CANCER GENOMICS DATA AND ANALYSIS CONSIDERATIONS

David Tamborero, PhD
Senior researcher
Karolinska Institutet, ScilifeLab

CANCER GENES

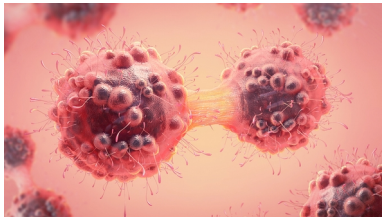
Cancer genes are those that can drive tumor phenotypes upon the occurrence of certain (genomic) alterations

Cancer genes are those that can drive tumor phenotypes upon the occurrence of certain (genomic) alterations

Oncogenes are **growth-promoting** genes

Genomic alterations cause that these genes **increase** their (tumor-promoting) activity without a physiological cause

These alterations are ‘dominant’ events referred as **gain-of-function***



** note that this concepts apply to any other genomic-driven disease*

How a DNA alteration can drive **gain-of-function** effects in oncogenes?

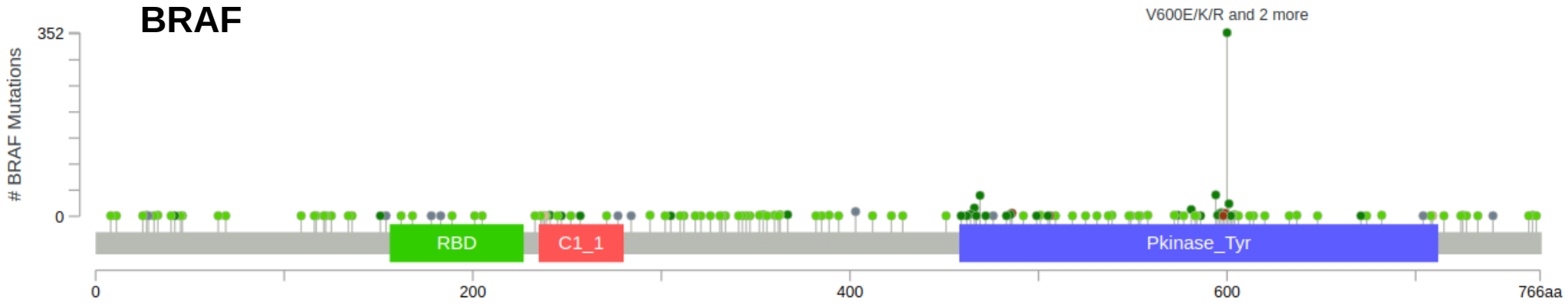
With very **specific** alterations that **up-regulate** the gene function
(e.g. a traslocation of a kinase domain with an over-active promoter of another gene)

With very **specific** mutations that change the properties of the oncogene
(e.g. a mutation leading to constitutive active form of the protein)

How a DNA alteration can drive **gain-of-function** effects in oncogenes?

With very **specific** alterations that **up-regulate** the gene function
(e.g. a traslocation of a kinase domain with an over-active promoter of another gene)

With very **specific** mutations that change the properties of the oncogene
(e.g. a mutation leading to constitutive active form of the protein)



Distribution of BRAF mutations across sequenced tumors

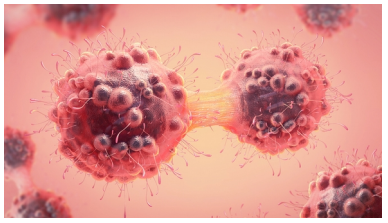
Tumors select mutations in certain protein residues of the oncogenes leading to very specific gain-of-function properties

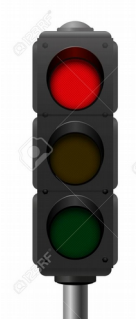
Cancer genes are those that can drive tumor phenotypes upon the occurrence of certain alterations in their DNA code

Tumor suppressors are **growth-controlling** genes

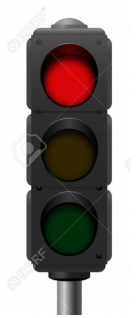
Gemomic alterations cause that these genes **decrease** their (cancer-control) activity without a physiological cause

These alterations are referred as **loss-of-function,** which must affect the two alleles to be functional (**two-hit**) except e.g. genes that demonstrate **haploinsufficiency (**or **dominant negative** events)





How a gene alteration can drive **loss-of-function** effects?

