

Mutational processes and clonal population structure in cancer genomes

Cancer genomics and transcriptomics course, 2023

Francesc Muyas Remolar, PhD
EMBL-EBI
11 – 07 - 2023



Outline

1. Clonal population structure
 - Clone / Subclone
 - Cancer cell fraction
2. Mutational signature analysis
 - Type of signatures
 - How to discover them
 - Signature refitting
 - Clinical implications and potential limitations

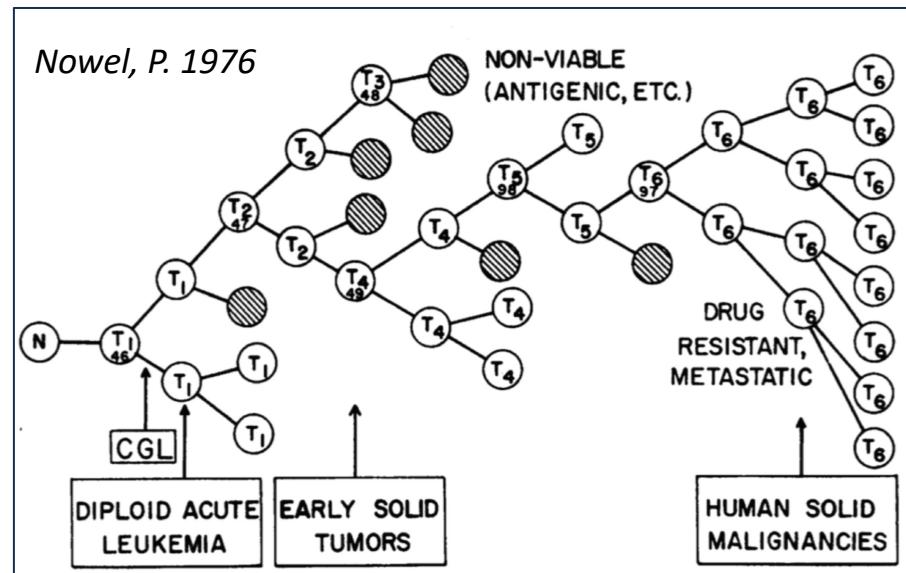
What do you know about these concepts?

- Clone / Subclone
- Purity
- Multiplicity
- Clone expansion
- Cancer cell fraction (CCF)
- Mutational signatures (MS)
- Mutation catalog
- Mutational processes
- De novo discovery of MS
- Refitting signatures
- NMF

Clonal population structure in cancer genomes

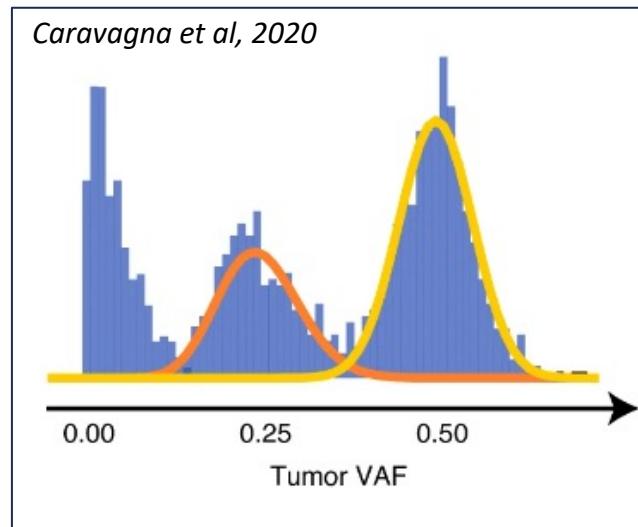
Definition of “clone”

- A cancer ‘clone’ remains a loosely defined entity, and its purest definition is ‘a group of cells within the tumour that share a common ancestor’
 - In phylogenetic terms, this would represent a monophyletic clade.
 - This implies that any ancestor in the entire phylogenetic tree of a tumour can be identified as the founder (clone), even though it may show no biological difference from the rest of the cancer cells.
 - This is why, in the field, we implicitly identify clones of interest, such as those that have growth/survival advantage

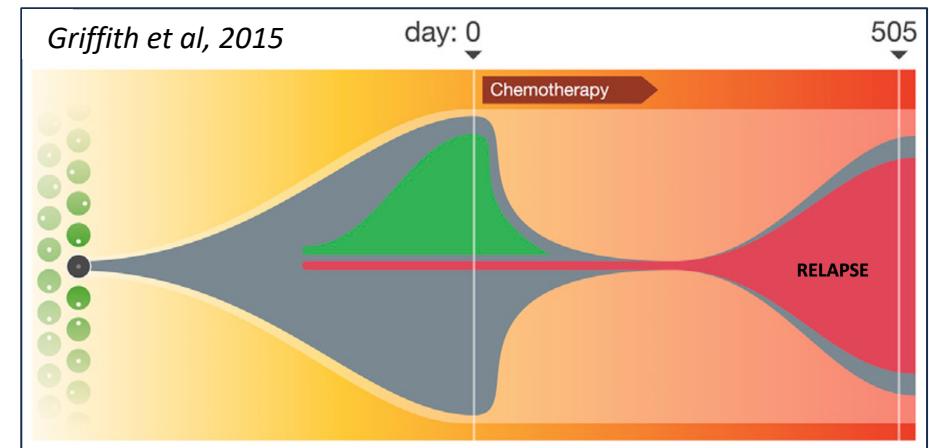


Clonal reconstruction

- Clonal reconstruction in bulk sequencing data refers to the process of **inferring the clonal composition of a tumour sample based** on sequencing data obtained from a bulk population of tumour cells
- Most cancer genomic data are generated from bulk samples composed of mixtures of cancer subpopulations, **as well as normal cells**
- Clonal reconstruction can provide insights into **tumour heterogeneity and evolutionary dynamics**
- However, it poses several **challenges due to the lack of single-cell resolution**



How from a plot with VAFs
(left) can we understand the
tumour heterogeneity?
*(unrelated figures)

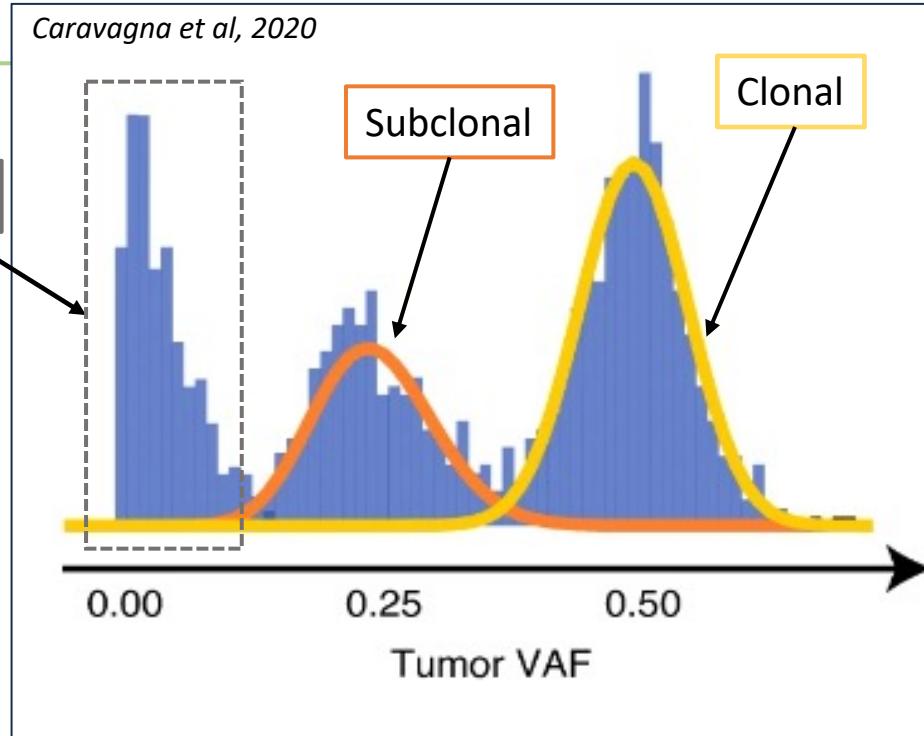


Subclonal heterogeneity

- Tumour samples often consist of multiple subclones with distinct genetic mutations and cellular populations.
- Bulk sequencing data provides an aggregated view of the genetic information, making it challenging to accurately identify and distinguish subclonal populations.
- Subclonal mutations that are present at low frequencies may be difficult to detect, leading to underestimation of clonal diversity.

