

# Class 17

## Mutational Signatures in Cancer



March 9<sup>th</sup>, 2023

Marcos Díaz-Gay

UC San Diego

 [marcos-diazg.github.io](https://github.com/marcos-diazg)

 [@mdiazgay](https://twitter.com/mdiazgay)

 [Alexandrov Lab](#)

# Today's agenda

Basics of cancer genomics: genomic sequencing data and somatic mutations identification

Exploring and obtaining tumor mutation data from **cBioPortal**

Characterization of the patterns of mutations in cancer

Mutational matrix generation using **Maftools**

Exploration of the biological processes generating mutations in different cancer types

Mutational signature analysis using **MutationalPatterns**

# Today's agenda

Basics of cancer genomics: genomic sequencing data and somatic mutations identification

Exploring and obtaining tumor mutation data from **cBioPortal**

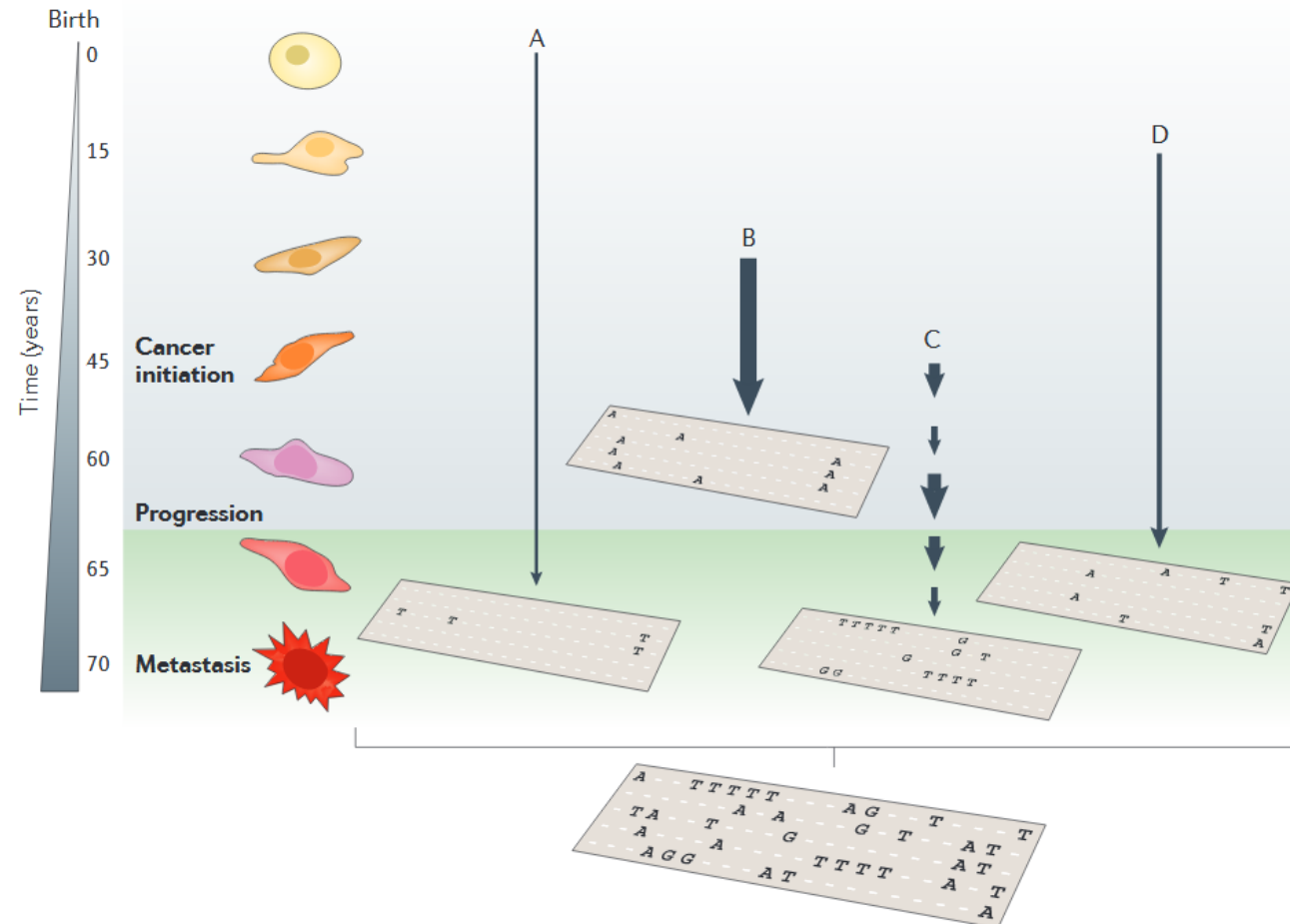
Characterization of the patterns of mutations in cancer

Mutational matrix generation using **Maftools**

**Exploration of the biological processes generating mutations in different cancer types**

Mutational signature analysis using **MutationalPatterns**

The mutational profile of a cancer patient is a mix of different processes characterized by specific mutational signatures

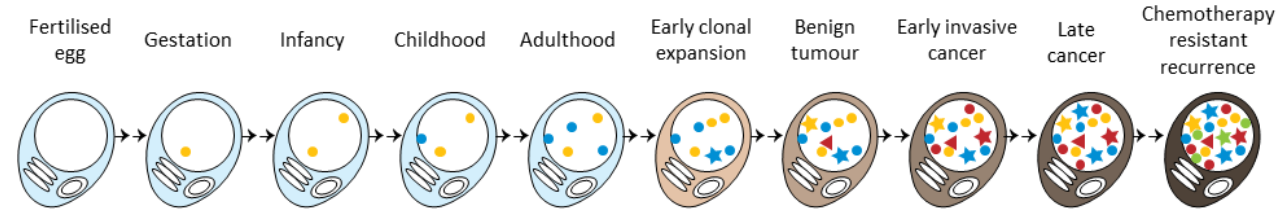


The final cancer genome represents an archaeological record of the effect of the different mutagenic and DNA repair processes

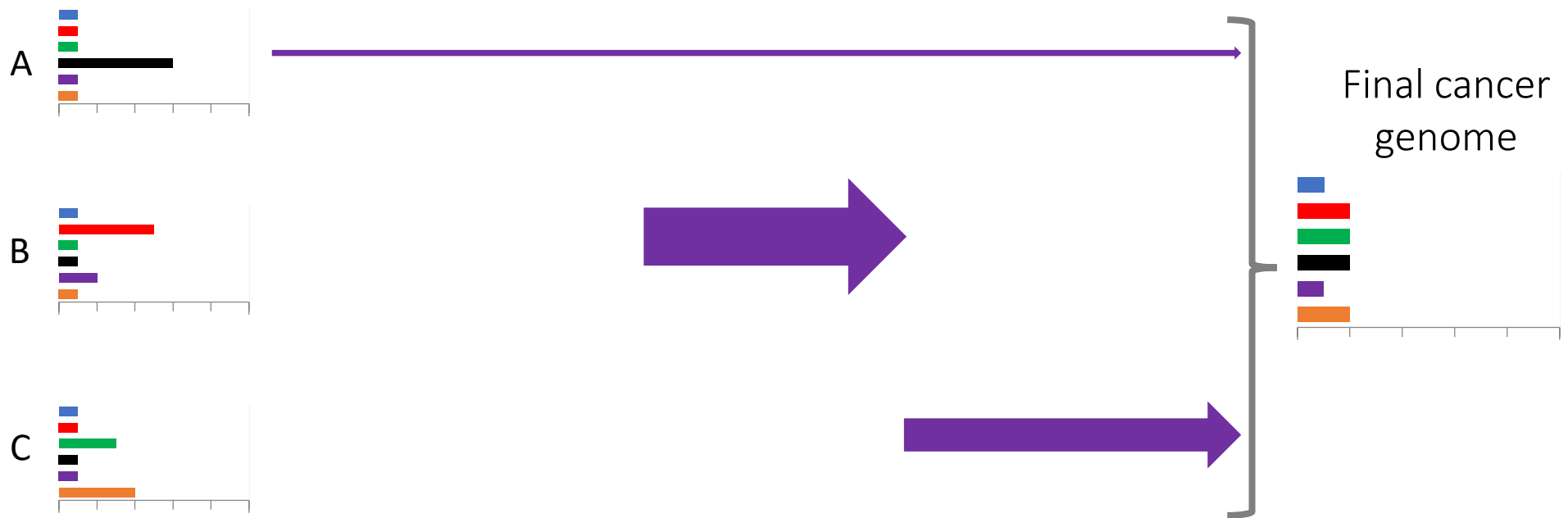
Chemotherapy  
resistant  
recurrence



# The final cancer genome represents an archaeological record of the effect of the different mutagenic and DNA repair processes