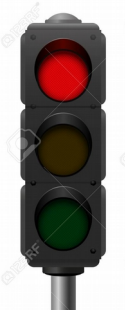


- With any alterations that **down-regulate** the gene function *(e.g. a traslocation that disrupts the open-reading frame, or a gene deletion)*

- With **specific** (missense) mutations able to cause loss-of-function effects

- With **any** mutation that **disrupts** the protein translation

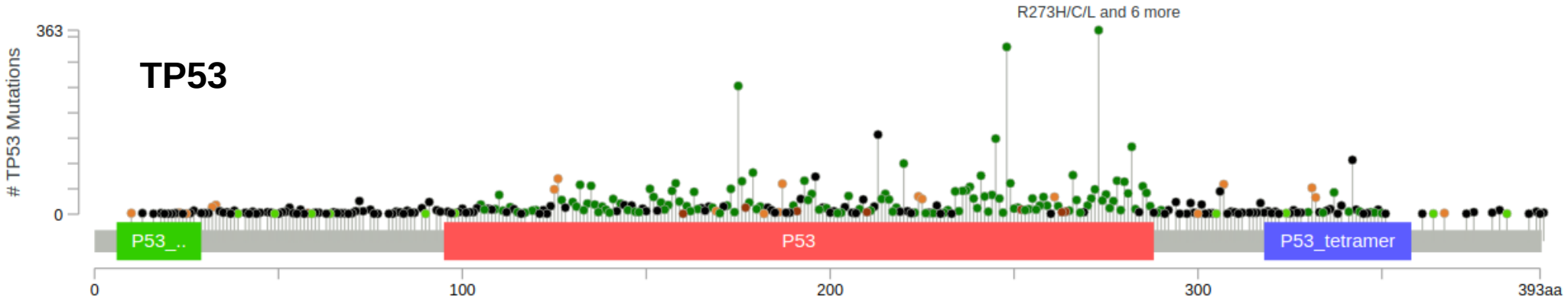
How a gene alteration can drive **loss-of-function** effects?



- With any alterations that **down-regulate** the gene function *(e.g. a traslocation that disrupts the open-reading frame, or a gene deletion)*

- With **specific** (missense) mutations able to cause loss-of-function effects

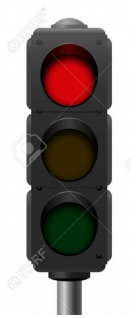
- With **any** mutation that **disrupts** the protein translation



Distribution of TP53 mutations across sequenced tumors

Very specific mutations that disrupts wildtype function (e.g. affecting protein binding residues)

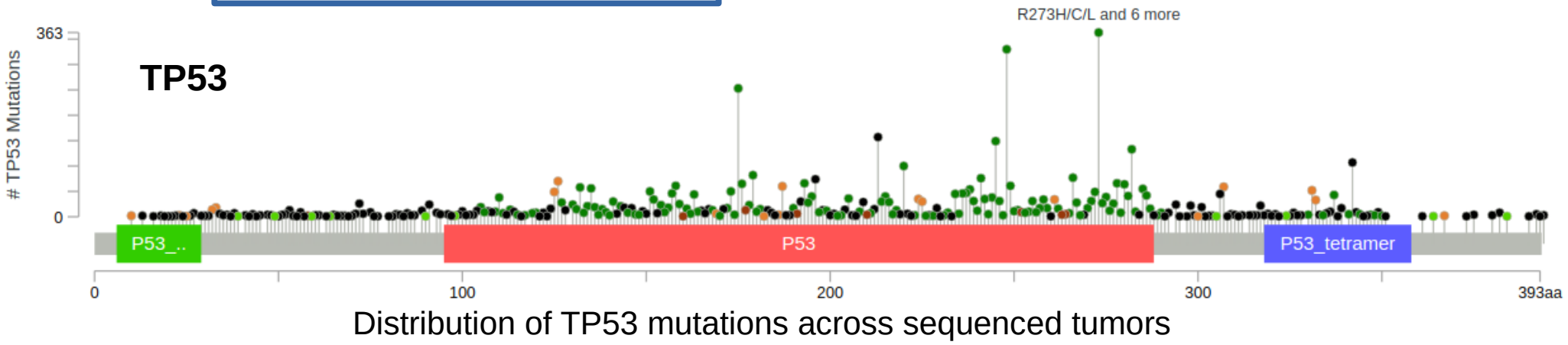
How a gene alteration can drive **loss-of-function** effects?