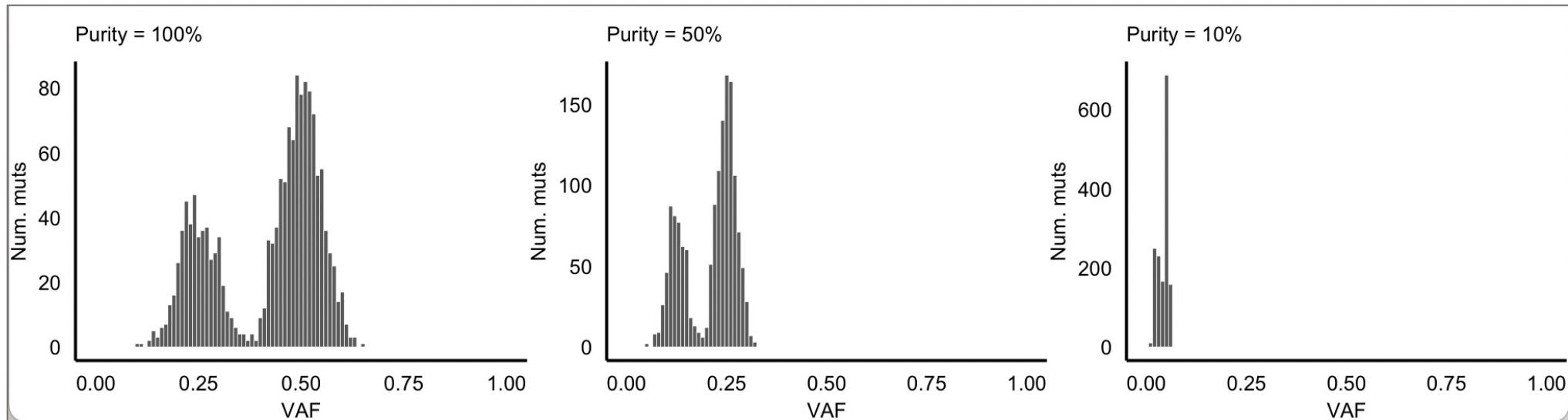
Tail (not all from the same clone)

- Very hard to distinguish
- New mutations (majority passenger) occurring in every cell division
- The accuracy of the whole reconstruction **massively depends on the depth of coverage**
- Very subclonal mutations **heavily affected by sequencing errors**



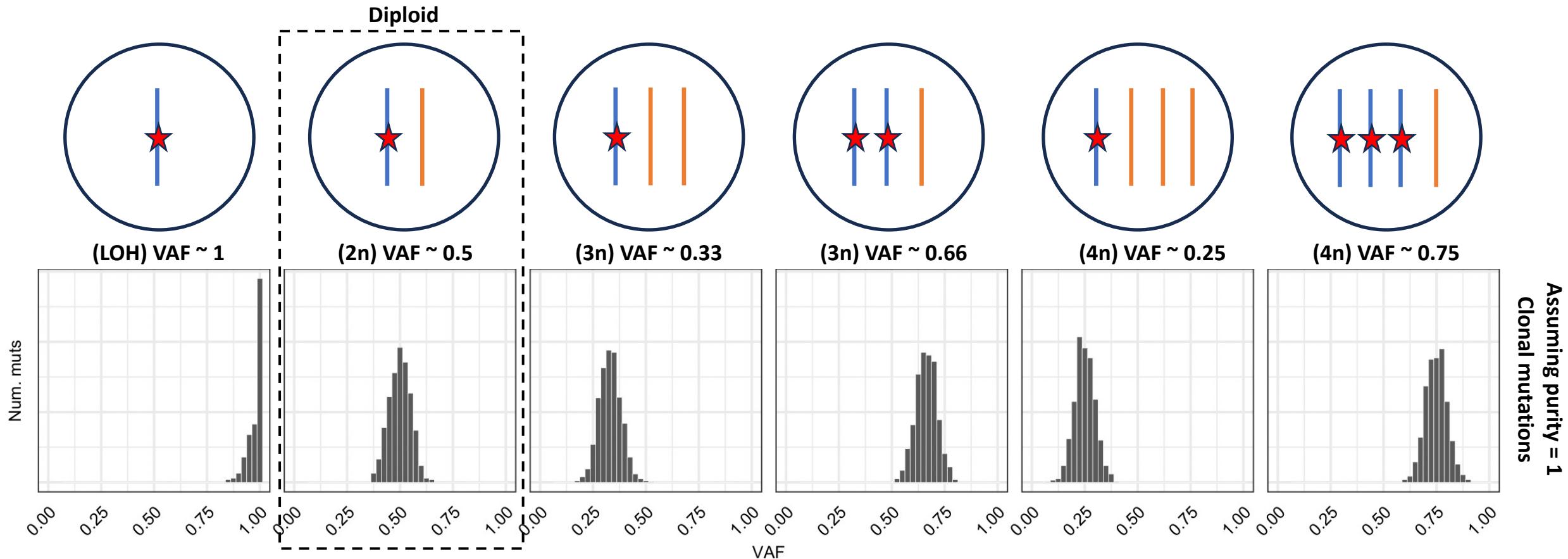
Tumour purity and contamination

- Tumour samples may contain normal tissue contamination or immune cell infiltration, leading to mixed signals in the sequencing data.
- The presence of non-tumour cells can confound clonal reconstruction efforts, obscuring the true clonal architecture.
- Accurate estimation of tumour purity and careful consideration of contamination sources are necessary to achieve reliable clonal reconstruction.



Copy number changes affect the VAF distribution

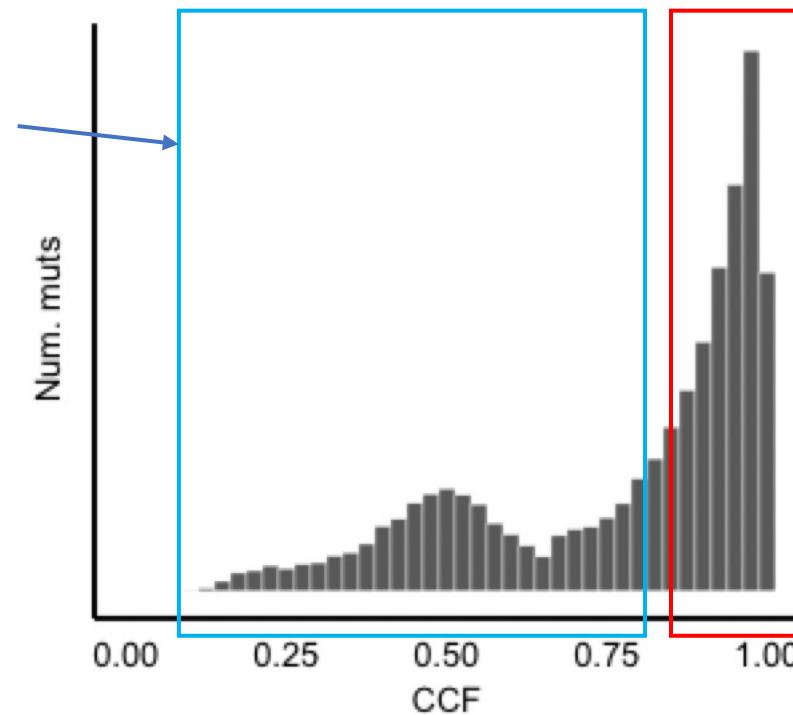
- Variant allele frequency refers to the proportion of reads or molecules in a sequencing sample that contain a specific genetic variant.
- Copy number changes** can have a significant impact on variant allele frequency (VAF) values.



Cancer cell fraction (CCF)

- The cancer cell fraction (CCF) is a measure that represents the proportion of cancer cells carrying a specific mutation within a tumour sample.
- It provides an estimation of the clonal composition of the tumour and helps understand the subclonal architecture and heterogeneity of cancer cells within a tumour.

Subclonal mutations are alterations that are present in only a subset of cancer cells, representing later events or subclonal expansions

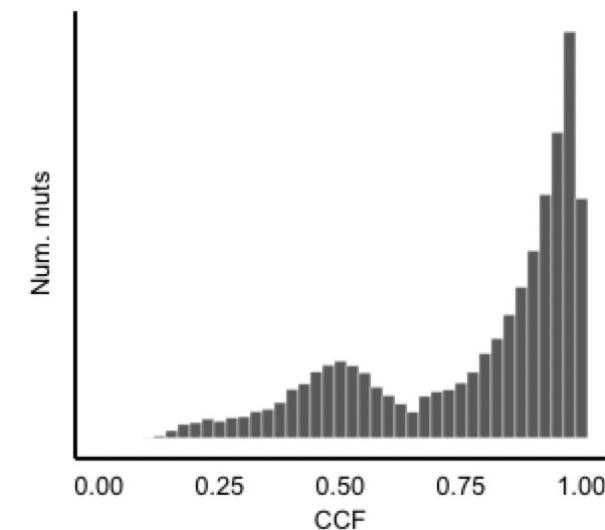
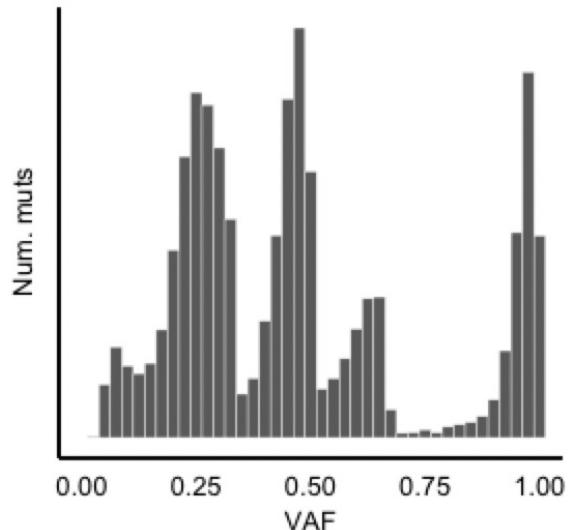


Clonal mutations are genetic alterations that are present in all cancer cells within the tumour.

