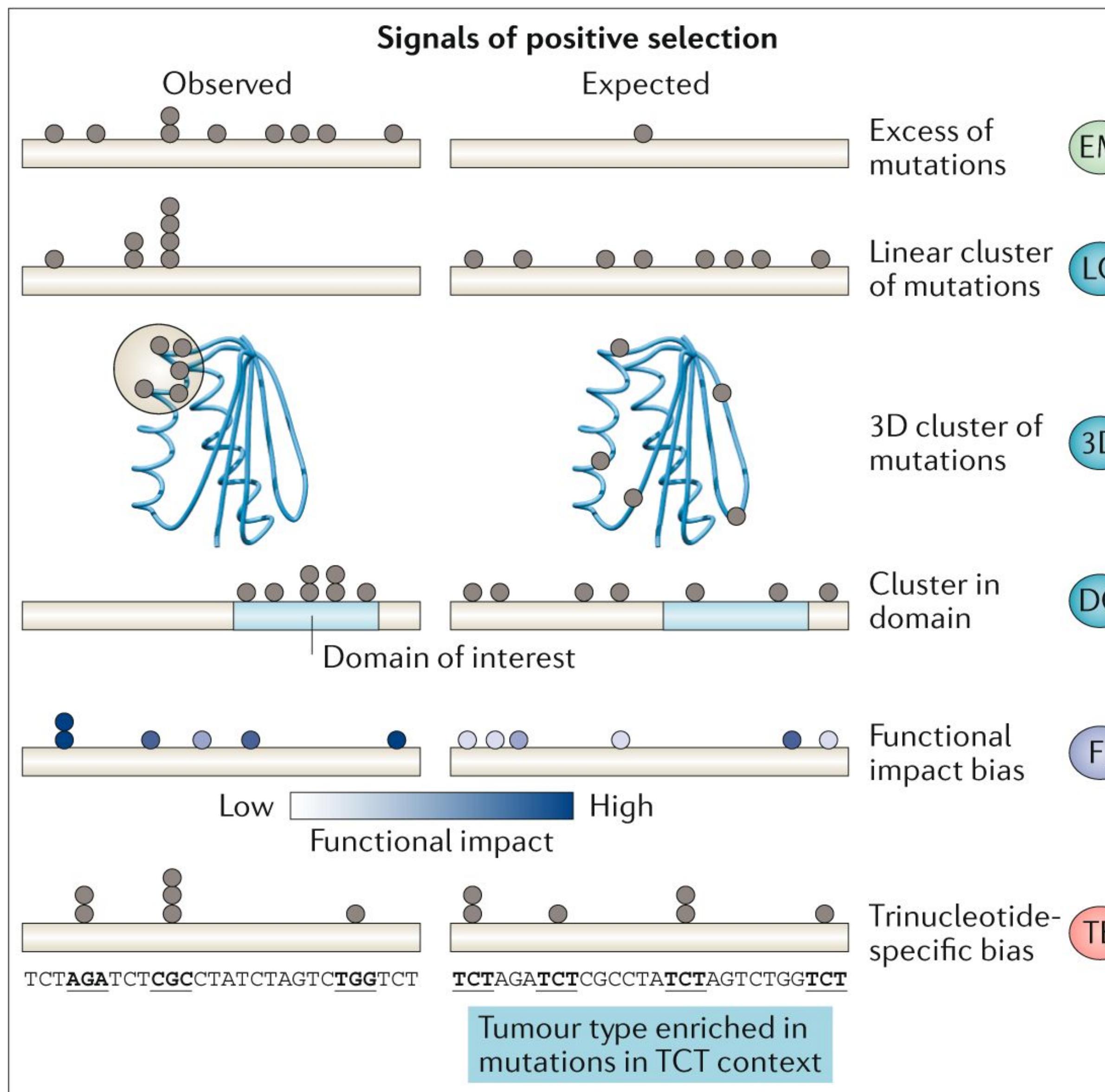
Martínez-Jiménez et al



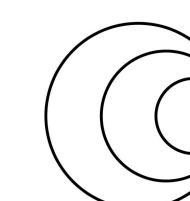
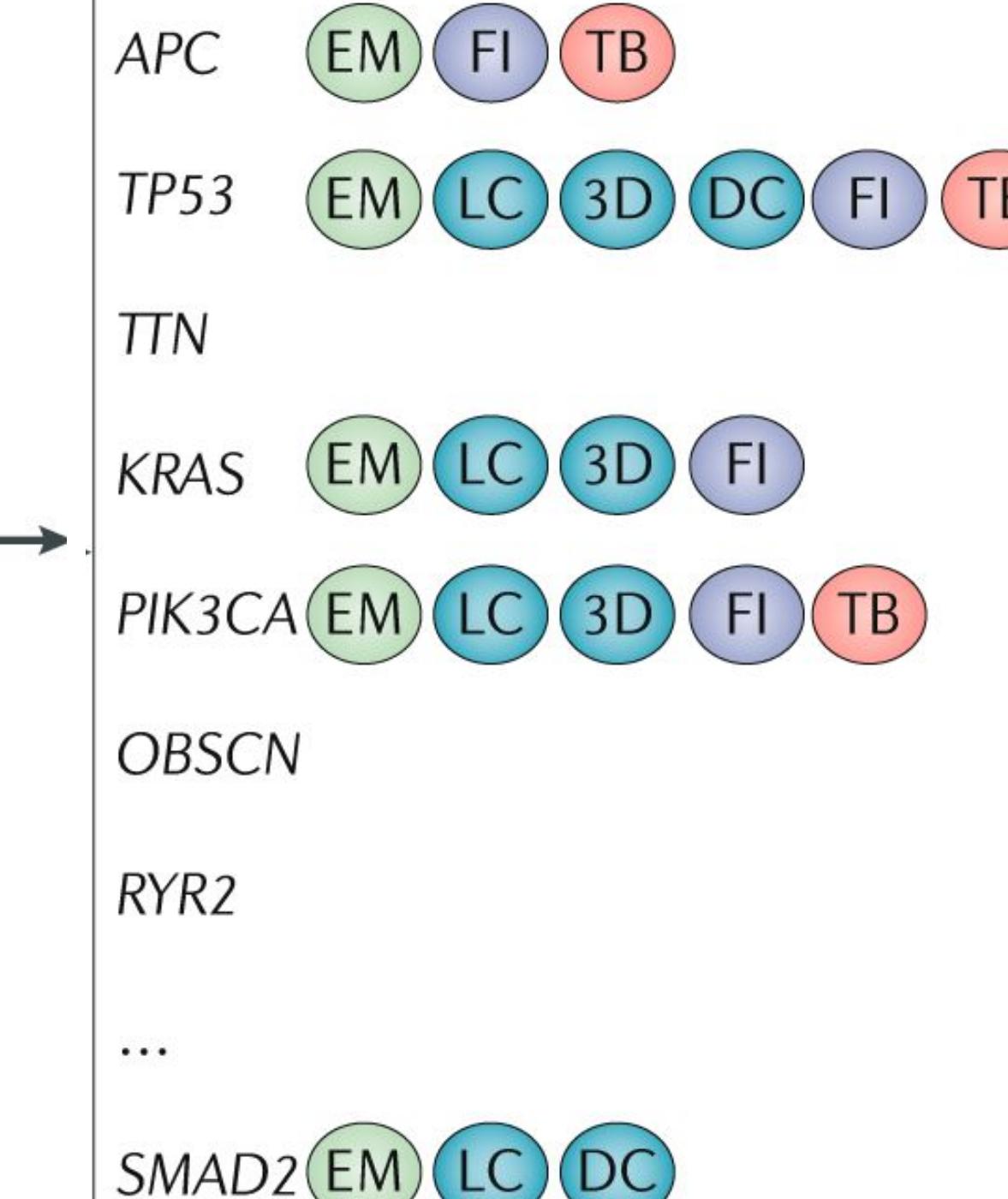
Alignment forum