Mutational processes



# Mutational signatures can be determined based on mutational profiles across a set of individuals

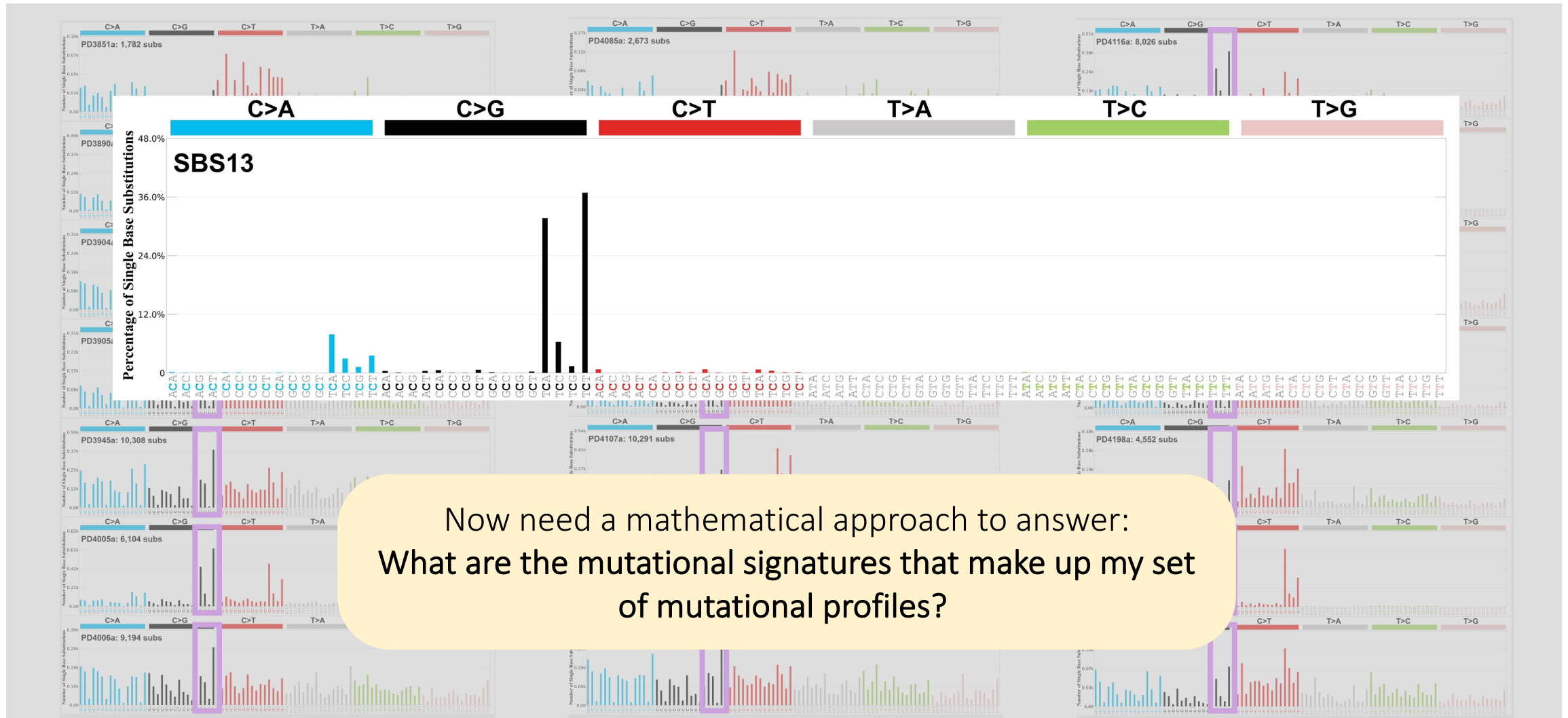


# Mutational signatures can be determined based on mutational profiles across a set of individuals



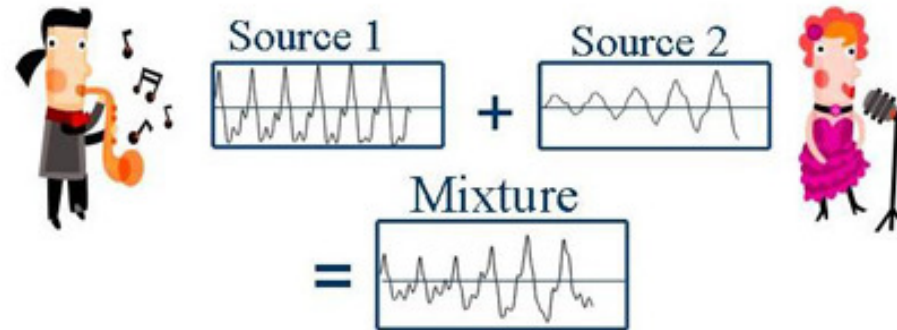


# Mutational signatures can be determined based on mutational profiles across a set of individuals



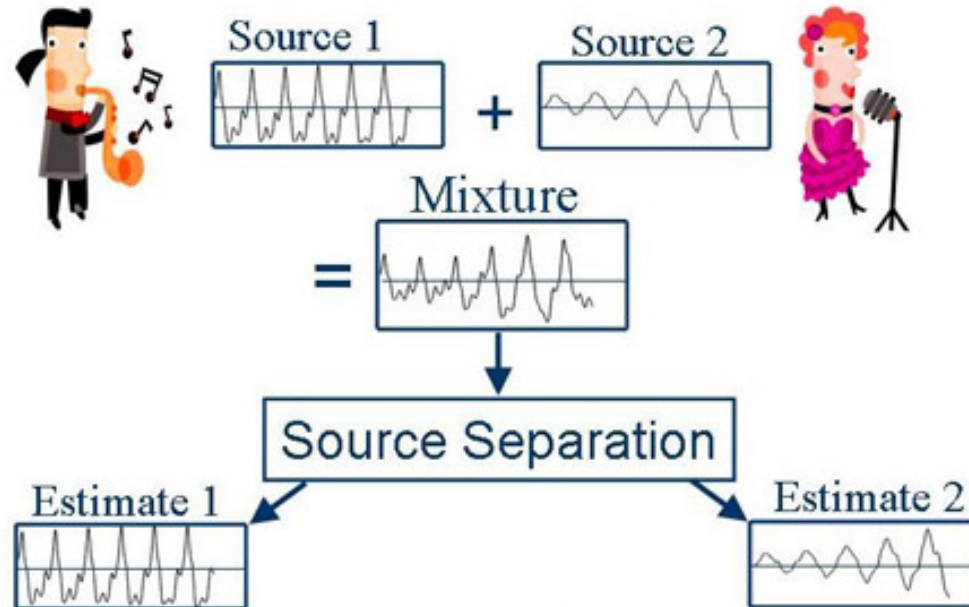
# Mathematical models allows the *un-mixing* and the extraction of mutational signatures

Blind signal separation problem



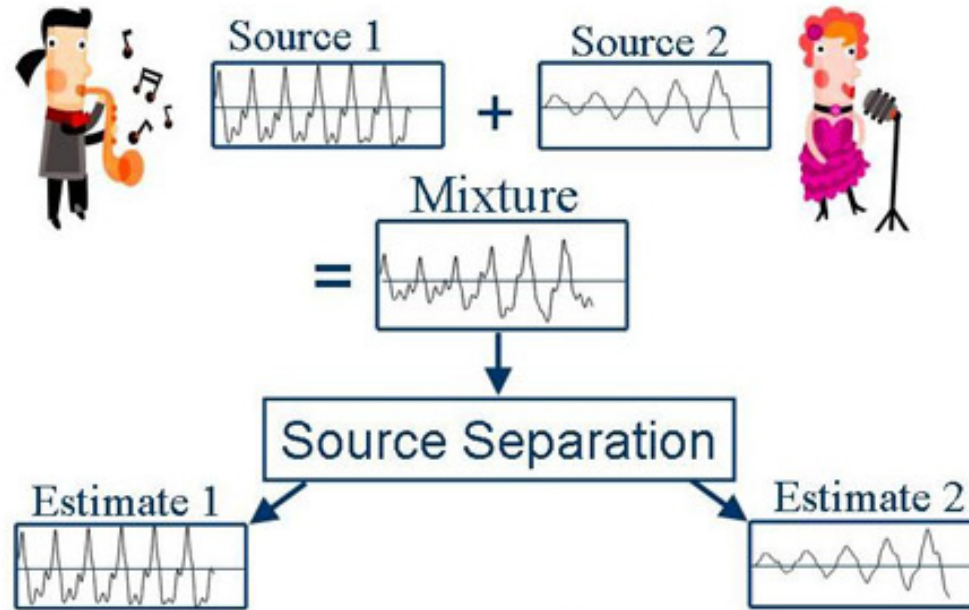
Mathematical models allows the *un-mixing* and the extraction of mutational signatures

Blind signal separation problem



Mathematical models allows the *un-mixing* and the extraction of mutational signatures

Blind signal separation problem



Non-negative matrix factorization

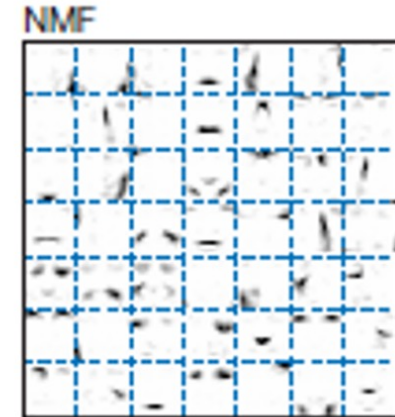
Mathematical models allows the *un-mixing* and the extraction of mutational signatures

.....  
**Learning the parts of objects by non-negative matrix factorization**

**Daniel D. Lee\* & H. Sebastian Seung\*†**

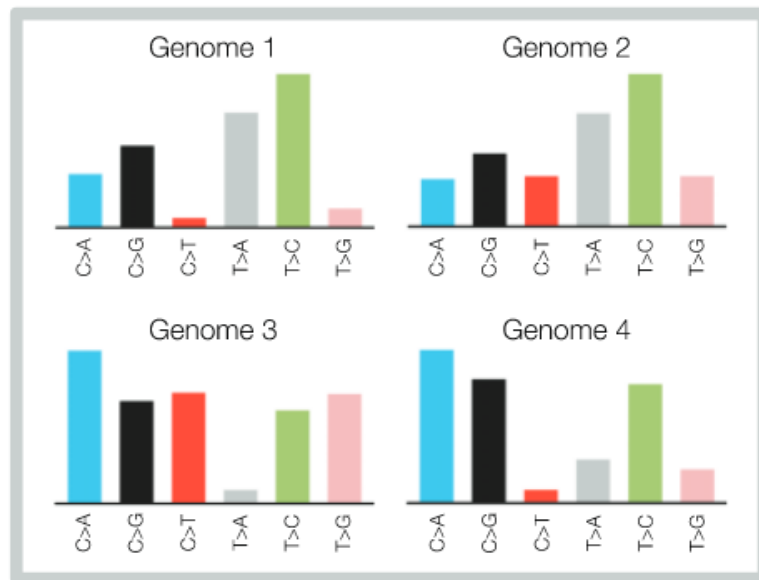
\* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA



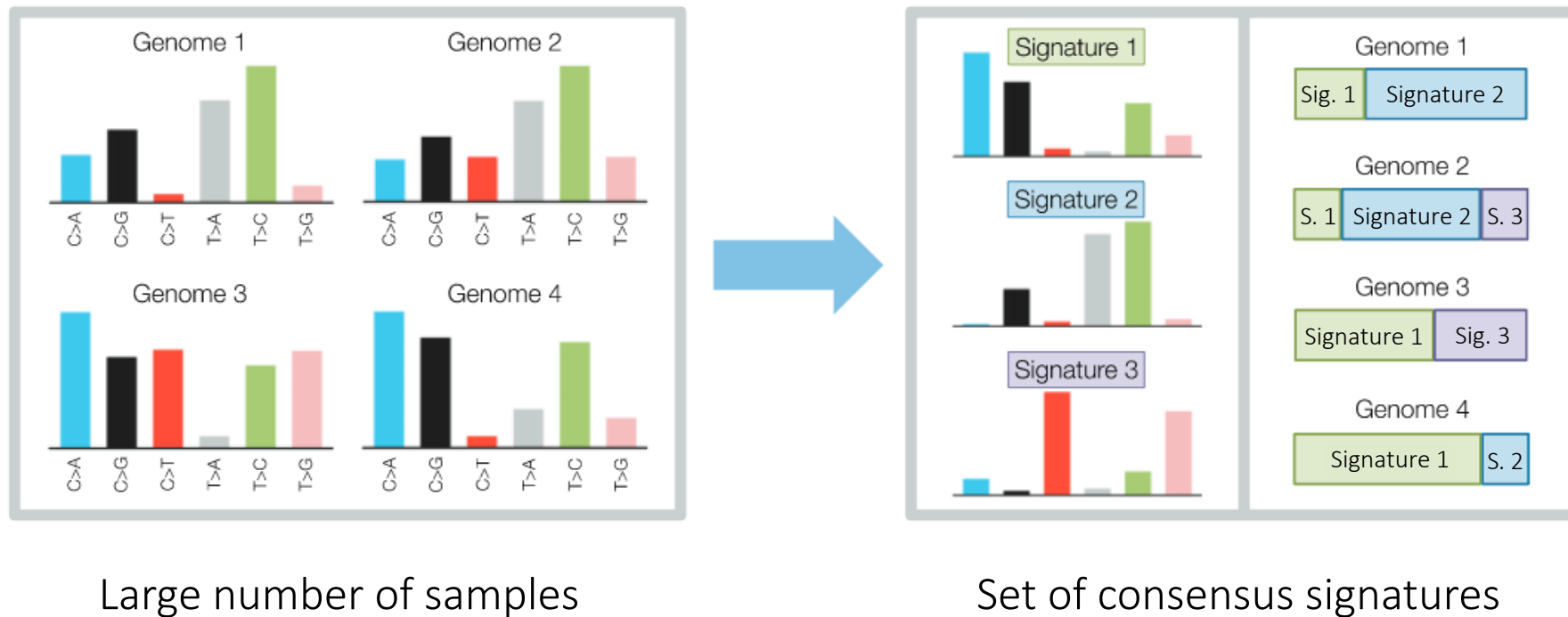
$$\begin{array}{ccc} \text{Mut. contexts} & & \text{Signatures} \\ \times & & \times \\ \text{Samples} & \leftarrow \mathbf{M} \approx \mathbf{S} \times \mathbf{A} \rightarrow & \text{Samples} \\ & \downarrow & \\ & \text{Mut. contexts} & \\ & \times & \\ & \text{Signatures} & \end{array}$$

