Technical University of Denmark

Sequencing in cancer genomics

Dr. Elena Papaleo
Associate Professor, Cancer Systems Biology

Email: elpap@dtu.dk

DTU Bioinformatics
Department of Bio and Health Informatics

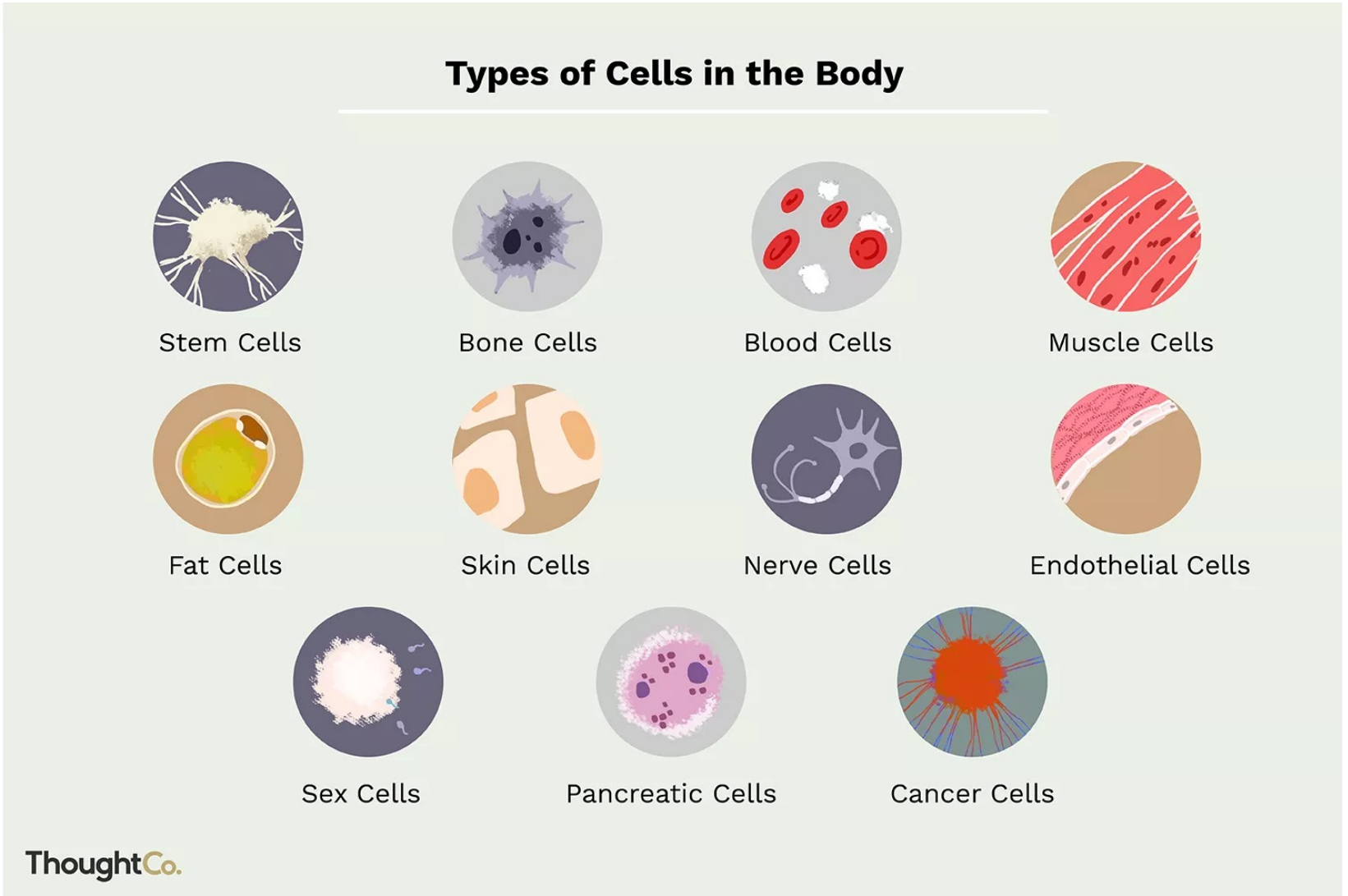
Outlines

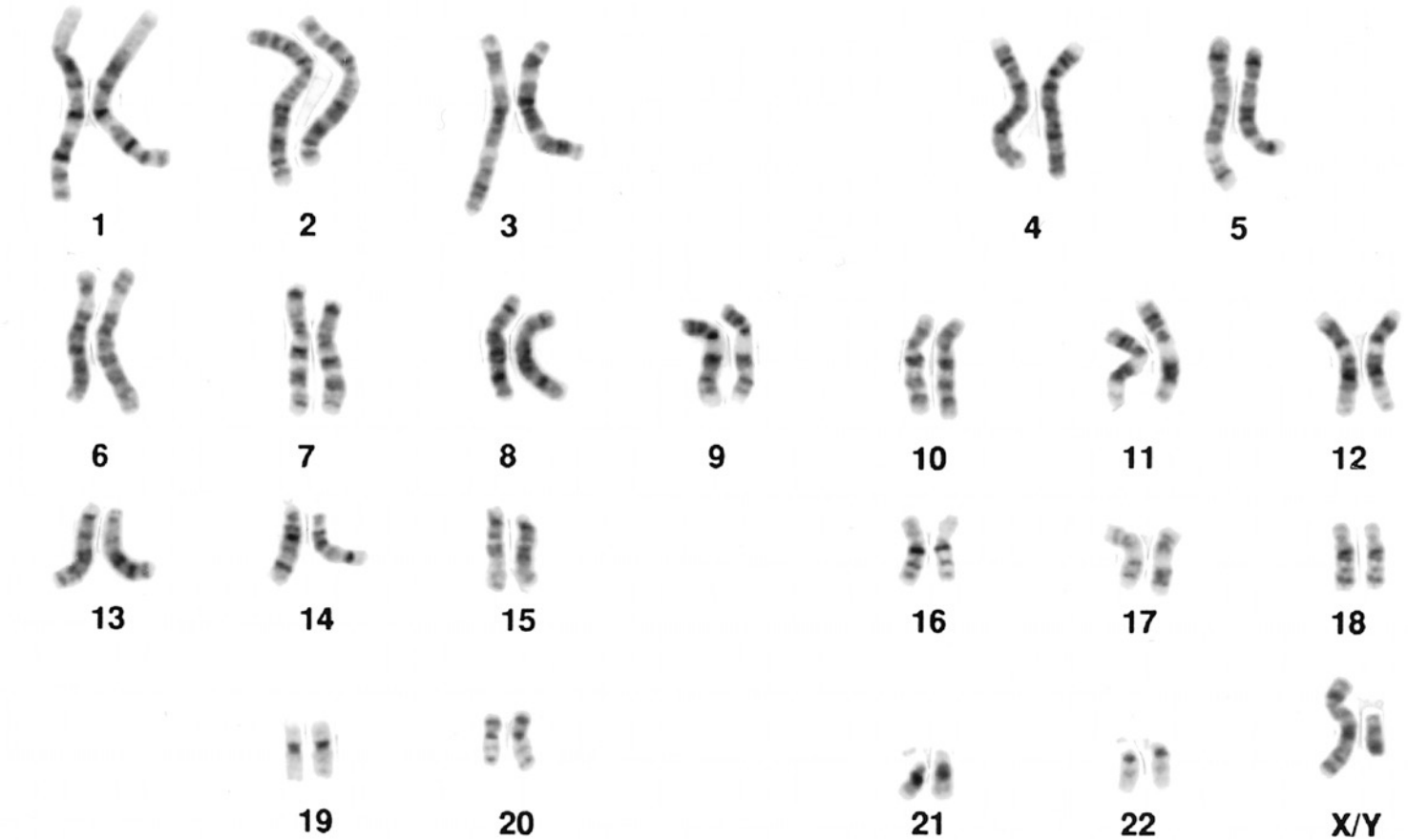
- Cancer
- Hallmarks of cancer
- Driver and passenger mutations
- Driver genes
- Exome and Whole Genome Sequencing: somatic variants
- Examples of RNA Sequencing applied to cancer research
- Databases for cancer research
- Cancer Heterogeneity

Learning objectives

- To define how cancer develops and evolves
- To understand that cancer is not just one unique disease
- To identify alterations in the cancer genome
- To analyze cancer samples using sequencing approaches (with a practical exercise)