# Signals of positive selection to identify cancer driver genes

## Detection of positively selected genes



## Cohort-specific catalogue of driver genes and their signals



wellcome  
connecting  
science

<https://www.intogen.org/>



e.g. Mutation distribution of BRAF in breast cancer

Search example | Show more examples

Release 2023-05-31

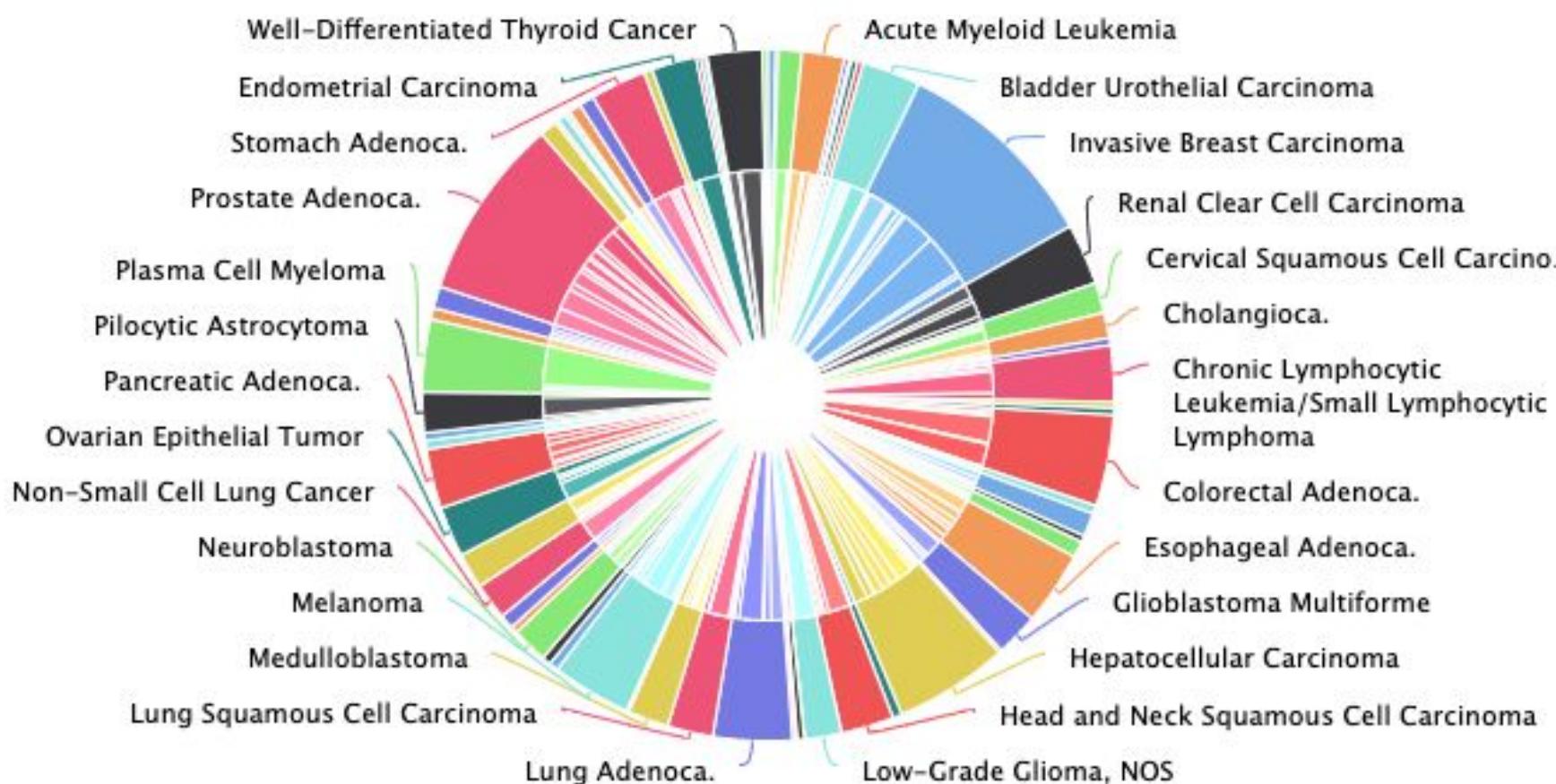
Plot

Table



### IntOGen Samples

Cancer types and cohorts chart



Cancer types

73

Cohorts

266

Samples

33,019

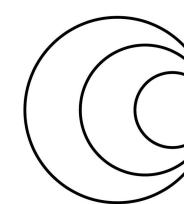
Mutations

252,486,809

Driver genes

619

Martínez-Jiménez et al., Nature Reviews Cancer, 20, 555–572 (2020)



wellcome  
connecting  
science

# Identifying driver mutations: Background mutation rate

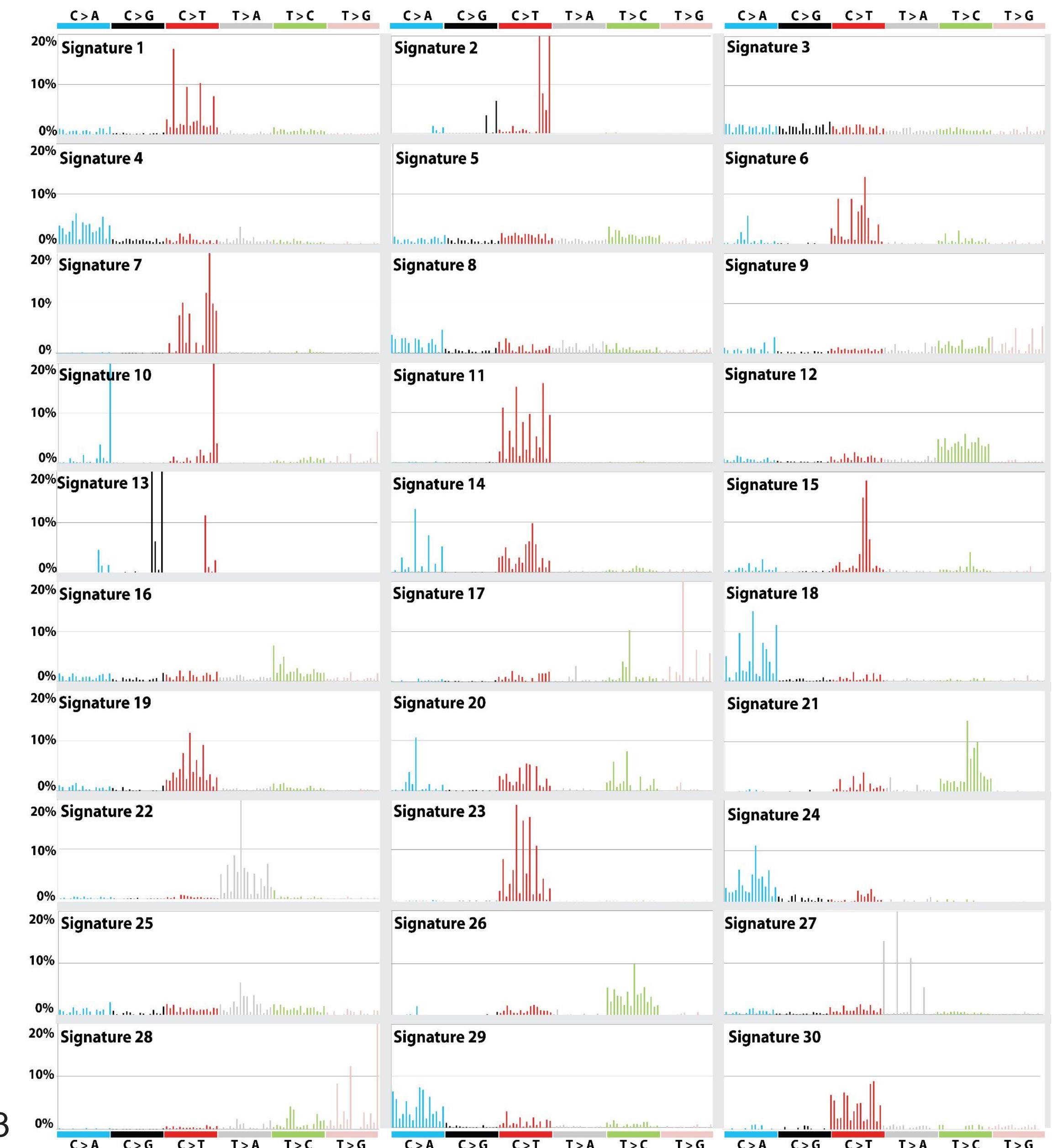
Most SNP mutational processes can be described using the tri-nucleotide context

Driver identification: Like finding needles in a haystack  
Key: proper modelling of mutational process  
Background mutation rate: Null (“expected”) model

- Null model difficult for structural variants:
  - CNAs (Gistic)
  - Gene fusions
  - Epigenetic alterations
- Better understood for point mutations
  - SNPs
  - Indels (difficult)

Background mutation rate depends on:

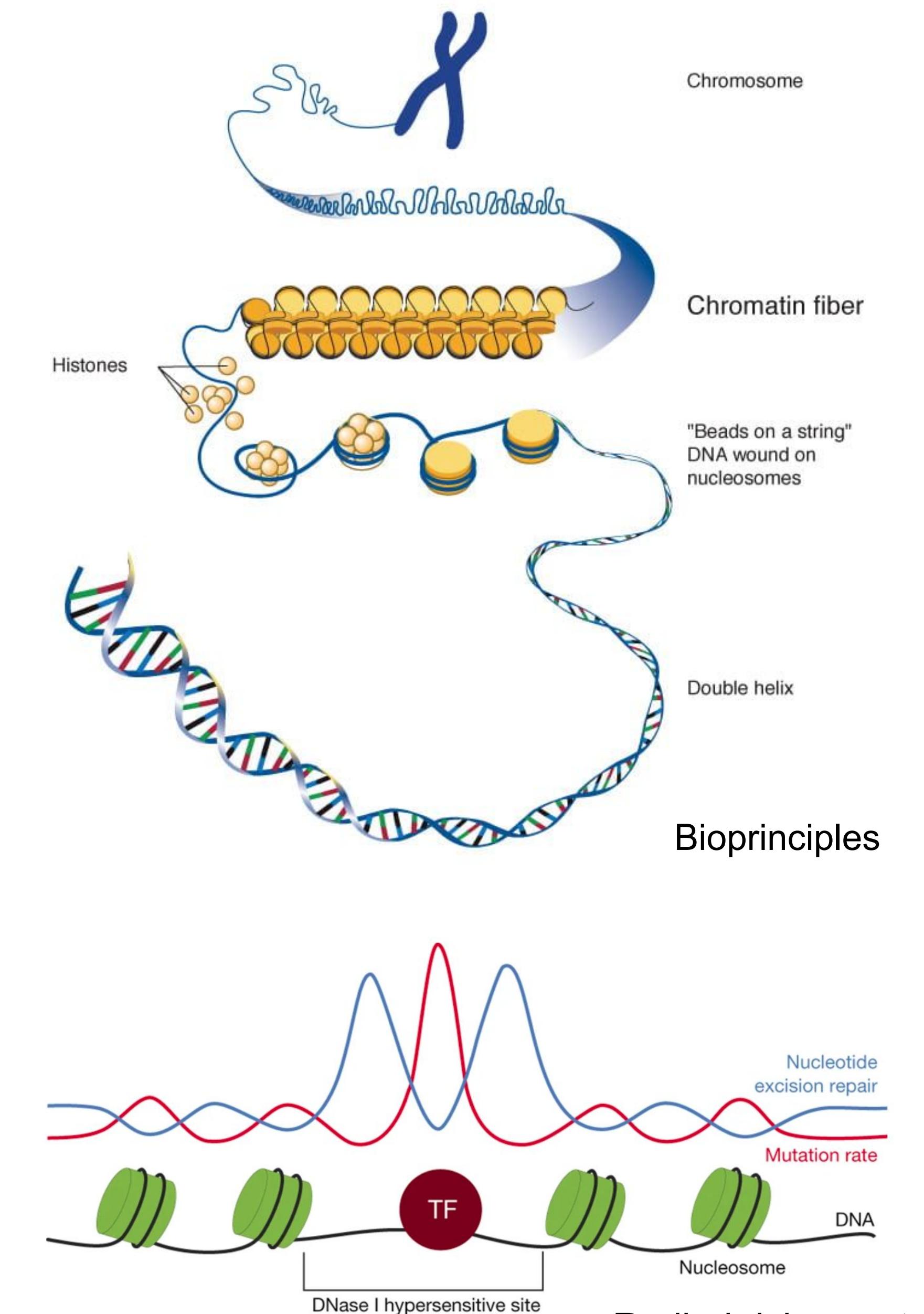
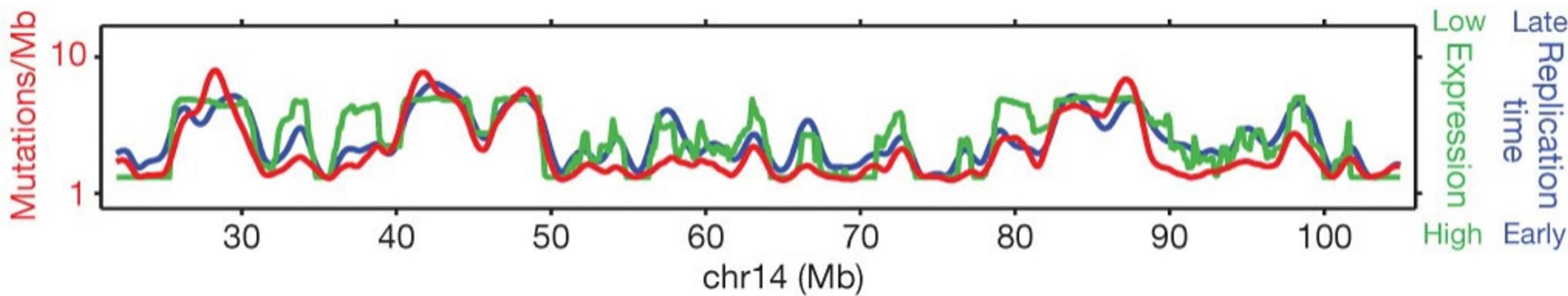
- The nucleotide sequence
- Active mutational processes
- Identity of the cell/tissue