Mathematical models allows the *un-mixing* and the extraction of mutational signatures

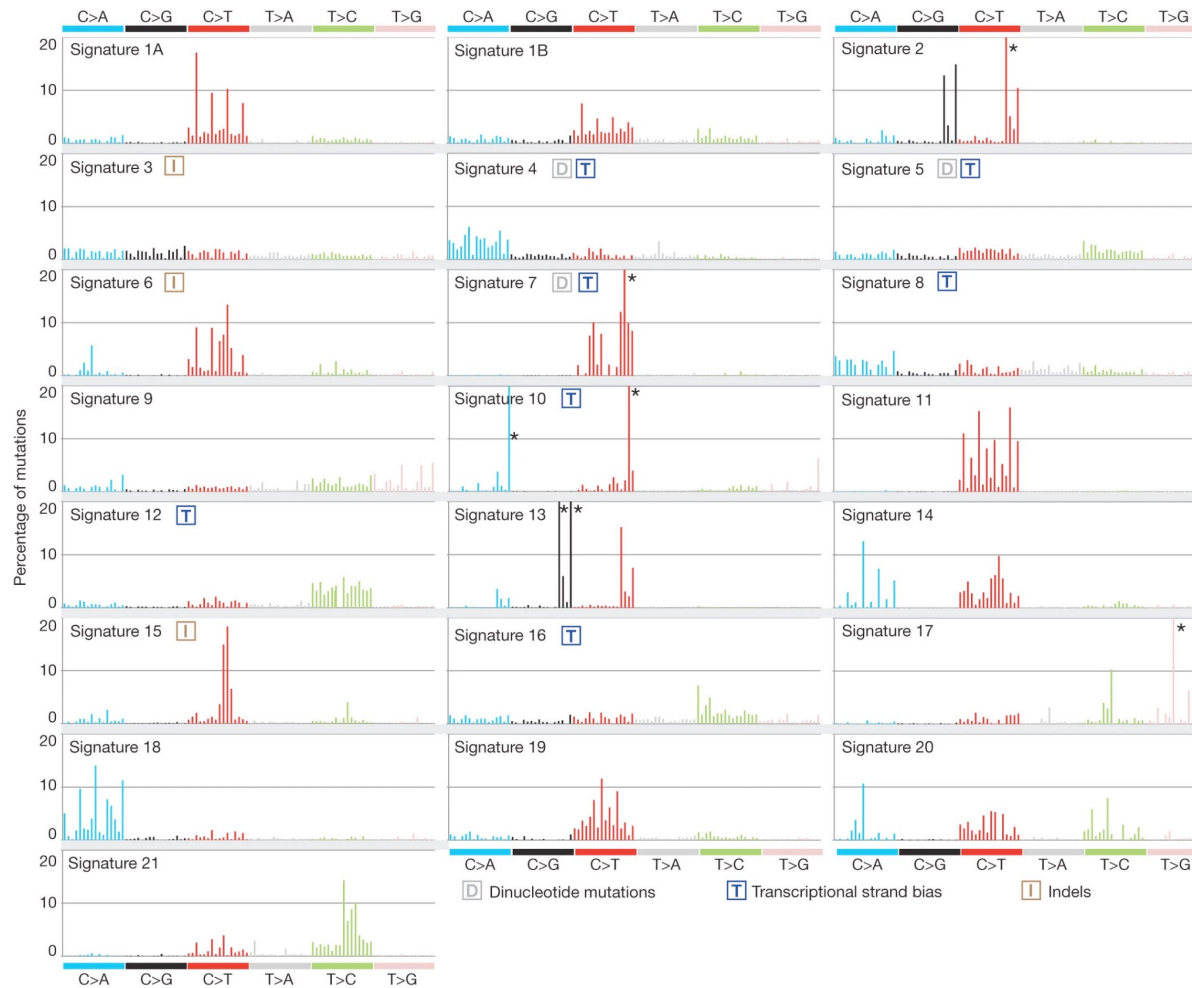


Large number of samples

# Mathematical models allows the *un-mixing* and the extraction of mutational signatures



# Reference mutational signatures have been extracted from thousands of samples



v1 (August 2013)

- 21 SBS signatures



Reference mutational signatures have been extracted from thousands of samples



v1 (August 2013)

- 21 SBS signatures

v2 (March 2015)

- 30 SBS signatures

# Reference mutational signatures have been extracted from thousands of samples



v3 (May 2019)

- 67 SBS signatures
- 11 DBS signatures
- 17 ID signatures

# Reference mutational signatures have been extracted from thousands of samples

**COSMIC**  
Catalogue Of Somatic Mutations In Cancer

Projects ▾ Data ▾ Tools ▾ News ▾ Help ▾ About ▾ Search COSMIC... SEARCH

## Mutational Signatures (v3.3 - June 2022)

### Introduction

**Somatic mutations are present in all cells of the human body and occur throughout life. They are the consequence of multiple mutational processes, including the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA and defective DNA repair. Different mutational processes generate unique combinations of mutation types, termed "Mutational Signatures".**

In the past few years, large-scale analyses have revealed many mutational signatures across the spectrum of human cancer types, including the latest effort by the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Network (Alexandrov, L.B. et al., 2020) using data from more than 23,000 cancer patients.

### About

COSMIC Mutational Signatures is a resource curated in partnership with COSMIC and Cancer Grand Challenges, and in close association with our collaborators at Wellcome Sanger Institute, the Pillay lab at University College London and the Alexandrov lab at University of California.

**wellcome sanger institute** **CANCER GRAND CHALLENGES** **COSMIC**

### Signature-based websites

At COSMIC Signatures we identify signatures from analysis of the PCAWG dataset and through curation of specific papers. Papers are looked at particularly (but not exclusively) when there is a specific exposure which captures signatures not present in the PCAWG dataset. Please note that this catalogue of signatures is not exhaustive or a final set, but a reference set of high confidence signatures that have been curated by experts in the field. We aim to update as comprehensively as possible as new data become available and improvements are made to extraction methodologies.

This summary includes the mutational profile, proposed aetiology and tissue distribution of each signature, as well as potential associations with other mutational signatures and how the signature has changed during iterations of analysis.

Currently, four different variant classes are considered, resulting in the following sets of mutational signatures.

**SBS Signatures** **DBS Signatures** **ID Signatures** **CN Signatures**

### Data downloads

Download current COSMIC Mutational Signatures version 3.3 and previous releases here.

**Downloads**

### Versions

COSMIC Mutational Signatures version 3.3 is the latest release.

Version 3 was released as part of COSMIC release v89 (May 2019), updated to version 3.1 in COSMIC release v91 (June 2020), to version 3.2 in COSMIC release v93 (March 2021) and most recently version 3.3 in COSMIC v95 (May 2022).

Version 2 signatures (March 2015) were part of earlier COSMIC releases can still be consulted:

**Version 2**

### SigProfiler tools

The current set of mutational signatures has been extracted using SigProfiler, a compilation of publicly available bioinformatic tools addressing all the steps needed for signature identification. SigProfiler functionalities include mutation matrix generation from raw data and signature extraction, among others.

**SigProfiler Tools**

### Mutational signatures as a collection of operative mutational processes

Mutational processes from different aetiologies are active during the course of cancer development. They can be identified using mutational signatures, due to their unique mutational pattern and specific activity on the genome.

This is illustrated in the figure below using a framework of 6 classes of single base substitutions, and three distinct mutational processes, whose respective strengths vary throughout a patient's life. At the beginning, all mutations were due to the activity of the endogenous mutational process. As time progresses, the other processes get activated and the mutational spectrum of the cancer genome continues to change.

**Endogenous mutational process**  
Clock-like signature

**Strong exogenous mutational process**  
Tobacco smoking

**Moderate mutational process activated at different times**  
APOBEC deaminase activity

**Mutational spectrum of final cancer genome**

**Number of mutations**

**Time**

**Signature activity**



Current set (v3.3)

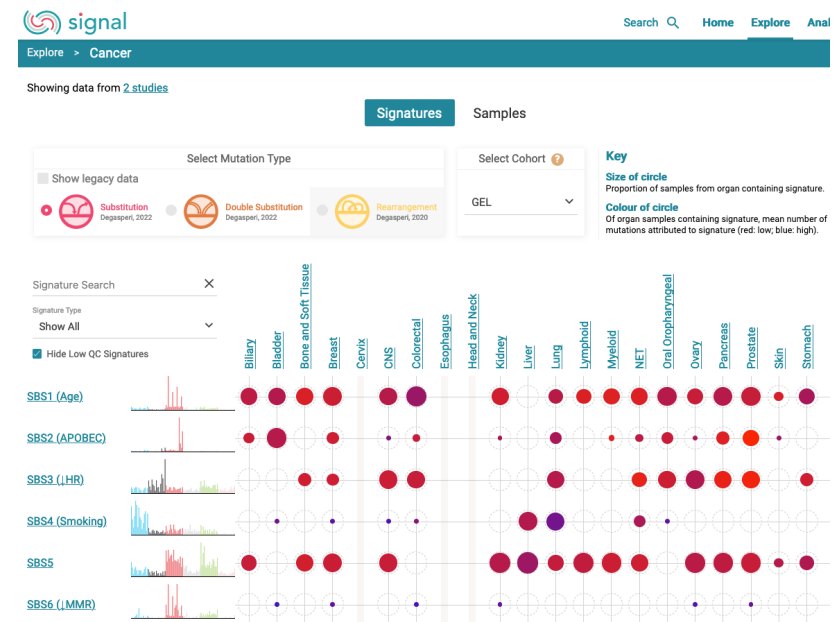
- 78 SBS signatures
- 11 DBS signatures
- 18 ID signatures
- 21 CN signatures