- With any alterations that **down-regulate** the gene function *(e.g. a traslocation that disrupts the open-reading frame, or a gene deletion)*

- With **specific** (missense) mutations able to cause loss-of-function effects

- With **any** mutation that **disrupts** the protein translation



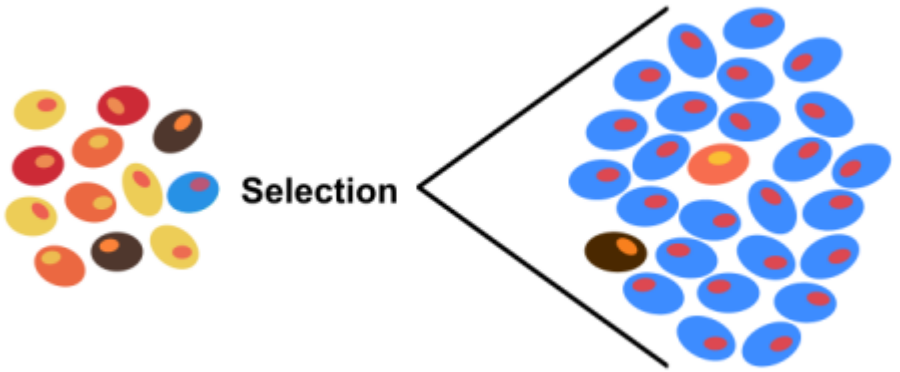
Distribution of TP53 mutations across sequenced tumors

● Truncating mutation*

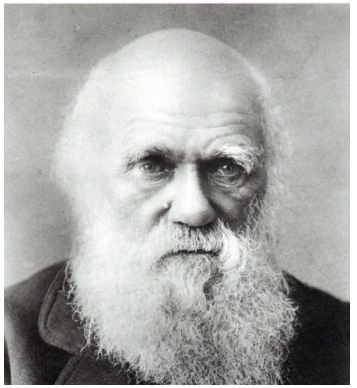
e.g. mutations leading to premature stop codons
(leading to a similar result at almost any protein position)

DRIVER MUTATIONS

Mutations accumulate in tumor cells following a Darwinian process



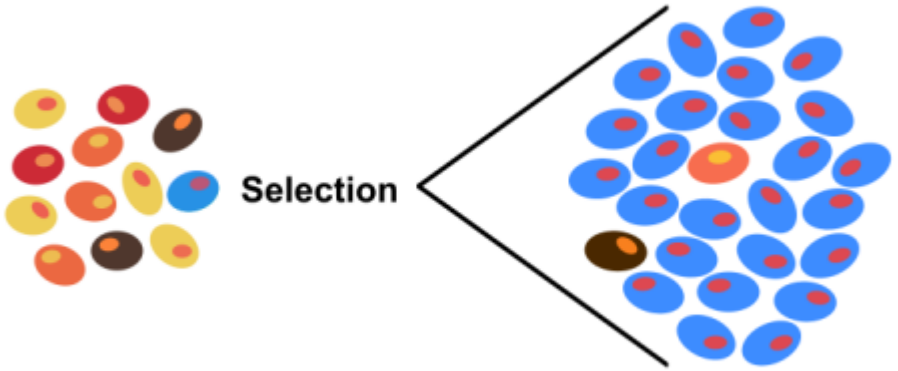
‘Bulk’ tumor is composed by **clone populations** with distinct alterations



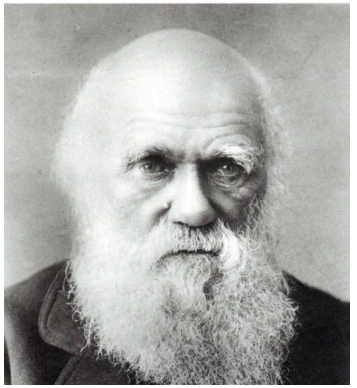
- In each **replication** of the tumor cells, **new (somatic) mutations** may appear

- If these mutations confer **selective advantages** (*e.g. avoid immune-survelliance*) they are more likely to become “fixed” in the population

Mutations accumulate in tumor cells following a Darwinian process



'Bulk' tumor is composed by **clone populations** with distinct alterations



- In each **replication** of the tumor cells, **new (somatic) mutations** may appear