# Identifying driver mutations: Background mutation rate

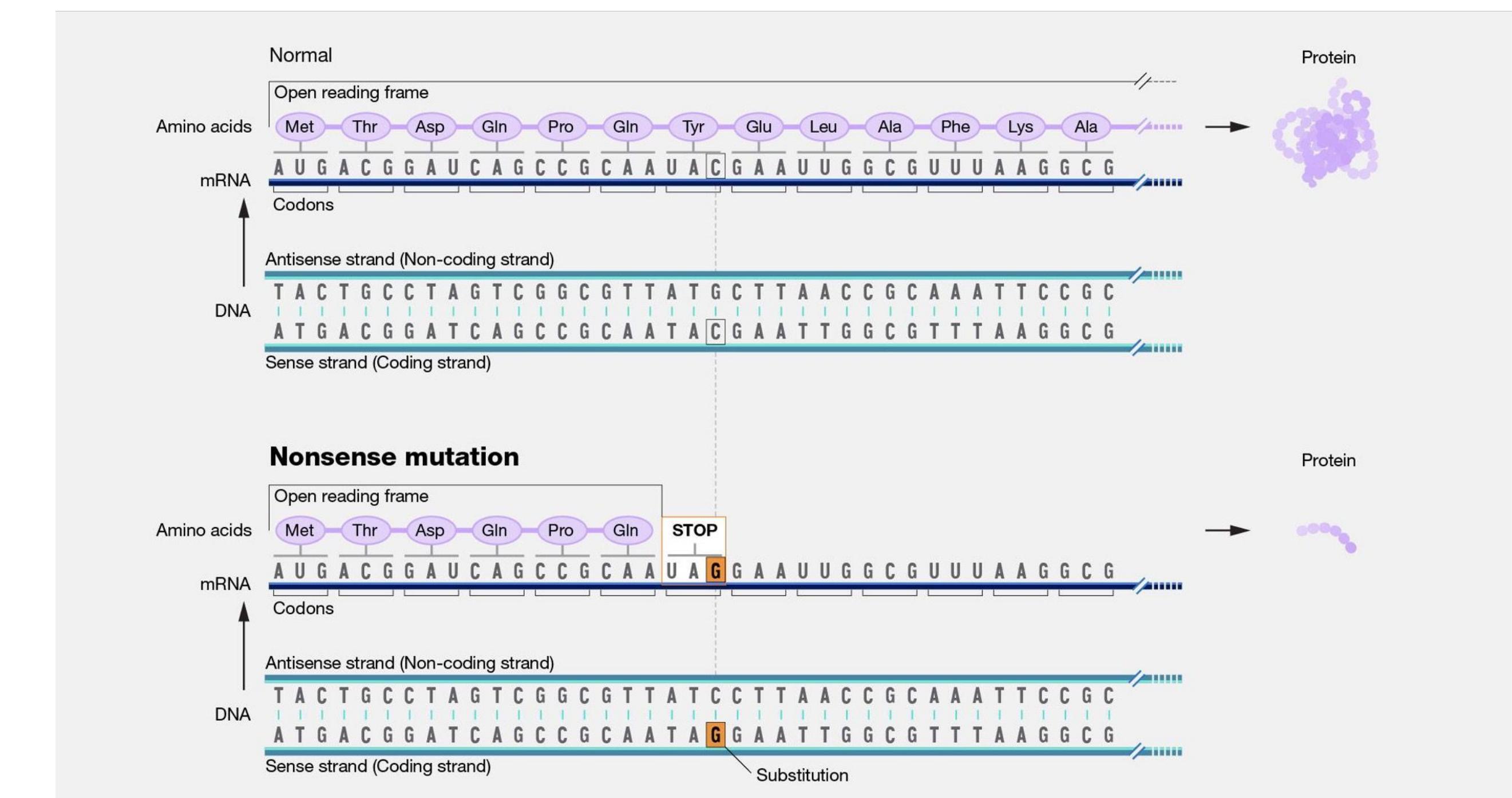
Background mutation rate on:

- The nucleotide sequence
- Active mutational processes
- Identity of the cell/tissue
- Replication timing
- Level of chromatin compaction
- Level of gene expression
- Occupancy by nucleosomes and other proteins
- Distribution of chromatin marks
- Formation of non-B-DNA structures



# Concept Review: Types of point mutations – substitutions and indels

		Second letter					
		U	C	A	G		
First letter	U	UUU Phe UUC UUA UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G	Third letter
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA CAG	CGU Arg CGC CGA CGG	U C A G	
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA AAG Lys	AGU Ser AGC AGA AGG	U C A G	
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA GAG Glu	GGU Gly GGC GGA GGG	U C A G	



Nonsense substitution, e.g. TAT > TAA (Tyr → Stop\*)

Synonymous substitution, e.g. TAT > TAC (Tyr → Tyr)

Missense substitution, e.g. TAT > TGT (Tyr → Cys)

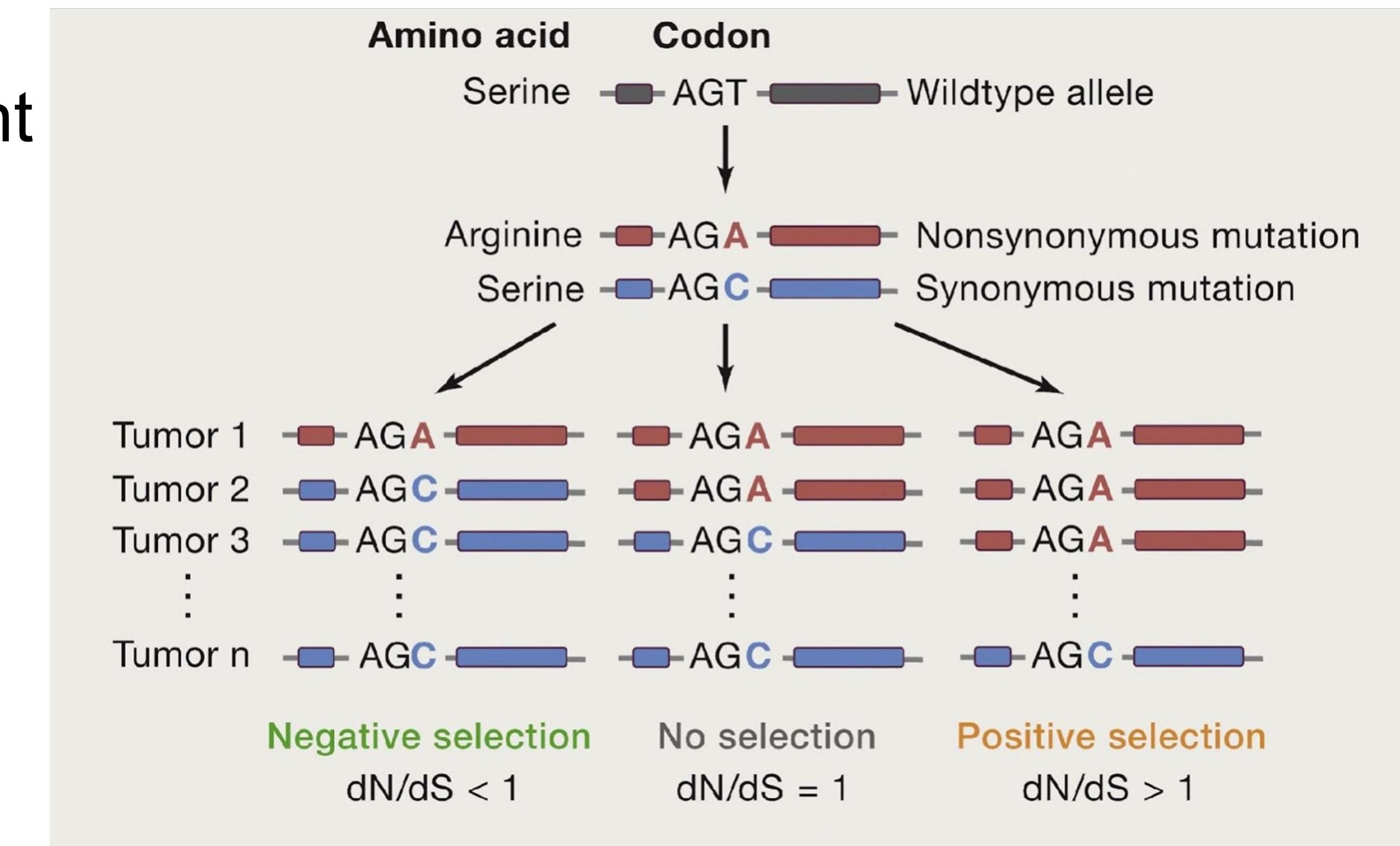
Indels: insertions/deletions, in frame vs out of frame