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

# Reference mutational signatures have been extracted from thousands of samples

Other reference databases exist that include different variant classes

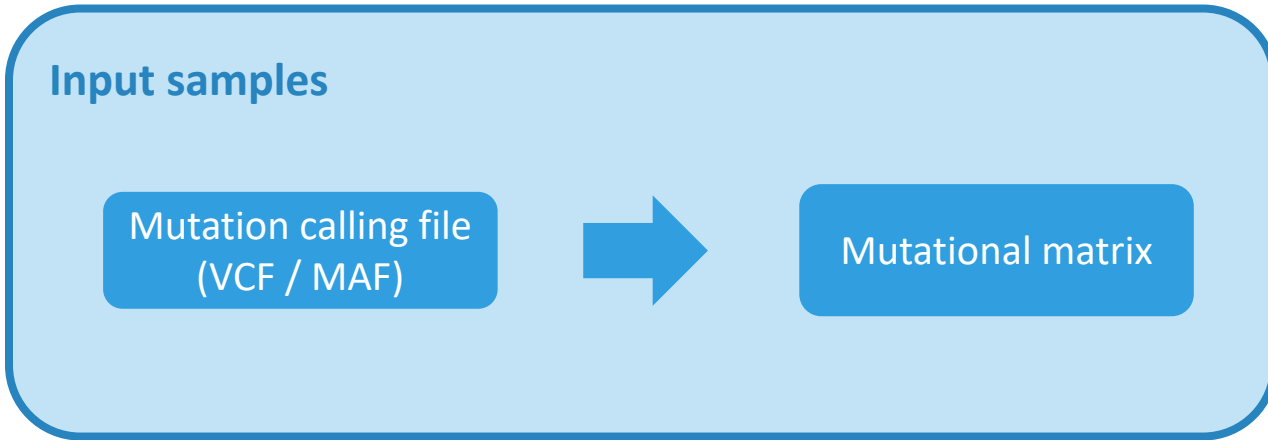
# MUTAGENE



<https://www.ncbi.nlm.nih.gov/research/mutagene/>

<https://signal.mutationalsignatures.com/>

# Overview of mutational signature assignment analysis

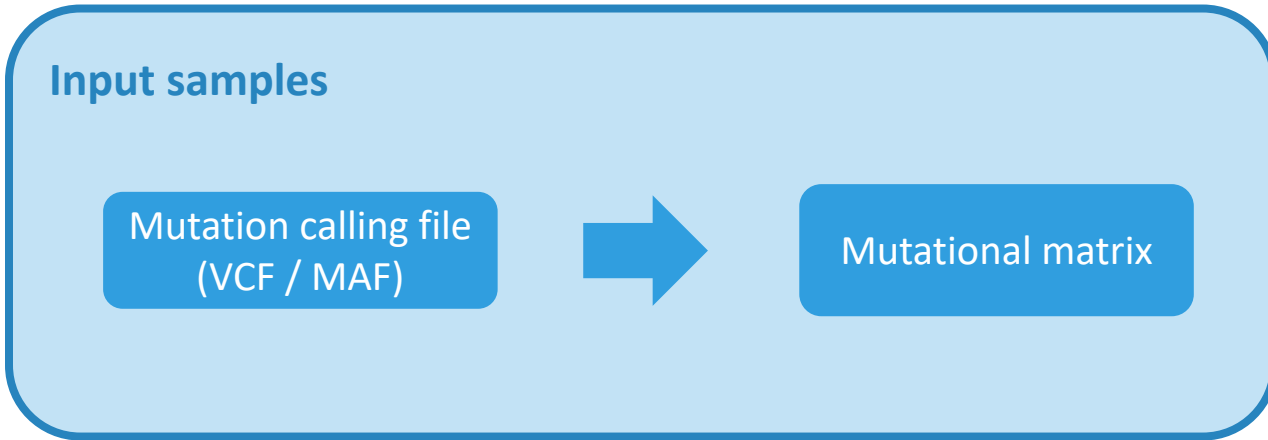


**Reference mutational signatures**

**COSMIC**  
Catalogue Of Somatic Mutations In Cancer

- v1, v2, v3, v3.1, v3.2, v3.3
- GRCh37/38, mm9/10, rn6
- Genome / Exome

# Overview of mutational signature assignment analysis



$$\mathbf{M} = \mathbf{S} \times \mathbf{A}$$

$t \times n$        $t \times k$        $k \times n$

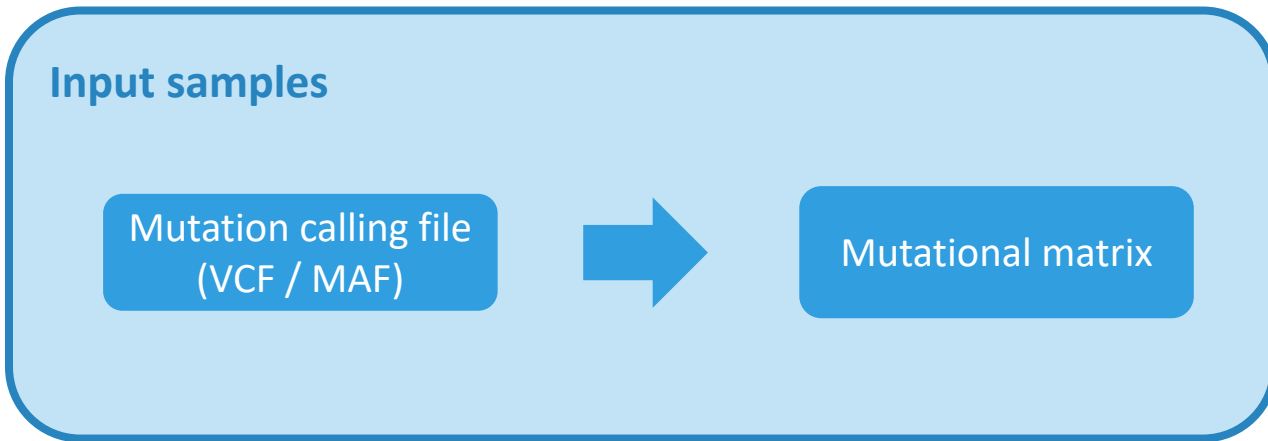
**Reference mutational signatures**

v1, v2, v3, v3.1, v3.2, v3.3

GRCh37/38, mm9/10, rn6

Genome / Exome

# Overview of mutational signature assignment analysis



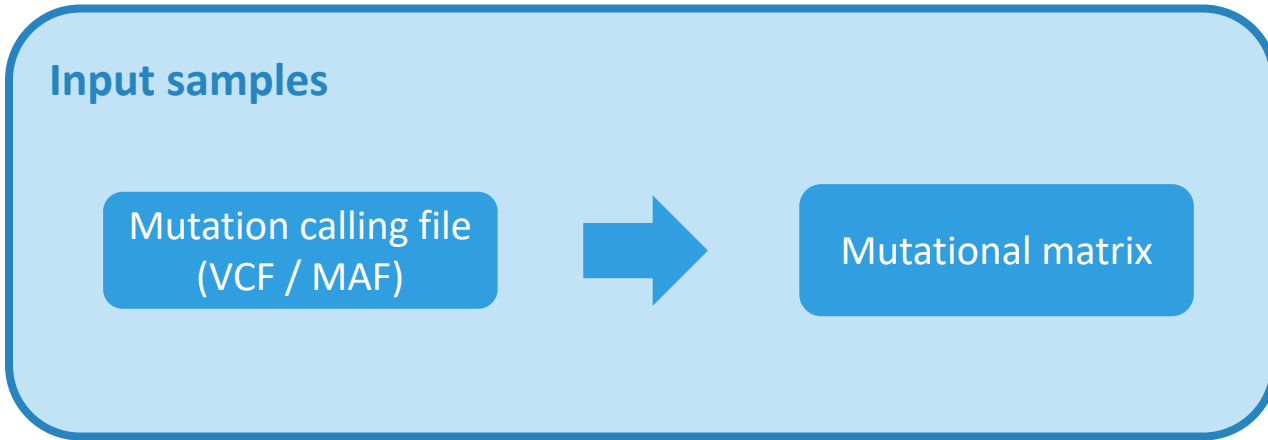
$$\mathbf{M} = \mathbf{S} \times \mathbf{A}$$

$t \times n$        $t \times k$        $k \times n$



$t$  mutational contexts  
 $n$  samples  
 $k$  signatures

# Overview of mutational signature assignment analysis



$$\mathbf{M} = \mathbf{S} \times \mathbf{A}$$

$t \times n$        $t \times k$        $k \times n$

**Reference mutational signatures**

 COSMIC  
Catalogue Of Somatic Mutations In Cancer

v1, v2, v3, v3.1, v3.2, v3.3

GRCh37/38, mm9/10, rn6

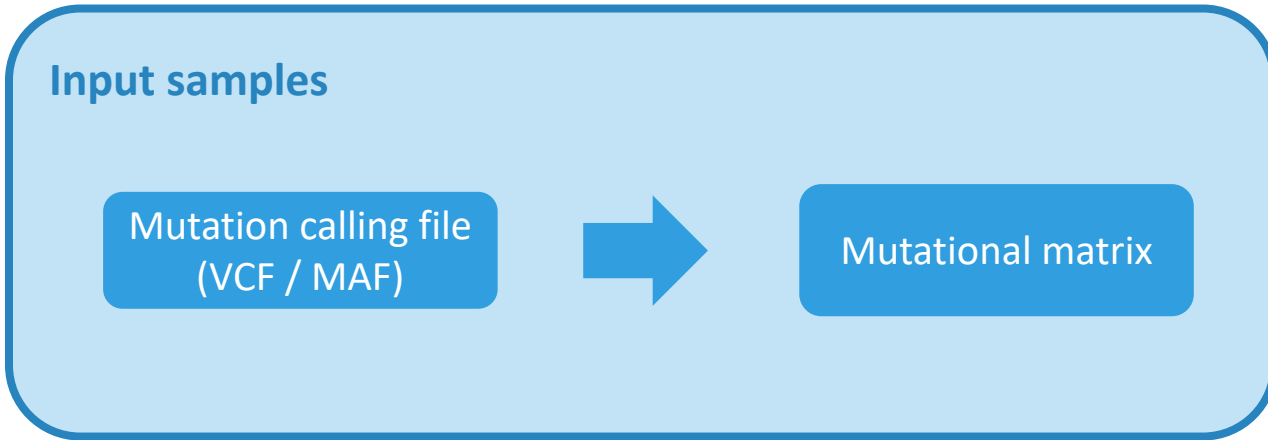
Genome / Exome



t mutational contexts  
n samples  
k signatures



# Overview of mutational signature assignment analysis



$$\begin{matrix} \mathbf{M} \\ t \times n \end{matrix} = \begin{matrix} \mathbf{S} \\ t \times k \end{matrix} \times \begin{matrix} \mathbf{A} \\ k \times n \end{matrix}$$