Cancer cell fraction (CCF) calculation

CCF calculation takes into account:

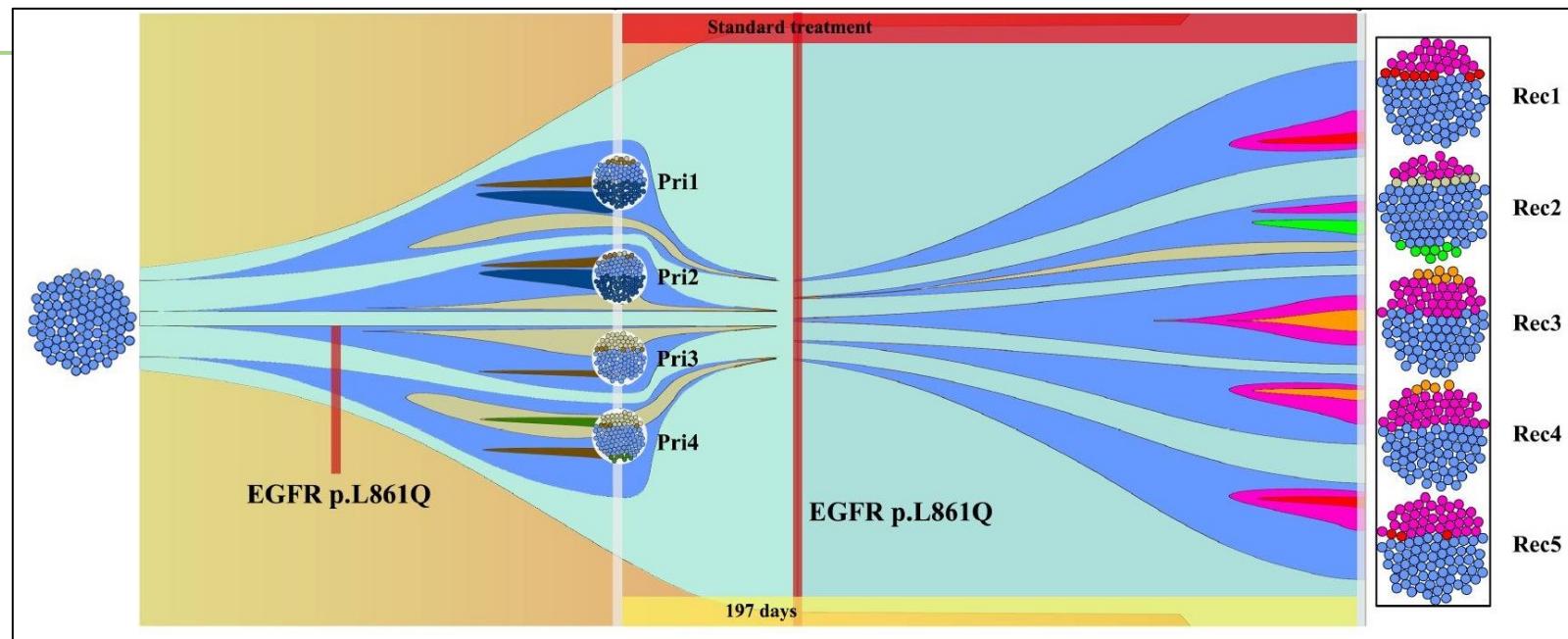
- VAF ~ the allele frequencies of mutations
- Copy number alterations ~ considers the genomic gains and losses (usually estimated with copy number algorithms)
- Multiplicity ~ how many genomic segments are harbouring the mutation
- Tumour purity ~ fraction of cancer cells (usually estimated with copy number algorithms)



Check [Dentro et al 2021 \(Cell\)](#) for CCF and multiplicity formulas

Limited temporal resolution

- Bulk sequencing data provides a snapshot of the tumour's genetic landscape at a specific time point.
- It may not capture the full evolutionary history or temporal ordering of genetic events during tumour progression.
- Inferring the timing and order of clonal events based solely on bulk sequencing data can be challenging and often requires additional information, such as multi-region sequencing or longitudinal sampling.



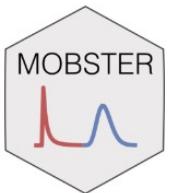
Yang et al. 2019

Computational complexity

- Clonal reconstruction from bulk sequencing data involves complex computational algorithms and statistical models.
- These methods often rely on assumptions and heuristics to infer clonal populations and their frequencies.
- Accurate interpretation of the results and validation of the inferred clonal architecture require careful consideration of the algorithmic choices and robust statistical analysis.
- They rely massively on the copy number calls, which are hard to get with low coverage or WES data

mobster

mobster is a package that implements a model-based approach for *subclonal deconvolution* of cancer genome sequencing data ([Caravagna et al; PMID: 32879509](#)).



Roth-Lab / pyclone-vi Public

genome / sciclone Public

The importance of clonal reconstruction analysis

Article | [Open Access](#) | Published: 12 April 2023

The evolution of lung cancer and impact of subclonal selection in TRACERx

Alexander M. Frankell, Michelle Dietzen, Maise Al Bakir, Emilia L. Lim, Takahiro Karasaki, Sophia Ward, Selvaraju Veeriah, Emma Collier, Ariana Huebner, Abigail Bunkum, Mark S. Hill, Kristiana Grigoriadis, David A. Moore, James R. M. Black, Wing Kin Liu, Kerstin Thol, Oriol Pich, Thomas B. K. Watkins, Cristina Naceur-Lombardelli, Daniel E. Cook, Roberto Salgado, Gareth A. Wilson, Chris Bailey, Mihaela Angelova, TRACERx Consortium, ... Charles Swanton  + Show authors

[Nature](#) **616**, 525–533 (2023) | [Cite this article](#)

- Analysis of 1,644 tumor regions from 421 patients reveals insights into lung cancer evolution and intratumor heterogeneity.
- Significant subclonal selection observed in 22 out of 40 common cancer genes, including TP53 and KRAS.
- Identification of clonal expansions, WGD, and CN heterogeneity as key factors impacting disease-free survival and patterns of relapse in non-small cell lung cancer.

Mismatch repair deficiency is not sufficient to increase tumor immunogenicity

 Peter M K Westcott,  Francesc Muyas,  Olivia Smith, Haley Hauck, Nathan J Sacks, Zackery A Ely, Alex M Jaeger, William M Rideout III, Arjun Bhutkar, Daniel Zhang,  Mary C Beytagh, David A Canner,  Roderick T Bronson, Santiago Naranjo, Abbey Jin, JJ Patten, Amanda M Cruz,  Isidro Cortes-Ciriano,  Tyler Jacks

doi: <https://doi.org/10.1101/2021.08.24.457572>

- MMRd in human cancer correlates with high tumour mutational burden (TMB) and response to immune checkpoint blockade (ICB) therapy
- Nevertheless, about half of MMRd tumors do not respond to ICB for unclear reasons.
- Using mouse MMRd models for lung and colon cancer showed that MMRd alone doesn't increase immunogenicity ICB response.
- Intratumoral heterogeneity (ITH) plays a crucial role in immune evasion, shaping the clonal architecture of neoantigens and impacting immunosurveillance in MMRd tumors.

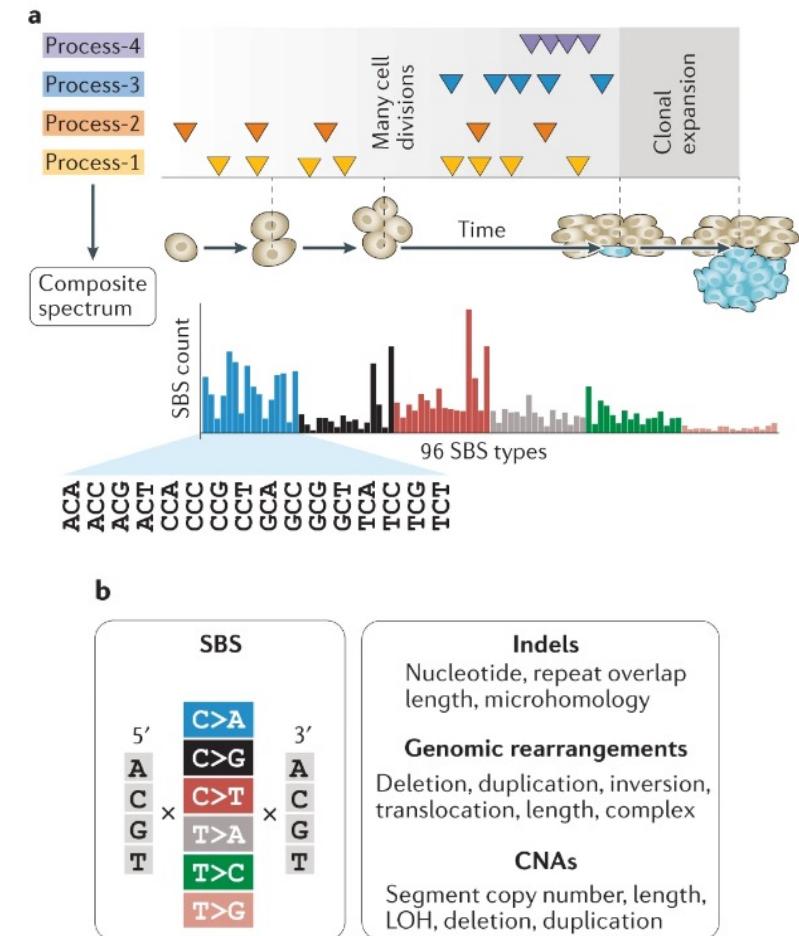
Conclusions

- Clonality analysis using bulk sequencing can help to determine the presence and abundance of different tumour clones within a sample.
- The cancer cell fraction is a nice way to check clonality
- It helps understand tumour evolution, heterogeneity, and response to treatment.
- It aids in identifying subclones, driver mutations, and clonal expansions.
- Bulk sequencing has limitations in capturing rare or minor subclones and may not provide precise spatial or temporal information.