# Identify driver genes: Estimate Selection Coefficients

To formalize the observed versus expected test, estimate the coefficient selection  $W=dN/dS$  where:

$dN$  = number of **non-synonymous** substitutions

$dS$  = number of **synonymous** substitutions

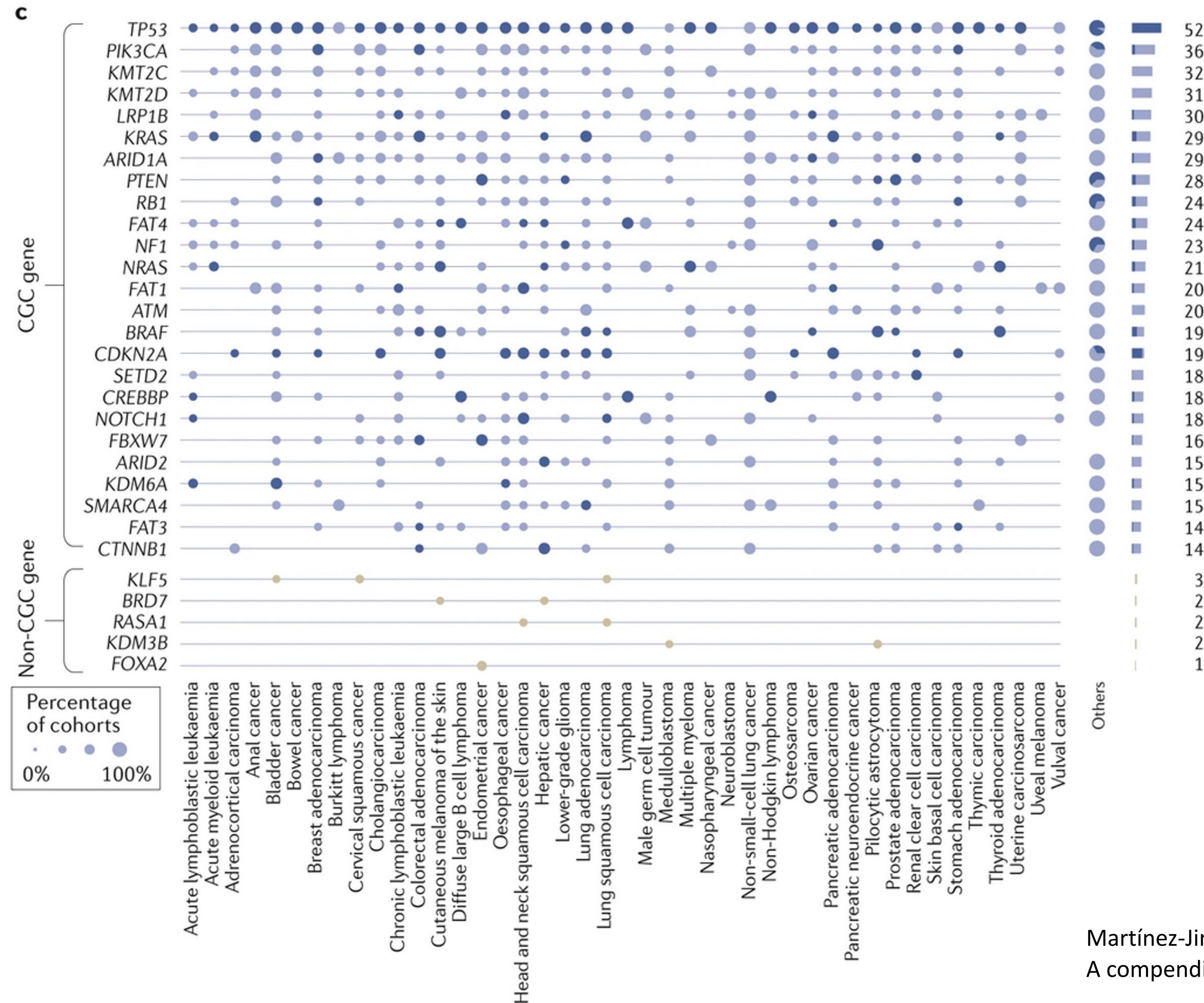


*Non-synonymous substitutions:*

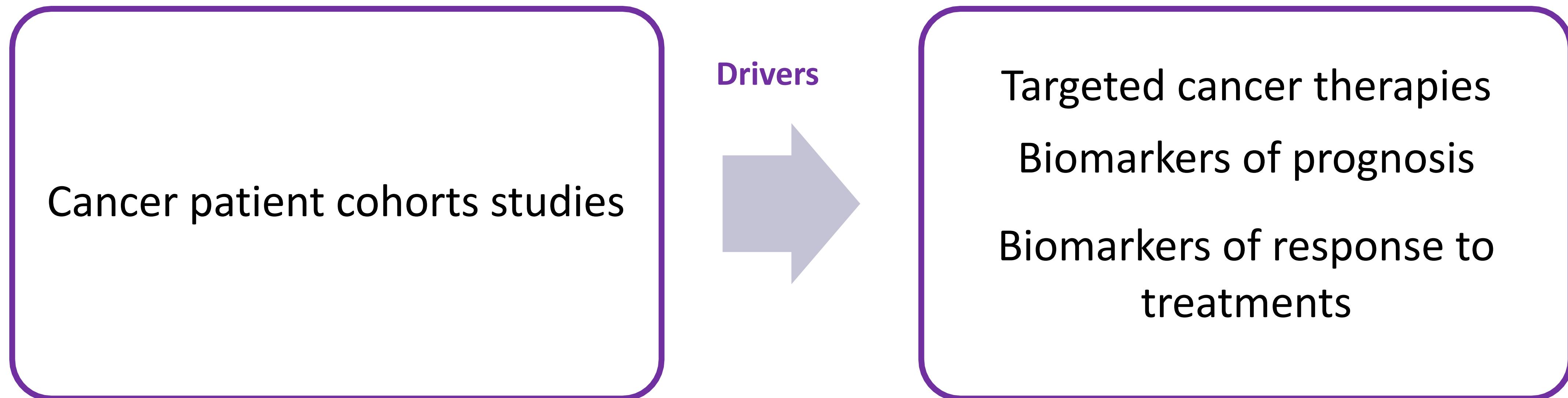
- *Missense* (e.g. *Leu* → *Pro*)
- *Nonsense* (e.g. *Ser* → *Stop*)
- *Splice sites*

Bakhoum SF, Landau DA. Cell. 2017 Nov 16;171(5):987-989.

# Mutational driver genes



# Research on cancer drivers



# Finding drivers in cohorts

- Recurrence --> signature of positive selection
- Requires proper modelling of the mutational process.

**The null (expected) model**

Difficult for --> structural variants

Better for --> point mutations (substitutions  
and small indels)

# Finding drivers in cohorts

- $dN/dS \rightarrow$  non-synonymous / synonymous substitutions  $\rightarrow$  observed / expected
  - $dN/dS < 1$  Negative selection
  - $dN/dS = 1$  Neutral selection
  - $dN/dS > 1$  Positive selection
- Interpretation
  - $dN/dS$  of 1  $\rightarrow$  All observed non-syn are expected
  - $dN/dS$  of 2  $\rightarrow$  50% of non-syn are selected
  - $dN/dS$  of 10  $\rightarrow$  90% of non-syn are selected
  - $dN/dS$  of 100  $\rightarrow$  99% of non\_syn are selected

$w = dN/dS$  (coef. of  
selection)  
 $(w-1)/w =$  fraction  
mutations selected

## dndscv R package

**Identifies positively selected genes  
using dN/dS model**

**Mutational model:**

Trinucleotide frequencies

Covariates for regional variation

**Generates selection estimates at**

- Gene level (or domains)
- Global
- Sites and codons
- Not only for cancer

# Hotspots

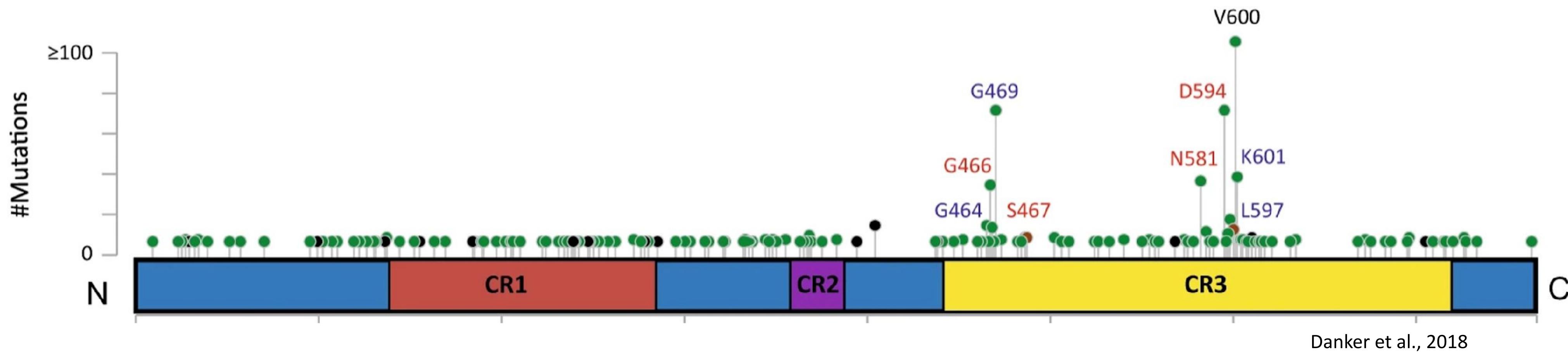
Recurrently mutated positions (nucleotide or amino acid).

Examples:

*BRAF* V600

Melanoma  
Colorectal cancer  
Lymphoma

A **BRAF**



# Hotspots

Recurrently mutated positions (nucleotide or amino acid).