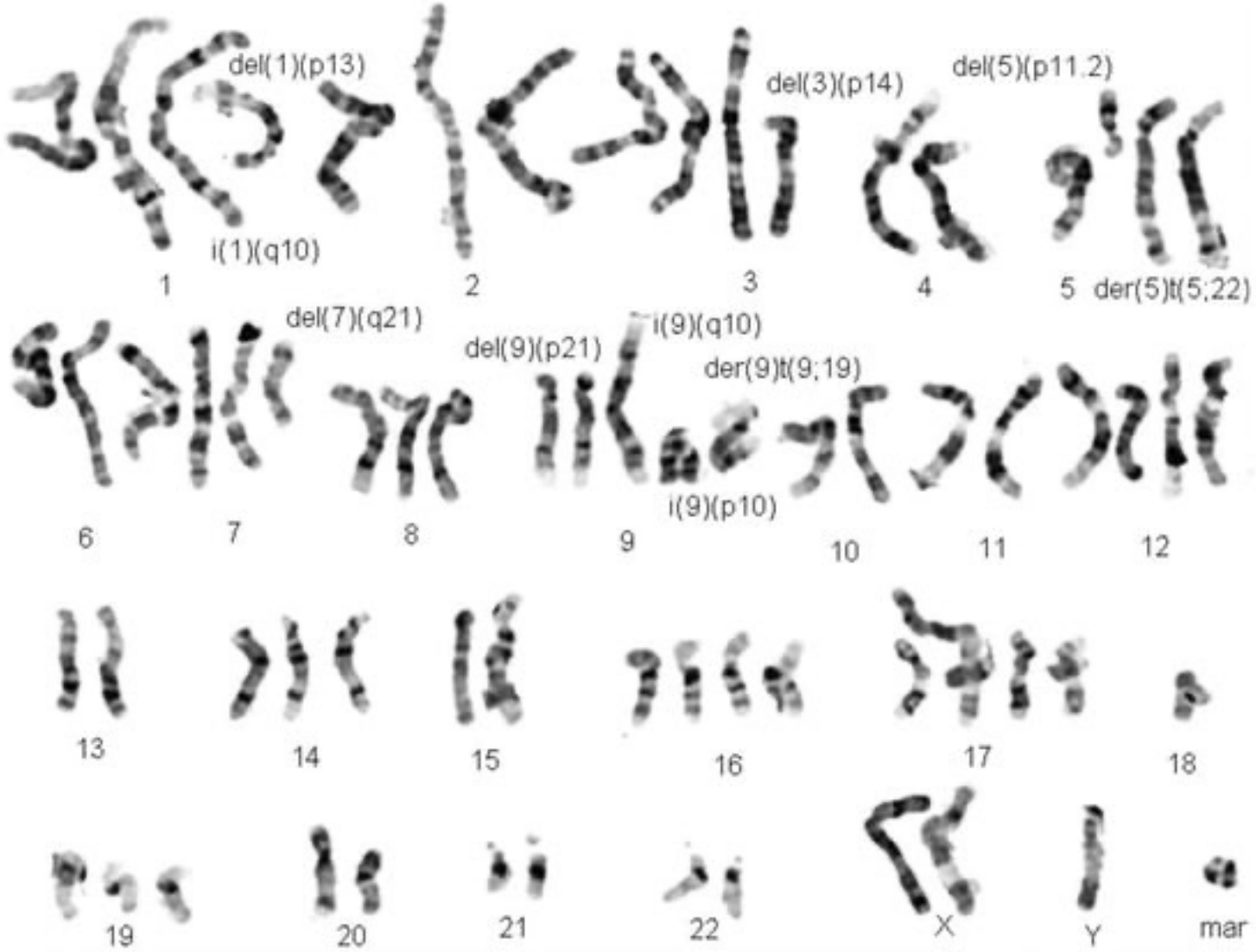
We are all made of cells





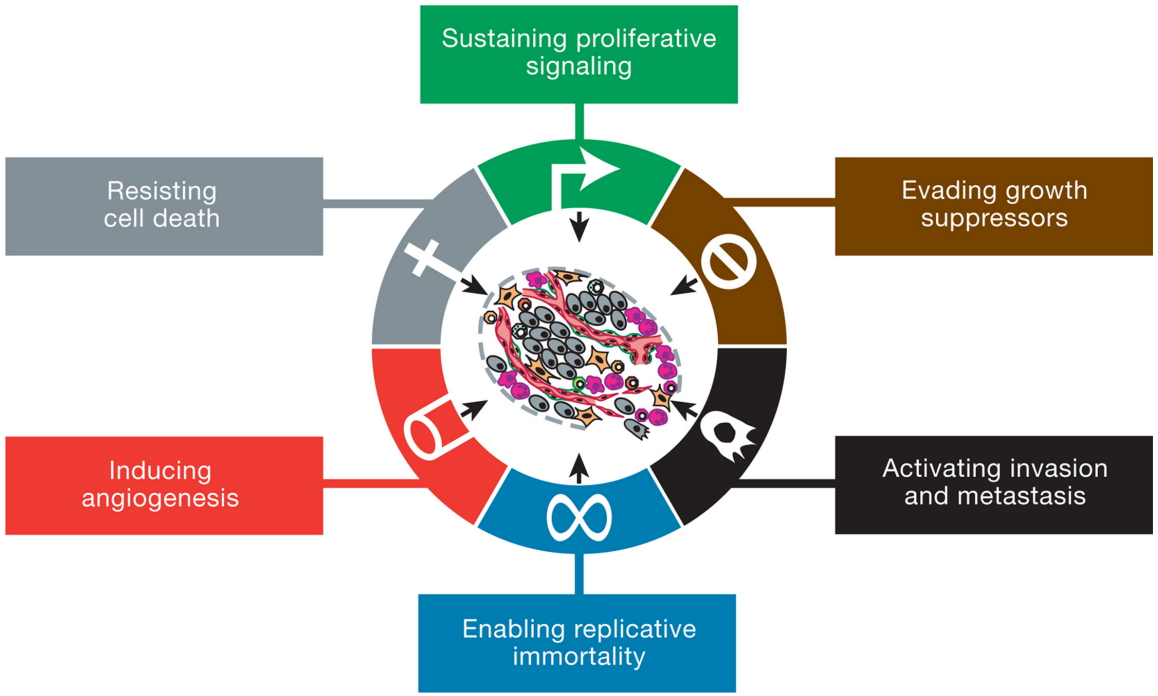
Normal Karyotype

National Cancer Institute, AV-9700-4394, 2001

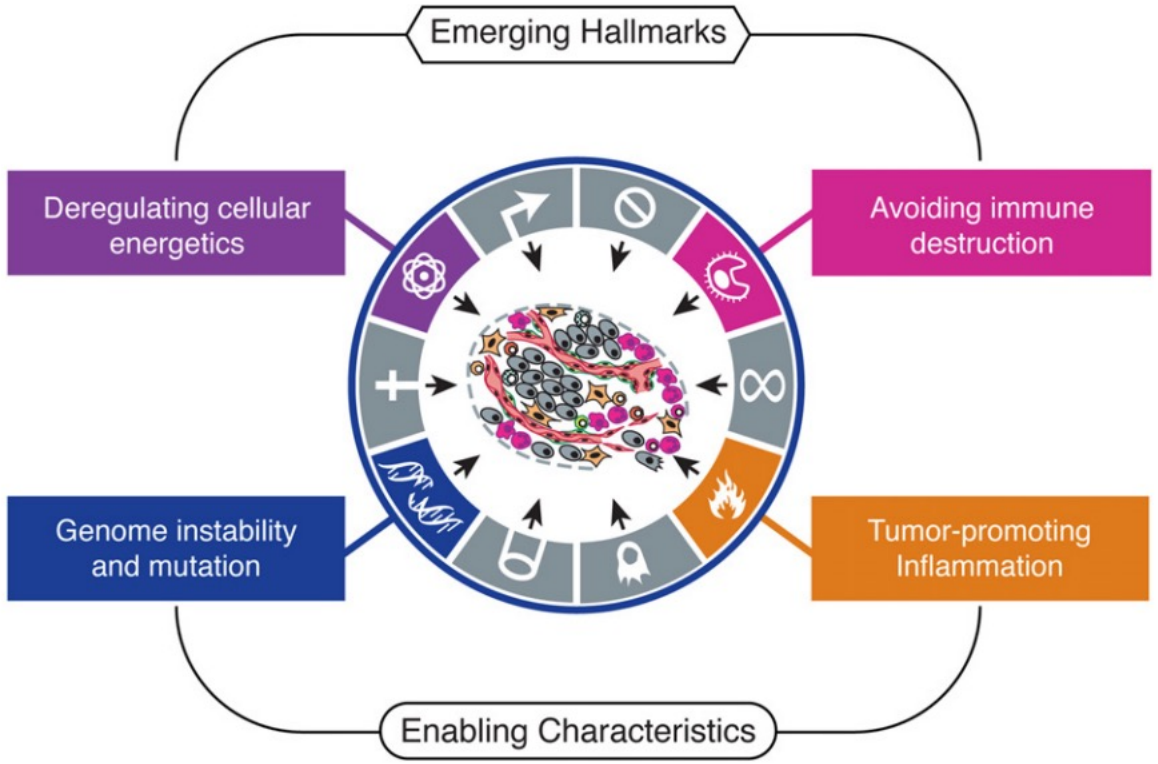


Karyogram of a representative G-banded Head and Neck Cancer metaphase (from Ribeiro et al. *Int. J. Mol. Sci.* **2019** ; https://doi.org/10.3390/ijms20194711)

Hallmarks of cancer



Weinberg and Hanahan Cell, 2000



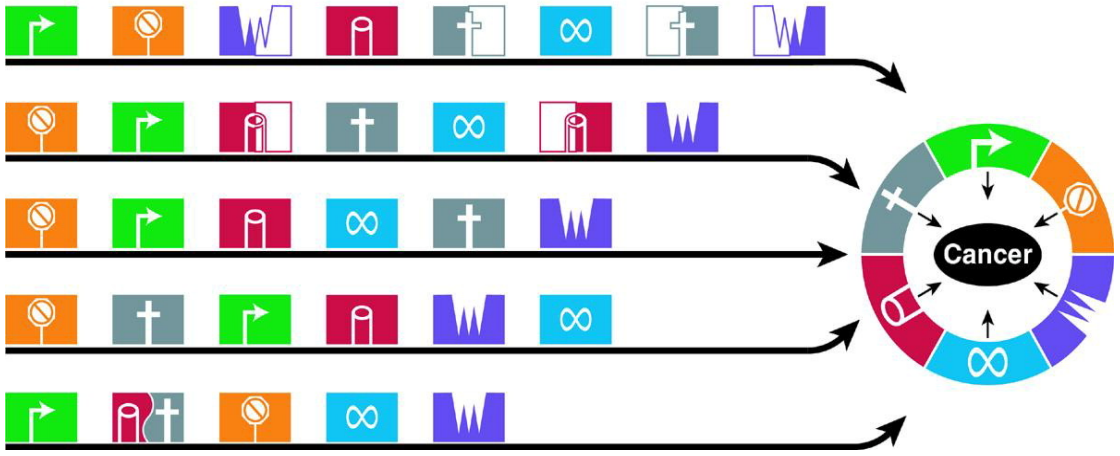
Weinberg and Hanahan Cell, 2011
Hanahan, Cancer Discovery, 2022