t mutational contexts  
n samples  
k signatures

**Reference mutational signatures**

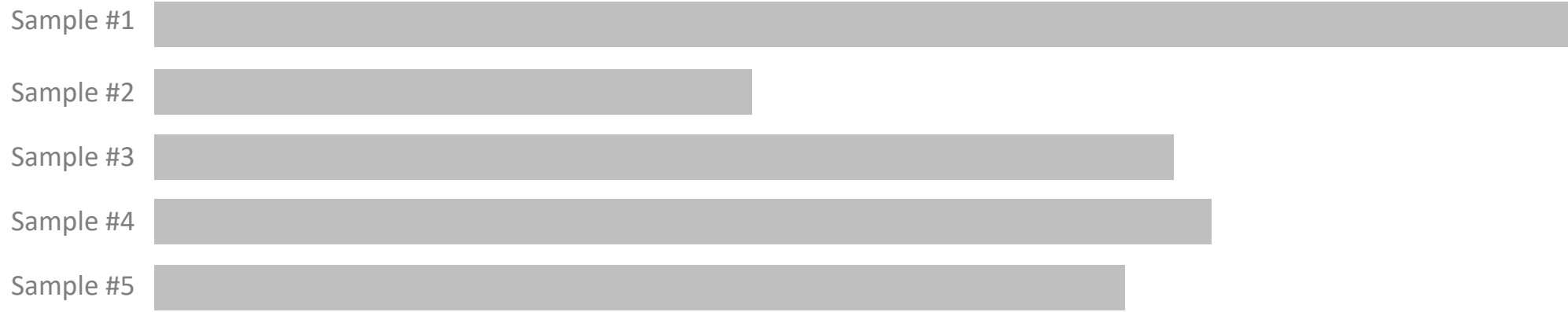
v1, v2, v3, v3.1, v3.2, v3.3

GRCh37/38, mm9/10, rn6

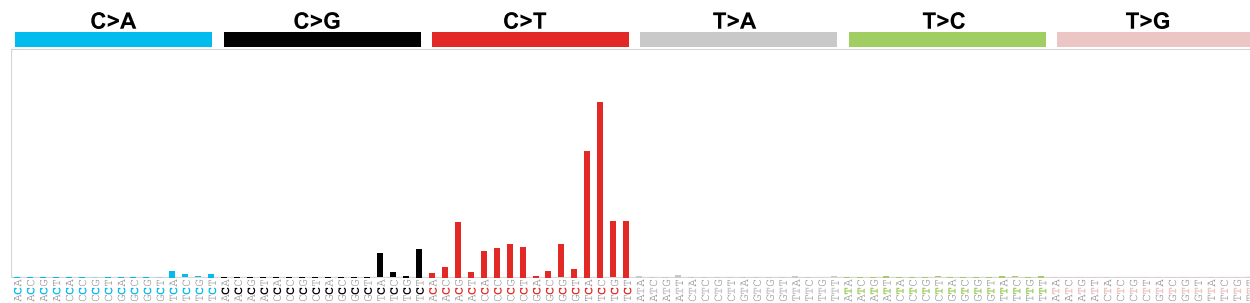
Genome / Exome

# Assigning reference signatures to individual samples

# M



Number of mutations



Mutational profile

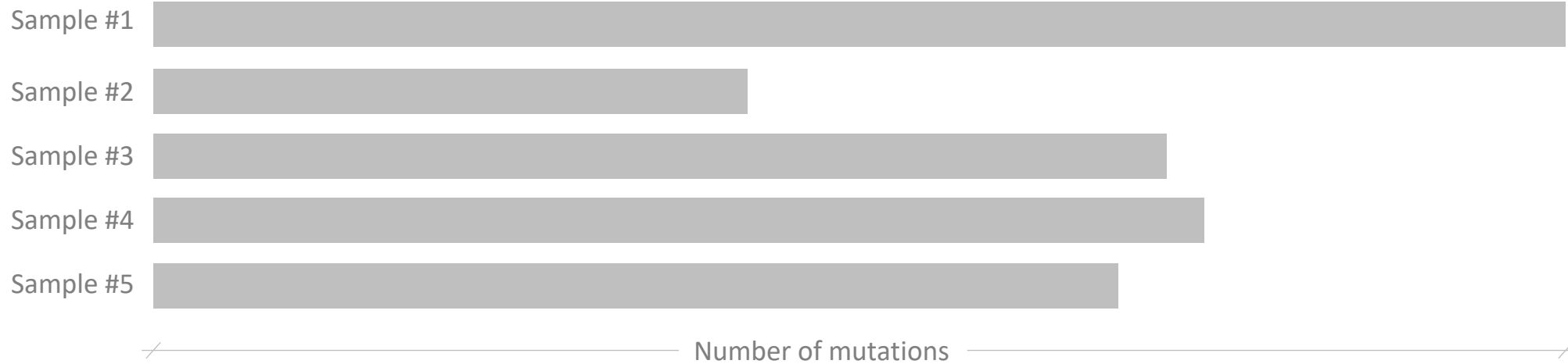
#CHROM	POS	FILTER	REF	ALT
1	809687	PASS	G	C
1	819245	PASS	G	T
1	1911011	PASS	C	G
1	2112413	PASS	T	C
1	2927666	PASS	A	G
1	3359791	PASS	C	T
1	4347912	PASS	G	A
1	4961889	PASS	G	C
1	5949138	PASS	C	T
1	7806339	PASS	A	C
1	9648435	PASS	G	A
1	9705025	PASS	C	T

Mutation calling file

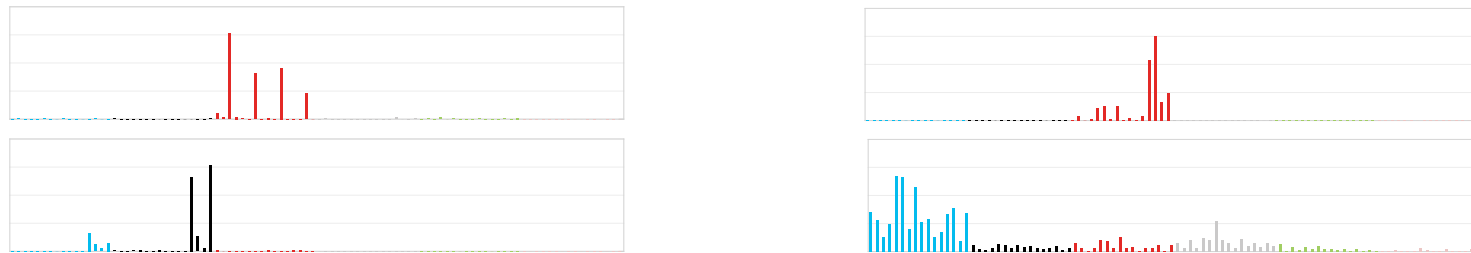


# Assigning reference signatures to individual samples

## M

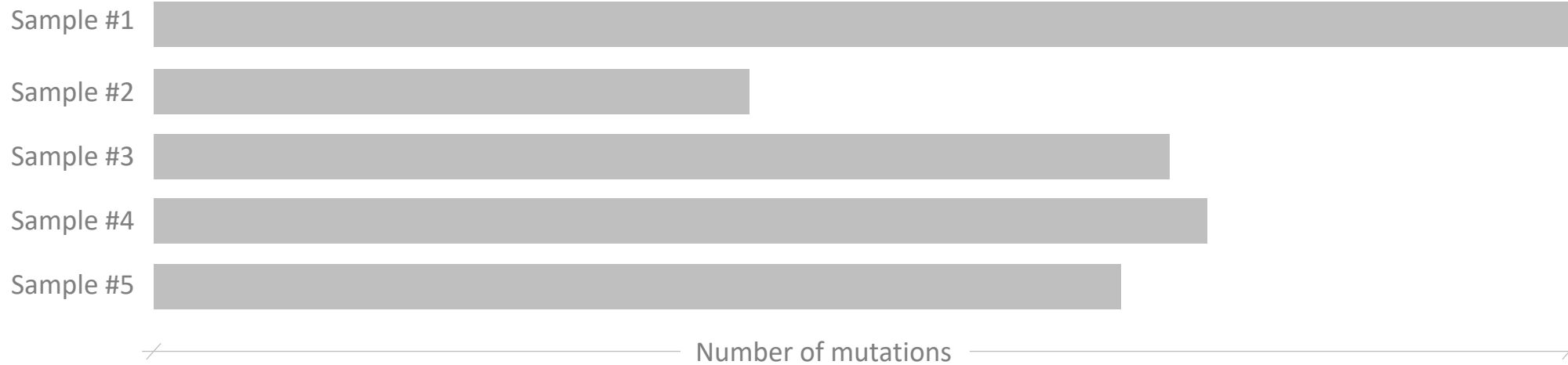


## S

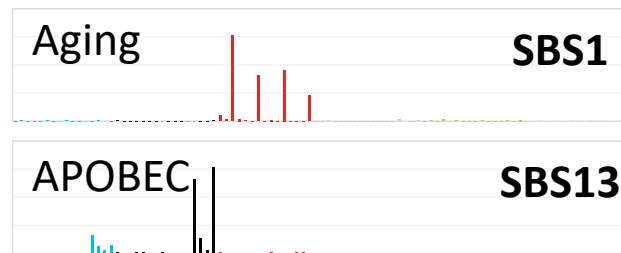


# Assigning reference signatures to individual samples

## M

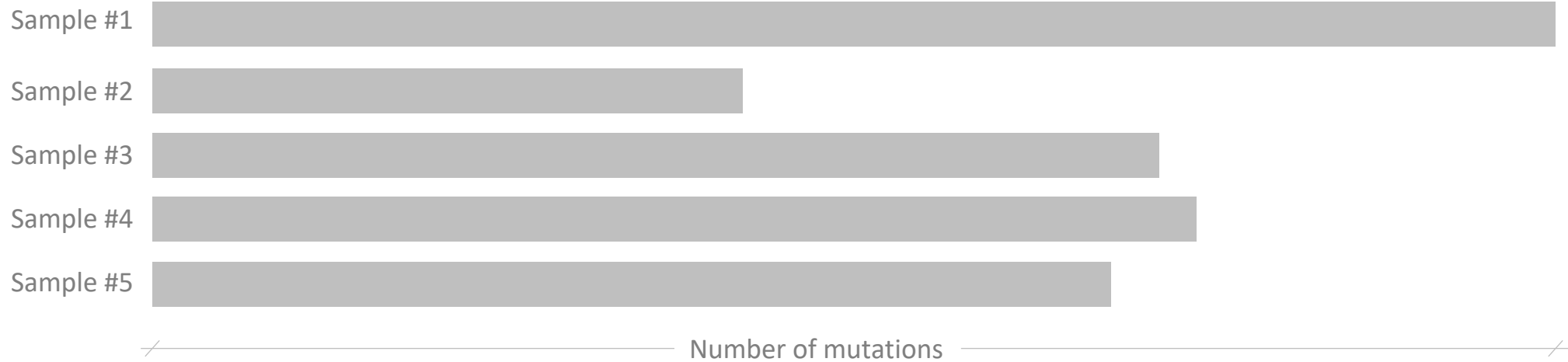


## S

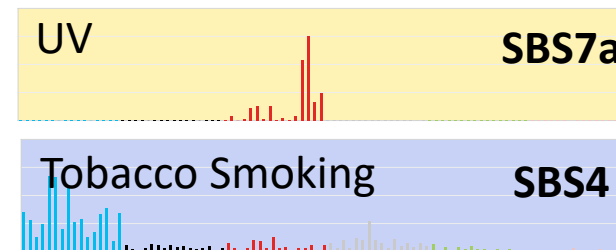
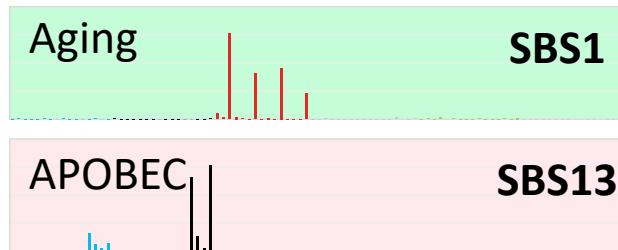


# Assigning reference signatures to individual samples

## M

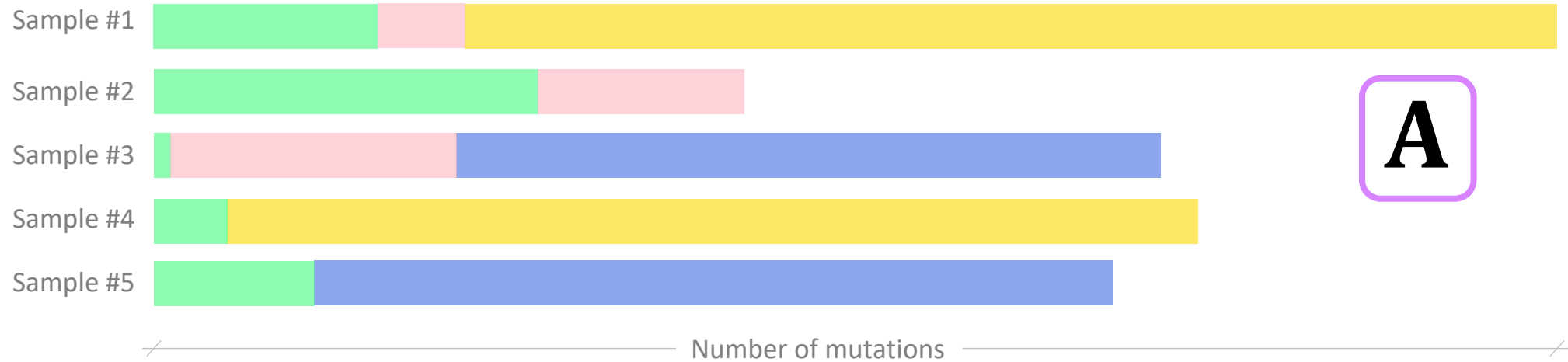


## S

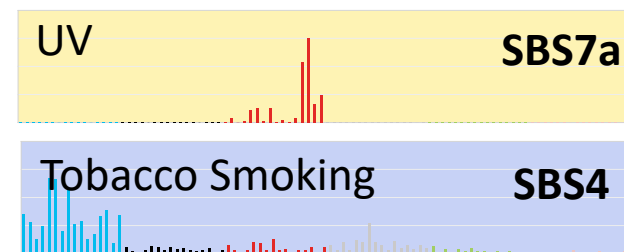
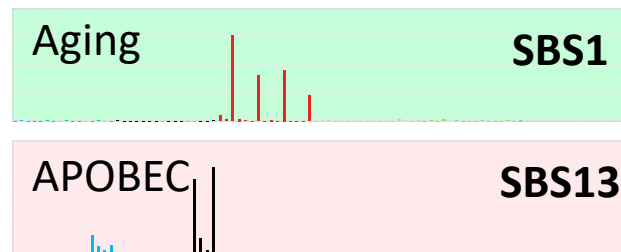