Mutational Signatures Analysis

Definition of “mutational signatures”

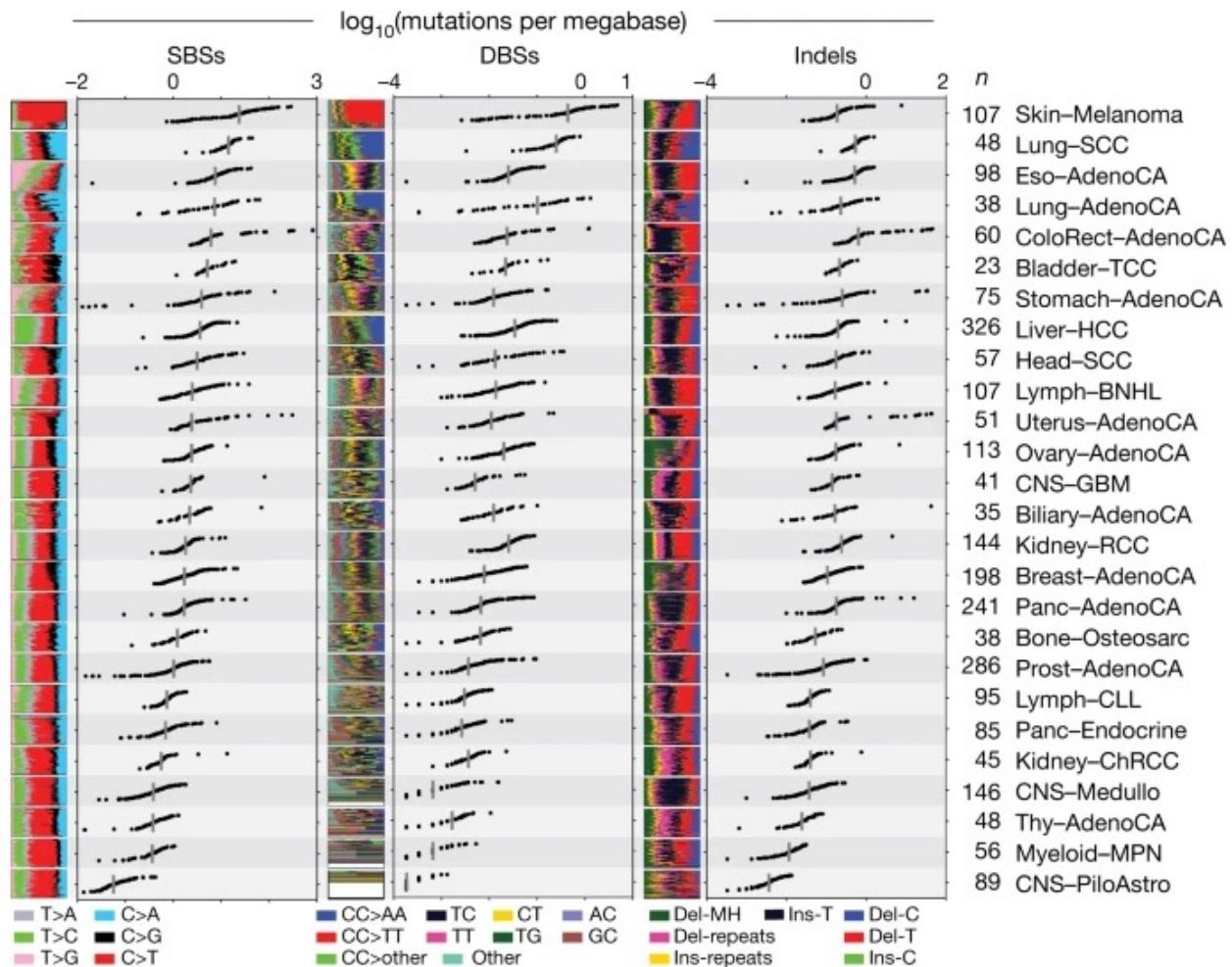
- Mutational signature analysis is a technique used in cancer genomics to identify and characterize patterns of DNA mutations in cancer genomes.
- These patterns can arise from various mutational processes that lead to changes in DNA sequences.
- Each of these processes leaves a distinct mark on the genome, resulting in specific patterns of mutations.
- Mutations can occur due to a variety of factors
 - DNA replication errors
 - Exposure to mutagenic agents (e.g., ultraviolet radiation, chemical carcinogens)
 - Defects in DNA repair mechanisms.



Cortes-Ciriano et al 2022

Definition of “mutational signatures”

- Typically identified by **analysing large-scale genomic sequencing data** from a diverse set of samples
- By comparing the frequencies and types of mutations across samples, researchers can uncover common patterns and classify them into different mutational signatures.



Alexandrov et al 2020

Mutational signatures in human cancer

Large dataset analysed

- 23,829 samples (most types of cancer):
 - 2,780 PCAWG whole genomes
 - 1,865 additional whole genomes
 - 19,184 exomes
- 79,793,266 somatic **SBSs**
- 814,191 doublet-base substitutions (**DBSs**)
- 4,122,233 small **indels**

Article

The repertoire of mutational signatures in human cancer

<https://doi.org/10.1038/s41586-020-1943-3>

Received: 18 May 2018

Accepted: 18 November 2019

Published online: 5 February 2020

Open access

Ludmil B. Alexandrov^{1,920}, Jaegil Kim^{2,920}, Nicholas J. Haradhvala^{2,3,920}, Mi Ni Huang^{4,5,920}, Alvin Wei Tian Ng^{4,5}, Yang Wu^{4,5}, Arnoud Boot^{4,5}, Kyle R. Covington^{6,7}, Dmitry A. Gordenin⁸, Erik N. Bergstrom¹, S. M. Ashiquul Islam¹, Nuria Lopez-Bigas^{9,10,11}, Leszek J. Klimczak¹², John R. McPherson^{4,5}, Sandro Morganella¹³, Radhakrishnan Sabarinathan^{10,14,15}, David A. Wheeler^{6,16}, Ville Mustonen^{17,18,19}, PCAWG Mutational Signatures Working Group²⁰, Gad Getz^{2,3,21,22,921}, Steven G. Rozen^{4,5,23,921*}, Michael R. Stratton^{13,921*} & PCAWG Consortium²⁴

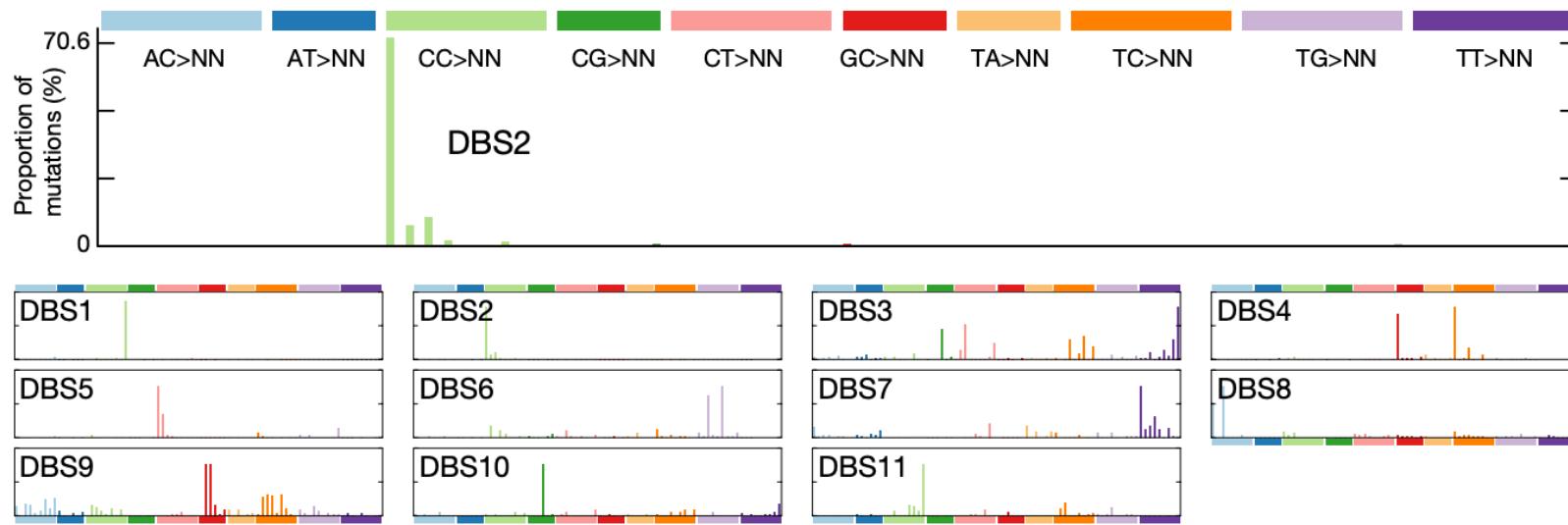
Different types of mutational signatures