Multiple routes to cancer

A

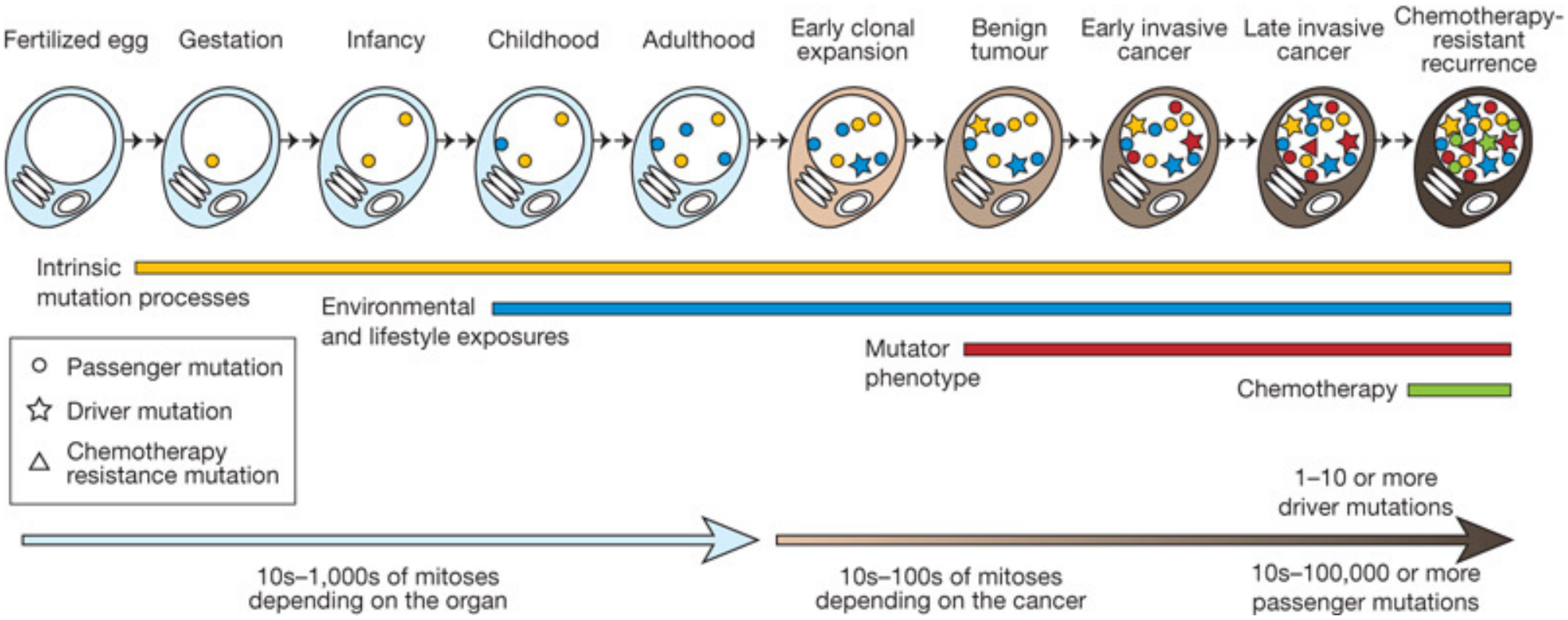
Component	Acquired Capability	Example of Mechanism
	Self-sufficiency in growth signals	Activate H-Ras oncogene
	Insensitivity to anti-growth signals	Lose retinoblastoma suppressor
	Evading apoptosis	Produce IGF survival factors
	Limitless replicative potential	Turn on telomerase
	Sustained angiogenesis	Produce VEGF inducer
	Tissue invasion & metastasis	Inactivate E-cadherin

B



Point Mutations
Insertions/ Deletions
Copy Number Variations
Structural variations
Gene Fusions
Changes in gene expression levels
Epigenetic modifications …

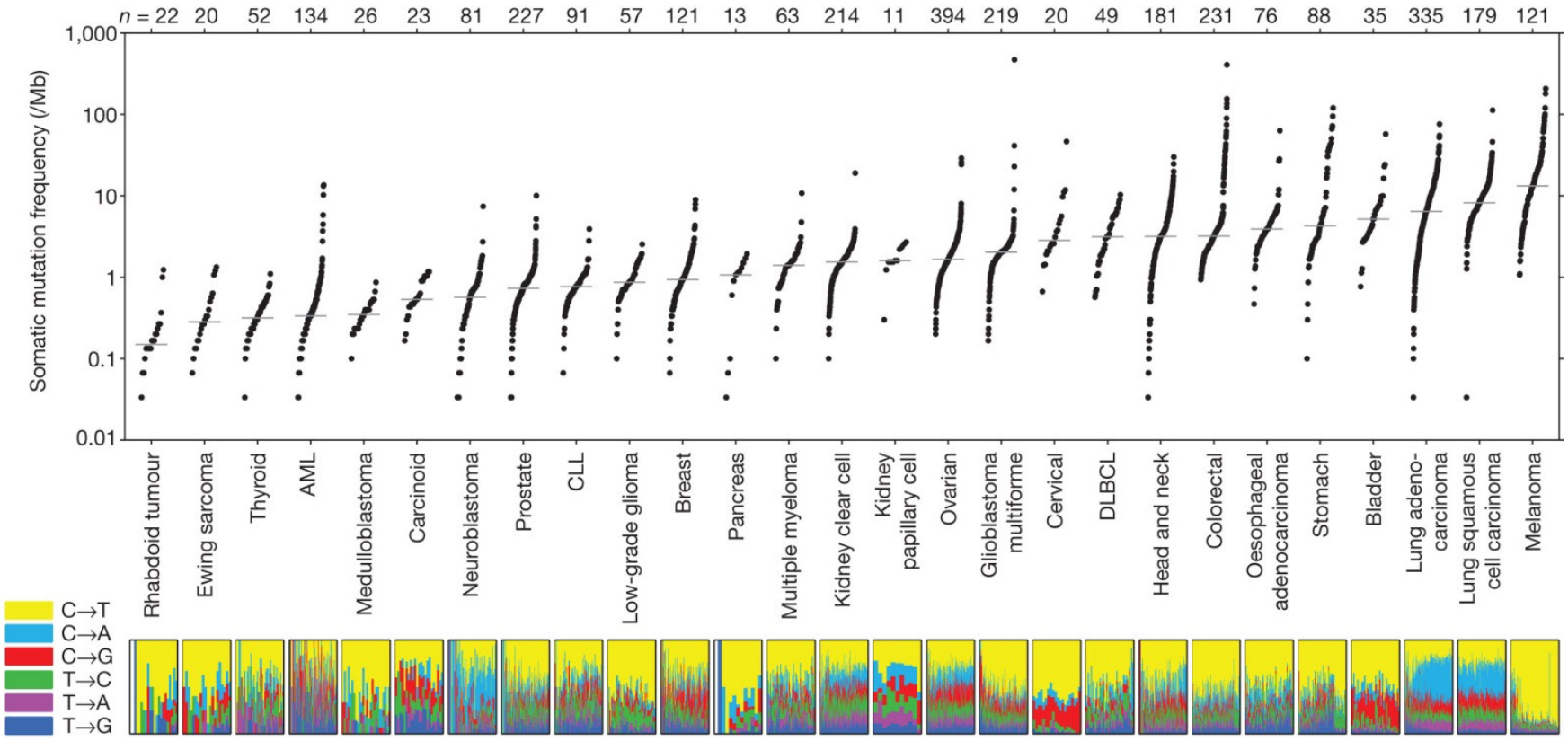
Driver and passenger mutations in cancer



MR Stratton *et al. Nature* **458**, 719-724 (2009) doi:10.1038/nature07943

- Combination of germline and somatic mutations
- Driver and passenger mutations
- Reccurence of mutations

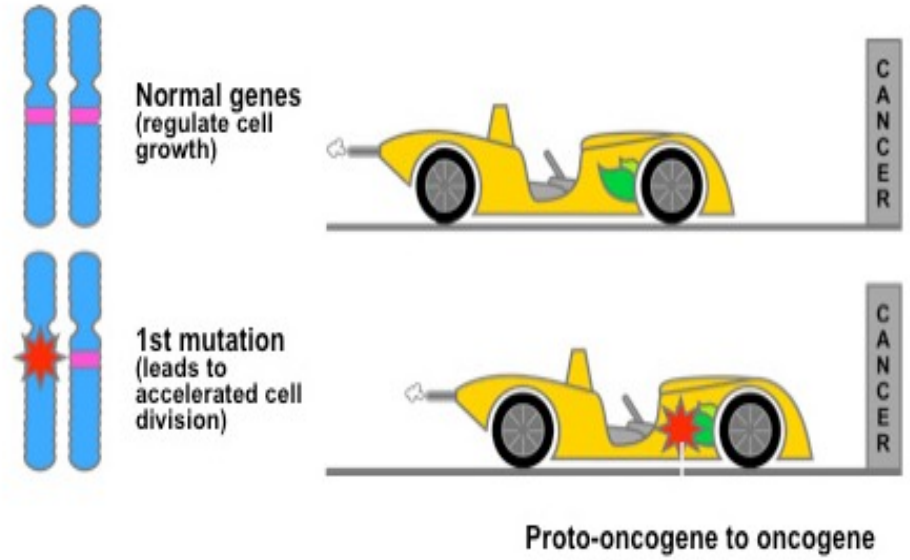
Mutation burden in different cancer types and mutational signature



MS Lawrence *et al. Nature* (2013) doi:10.1038/nature12213

Need for speed: **Oncogenes** vs Tumour suppressors

Oncogenes:
- Mutated proto-oncogenes
- Turn abnormal cell growth on
- 245 known proto-oncogenes
- gain of function genes
- primarily somatic activated
- [throttle pedal in a car]



"Oncogenes are mutated genes whose PRESENCE can stimulate the development of cancer"

Examples: HER-2/neu. RAS, MYC, SRC

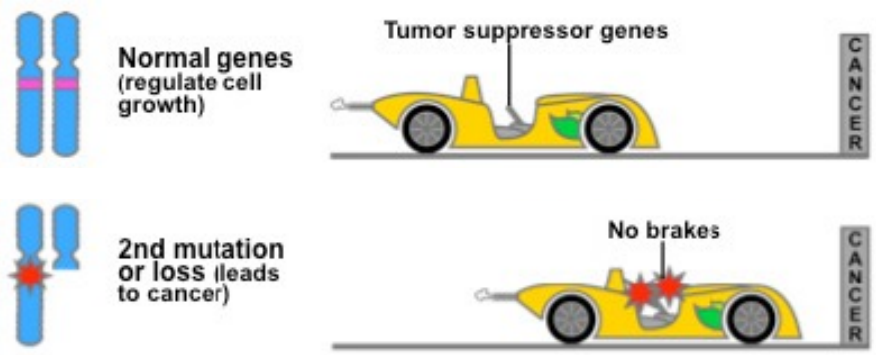
Many oncogenes are protein kinases ➔ Cell cycle regulation

(Adapted from Nat. Cancer Inst.)