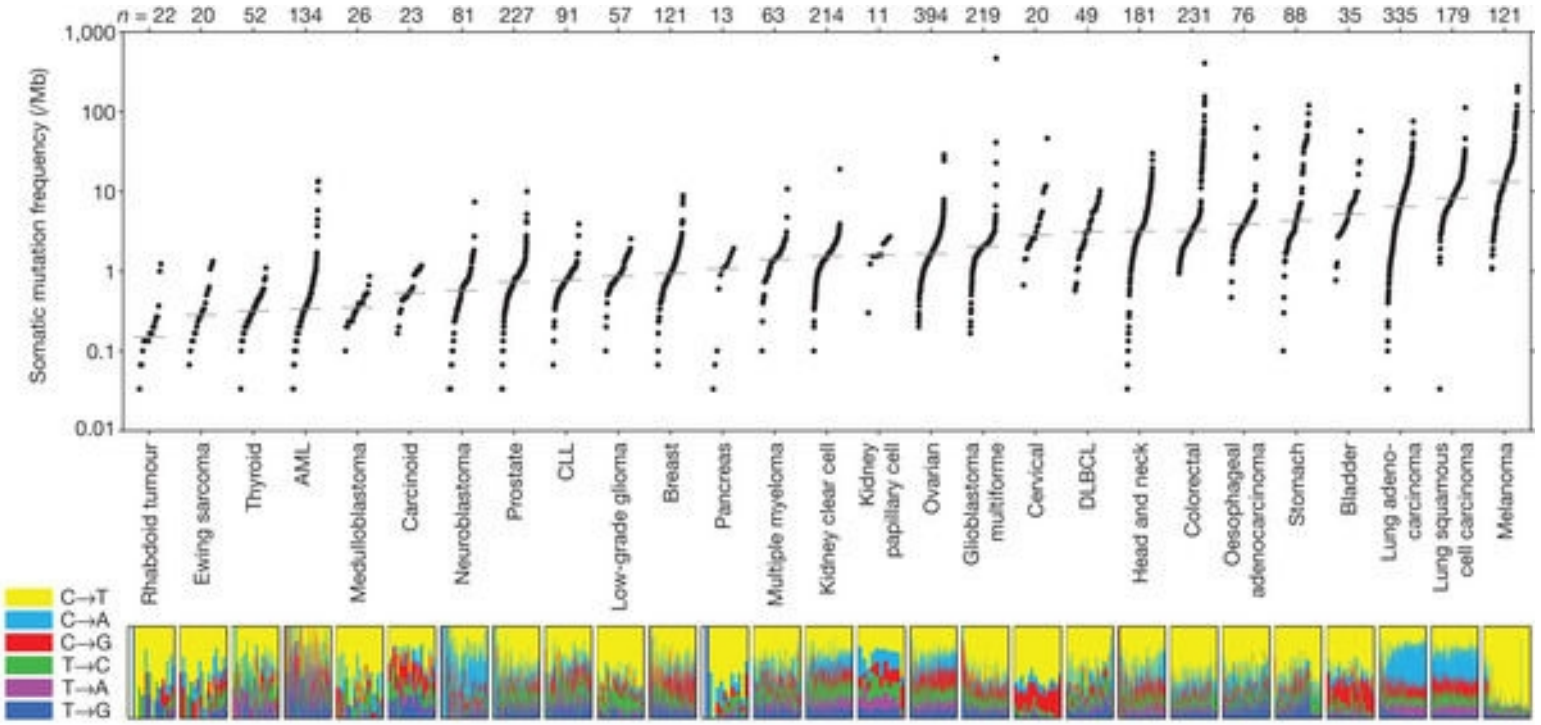
- If these mutations confer **selective advantages** (*e.g. avoid immune-survelliance*) they are more likely to become "fixed" in the population

Driver mutations — Mutations (in cancer genes) that provides selective advantage to the tumor cells *(i.e. promote cancer hallmark(s))*

Passenger mutations — Functionally ~neutral bystanders *(i.e. consequence not cause)*

Mutations accumulate in tumor cells following a Darwinian process

- Only **a subset of the mutations** observed in tumors are (likely) **driver events**



- Relevance of driver mutations is shaped by changes in the **selective forces** during the disease progression