Examples:

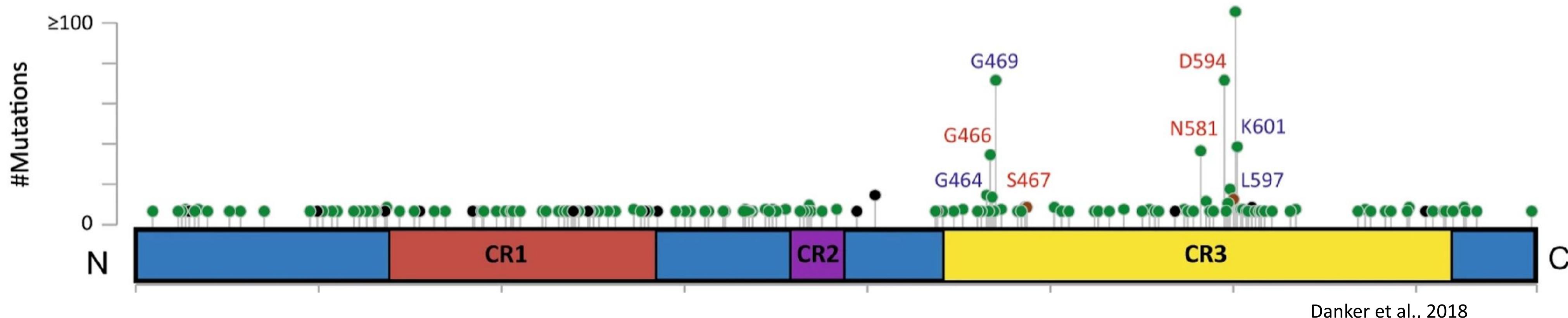
*BRAF* V600

Melanoma  
Colorectal cancer  
Lymphoma

Vemurafenib

Dabrafenib

A **BRAF**



# Structural drivers

- Copy number gains / loses
- Gene fusions
- Rearrangements

## Challenges

- Modelling the background
- Involve multiple chromosomes
- Require long read sequencing

# Non-coding somatic drivers

- lncRNAs
- UTRs/promoters
- tRNAs
- small RNAs
- micro RNAs

Example:

*TERT* promoter

## Challenges

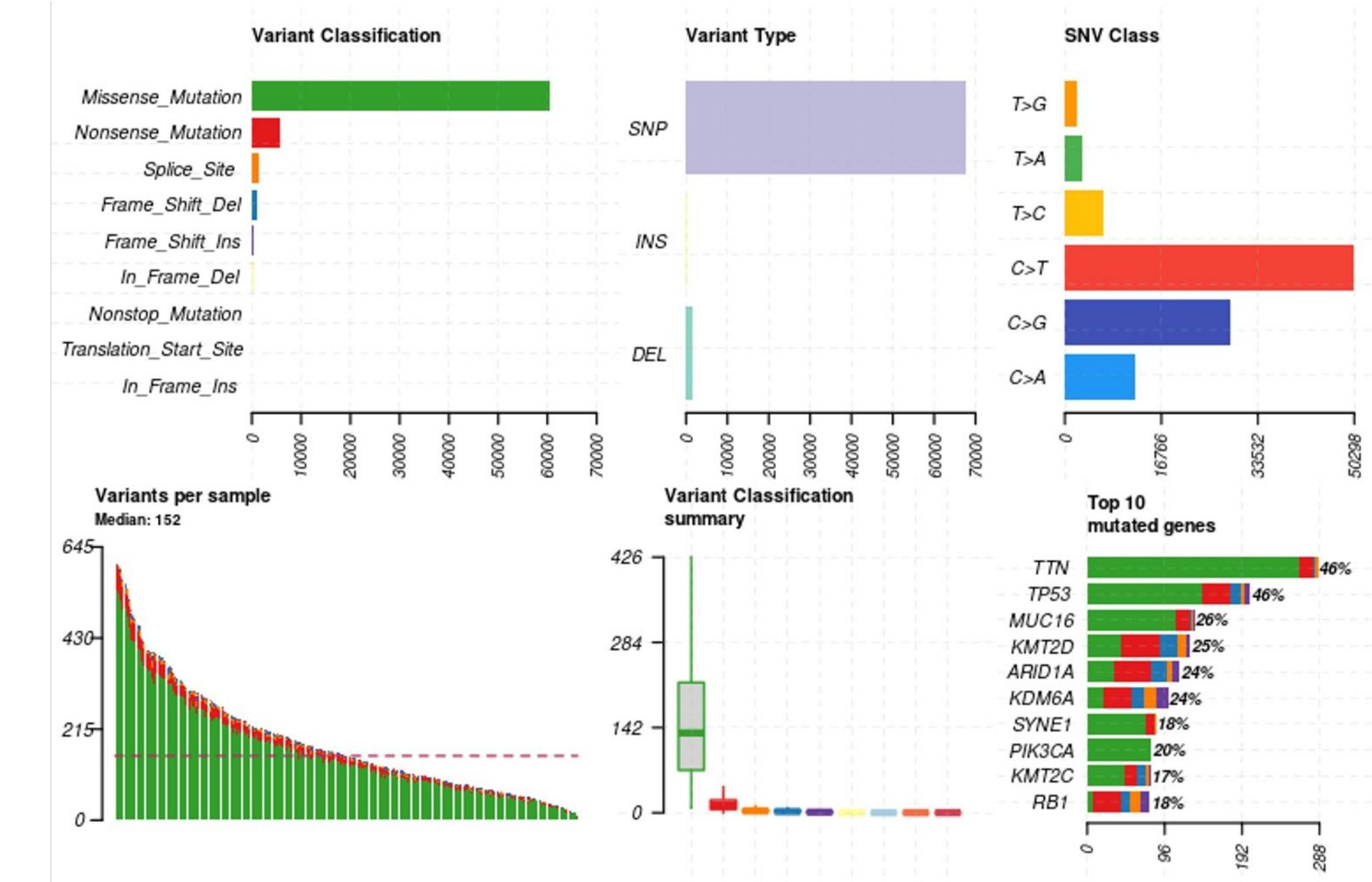
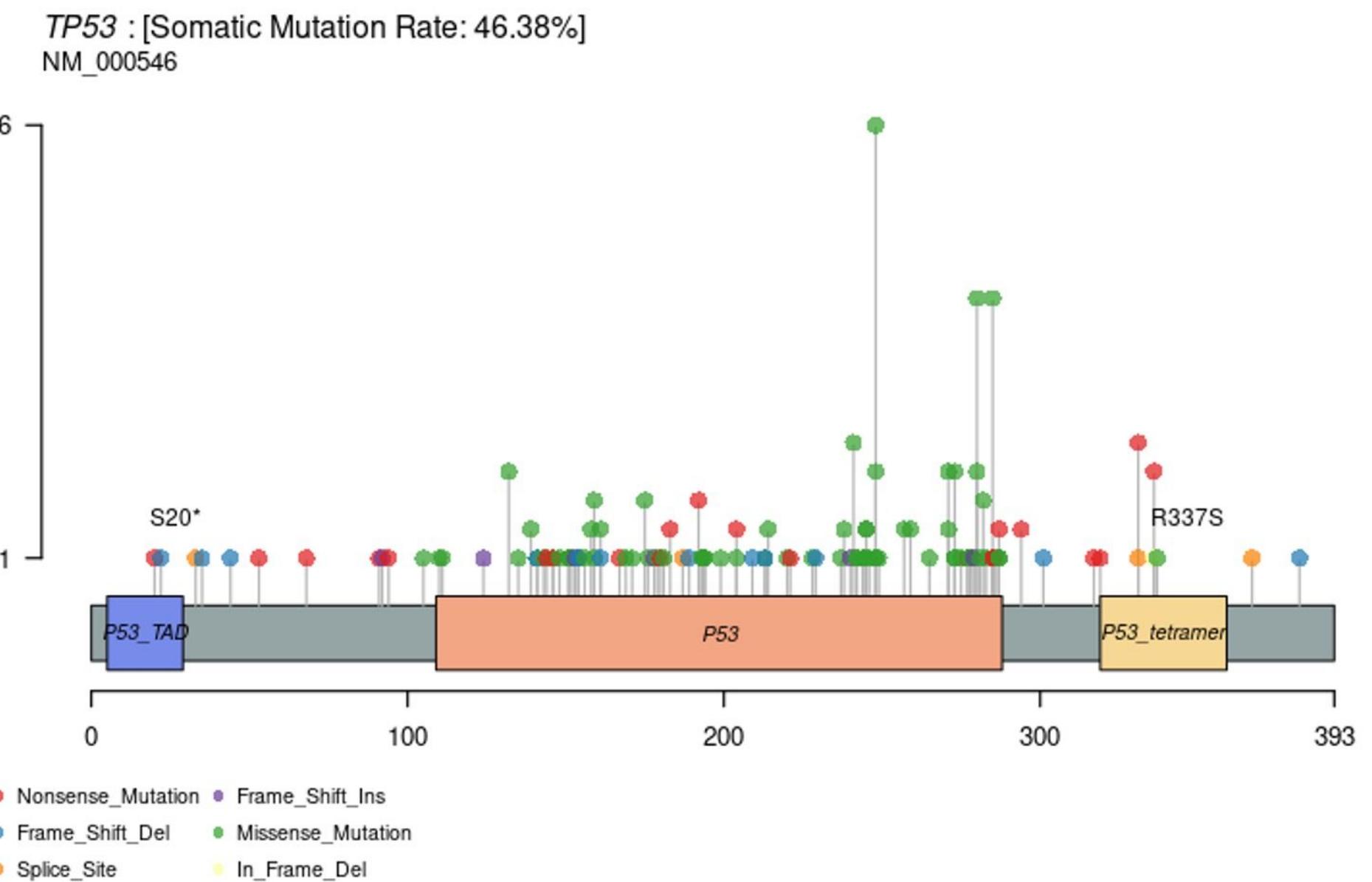
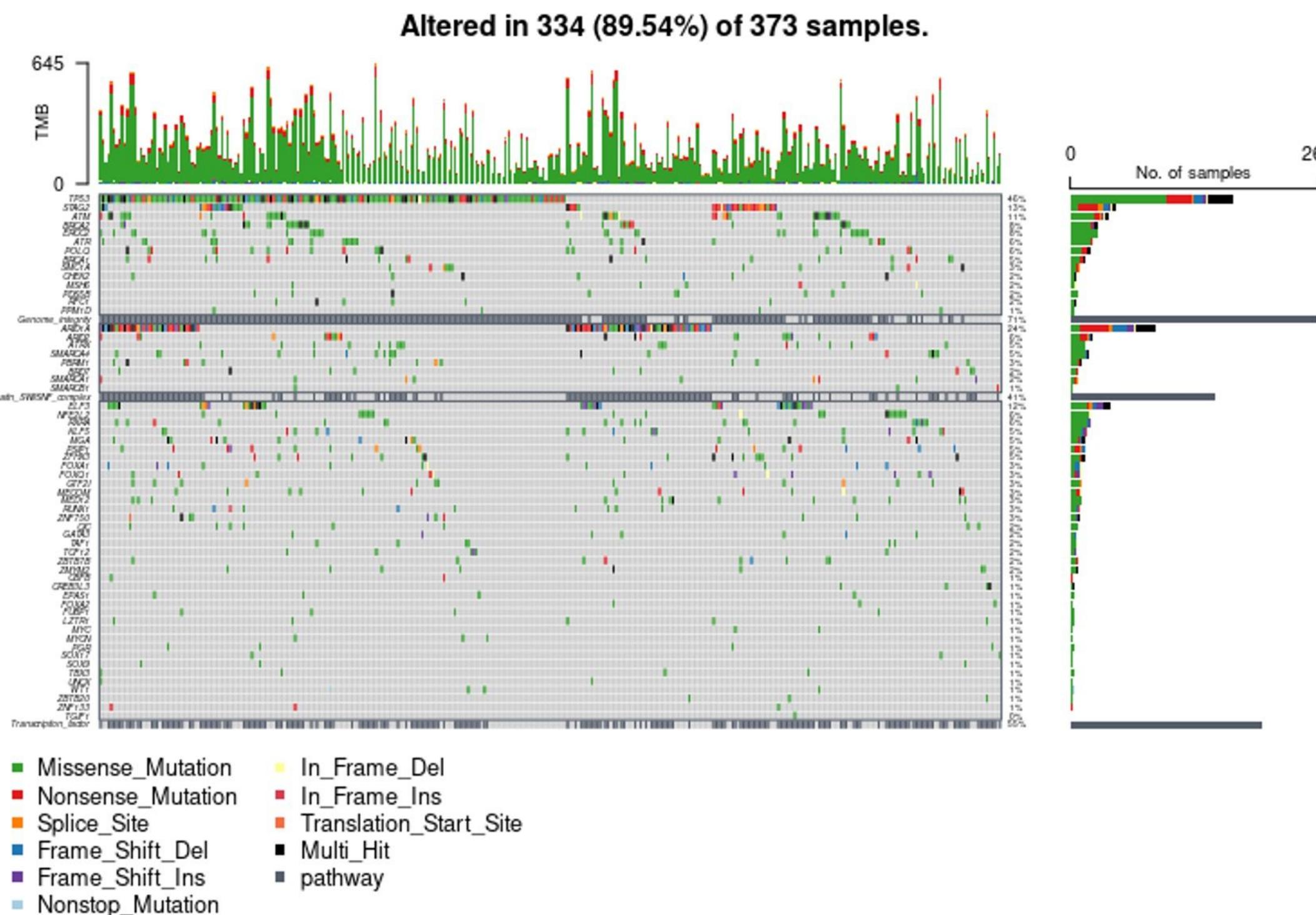
- Sequencing and mapping artefacts
- Incomplete annotation of regulatory regions
- Unknown functional effects

