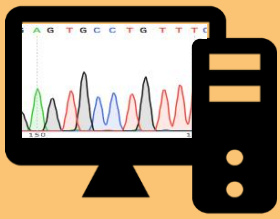
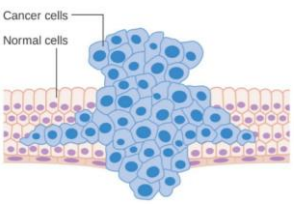
What can we learn from cancer genomes?



Introduction to cancer genomics

What is cancer?

Cancer is a group of diseases characterized by the **abnormal growth** and **spread of cells**.



Most cancers are caused by genetic alterations that can be:

- Inherited → germline
- Acquired during life → somatic

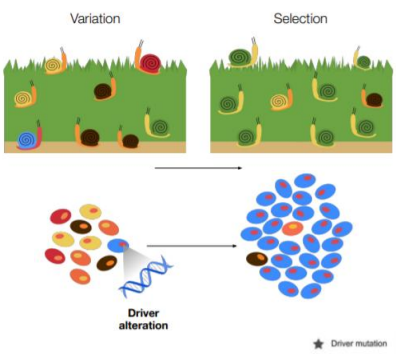
Most cancers are driven by somatic mutations.

Can Darwin teach us something about cancer?

YES!

Cancer evolves as a Darwinian evolutionary process.

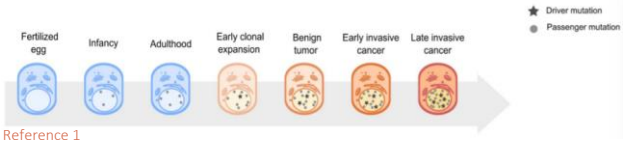
Here snails and cells behave the same way.
The more adapted the more survive.



IMPORTANT

Cancer drivers are the genomic alterations that confer selective advantage to the cell.

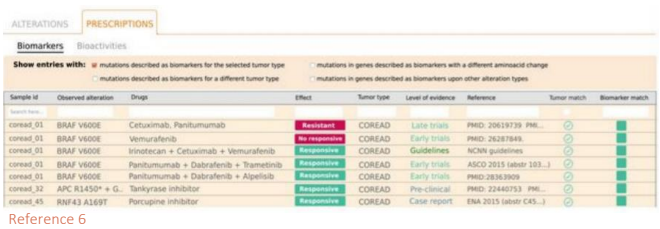
Driver genes are those that carry driver mutations.



Reference 1

Clinical application: Cancer Genome Interpreter

Cancer Genome Interpreter is a algorithm very useful to clinics, because of the **easy interpretation of the mutations** of an specific patient. Clinics can introduce the mutations that the patients have in their tumours and Cancer Genome Interpreter **identifies which ones are drivers and passengers**. That knowledge allows to choose **the most effective therapy** to each patient (personalize therapy).

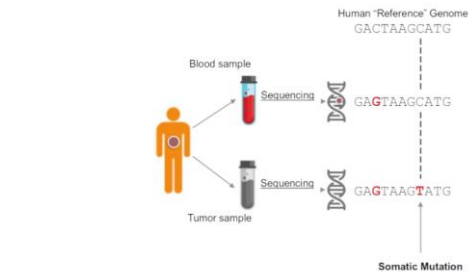


Reference 6

The identification of cancer drivers

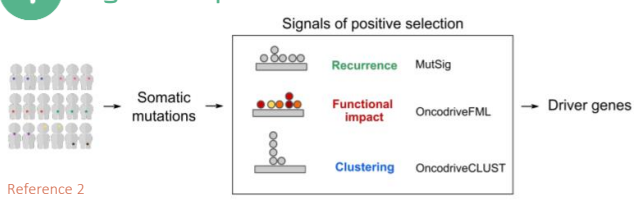
HOW DO WE IDENTIFY CANCER DRIVERS?

Sequencing of tumour samples to identify cancer somatic mutations



The main problem is that **tumours genomes have thousand of somatic** mutations, and only an **average on 4.6 of the are drivers**.

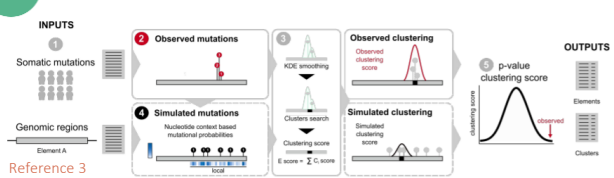
1 Signals of positive selection



Reference 2

Driver genes can be detected through **signals of positive selection** (special patterns in the distribution of mutations that appear as a result of tumorigenesis → **selective advantage**). That signals are **complementary** (not all driver genes have the 3 signals). To identified them we test if **what we observe is expected to occur by chance** (neutral evolution)

2 A method in depth: Oncodrive CLUSTL



Reference 3

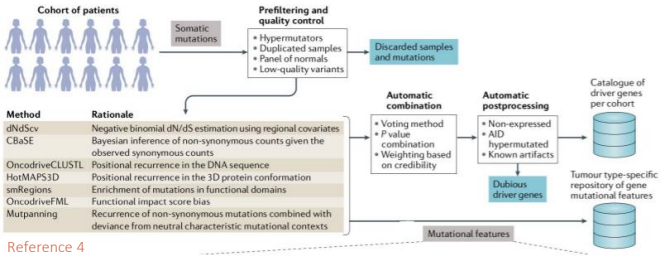
OncodriveCLUSTL is a sequence-based clustering method that **identifies drivers in protein-coding regions**.

3 The pipeline: IntOGen + boostDM

Since the signals are complementary, the idea would be to merge all these methods.

IntOGen: One method to rule them all

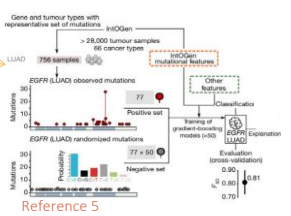
Using sequences of patients of different projects, **IntOGen uses 7 methods** (2 based on recurrence, 3 based on clustering and some based on functional impact) to identify driver genes.



Reference 4

BUT... **not all mutations in driver genes are driver mutations!**

They are detected by **boostDM**, a machine learning algorithm that can predict driver mutations in human cancers.



Reference 5

PERSPECTIVES

The number of sequencings of different types of cancers will increase, which will allow identify driver mutations in less common cancers.

REFERENCES

Ref 1. Adapted from Stratton et al., Nature 2009
Ref 2. Tamborero et al, Sci. Rep. 2013; Porta-Pardo et al, Nat. Methods 2017
Ref 3. Arnedo-Pac et al., Bioinformatics 2019
Ref 4. Martinez-Jimenez et al., Nat Rev Cancer 2020
Ref 5. Muiños et al., Nature 2021
Ref 6. Tamborero et al., Genome Medicine 2018

UAB Universitat Autònoma de Barcelona
BBGLab