SBS signatures - 49 SBS signatures identified



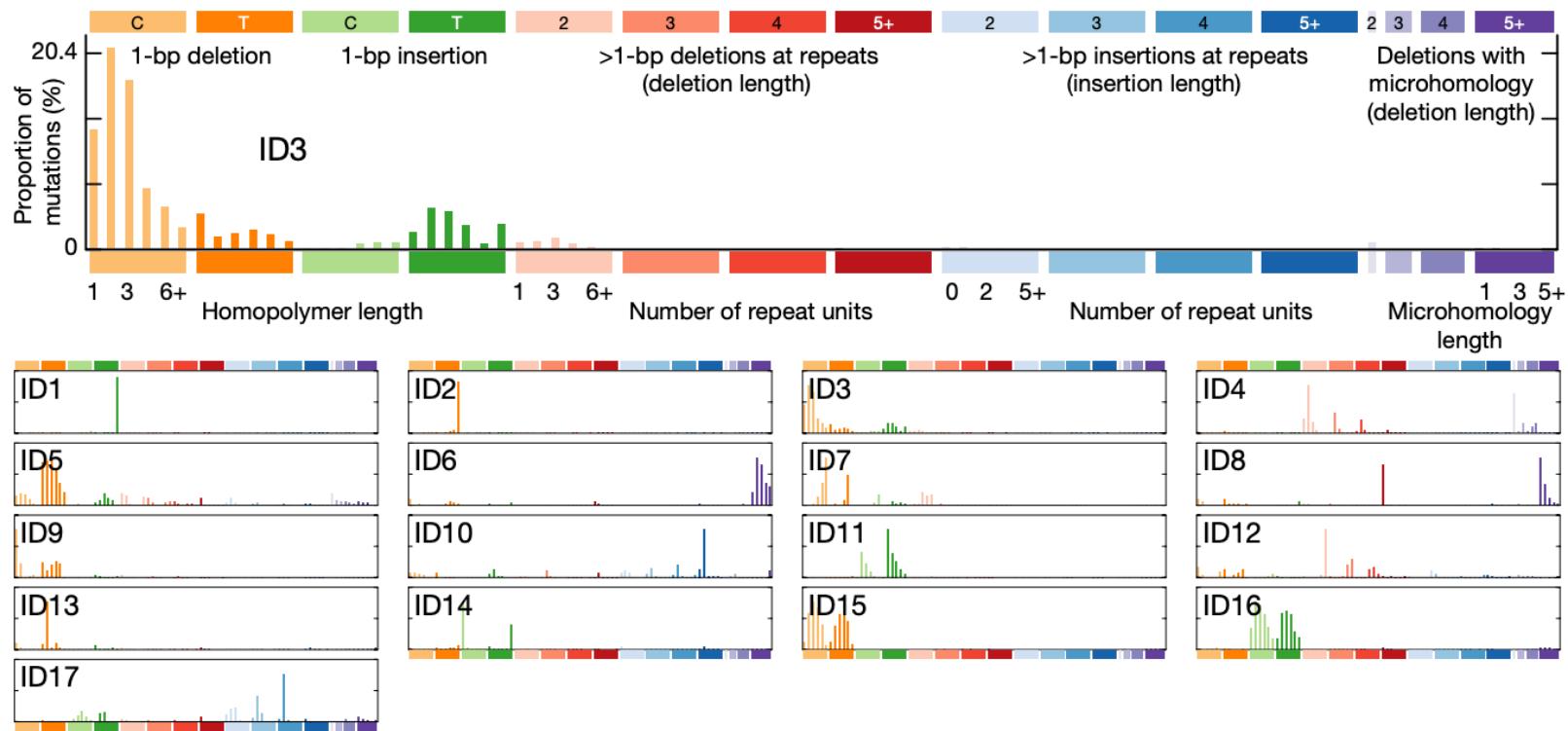
Different types of mutational signatures

DBS signatures – 11 DBS signatures identified



Different types of mutational signatures

Indels signatures – 17 small indel signatures identified

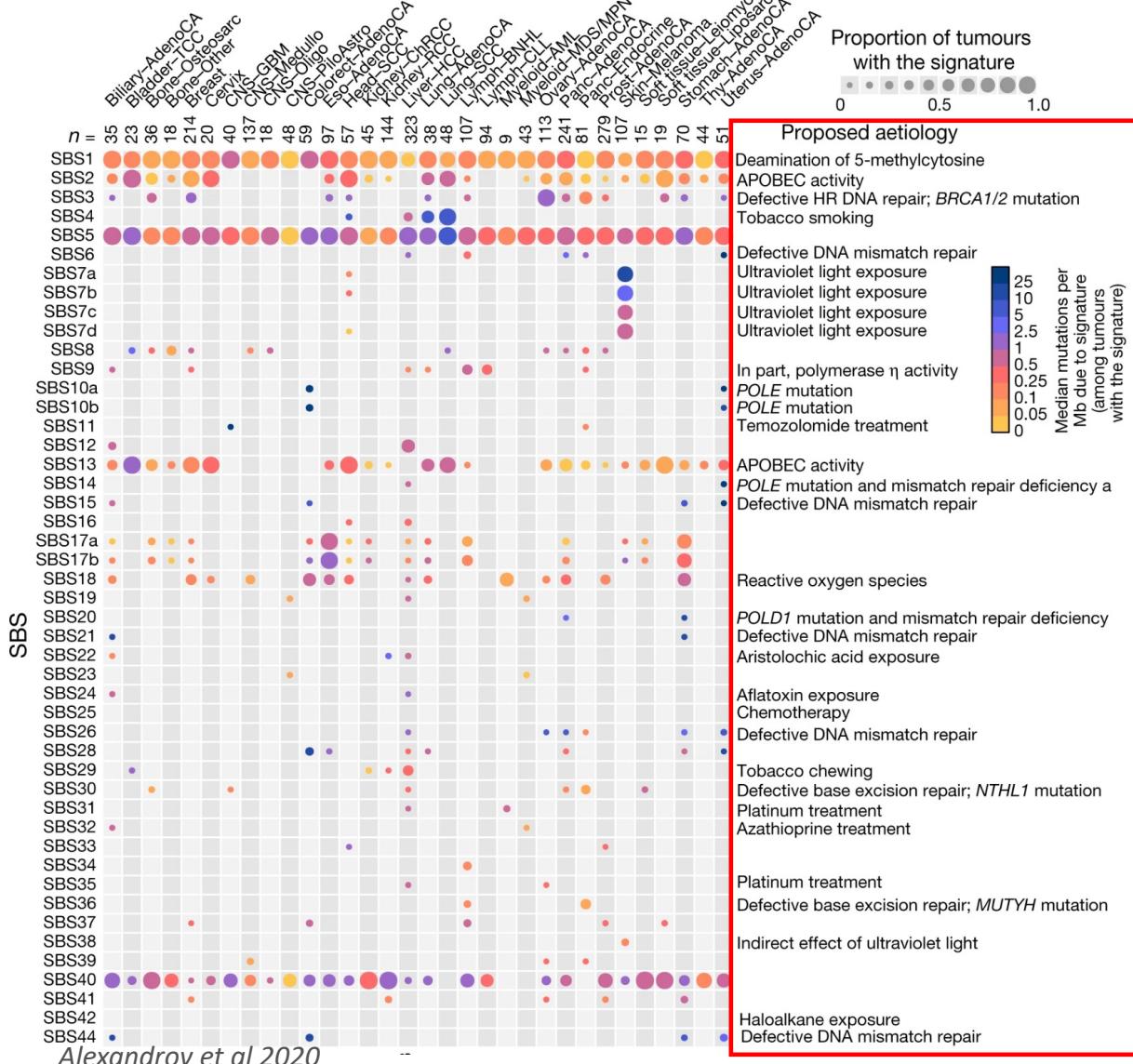


Classification of mutational signatures based on their underlying biological processes

Cancer Aetiology Investigation

- Mutational signatures provide insights into the underlying causes mutational processing undergoing on cancer (and non-cancer samples)
- All included in the new COSMIV (v3) version of mutational signatures

<https://cancer.sanger.ac.uk/signatures/>



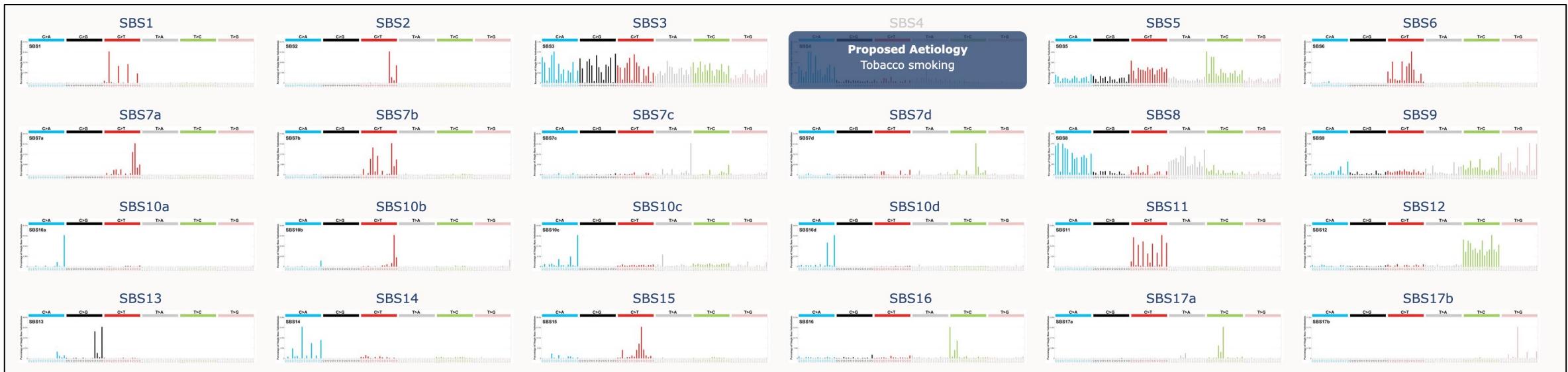
COSMIC – Catalogue of Somatic Mutations in Cancer

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

- You can check the last updates on mutational signatures
- Download mutational profiles for different reference genomes (human, mouse...)