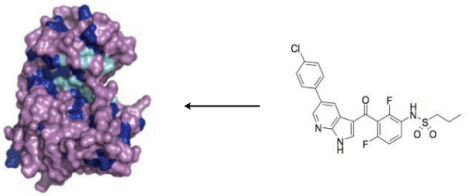
  e.g. onset versus metastatic processes

Example of the adaptative capabilities of the tumor cells: resistance to therapy

Melanoma bearing
BRAF V600E

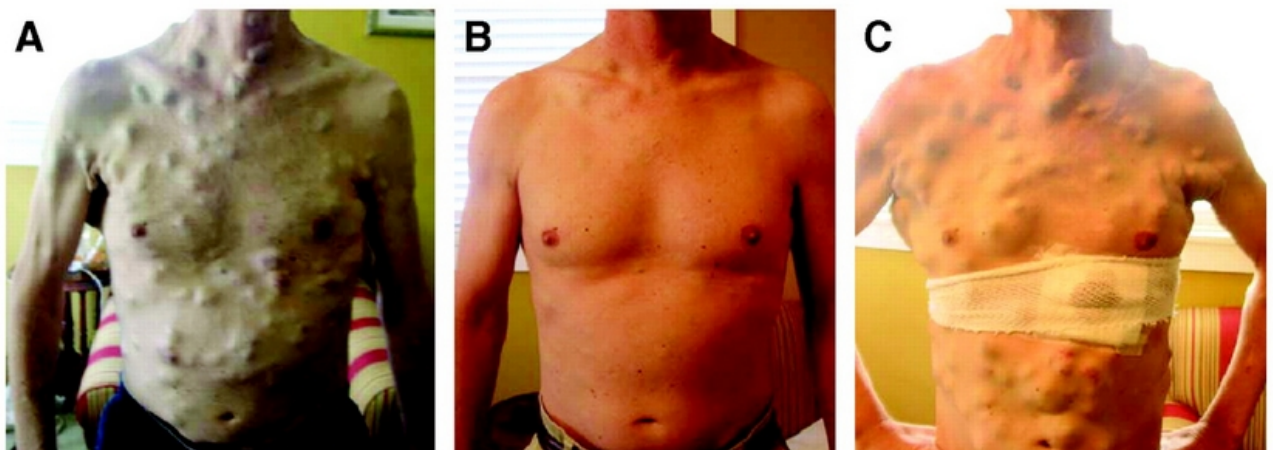
Example of the adaptative capabilities of the tumor cells: resistance to therapy

2w treatment with vemurafenib



BRAF (V600E) Vemurafenib

Example of the adaptative capabilities of the tumor cells: resistance to therapy



Relapse due to
resistant clones

Tumor cells bearing mutations* that confer resistance** to the drug ‘repopulate’ the tumor bulk and drive the relapse

* *de novo* arisen or pre-existing

** e.g. mutations that disrupt the drug-protein binding

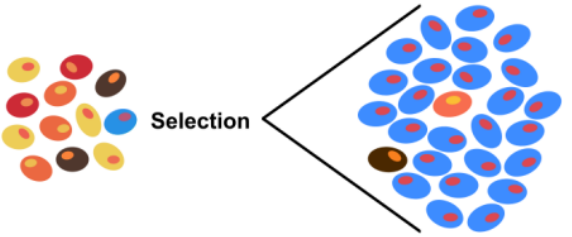
Cancer is a evolving process of the somatic cells

- **Germline** variants can predispose to cancer

  (e.g. all the cells 'start' with the loss-of-function of one allele in a tumor suppressor → only one additional hit is required)

- **Somatic** mutations that lead to the gain-of-function of an **oncogene** or loss-of-function of a **tumor suppressor** are **selected**

  ("brute force", as cells bearing them take over the rest)



- **Somatic** mutations occur by the interaction of intrinsic and extrinsic **mutational processes**

- The **selective pressures** that select them can change in time and space

  e.g. mutations involved in onset vs metastasis

  e.g. mutations favoring drug resistance mechanisms

CANCER DATA REPOSITORIES

- Big **international projects** have characterized the molecular profiles of tumors across large cohorts of different cancer types

- These data have been made **available for the community**, being a goldmine for cancer research during the last years

- In addition, an **increasing number** of “private” projects also make the data available through different means*

  ** e.g. when publishing a manuscript, the data is made available in the publication*

- Differences in the tumor data are related to:

  - cancer type (disease type and status)

  - sample (e.g primary vs met)