Need for speed: Oncogenes vs **Tumor suppressors**

Tumour suppressor genes:
- Stop the cell cycle, G1 phase
- Slow the cell cycle before S phase
- Can induce apoptosis
- 247 known tumour suppressors

- primarily somatic de-activated
- loss-of-function mutations
- [brake pedal in a car]



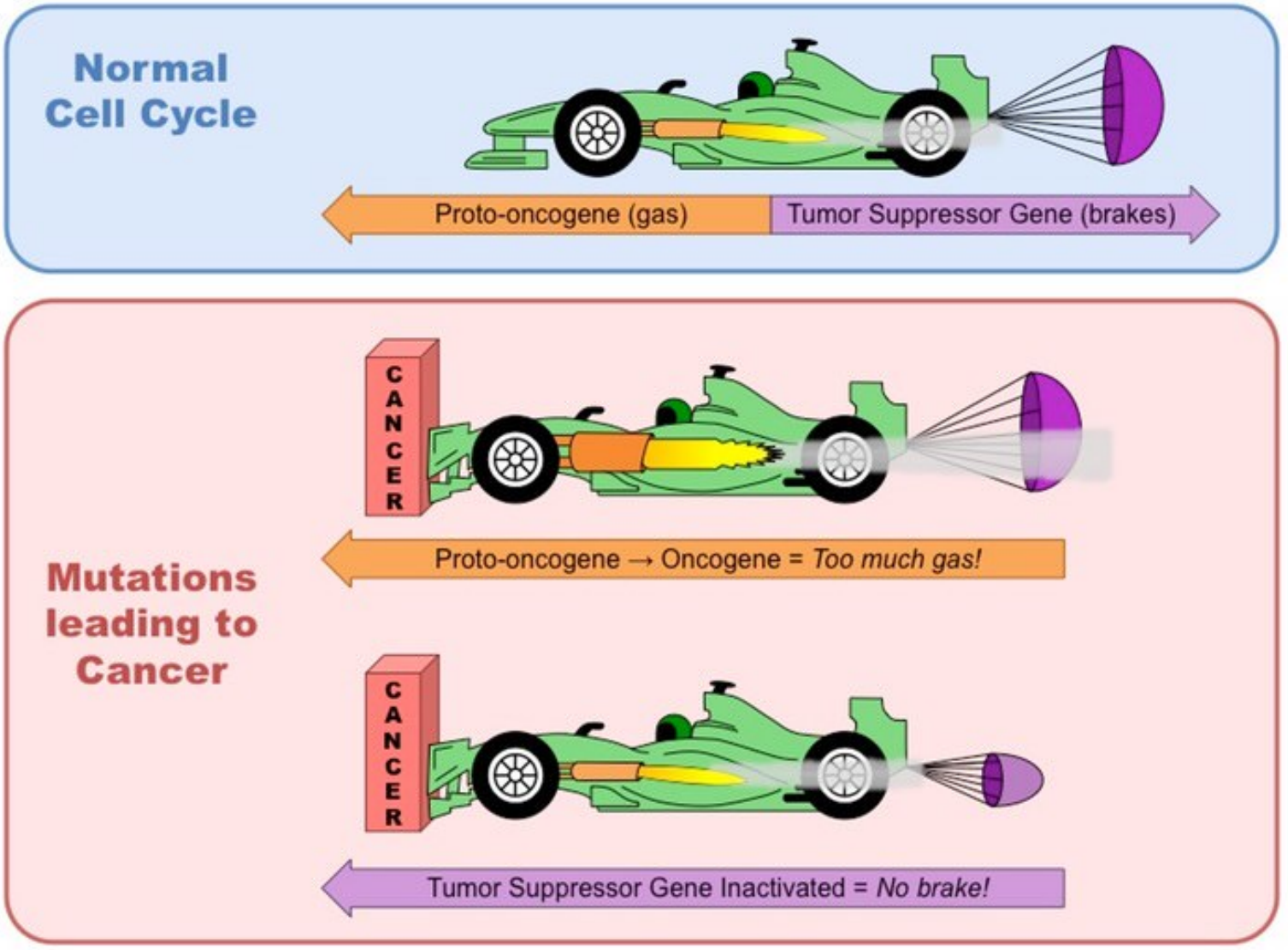
"Tumour suppressors are normal genes whose ABSENCE can stimulate the development of cancer"

Examples: p53, Rb, APC

A single functional copy (heterozygous) is enough to prevent cancer

(Adapted from Nat. Cancer Inst.)

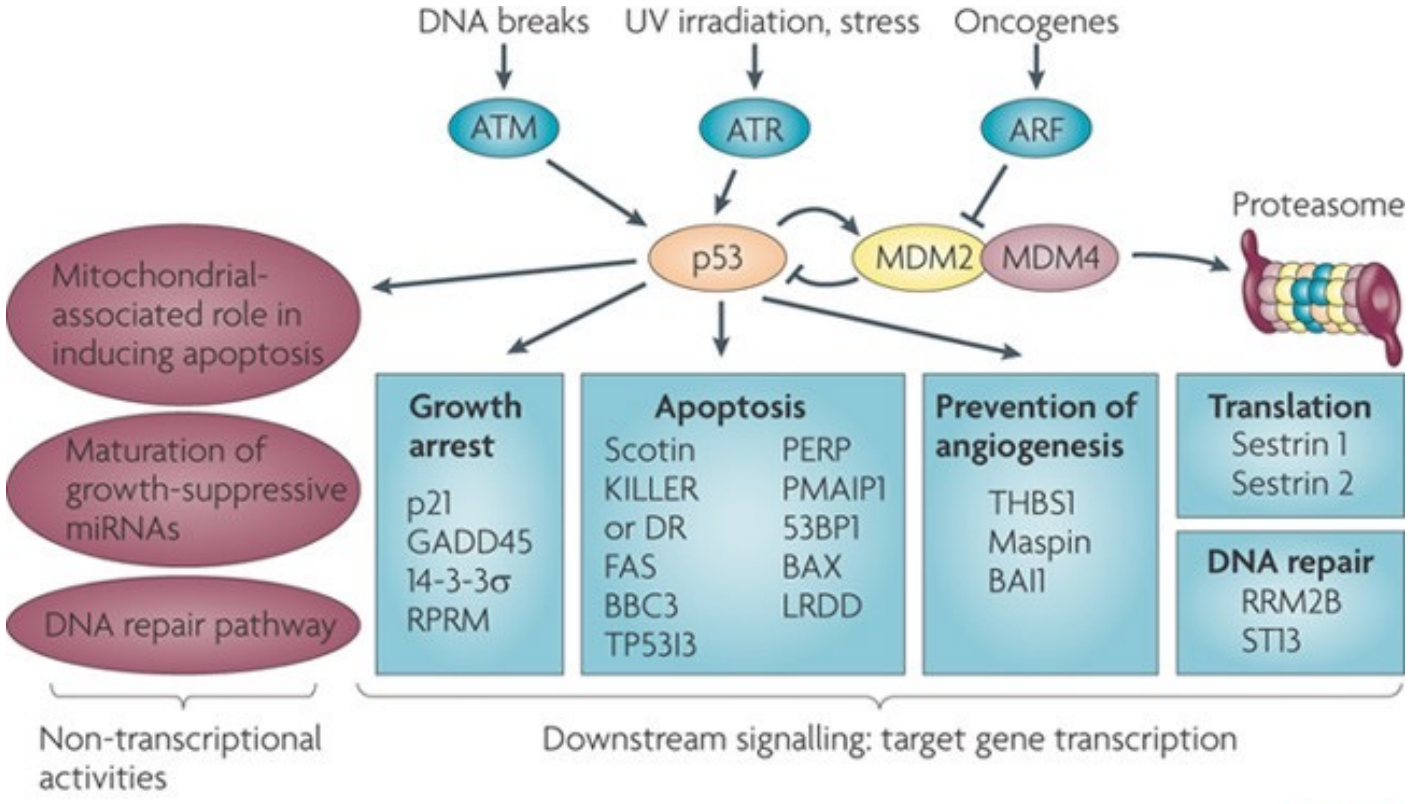
Oncogenes vs Tumor suppressors



(ib.bioninja.com.au)