Less frequent compared to protein coding  
drivers (Rheinbay et al., 2020)

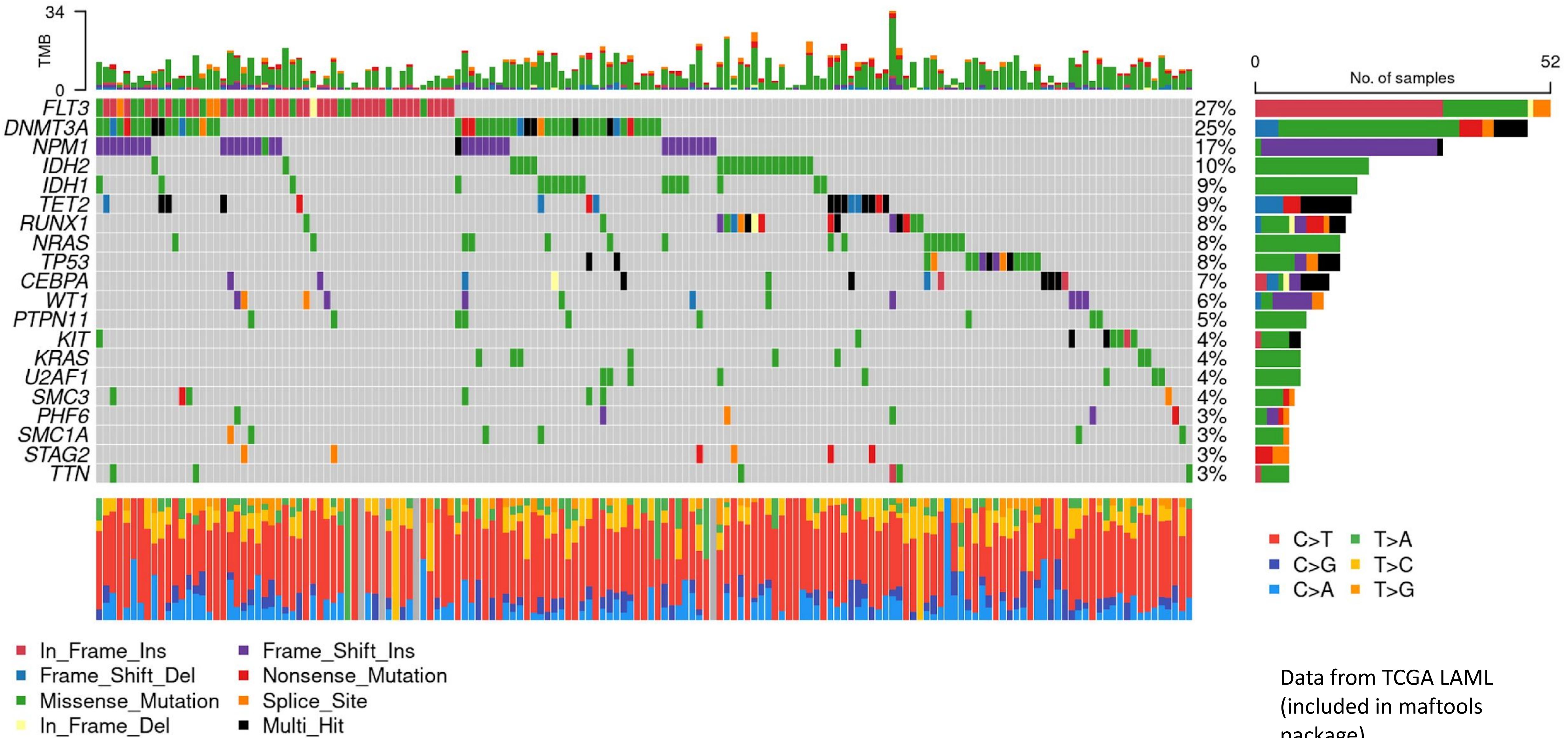
# Oncoplots

# maftools

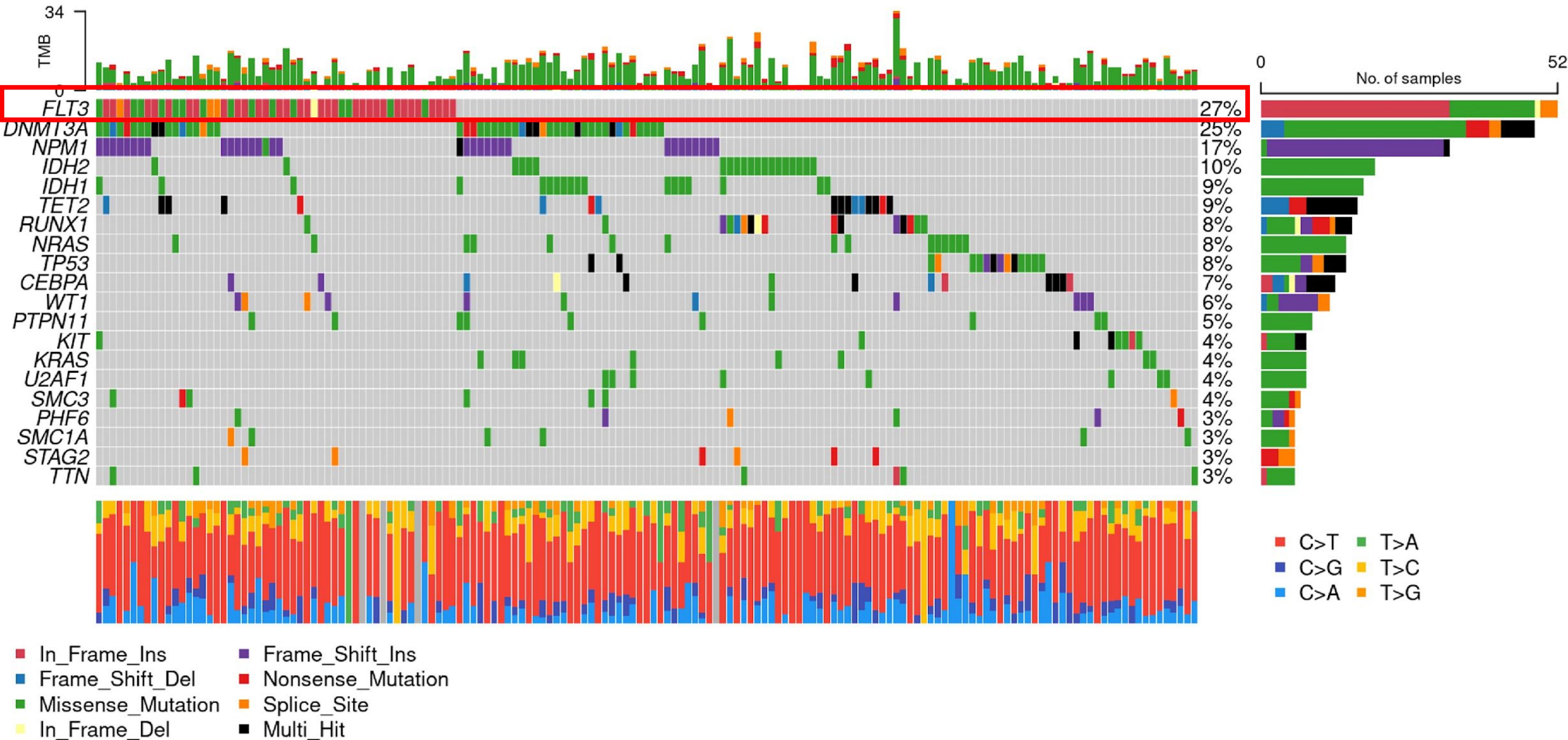
- R package
- Work with data stored in MAF
- Different type of visualizations