# Assigning reference signatures to individual samples

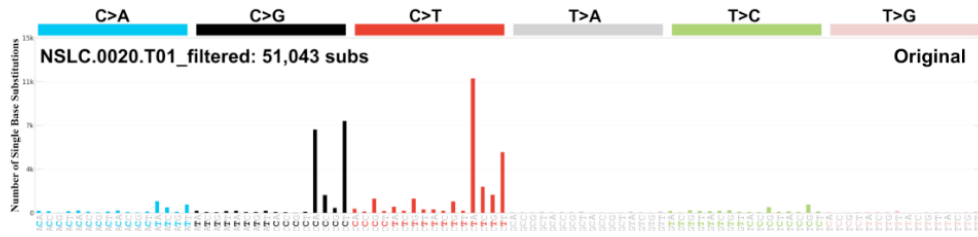
## M



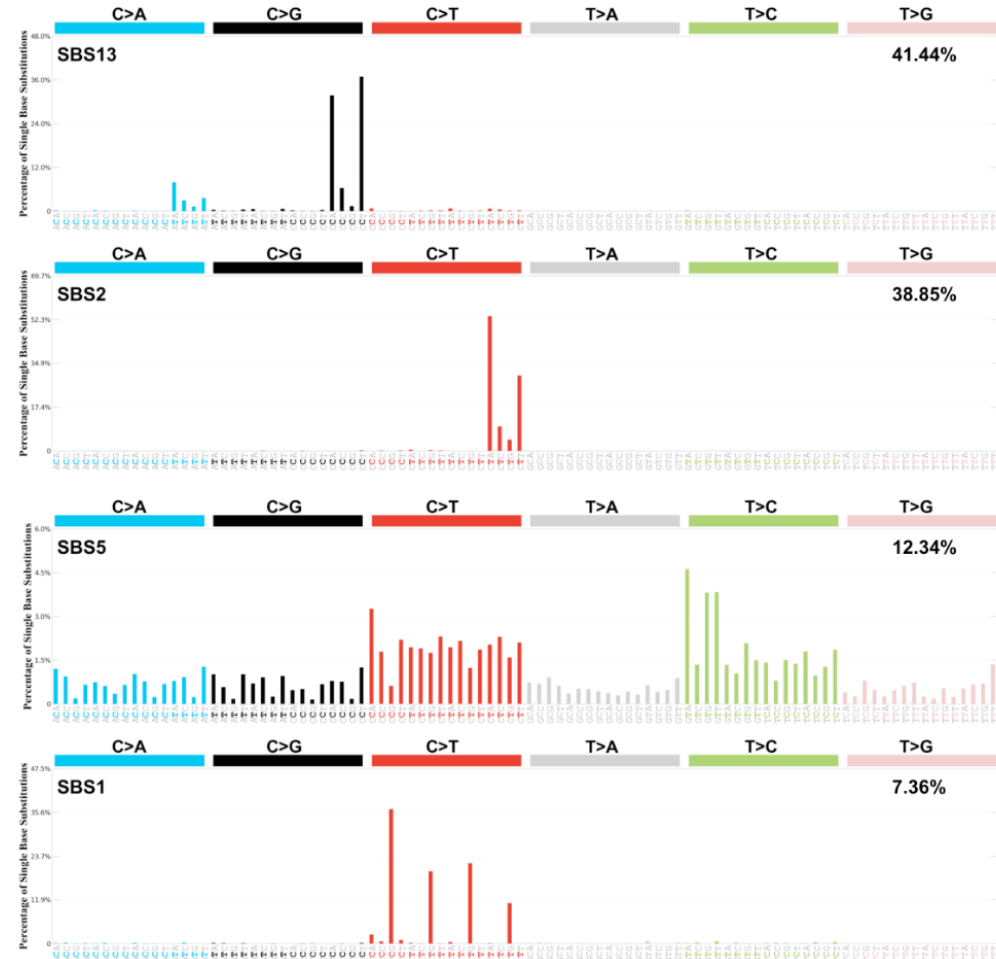
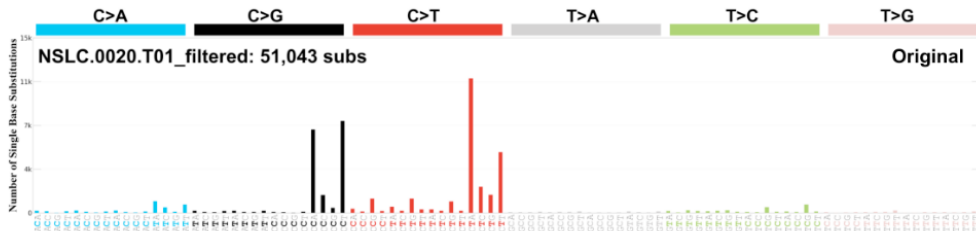
## S



# Assigning reference signatures to individual samples

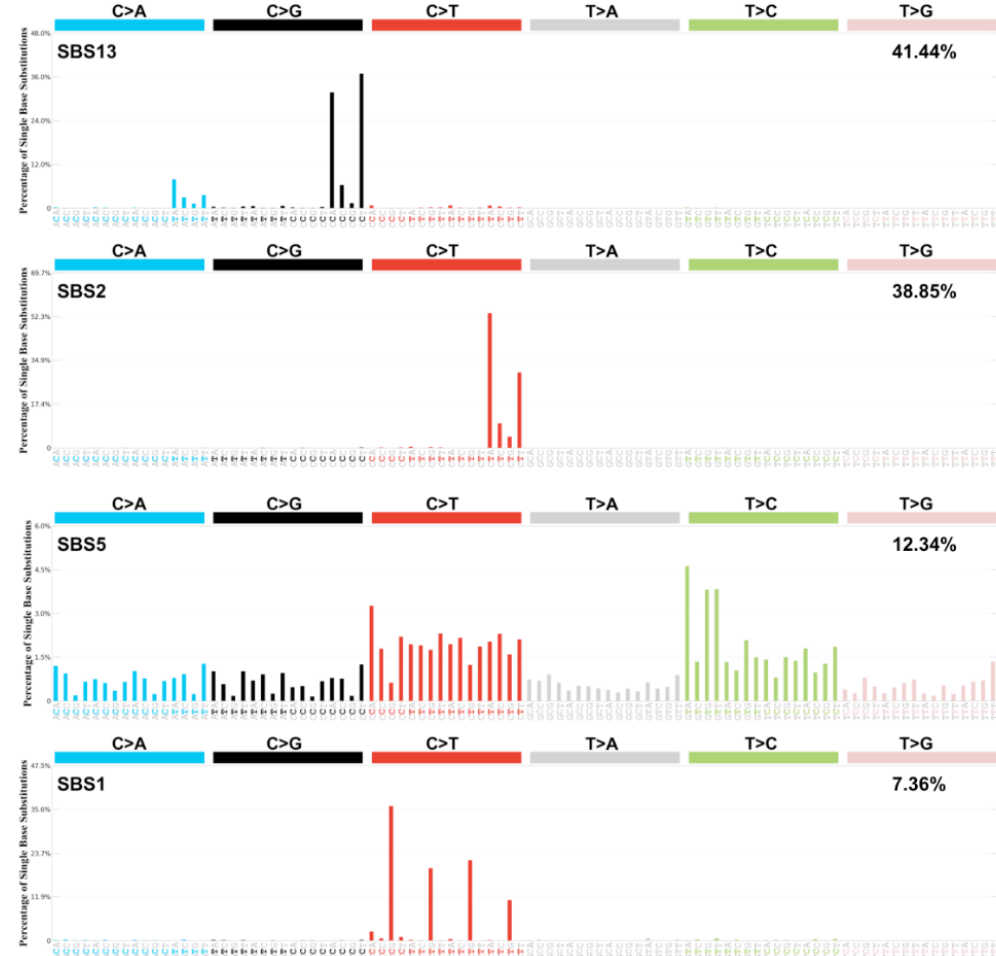
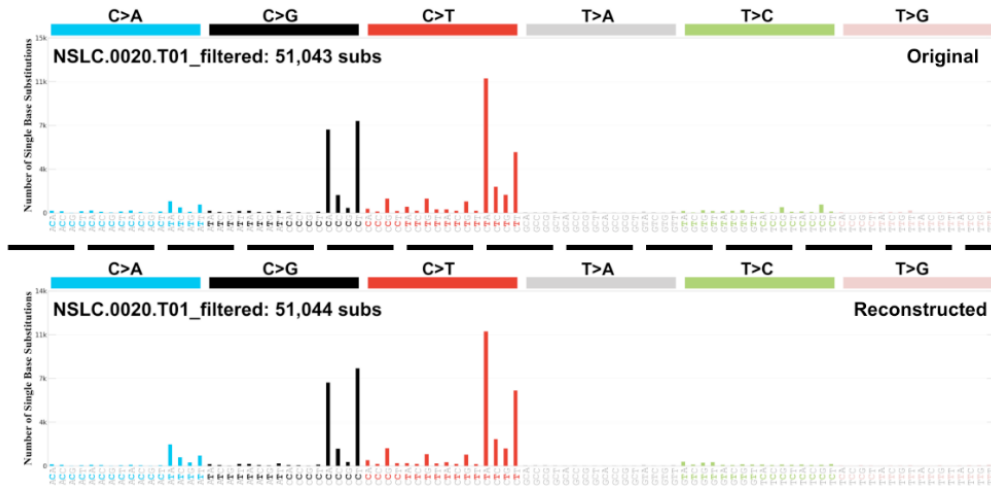