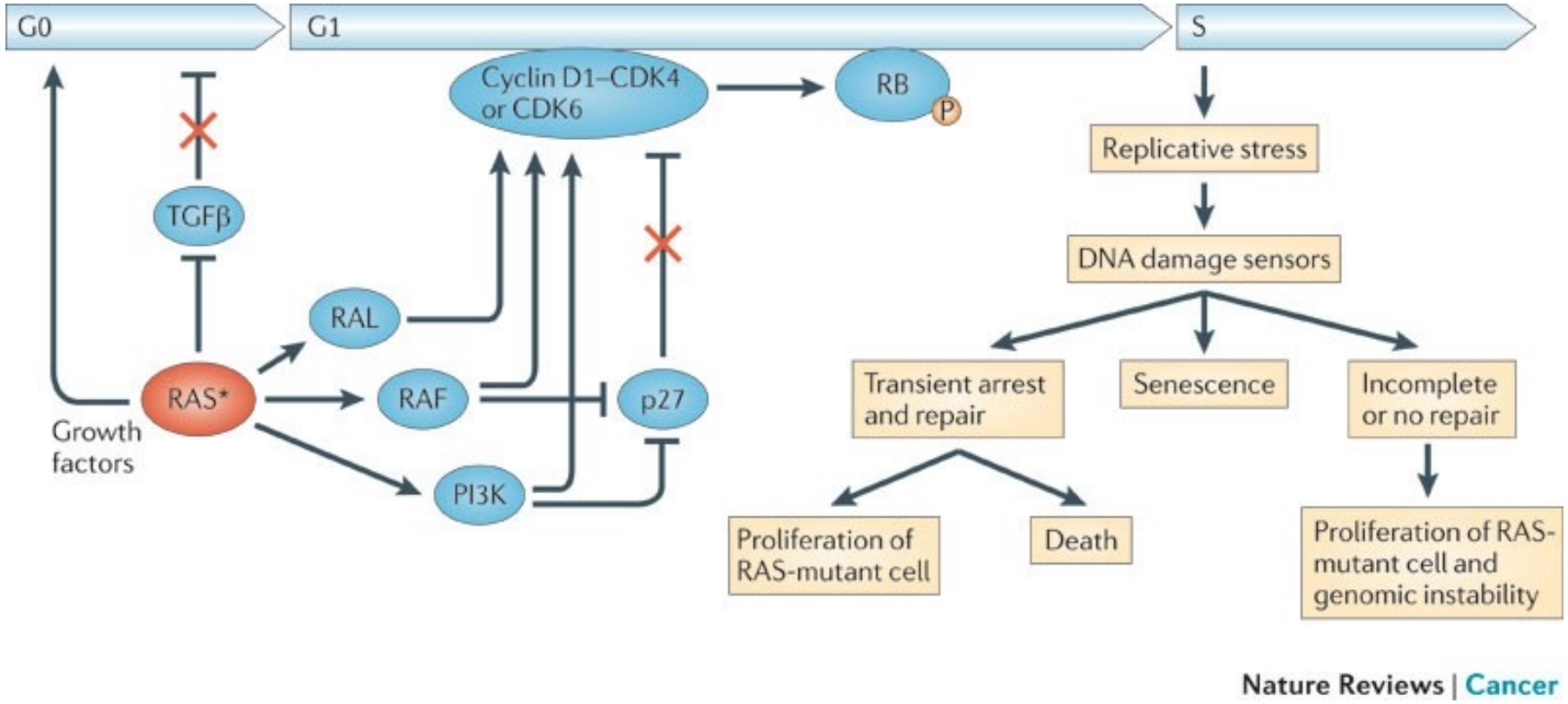
EXAMPLE OF TUMOUR SUPPRESSOR: P53

- Guardian of the human genome
- Key link between DNA damage and repair/apoptosis
- Mutations cause loss of function and promotes tumour emergence and growth



Brown CJ et al. 2009, Nature Reviews, https://doi.org/10.1038/nrc2763

EXAMPLE OF ONCOGENES: RAS family

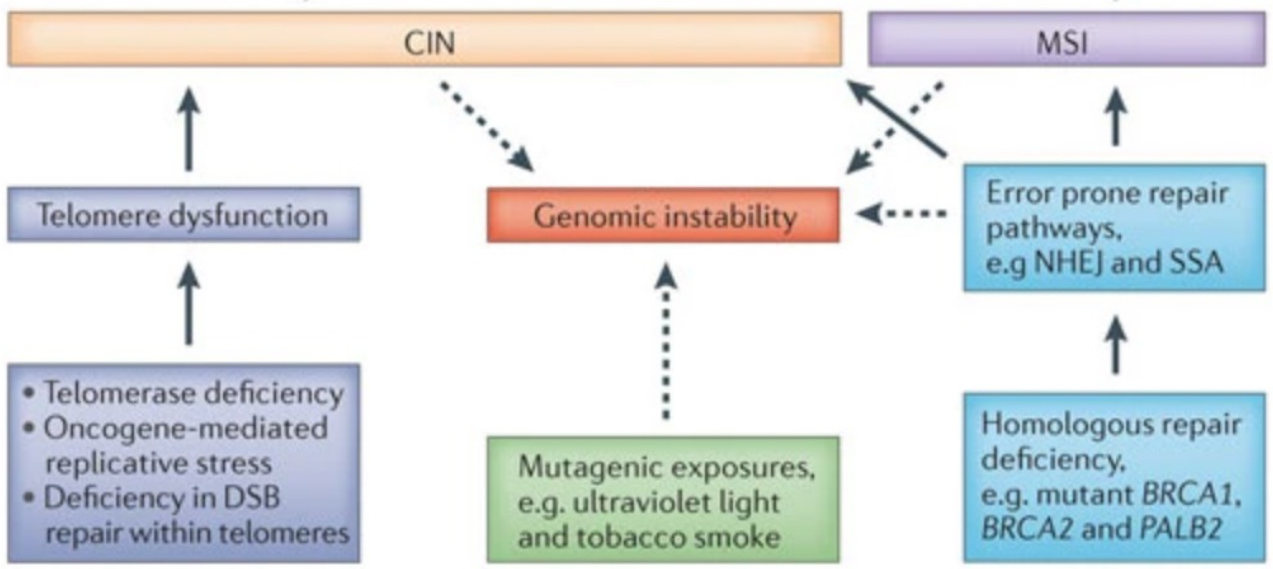


Small GTPases that activate a cascade of protein kinases involved in cell cycle and cell-growth pathways

Mutations can make them more active and overcome their regulation, sustaining continuous growth and proliferation

Mutator genes: master mutational switches

- Mutations in genes that decrease repair efficiency
- Increase of tumour mutation rate
- Involved in DNA repair, or chromatin stability

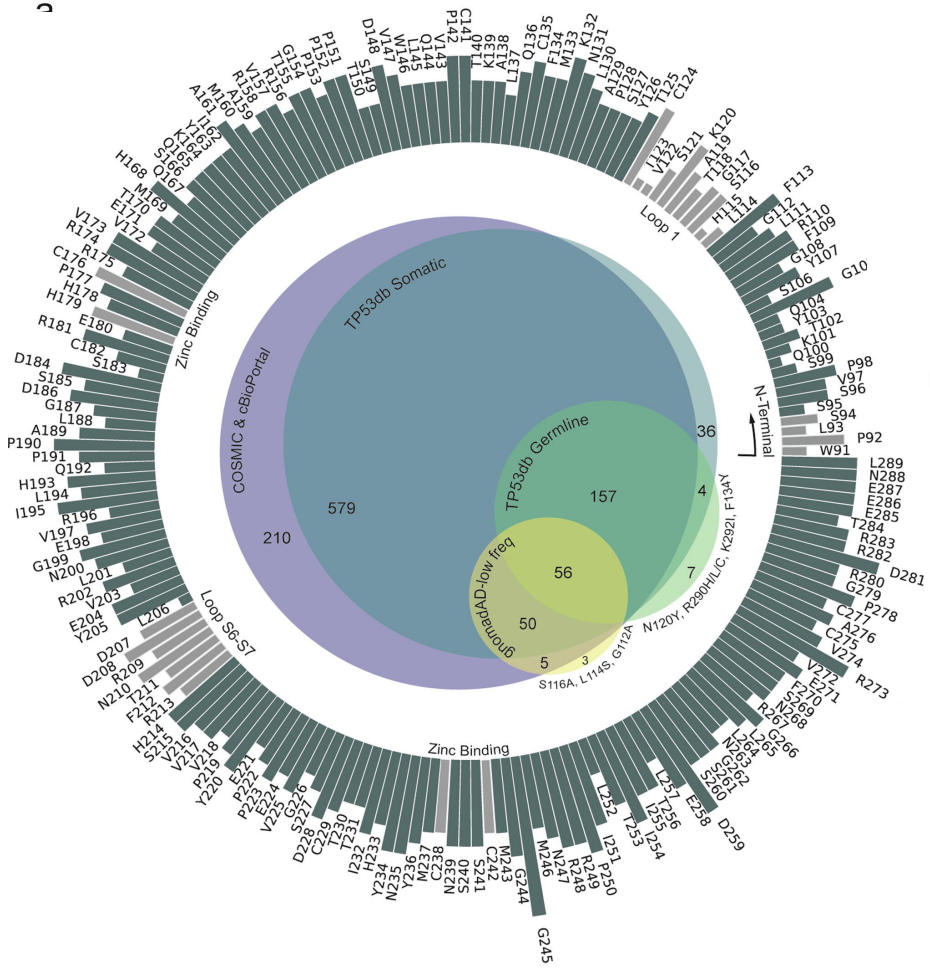


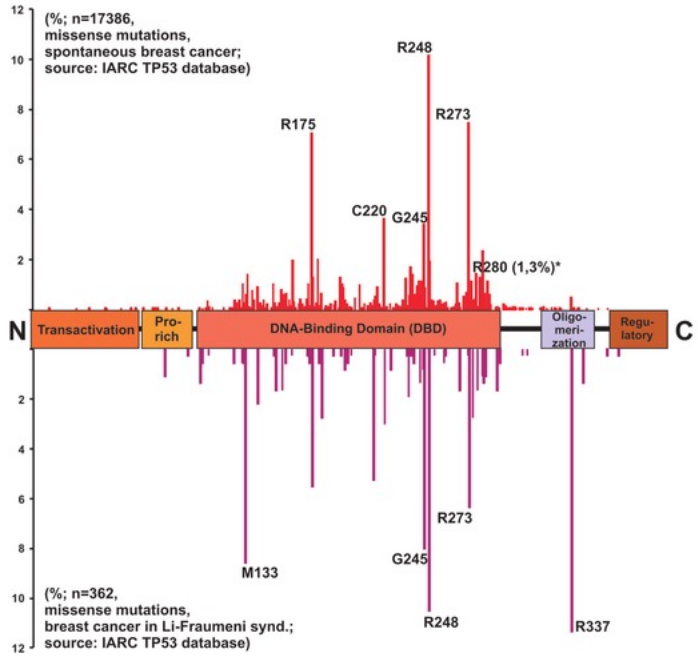
CI: Chromosome Instability
MSI: Microsatellite Instability