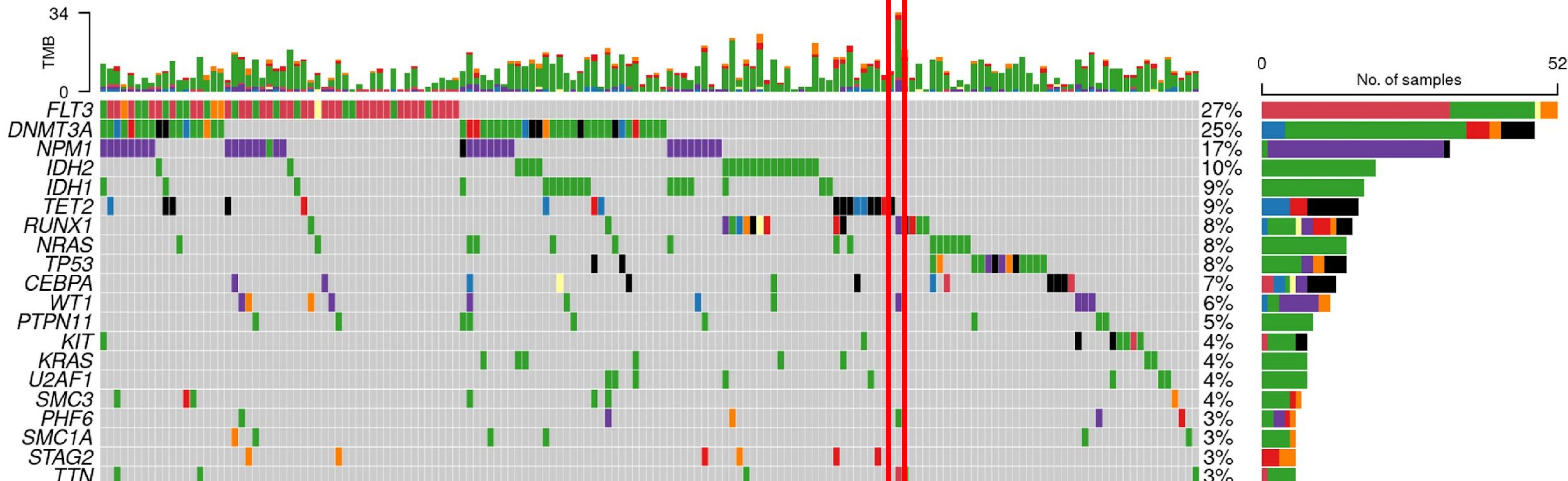
# Altered in 159 (82.38%) of 193 samples.



# Altered in 159 (82.38%) of 193 samples.



# Altered in 159 (82.38%) of 193 samples.



- In\_Frame\_Ins      ■ Frame\_Shift\_Ins
- Frame\_Shift\_Del   ■ Nonsense\_Mutation
- Missense\_Mutation   ■ Splice\_Site
- In\_Frame\_Del      ■ Multi\_Hit

- C>T      ■ T>A
- C>G      ■ T>C
- C>A      ■ T>G

# Practical - dndscv

## Input file

Five column table of mutations

- 1) sampleID
- 2) chromosome
- 3) position
- 4) reference
- 5) mutant

```
##   sampleID chr      pos ref mut
## 1 Sample_1   1 871244   G   C
## 2 Sample_1   1 6648841  C   G
## 3 Sample_1   1 17557072  G   A
## 4 Sample_1   1 22838492  G   C
## 5 Sample_1   1 27097733  G   A
## 6 Sample_1   1 27333206  G   A
```

# Outputs

- List of objects in R
  - Usually the most important object is the neutrality test results (**dndsout\$sel**). This table has information on the number of substitutions for each class, dN/dS ratios and global p and q values.

```
sel_cv = dndtout$sel_cv  
print(head(sel_cv), digits = 3)
```

	gene_name	n_syn	n_mis	n_non	n_spl	n_ind	wmis_cv	wnon_cv	wspl_cv	
##	18057	TP53	1	43	5	4	5	113.66	221.8	221.8
##	12977	PIK3CA	3	34	0	0	3	29.75	0.0	0.0
##	9225	KRAS	1	21	0	0	0	125.98	0.0	0.0
##	18924	VHL	3	9	1	0	4	24.75	38.3	38.3
##	1296	APC	2	8	10	0	6	2.82	31.2	31.2
##	1465	ARID1A	1	7	10	0	3	3.30	53.5	53.5
		wind_cv	pmis_cv	ptrunc_cv	pallsubs_cv	pind_cv	qmis_cv	qtrunc_cv		
##	18057	138.8	0.00e+00	1.11e-16	0.00e+00	1.73e-09	0.00000	2.23e-12		
##	12977	30.7	0.00e+00	5.68e-01	0.00e+00	2.12e-04	0.00000	9.47e-01		
##	9225	0.0	0.00e+00	8.44e-01	0.00e+00	1.00e+00	0.00000	9.47e-01		
##	18924	204.5	7.46e-09	2.23e-02	7.97e-09	1.32e-08	0.00003	9.47e-01		
##	1296	23.1	4.96e-02	2.81e-10	2.12e-09	1.81e-06	0.77064	1.41e-06		
##	1465	14.4	3.64e-02	2.96e-12	2.38e-11	1.84e-03	0.77064	2.97e-08		
		qallsubs_cv	pglobal_cv	qglobal_cv						
##	18057	0.00e+00	0.00e+00	0.00e+00						
##	12977	0.00e+00	0.00e+00	0.00e+00						
##	9225	0.00e+00	0.00e+00	0.00e+00						
##	18924	2.00e-05	4.00e-15	2.01e-11						
##	1296	6.07e-06	1.31e-13	5.26e-10						
##	1465	1.20e-07	1.39e-12	4.67e-09						

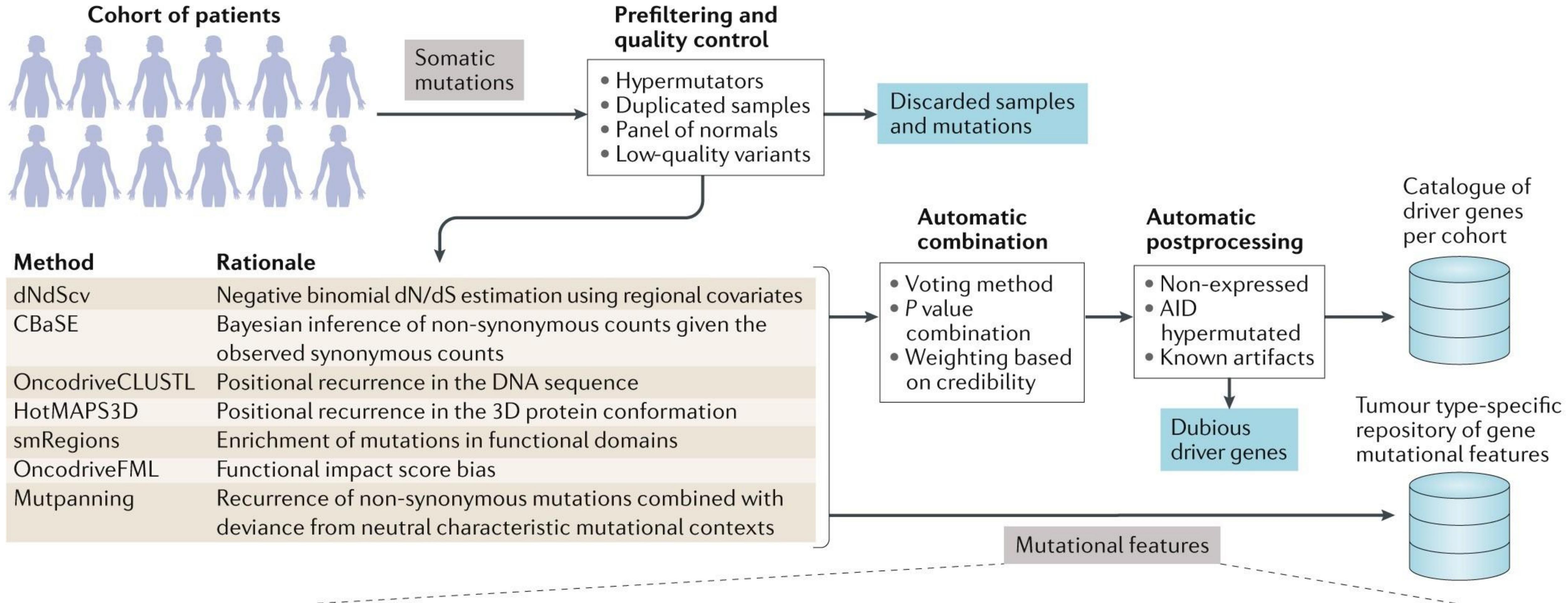
# Take Home Messages

- Oncogenes and tumor suppressors behave very differently – no clear cut rules of thumb in Biology
- Passengers >>>>> drivers (hypermutators are particularly problematic)
- Recurrence = positive selection (obs > exp)
- Most drivers are protein coding (+ Tert) - ~ 1% of the genome
- Not all non-synonymous mutations are drivers (dN/dS)
- Structural and epigenetic alterations can be drivers too