  - technology (e.g panel vs WGS/WES)

LARGE INTERNATIONAL CANCER SEQUENCING INITIATIVES

(updated ~2020)

International Cancer Genome Consortium

25,000 **WES/WGS** across **24 tissues**
Mainly **adult tumors at diagnosis**

https://dcc.icgc.org/

https://icgc.org/

The Cancer Genome Atlas

37,000 **WES** across **27 tissues**
Mainly **adult tumors at diagnosis**

https://portal.gdc.cancer.gov/

6,000 **WGS** across **11 cancer types**
Pediatric cancers

https://ocg.cancer.gov/programs/target/data-matrix

AACR American Association for Cancer Research FINDING CURES TOGETHER

PROJECT GENIE Genomics Evidence Neoplasia Information Exchange

70,000 **panel data** across **80 cancer types**
Adult advanced cancers

https://www.aacr.org/RESEARCH/RESEARCH/PAGES/AACR-PROJECT-GENIE.ASPX

Memorial Sloan Kettering Cancer Center

11,000 **panel data** across **65 cancer types**
Adult advanced cancers (enriched by mets)

https://www.cbioportal.org/study/summary?id=msk_impact_2017

Hartwig MEDICAL FOUNDATION

4,000 **WGS** across **several cancer types**
Adult advanced cancers (only mets)

https://www.hartwigmedicalfoundation.nl/en/

LARGE INTERNATIONAL CANCER SEQUENCING INITIATIVES

- All these projects make the data **public**
- In addition to the data, the websites of these projects also include tools to **navigate the data**
- There are also '**aggregators**', merging the data of (some) of these projects plus curation of individual-center efforts

https://www.cbioportal.org/

https://cancer.sanger.ac.uk/cosmic#

LARGE INTERNATIONAL CANCER SEQUENCING INITIATIVES

cBioPortal FOR CANCER GENOMICS

Query | Quick Search **Beta!** | Download

Please cite: Cerami et al., 2012 & Gao et al., 2013

Select Studies for Visualization & Analysis:

0 studies selected (**0** samples)

Search...

You can select the study and download the available data

PanCancer Studies	8
Pediatric Cancer Studies	13
Immunogenomic Studies	8
Cell lines	3
Adrenal Gland	3
Ampulla of Vater	1
Biliary Tract	13
Bladder/Urinary Tract	17
Bone	2
Bowel	11
Breast	20
CNS/Brain	20
Cervix	2
Esophagus/Stomach	17

Quick select: TCGA PanCancer Atlas Studies | Curated se

PanCancer Studies

MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)	10945 samples
Metastatic Solid Cancers (UMich, Nature 2017)	500 samples
MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)	249 samples
SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)	141 samples
TMB and Immunotherapy (MSKCC, Nat Genet 2019)	1661 samples
Tumors with TRK fusions (MSK, Clin Cancer Res 2020)	106 samples
Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020)	24146 samples
China Pan-cancer (OrigiMed2020)	10194 samples