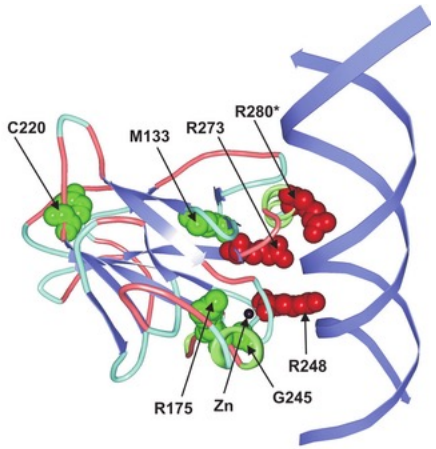
Yates and Campbell, Nat Rev Genetics, 2012, https://doi.org/10.1038/nrg3317

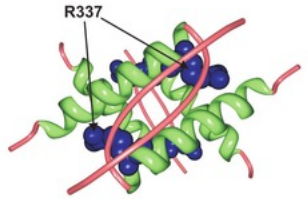
Coding variants– exome sequencing

- Example: p53 – majority of mutations are in the DNA binding domain







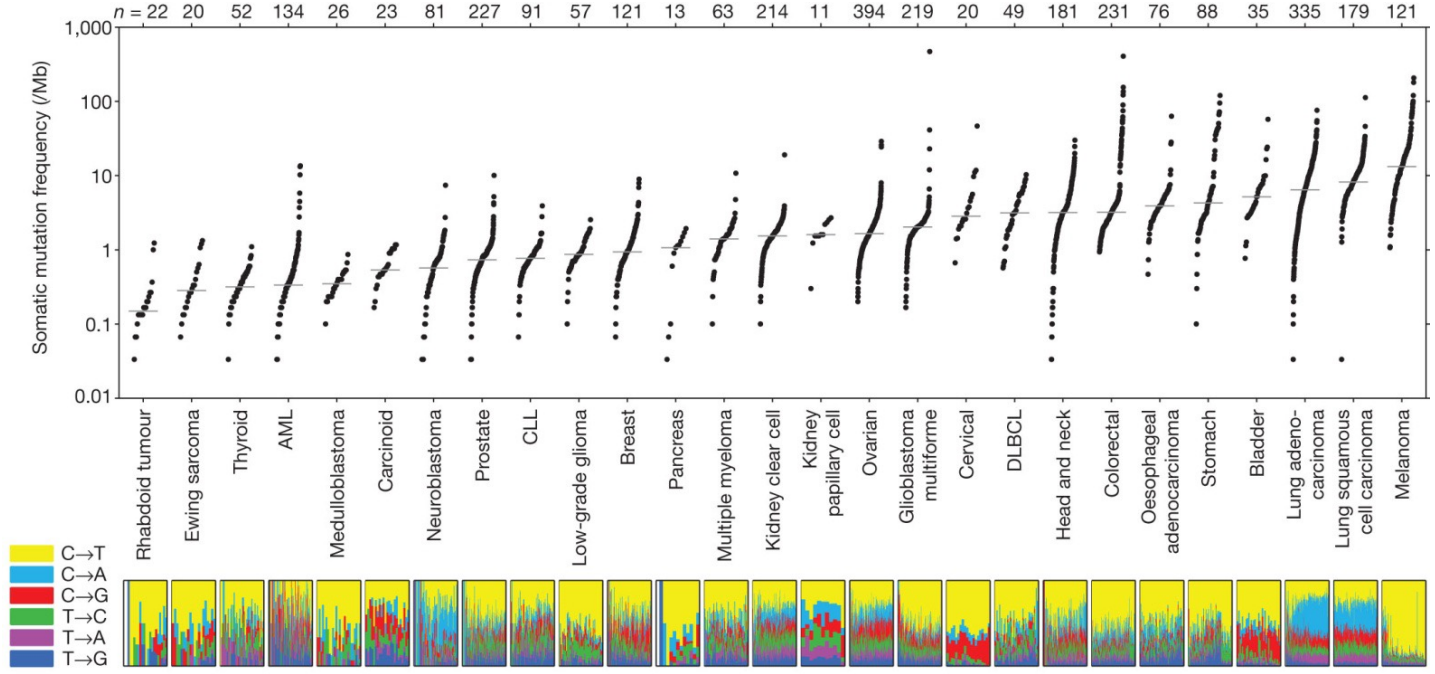


Somatic mutations and sequencing

- Paired normal and tumour samples
- Challenges due to tumour purity
- Often difficult to have sources of normal samples
- Recurrent mutations, hotspots implicated in cancer
- Clonal heterogeneity (different mutations in the same tumour)
- Challenges: low allele frequency, small fraction of reads with mutations
- Need sensitive variant calling algorithms
- MuTecT2 (bayesian framework)
- Functional annotations -> different algorithms for variant interpretation

Background mutation rate

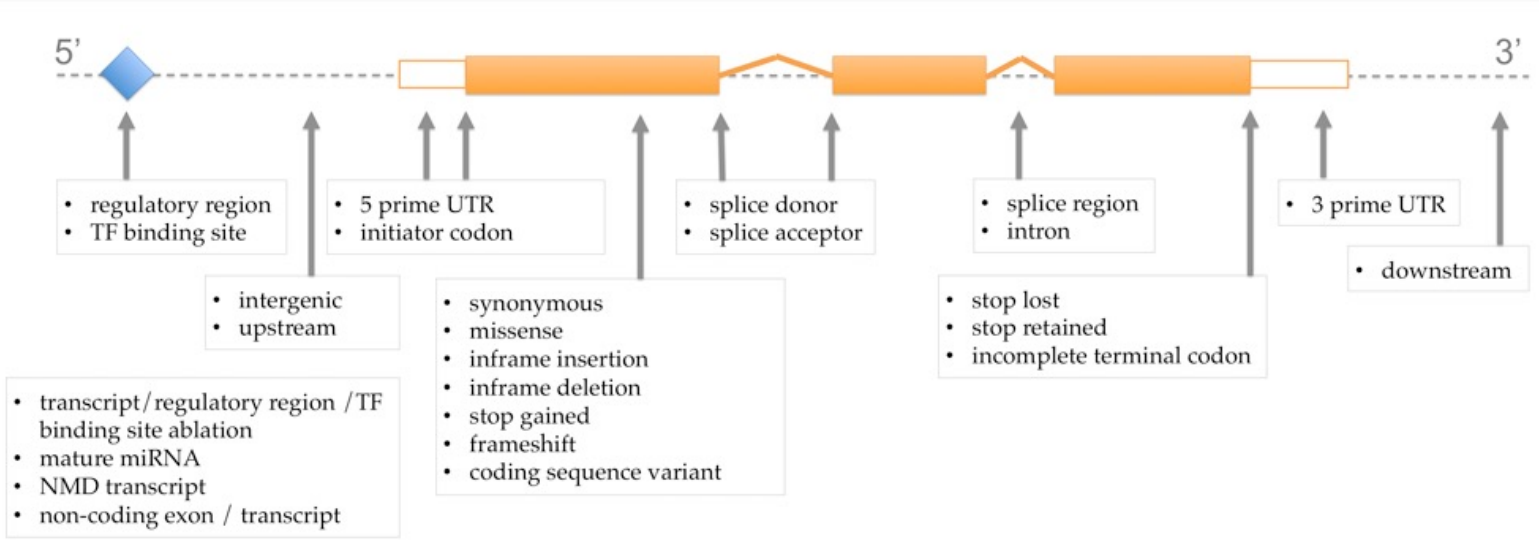
- adjust for background mutation rate
- different sources of differences in mutation rate
  - cancer-wise
  - patient-wise
  - genome region–wise



Non-coding variants and whole genome sequencing

LIFE 10e, In-Text Art, Ch. 4, p. 65
© 2014 Sinauer Associates, Inc.

Coding mutations: Affect regions that **translate** to protein
Non-coding mutations: Affect regions **that do not translate** to protein