Cancer Driver Callers <b>NOT EXHAUSTIVE</b>	Basis/Rationale	Reference
<a href="#">ddscnv</a>	Negative binomial dN/dS estimation using regional covariates	Martincorena, I. et al. <b>Universal Patterns of Selection in Cancer and Somatic Tissues.</b> Cell 171, 1029-1041.e21 (2017). doi: 10.1016/j.cell.2017.09.042
<a href="#">2020+</a>		Collin J. Tokheim, Nickolas Papadopoulos, Kenneth W. Kinzler, Bert Vogelstein, and Rachel Karchin. <b>Evaluating the evaluation of cancer driver genes.</b> PNAS 2016 ; published ahead of print November 22, 2016, doi:10.1073/pnas.1616440113
<a href="#">CBaSE</a>	Bayesian inference of non-synonymous counts given the observed synonymous counts	Weghorn D, Sunyaev S. <b>Bayesian inference of negative and positive selection in human cancers.</b> Nat Genet. 2017 Dec;49(12):1785-1788. doi: 10.1038/ng.3987. Epub 2017 Nov 6. PMID: 29106416.
<a href="#">HotSpot3D</a>	3D proximity tool can be used to identify mutation hotspots from linear protein sequence and correlate the hotspots with known or potentially interacting domains, mutations, or drugs.	Niu B, Scott AD, Sengupta S, Bailey MH, Batra P, Ning J, Wyczalkowski MA, Liang WW, Zhang Q, McLellan MD, Sun SQ, Tripathi P, Lou C, Ye K, Mashl RJ, Wallis J, Wendl MC, Chen F, Ding L. <b>Protein-structure-guided discovery of functional mutations across 19 cancer types.</b> Nat Genet. 2016 Aug;48(8):827-37. doi: 10.1038/ng.3586. Epub 2016 Jun 13. Erratum in: Nat Genet. 2017 Jul 27;49(8):1286. PMID: 27294619; PMCID: PMC5315576.
<a href="#">HotMAPS3D</a>	Positional recurrence in the 3D protein conformation	Tokheim C, et al. Exome-scale discovery of hotspot mutation regions in human cancer using 3D protein structure. Cancer research. 2016a;76:3719–3731. doi: 10.1158/0008-5472.CAN-15-3190
<a href="#">Mutpanning</a>	recurrence of non-synonymous mutations combined with deviance from neutral characteristic mutational contexts	Dietlein, F., Weghorn, D., Taylor-Weiner, A. et al. <b>Identification of cancer driver genes based on nucleotide context.</b> Nat Genet (2020). <a href="https://doi.org/10.1038/s41588-019-0572-y">https://doi.org/10.1038/s41588-019-0572-y</a>
<a href="#">OncodriveCLUSTL</a>	a sequence-based clustering method to identify significant clustering signals in nucleotide sequence	Arnedo-Pac C, Mularoni L, Muños F, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. Bioinformatics. 2019 Nov 1;35(22):4788-4790. doi: 10.1093/bioinformatics/btz501. Erratum in: Bioinformatics. 2019 Dec 15;35(24):5396. PMID: 31228182; PMCID: PMC6853674.
<a href="#">smRegions</a>	Enrichment of mutations in functional domains	Francisco Martínez-Jiménez, et al. <b>Disruption of ubiquitin mediated proteolysis is a widespread mechanism of tumorigenesis.</b> bioRxiv 2019. doi: <a href="https://doi.org/10.1101/507764">https://doi.org/10.1101/507764</a>
<a href="#">OncodriveFML</a>	Functional impact score bias	Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. <b>OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations.</b> Genome Biol. 2016 Jun 16;17(1):128. doi: 10.1186/s13059-016-0994-0. PMID: 27311963; PMCID: PMC4910259.

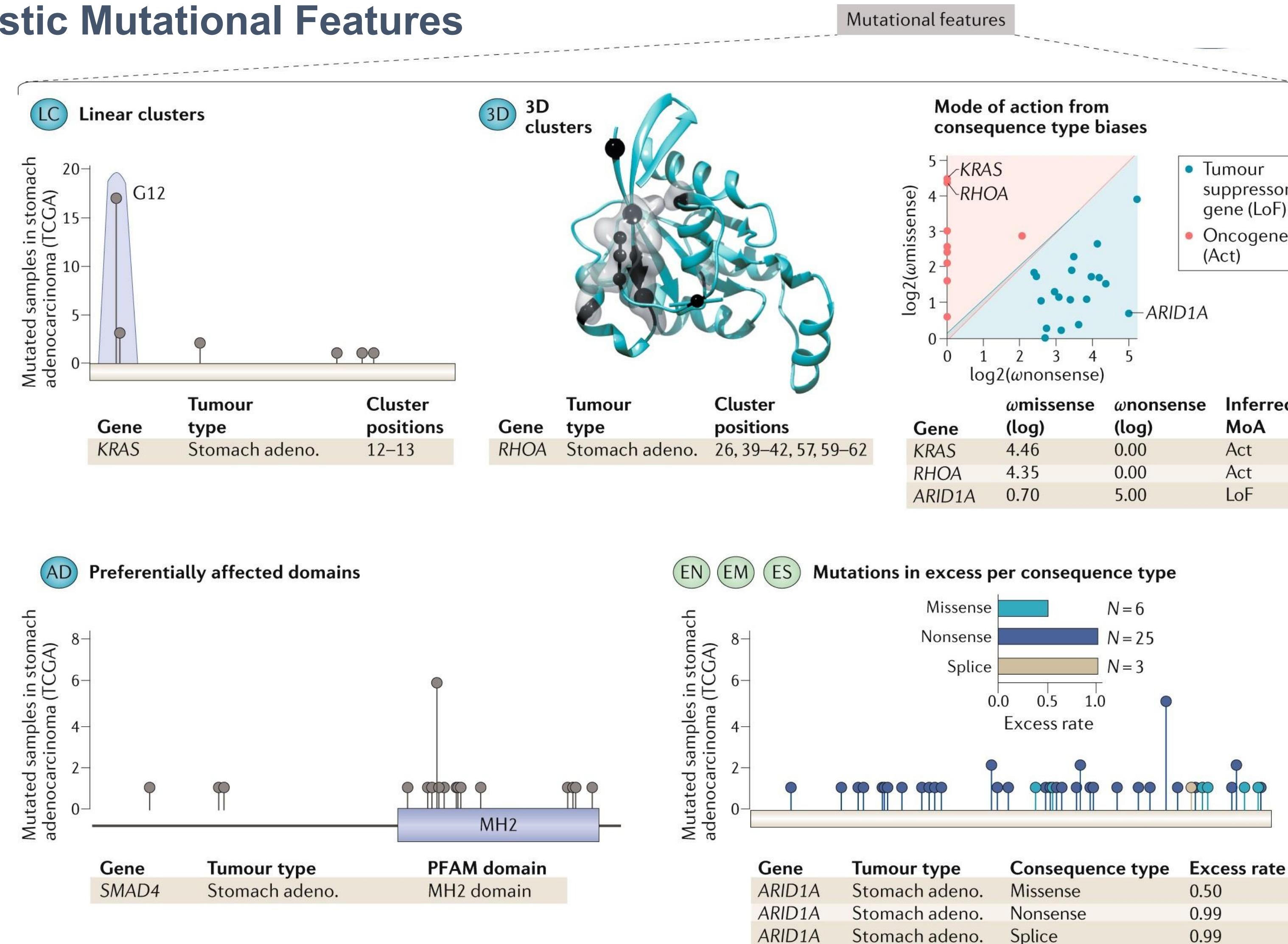
# Consensus Driver Identification Pipeline: IntOGen

## Schematic representation of the Integrative OncoGenomics (IntOGen) pipeline



# Consensus Driver Identification Pipeline: IntOGen

## Characteristic Mutational Features



# References

1. Nowell PC. **The clonal evolution of tumor cell populations.** Science. 1976 Oct 1;194(4260):23-8. doi: 10.1126/science.959840. PMID: 959840. *The first paper to describe the evolution dynamics in cancer.*
2. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, Mularoni L, Pich O, Bonet J, Kranas H, Gonzalez-Perez A, Lopez-Bigas N. **A compendium of mutational cancer driver genes.** Nat Rev Cancer. 2020 Oct;20(10):555-572. doi: 10.1038/s41568-020-0290-x. Epub 2020 Aug 10. PMID: 32778778.
3. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. **Universal Patterns of Selection in Cancer and Somatic Tissues.** Cell. 2018 Jun 14;173(7):1823. doi: 10.1016/j.cell.2018.06.001. Erratum for: Cell. 2017 Nov 16;171(5):1029-1041.e21. PMID: 29906452; PMCID: PMC6005233.
4. Rheinbay E,.....; PCAWG Drivers and Functional Interpretation Working Group; PCAWG Structural Variation Working Group; Weischenfeldt J, Beroukhim R, Martincorena I, Pedersen JS, Getz G; PCAWG Consortium. **Analyses of non-coding somatic drivers in 2,658 cancer whole genomes.** Nature. 2020 Feb;578(7793):102-111. doi: 10.1038/s41586-020-1965-x. Epub 2020 Feb 5. Erratum in: Nature. 2023 Feb;614(7948):E40. PMID: 32025015; PMCID: PMC7054214.
5. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. **IntOGen-mutations identifies cancer drivers across tumor types.** Nat Methods. 2013 Nov;10(11):1081-2. doi: 10.1038/nmeth.2642. Epub 2013 Sep 15. PMID: 24037244; PMCID: PMC5758042.
6. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, Tusquets I, Albanell J, Rodon J, Tabernero J, de Torres C, Dienstmann R, Gonzalez-Perez A, Lopez-Bigas N. **Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations.** Genome Med. 2018 Mar 28;10(1):25. doi: 10.1186/s13073-018-0531-8. PMID: 29592813; PMCID: PMC5875005.
7. Bailey MH,..... MC3 Working Group..... Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018 Apr 5;173(2):371-385.e18. doi: 10.1016/j.cell.2018.02.060. Erratum in: Cell. 2018 Aug 9;174(4):1034-1035. PMID: 29625053; PMCID: PMC6029450.