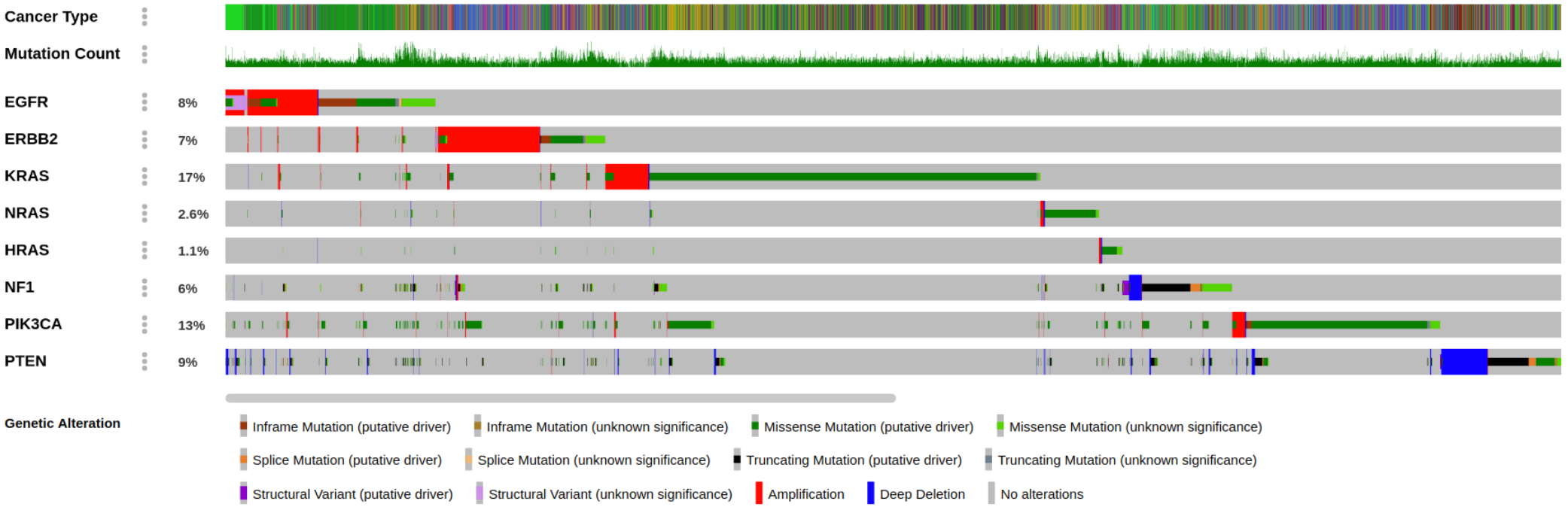
Pediatric Cancer Studies

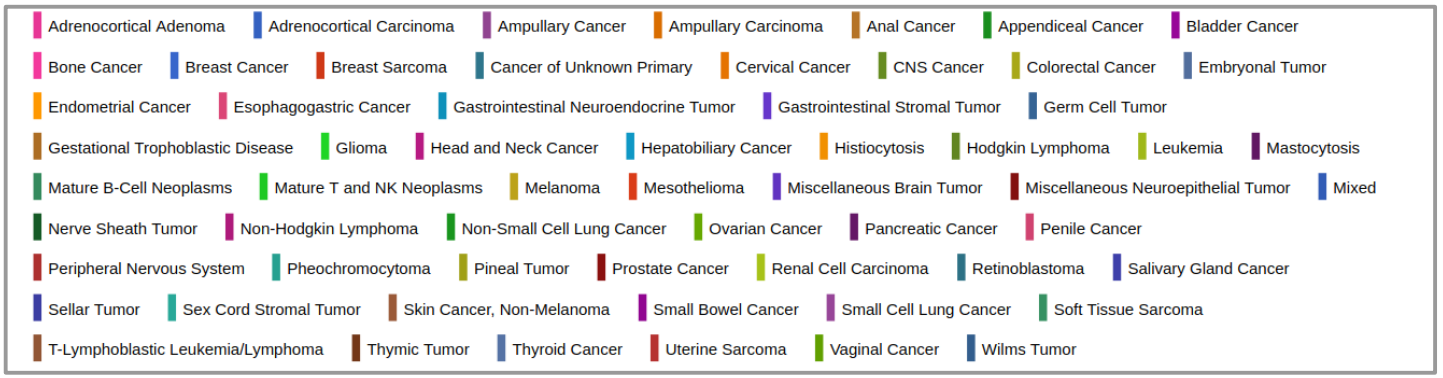
Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019)	261 samples
Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)	1978 samples
Pediatric Rhabdoid Tumor (TARGET, 2018)	72 samples
Pediatric Wilms' Tumor (TARGET, 2018)	657 samples
Pediatric Acute Myeloid Leukemia (TARGET, 2018)	1025 samples
Pediatric Neuroblastoma (TARGET, 2018)	1089 samples
Pediatric Pan-Cancer (DKFZ, Nature 2017)	961 samples
Pediatric Pan-cancer (Columbia U, Genome Med 2016)	103 samples
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)	73 samples
Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)	93 samples

LARGE INTERNATIONAL CANCER SEQUENCING INITIATIVES

These resources also provide tools to navigate
the accumulated data and perform some basic analyses

cBioPortal FOR CANCER GENOMICS





TAKE HOME MESSAGES

- **NGS technology** has revolutionized the study of the **molecular characteristics** of tumor cells

- **Data sharing** and **computational analyses** has (greatly) improved our understanding of the **mutational processes** and **biological hallmarks** of tumor cells

- The final aim is to **efficiently** translate this (evolving) knowledge in better **anti-cancer therapy** strategies under the paradigm of **(precision) oncology**