<https://cancer.sanger.ac.uk/signatures/>



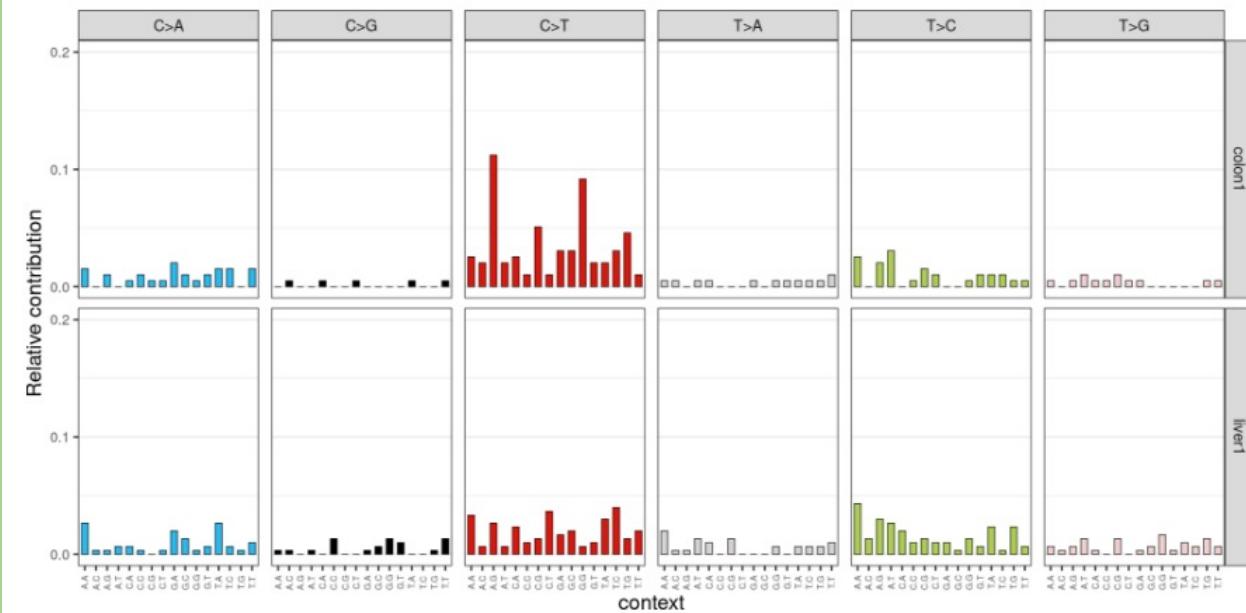
Explanation of how mutational signatures are derived from genomic data

Mutational signatures are derived from genomic data through a process known as mutational signature analysis.

1. Genomic Data Collection & Mutation calling:

- Large-scale genomic sequencing data is collected from a diverse set of cancer samples
- Somatic variant calling: high-quality set of mutations

2. Mutation characterization: coordinates and genomic context of the mutations, checking the strand, removing and detecting artefacts: creates the mutation catalog



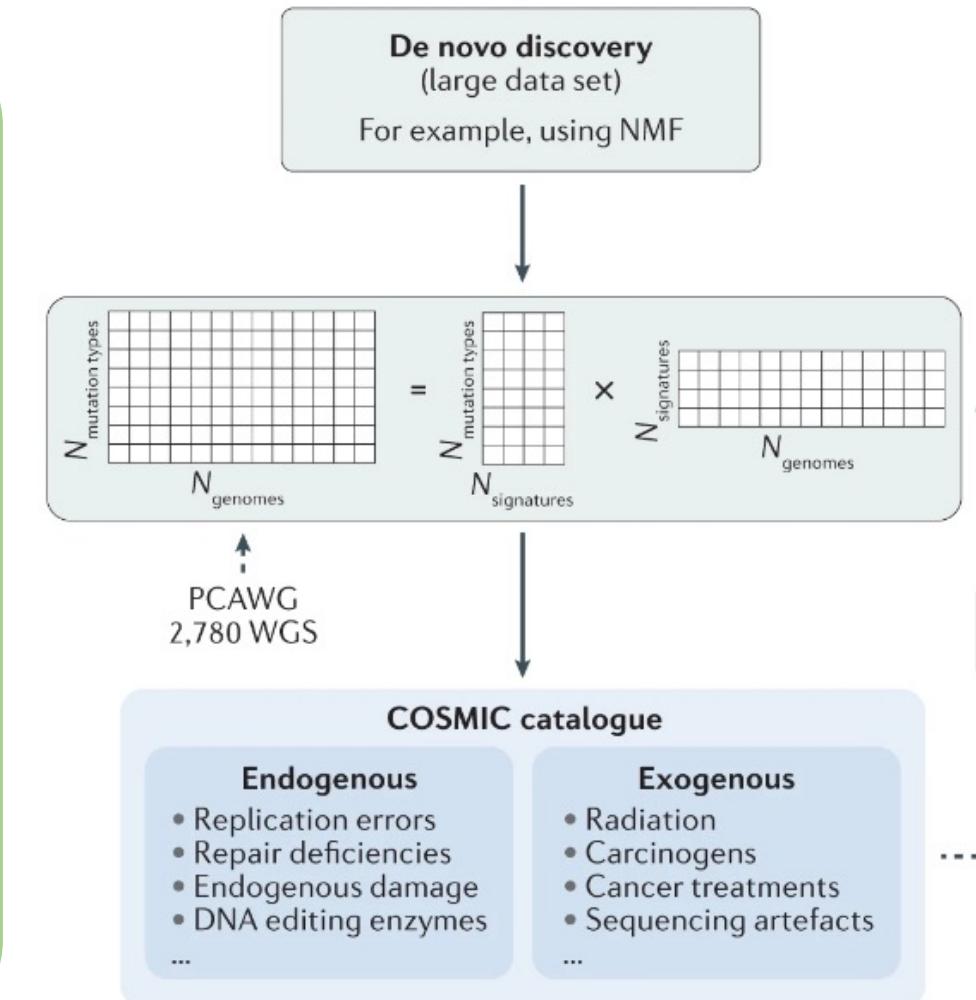
Explanation of how mutational signatures are derived from genomic data

3. Signature Extraction:

- Statistical algorithms and computational methods are applied to the pre-processed mutation catalog to extract mutational signatures.
- One commonly used method is non-negative matrix factorization (NMF): decomposes the mutation catalog into a set of underlying mutational signatures and their corresponding activities.

4. Signature Identification

- The extracted mutational signatures are then analysed to identify their characteristics and potential underlying biological processes.
- This involves comparing the extracted signatures with known reference signatures and determining the most similar matches based on various metrics (e.g., cosine similarity).



A bit more about signature extraction (*de novo* discovery)

- Once you have a high-quality set of mutations, we must create a mutation catalog with the mutation information:
 - Mutation type, genomic coordinates, genomic context... (into mutation matrixes)
- Signature extraction ~ statistical algorithms to extract mutational patterns from the mutation catalog (**require many samples**)
 - One commonly used method is non-negative matrix factorization (NMF), which decomposes the mutation matrix into a set of underlying mutational signatures
 - Alternative methods, such as independent component analysis (ICA) or probabilistic methods, can also be used

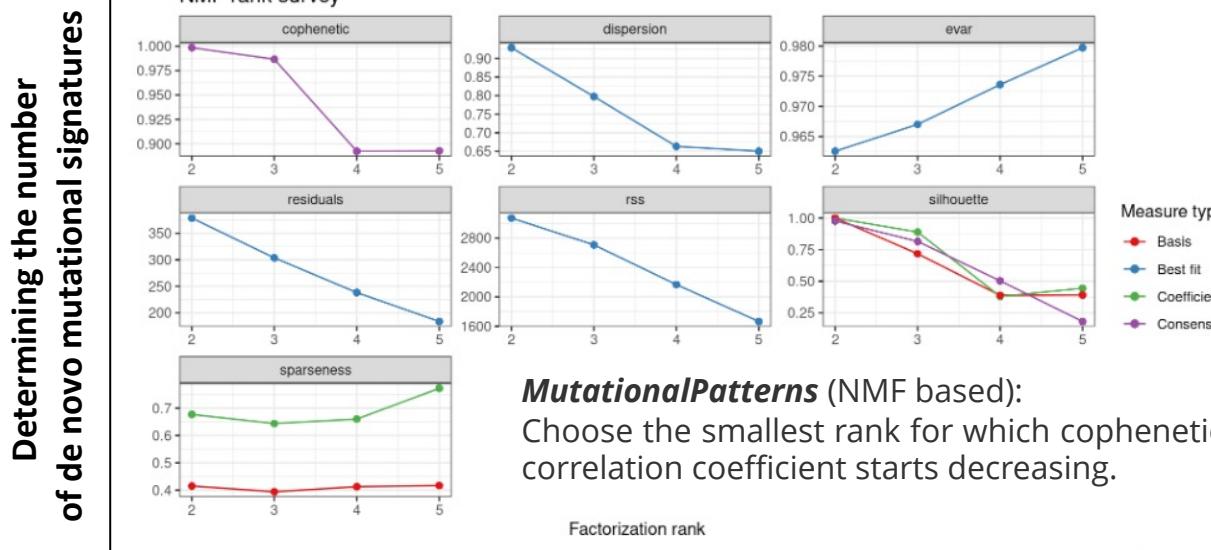


Figure: MutationalPatters

https://bioconductor.org/packages/release/bioc/vignettes/MutationalPatterns/inst/doc/Introduction_to_MutationalPatterns.html

*Different examples
(not comparable)

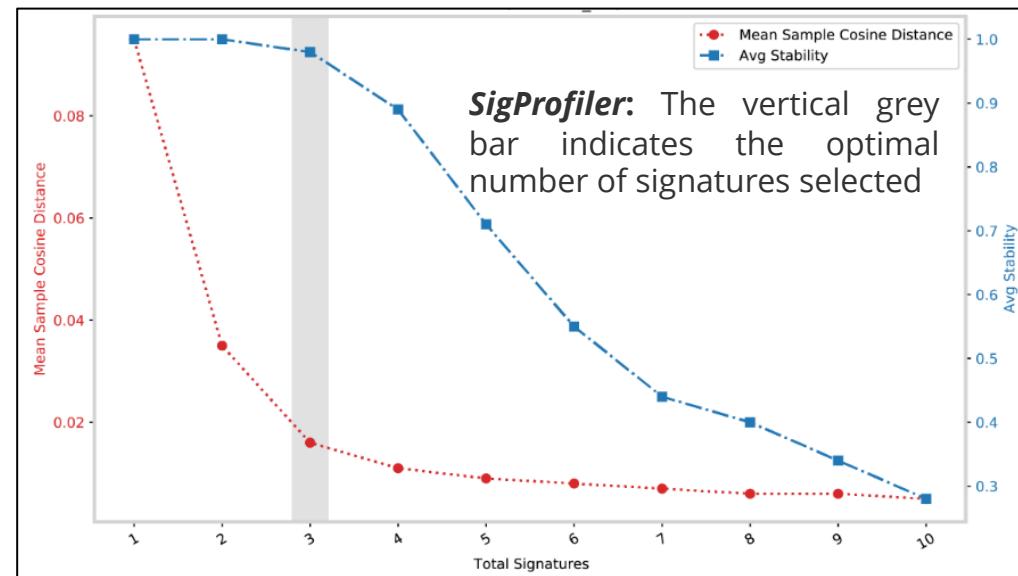


Figure: SigProfileExtractor <https://osf.io/t6j7u/>

A bit more about signature identification

- Analyse and interpret the extracted mutational signatures to identify their characteristics and potential underlying biological processes.
- Compare the extracted signatures with known reference signatures or databases, such as the COSMIC database, using metrics like cosine similarity.
- Identify the best matching or most similar reference signatures for each extracted signature.