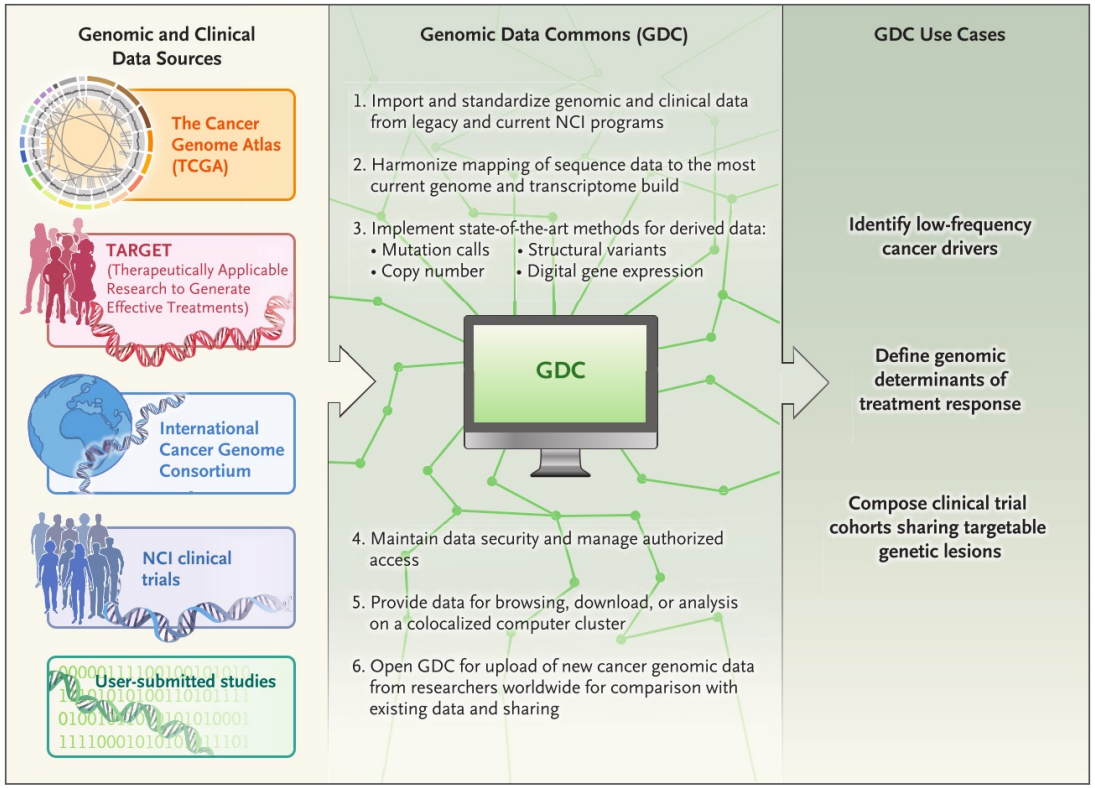


Non-coding variants and whole genome sequencing

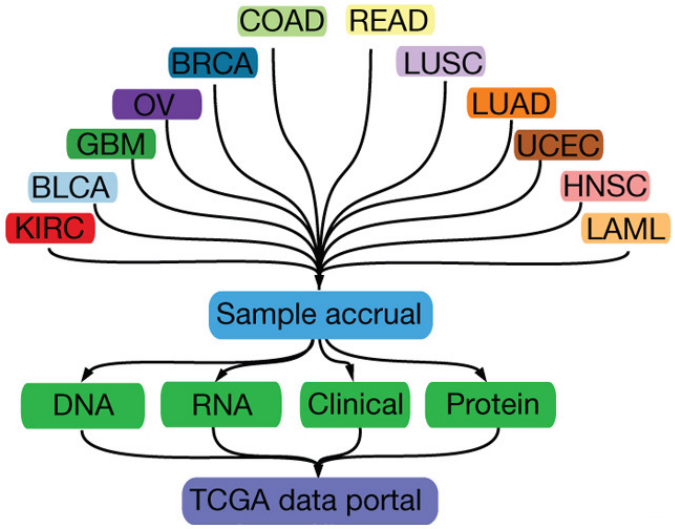
- Even more challenging to find drivers
- Recurrence principle can be used similar to coding
- Variable background mutational rates within non-coding regions
- Scattered mutations
- Little knowledge of their functional impact

(Pan)cancer genomic initiatives

THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute



NATIONAL CANCER INSTITUTE
GDC Data Portal



- A source of genomic profiling in cancer-patients across > 30 different cancer studies
- Similar approaches to other data from –omics initiatives
- Need of tools to understand the impact of the alterations at the molecular level
- Different layers of information

COSMIC
Catalogue Of Somatic Mutations In Cancer

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

COSMIC
The core of COSMIC, an expert-curated database of somatic mutations

Cell Lines Project
Mutation profiles of over 1,000 cell lines used in cancer research

COSMIC-3D
An interactive view of cancer mutations in the context of 3D structures

Cancer Gene Census
A catalogue of genes with mutations that are causally implicated in cancer

Cancer Mutation Census
Classification of genetic variants driving cancer

Actionability
Mutations actionable in precision oncology

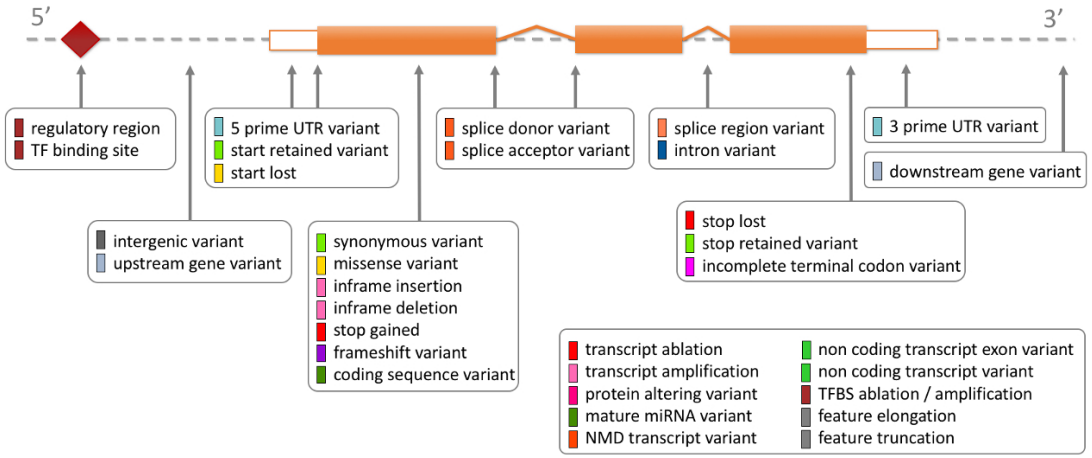
Pipelines for annotation of variants (Oncotator)

Table 1. Oncotator Datasources

Annotation Category	Resource	URL	Comments
Genomic	GENCODE	http://www.gencodegenes.org/	GENCODE/ENSEMBL transcripts and annotations for hg19
	ref_context		Can be used for artifact inference
	gc_content		Can be used for artifact inference
	Human DNA Repair Genes	http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html	Alteration in such genes can help explain higher overall mutation rates in specific samples
Protein	UniProt	http://www.uniprot.org/	Includes Drugbank & GO annotations
	dbNSFP	https://sites.google.com/site/jpopgen/dbNSFP	Contains pre-computed conservation scores, prediction classifications, and other information
Cancer Variant	COSMIC	http://www.sanger.ac.uk/genetics/CGP/cosmic/	
	Cancer Gene Census	http://www.sanger.ac.uk/genetics/CGP/Census/	
	CCLE	http://www.broadinstitute.org/ccle/home	Cancer cell line annotations. Can be used to identify cell line models containing variants of interest
	Familial Cancer Database	http://www.familialcancerdatabase.nl/	
	ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/	
Non-Cancer Variant	dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/	b142 release for human (9606)
	1000 Genomes	http://www.1000genomes.org/data	Phase 3 variant set
	NHLBI GO Exome Sequencing Project (ESP)	https://esp.gs.washington.edu/drupal/	

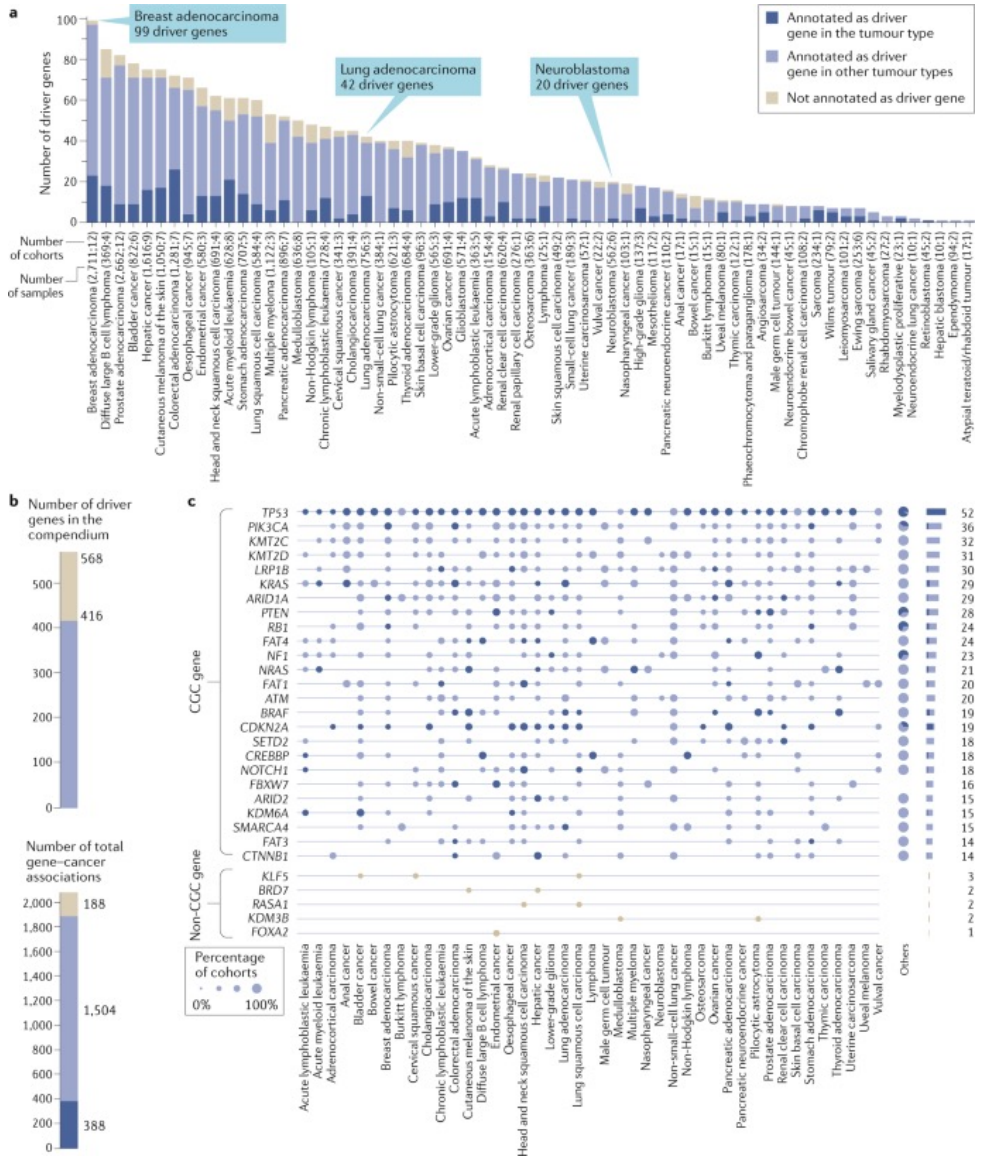
Ramos et al. 2015, Human Mutation, **https://doi.org/10.1002/humu.22771**