# Assigning reference signatures to individual samples

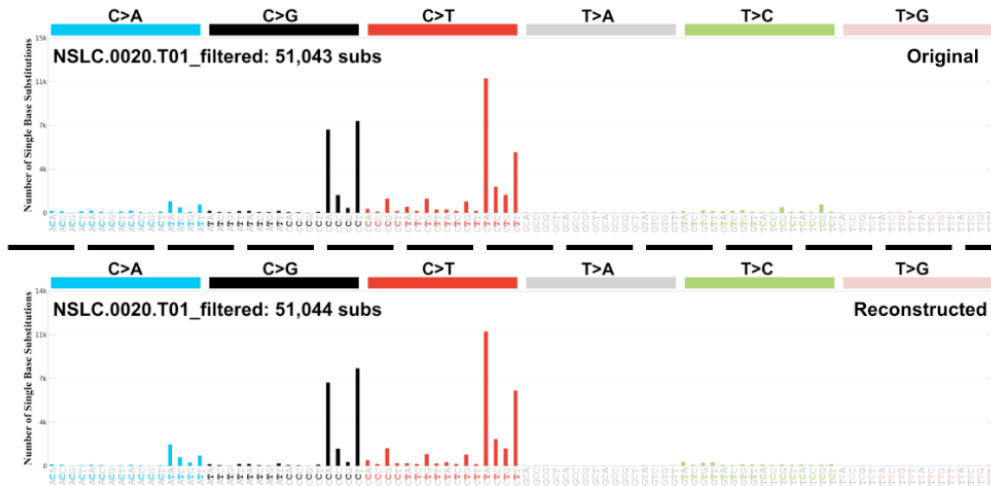




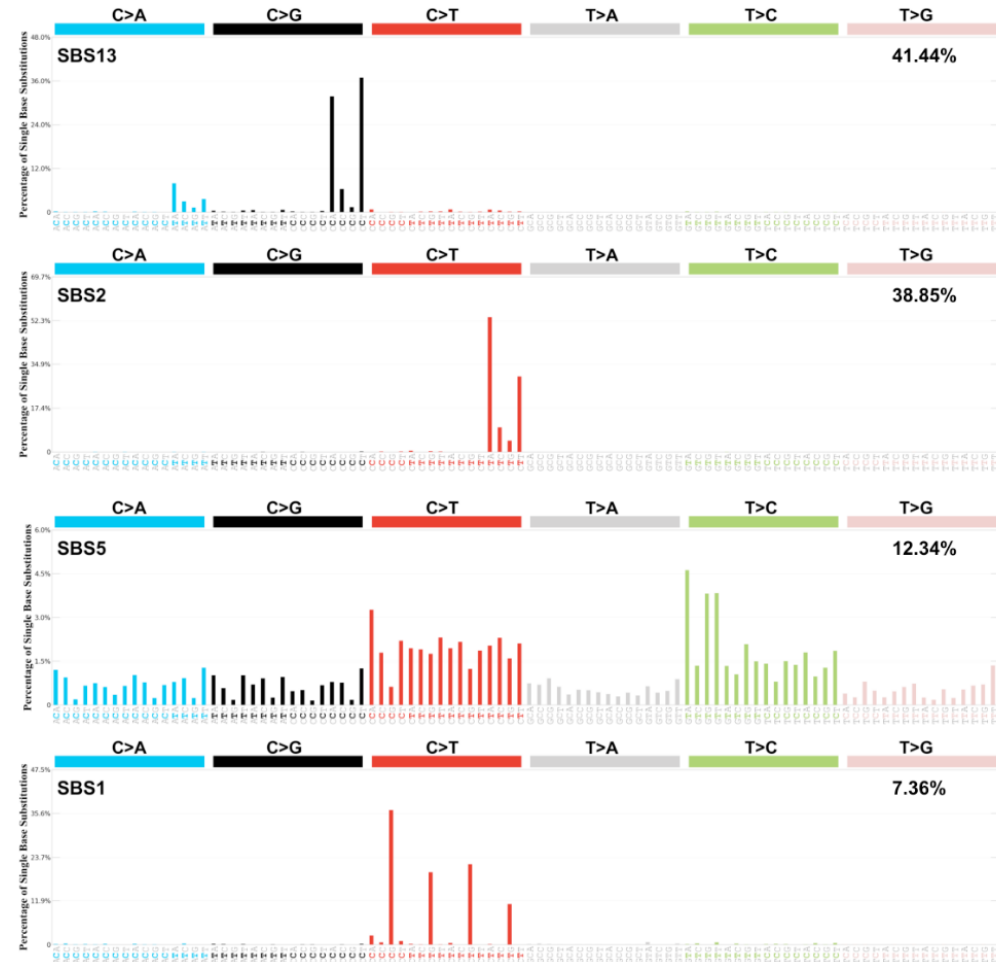
# Assigning reference signatures to individual samples



# Assigning reference signatures to individual samples



Cosine Similarity: 0.995    L1 Error %: 15.13%    KL Divergence: 0.052  
 Correlation: 0.994        L2 Error %: 10.37%    Signature Version: 3.3



# Mutational signature analysis marked a turning point in cancer diagnosis, prognosis and treatment

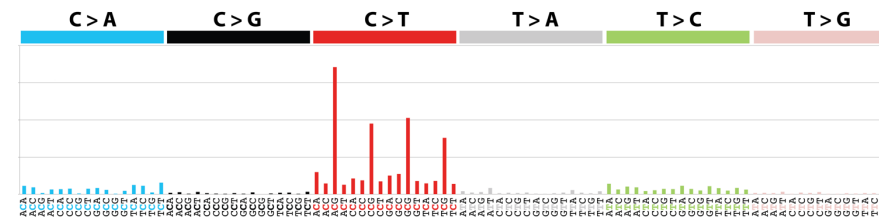


Cancer Cell  
Article

## Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype

Judith E. Grolleman,<sup>1,36</sup> Richarda M. de Voer,<sup>1,36,38,\*</sup> Fadwa A. Elsayed,<sup>2,36</sup> Maartje Nielsen,<sup>3,36</sup> Robbert D.A. Weren,<sup>1,36</sup> Claire Palles,<sup>4</sup> Marjolijn J.L. Ligtenberg,<sup>1,5</sup> Janet R. Vos,<sup>6</sup> Sanne W. ten Broeke,<sup>3</sup> Noel F.C.C. de Miranda,<sup>2</sup> Renske A. Kuiper,<sup>1</sup> Eveline J. Kamping,<sup>1</sup> Erik A.M. Jansen,<sup>1</sup> M. Elisa Vink-Börger,<sup>9</sup> Isabell Popp,<sup>7</sup> Alois Lang,<sup>8</sup> Isabel Spier,<sup>9,10</sup> Robert Hüneburg,<sup>10,11</sup> Paul A. James,<sup>12</sup> Na Li,<sup>13,14</sup> Marija Staninova,<sup>15</sup> Helen Lindsay,<sup>16</sup>

nature  
medicine



## HRDetect is a predictor of *BRCA1* and *BRCA2* deficiency based on mutational signatures

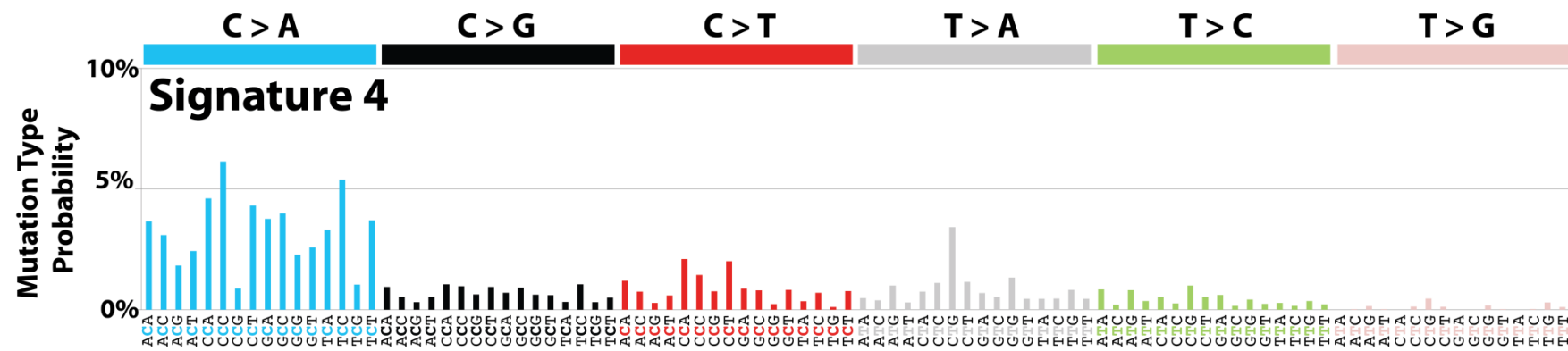
Helen Davies<sup>1,32</sup>, Dominik Glodzik<sup>1,32</sup>, Sandro Morganella<sup>1</sup>, Lucy R Yates<sup>1,2</sup>, Johan Staaf<sup>3</sup>, Xueqing Zou<sup>1</sup>, Manasa Ramakrishna<sup>1,4</sup>, Sancha Martin<sup>1</sup>, Sandrine Boyault<sup>5</sup>, Anieta M Sieuwerts<sup>6</sup>, Peter T Simpson<sup>7</sup>, Tari A King<sup>8</sup>, Keiran Raine<sup>1</sup>, Jorunn E Eyfjord<sup>9</sup>, Gu Kong<sup>10</sup>, Åke Borg<sup>3</sup>, Ewan Birney<sup>11</sup>, Hendrik G Stunnenberg<sup>12</sup>, Marc J van de Vijver<sup>13</sup>, Anne-Lise Børresen-Dale<sup>14,15</sup>, John W M Martens<sup>6</sup>, Paul N Span<sup>16,17</sup>, Sunil R Lakhani<sup>7,18</sup>, Anne Vincent-Salomon<sup>19,20</sup>, Christos Sotiriou<sup>21</sup>, Andrew Tutt<sup>22,23</sup>, Alastair M Thompson<sup>24</sup>, Steven Van Laere<sup>25,26</sup>, Andrea L Richardson<sup>27,28</sup>, Alain Viari<sup>29,30</sup>, Peter J Campbell<sup>1</sup>, Michael R Stratton<sup>1</sup> & Serena Nik-Zainal<sup>1,31</sup>

# Summary

- Different mutational processes generate somatic mutations, including endogenous and exogenous sources
- The pattern of mutations imprinted by a particular mutational process is known as mutational signature
- Reference mutational signatures have been identified and deposited in COSMIC after the analysis of thousands of cancer samples
- Leveraging these reference signatures, the contributions of the different mutational processes to a given tumor can be quantified
- Mutational signatures can be used clinically as biomarkers for cancer prevention, prognosis and treatment

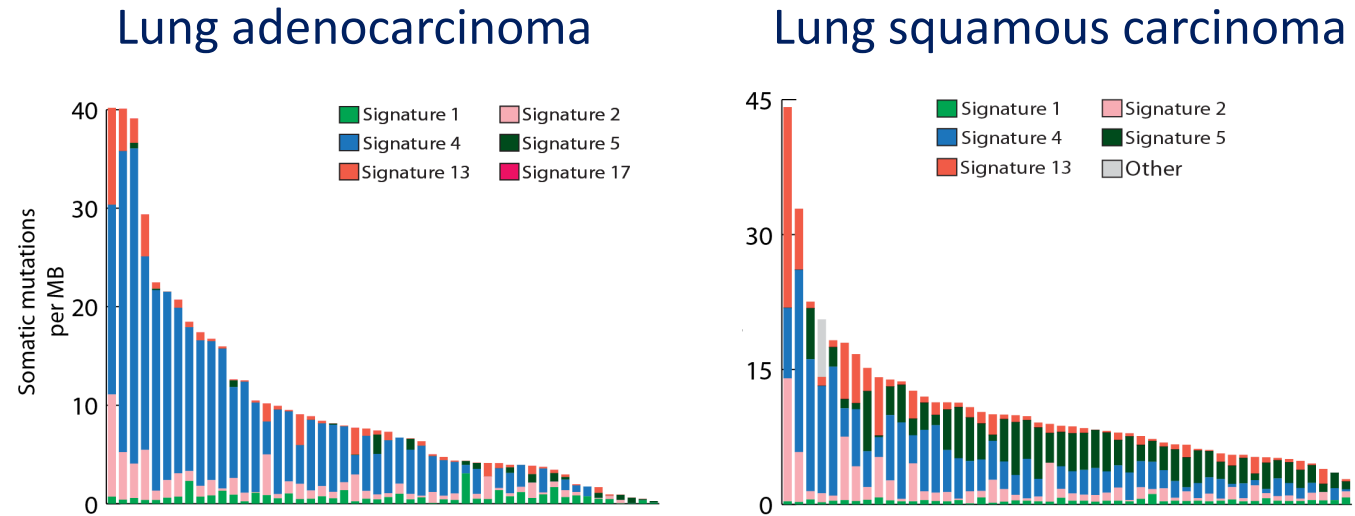
How do you know mutational signature etiologies?