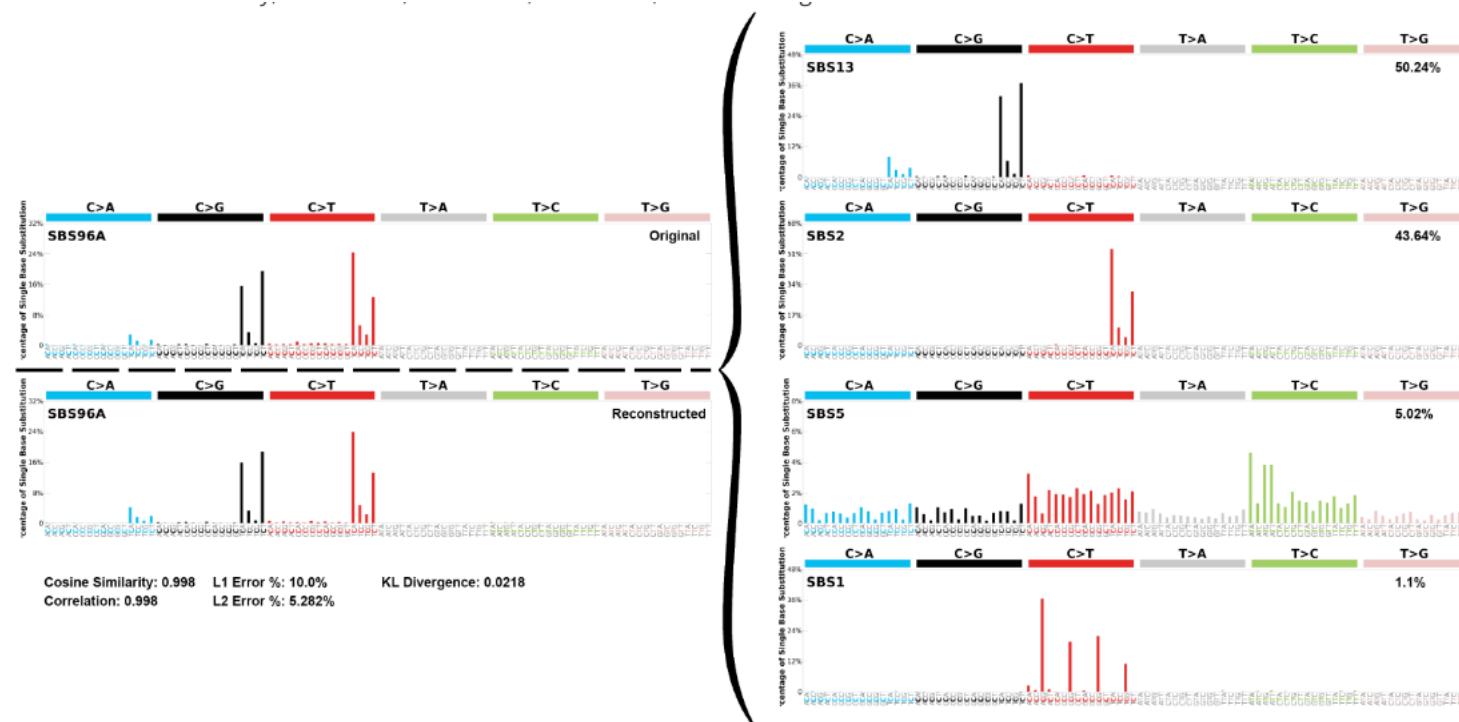


Figure: SigProfileExtractor <https://osf.io/t6j7u/>

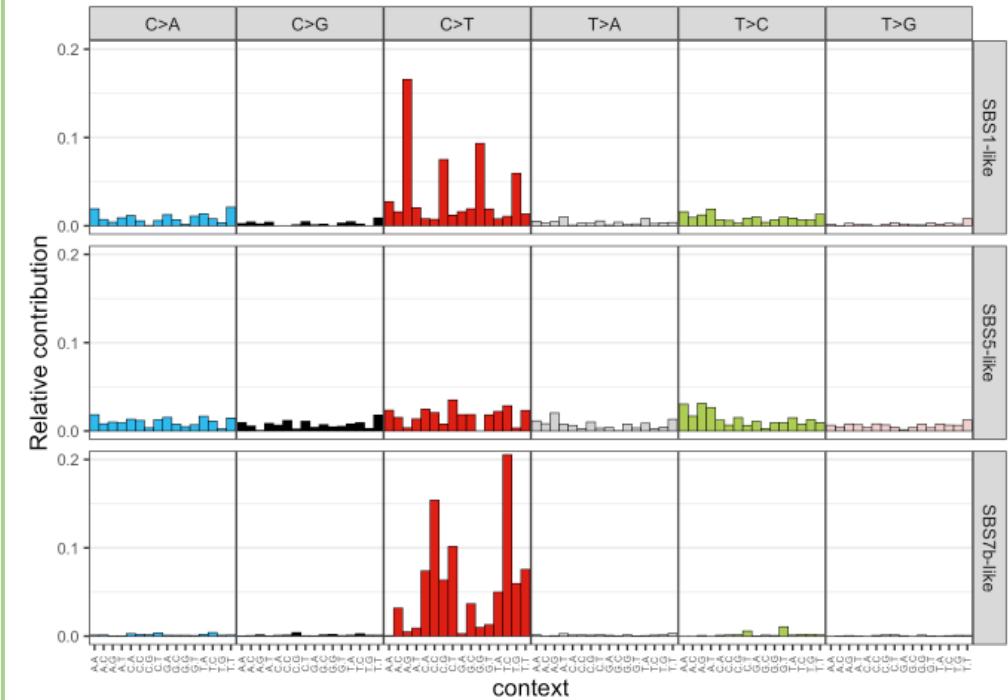
Explanation of how mutational signatures are derived from genomic data

5. Visualization and Interpretation

- The mutational signatures are visualized using different techniques, such as mutation spectra, signature plots, or cosine similarity matrices.