Ensembl Annotation for CONSEQUENCES



* SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice region variant	LOW
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Incomplete terminal codon variant	LOW
start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains	SO:0002019	Start retained variant	LOW

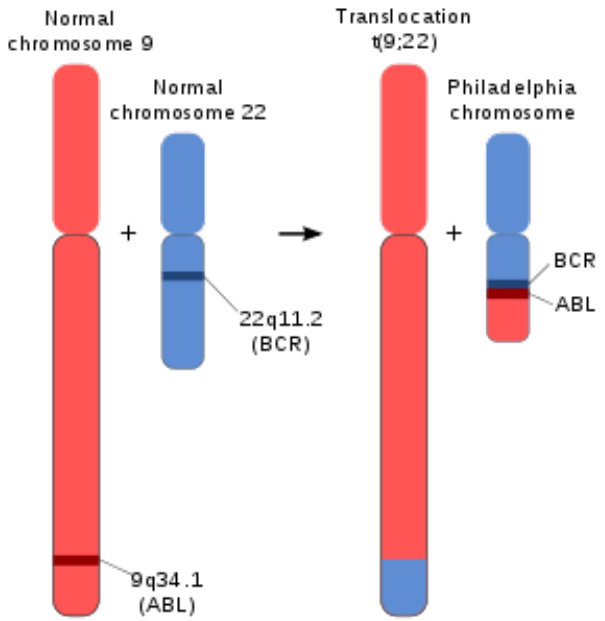
Large-scale cancer sequencing analyses and recurrent mutations



Martínez-Jiménez, F., Muiños, F., Sentís, I. *et al. Nat Rev Cancer* **20**, 555–572 (2020). https://doi.org/10.1038/s41568-020-0290-x

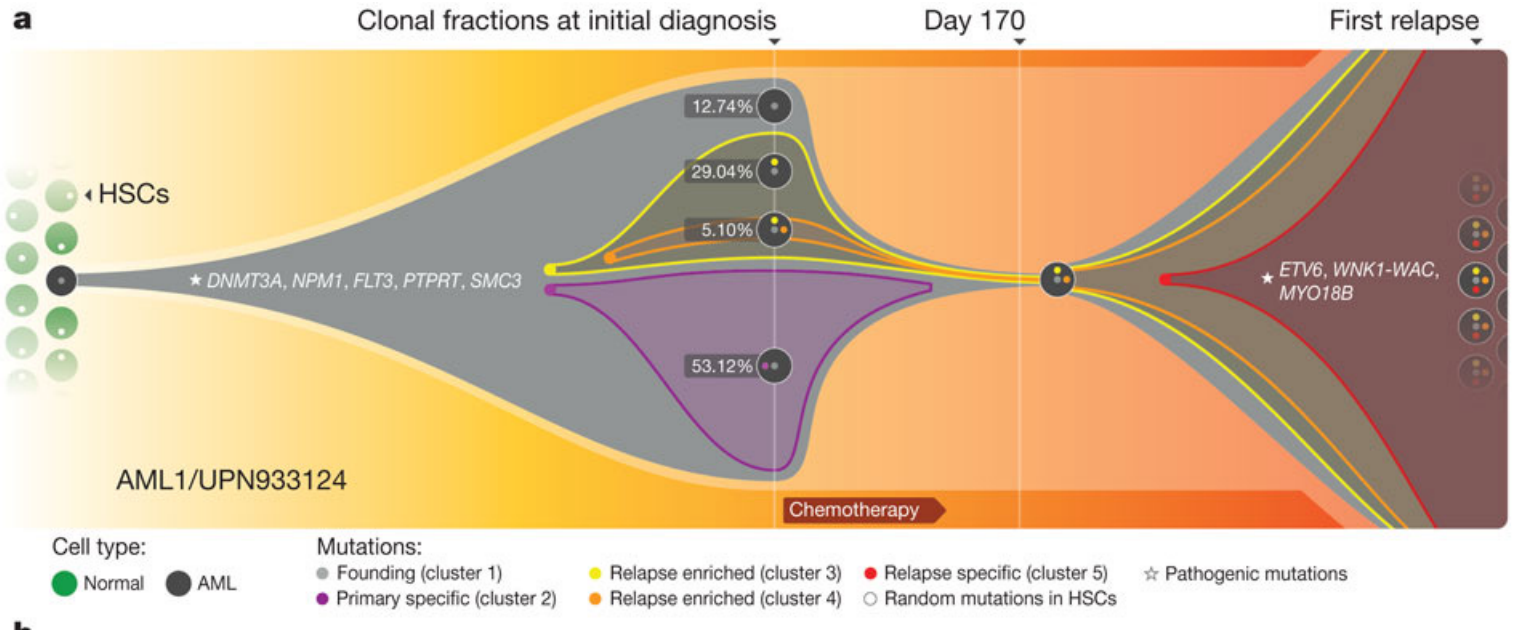
RNASEQ to identify gene fusions

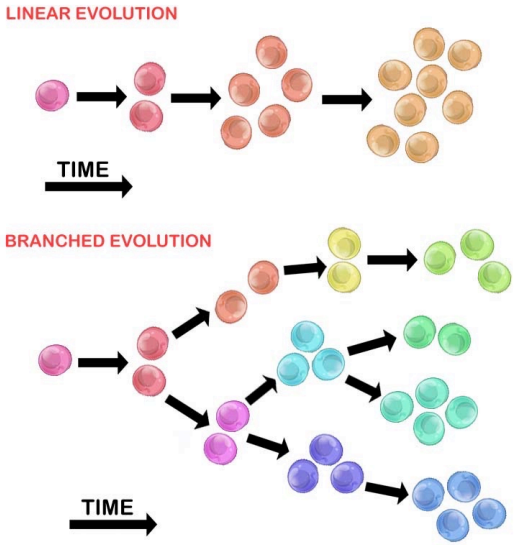
- BCR-ABL fusion and Philadephia chromosome in leukemia
- Tumours driven by recurrent fusion events that create chimeric proteins that serve as oncogenes
- Useful as markers or for decisions on therapy



Clonal genomic heterogeneity

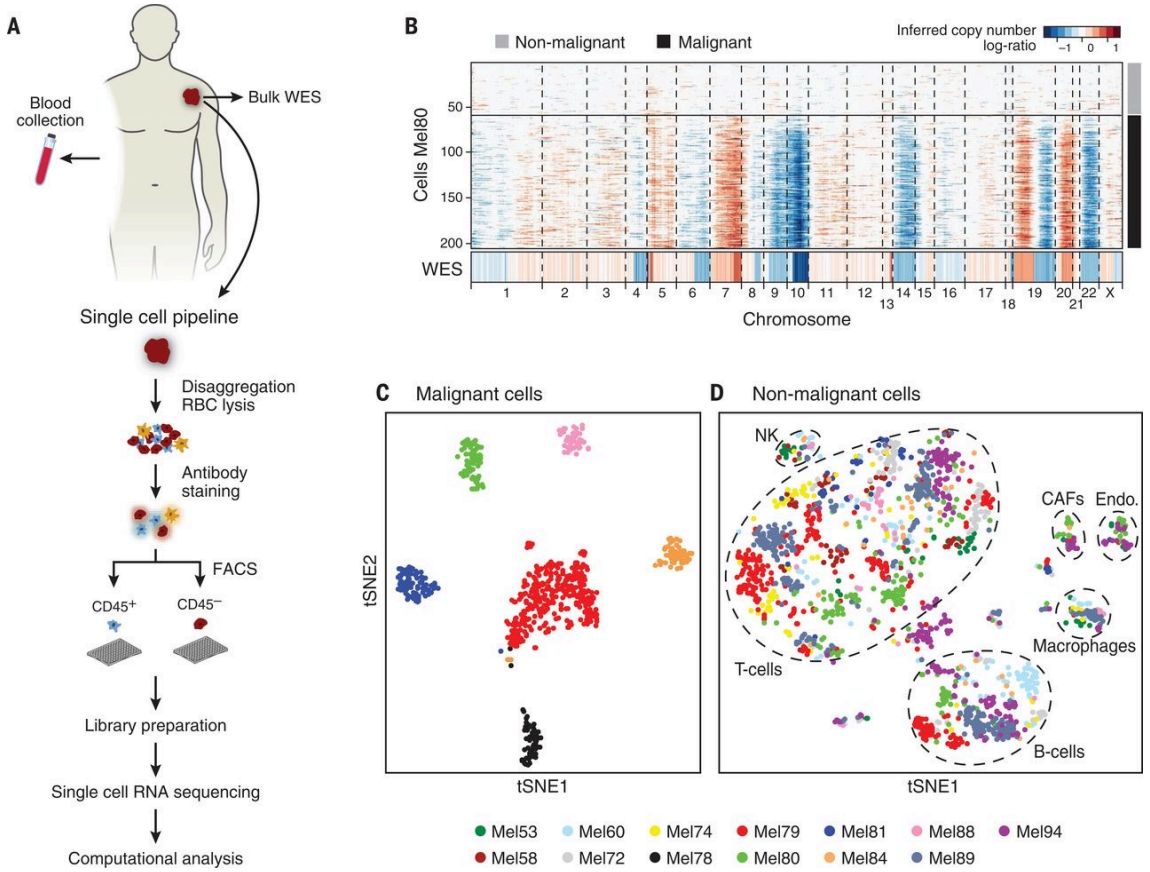
- Cancer is an evolutionary process driven by positive selection
- Large number of pre-cancerous cells constantly subjected to selection
- Many way a oncogenic pathway can be hit
- Genomic instability key source of heterogeneity
- Heterogeneity both in emergence of tumour and resistance to drugs





Ding et al. Nature, 2012, https://doi.org/10.1038/nature10738

Single cell techniques to study diversity of clonal groups



Tirosh et al. 2016 Science, [DOI: 10.1126/science.aad0501](https://doi.org/10.1126/science.aad0501)

Learning objectives

- To define how cancer develops and evolves
- To understand that cancer is not just one unique disease
- To identify alterations in the cancer genome
- To analyze cancer samples using sequencing approaches