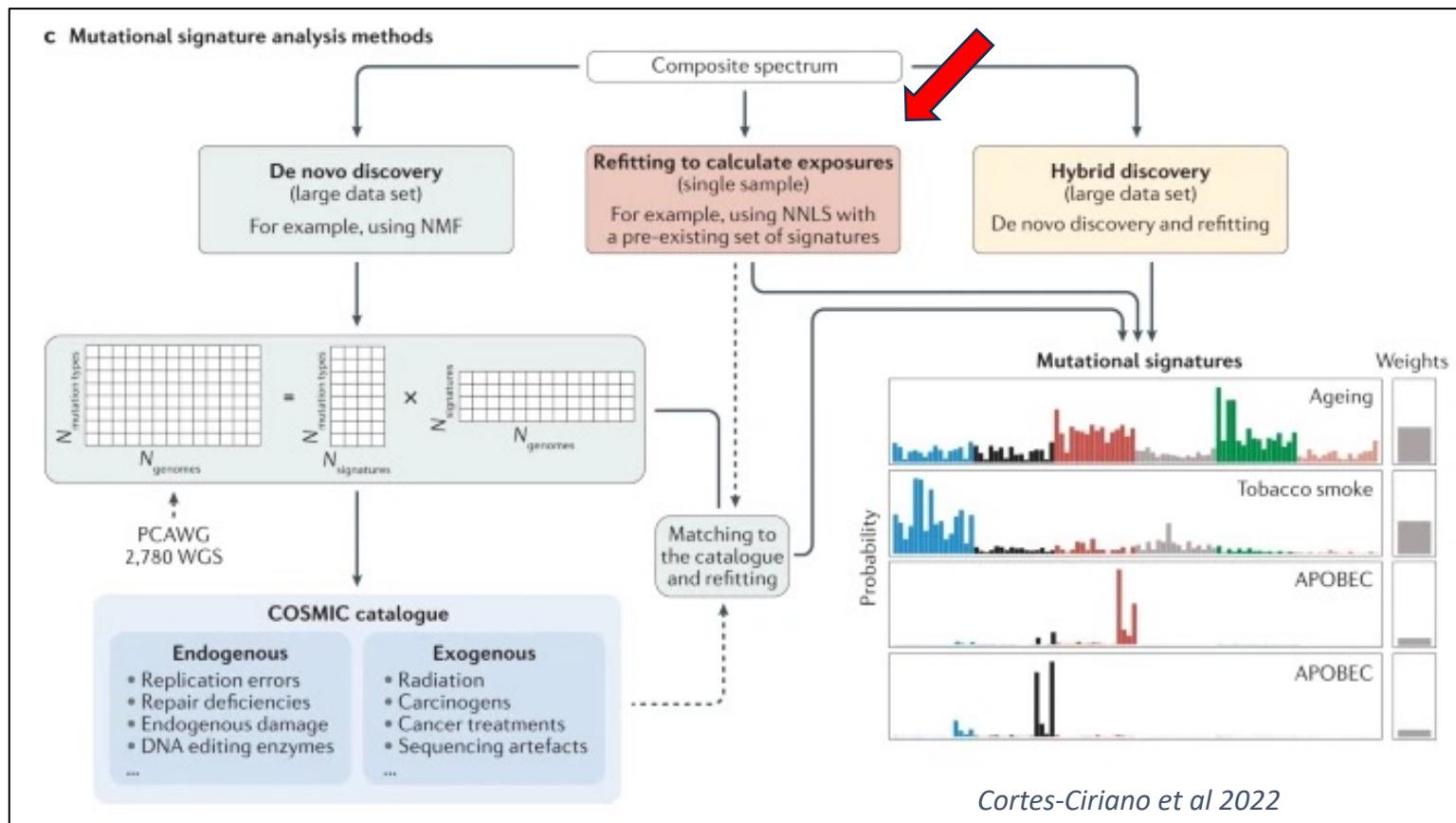
6. Validation and Replication

- The identified mutational signatures are validated and replicated using independent datasets.
- This step helps confirm the consistency and reproducibility of the identified signatures across different samples and cohorts.



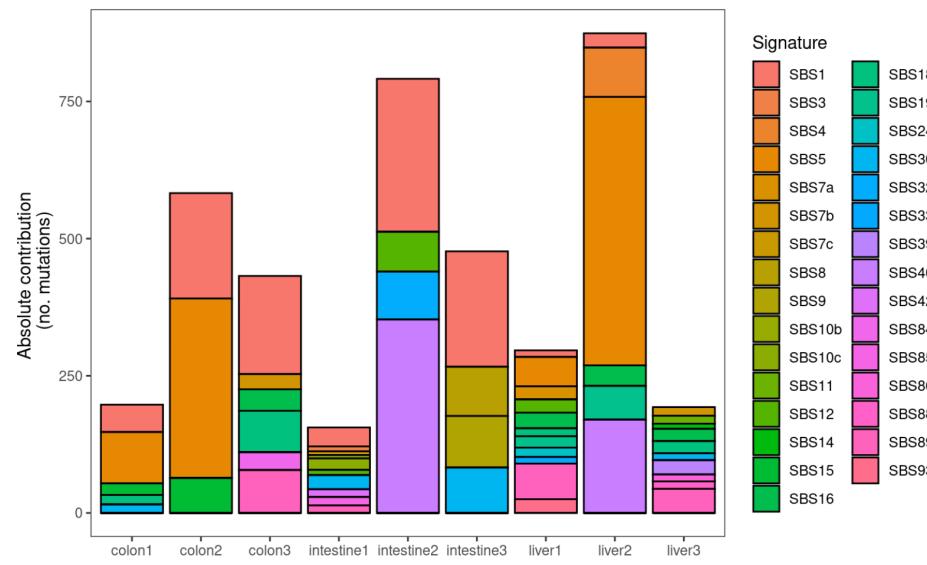
Signature refitting

- De novo discovery of mutational signatures is not always possible (low number of samples, mutations...)
- Signature refitting quantifies the contribution of any set of signatures to the mutational profile of a sample.

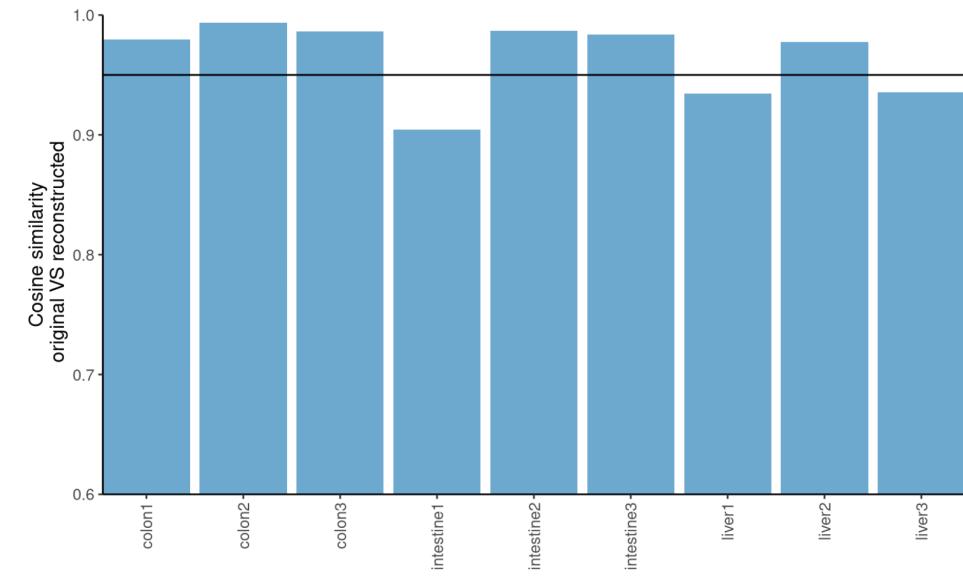


Signature refitting

- This is specifically useful for mutational signature analyses of small cohorts or individual samples, but also to relate own findings to known signatures and published findings.
 - Typical analysis: find mathematically optimal contribution of COSMIC signatures



Reconstruction quality
We can compute the cosine similarity to quantify the match between our mutational patterns and the decomposed COSMIC signatures



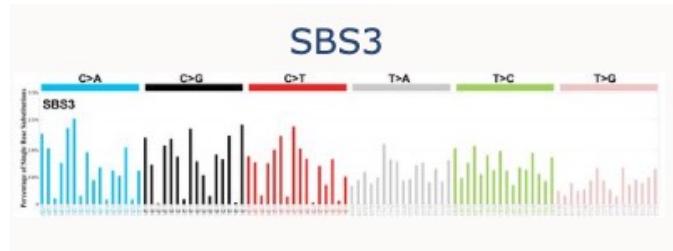
Clinical implications

- **Diagnosis and Subtyping:** Mutational signatures can help in the diagnosis and classification of different types of cancers.
- **Prognosis and Risk Assessment:** Certain mutational signatures are associated with different clinical outcomes.
- **Treatment Selection:** Mutational signature analysis can reveal mutations that are known to be associated with response or resistance to specific treatments.
- **Understanding Carcinogen Exposure:** Some mutational signatures are associated with specific mutagenic exposures, such as tobacco smoke, ultraviolet radiation, or certain chemicals.

Homologous recombination deficiency (HRD)

SBS3

PARP inhibitors

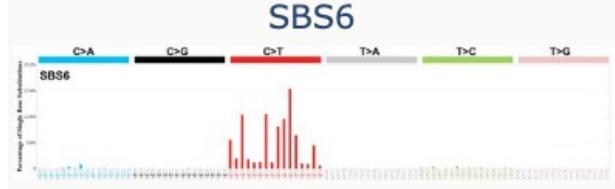


Mismatch repair deficiency (MMRd)

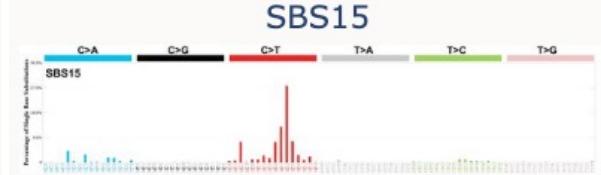
SBS6, 15, 26, 44...

Immune Checkpoint Inhibitors (ICI)

SBS6



SBS15



Limitations and Challenges

- **Signature Interpretation:** It can be difficult to assign a specific biological cause to each mutational signature identified.
- **Signature Heterogeneity:** Tumours can exhibit multiple mutational signatures, and the relative contributions of each signature can vary across different regions
- **Signature Identification:** Identifying novel mutational signatures can be a complex task as existing mutational signature databases may not cover the full
- **Sample Size and Data Quality:** Mutational signature analysis requires large sample sizes and high-quality sequencing
- **Technical Artifacts:** Mutational signature analysis can be affected by technical
- **Biological Complexity:** understanding their clinical implications can be complex.



Conclusions

- Mutational signature analysis identifies patterns of DNA mutations in cancer genomes.
- It can reveal exposure to mutagens and guide preventive measures.
- It provides insights into prognosis, treatment selection, and risk assessment.
- Challenges include interpretation, heterogeneity, sample size, technical artifacts, and limited functional understanding.

What is next?

Day two – Tuesday

11 July 2023

08:45 – 09:00	Arrival and registration	
09:00 – 09:45	Structural and copy-number variation analysis	Tobias Rausch
09:45 – 10:30	Interrogating cancer genomes using long reads	Tobias Rausch
10:30 – 11:00	Break	
11:00 – 12:00	Mutational processes and clonal population structure in cancer genomes	Francesc Muyas Remolar
12:00 – 13:00	Lunch	
13:00 – 15:00	Practicals: SV/CNV analysis using short-reads	Tobias Rausch
15:00 – 15:30	Break	
15:30 – 18:00	Practicals: Mutational signatures and clonal population structure analysis in cancer genomes	Francesc Muyas Remolar
18:00 – 18:30	Free time	
18:30	Dinner at Hinxton Hall Conference Centre	

Acknowledgements

Cortes-Ciriano group (EMBL-EBI)

Course organisers

All trainers

Isidro Cortes Ciriano, EMBL-EBI

Ajay Mishra, EMBL-EBI

Tobias Rausch, EMBL

Sophie Spencer, EMBL-EBI