Signature SBS4 is likely due to tobacco smoking



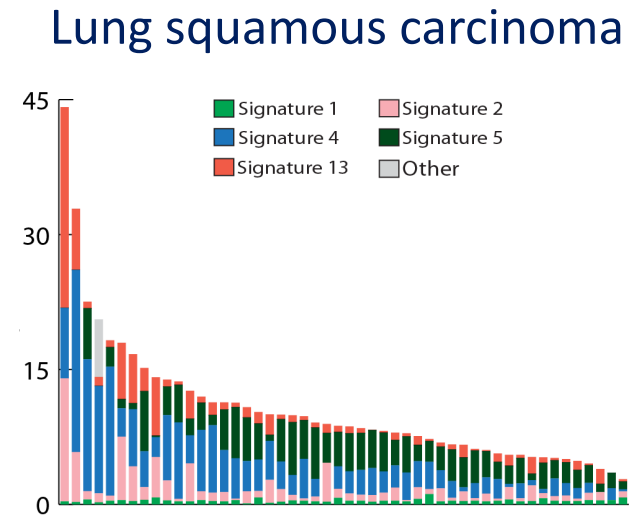
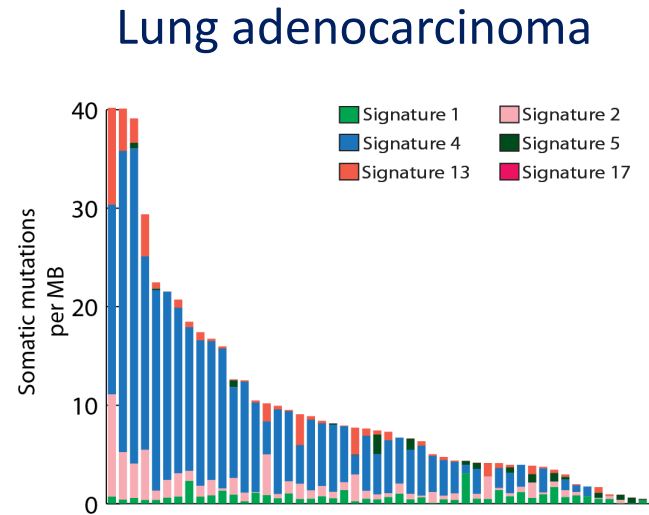
# Contributions of mutational signatures to smoking induced and non-smoking induced cancer types

Smoking induced cancer types

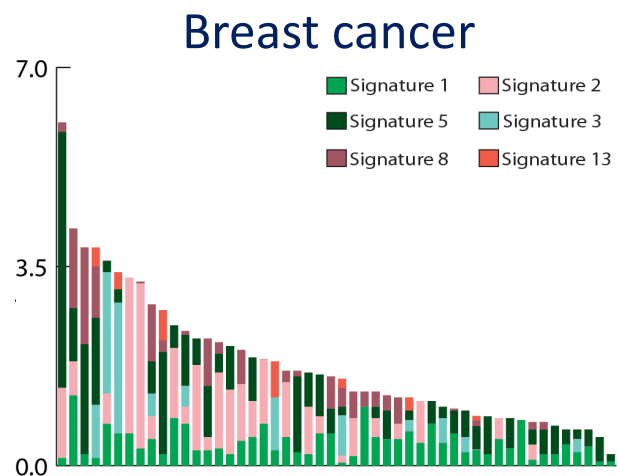
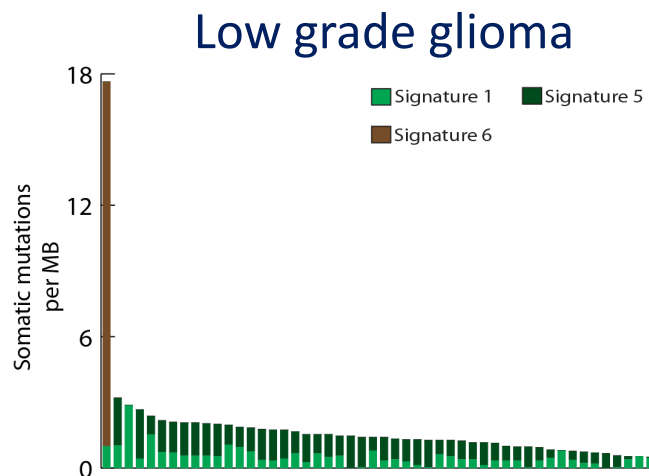


# Contributions of mutational signatures to smoking induced and non-smoking induced cancer types

Smoking induced cancer types



Non-smoking induced cancer types

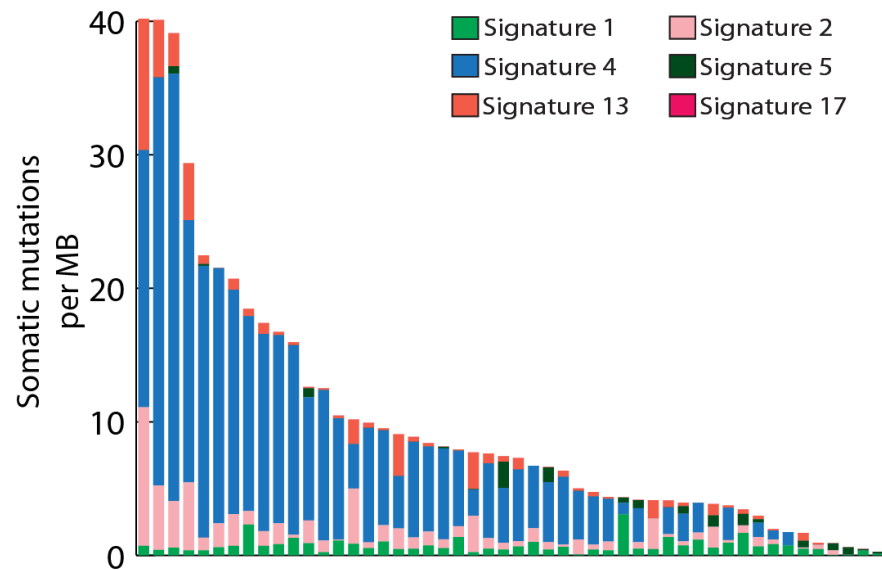




# Contributions of mutational signatures to lung adenocarcinomas

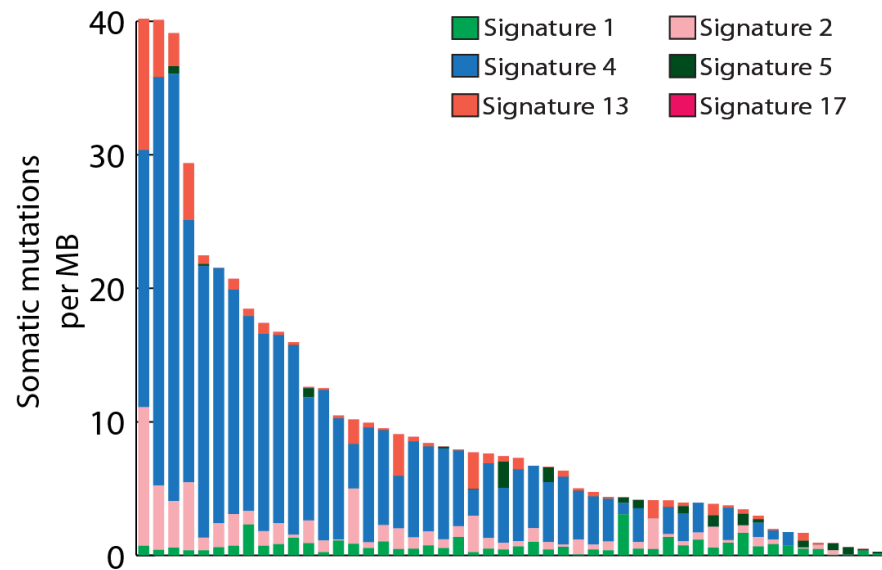
# Contributions of mutational signatures to lung adenocarcinomas

## Tobacco smokers

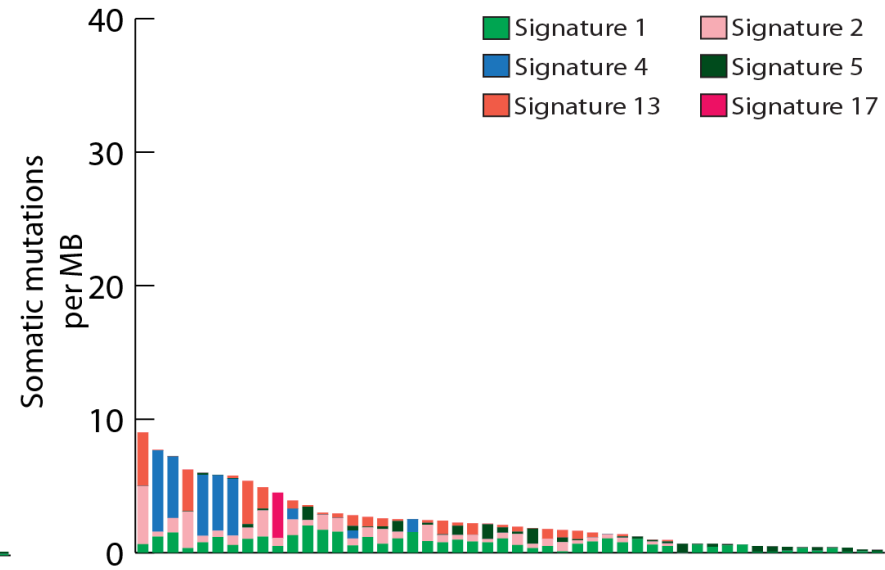


# Contributions of mutational signatures to lung adenocarcinomas

## Tobacco smokers

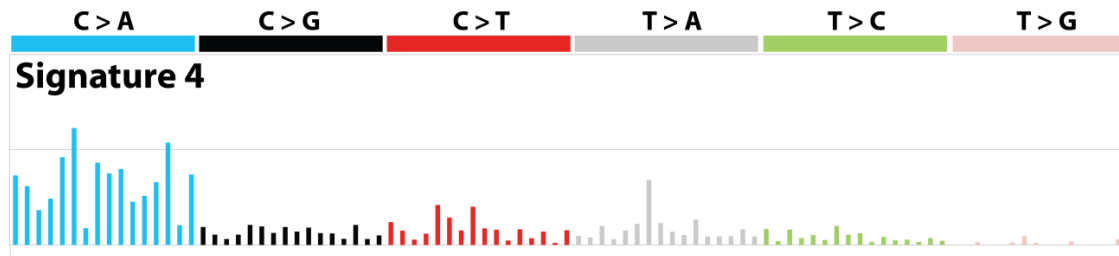


## Life-long non-smokers



The mutational signature of in vitro benzo[a]pyrene exposure is similar to signature 4

Signature 4 extracted from human cancers





# Evidence for the aetiology of signature 4

# Evidence for the aetiology of signature 4

Identified only in cancer types epidemiologically known to be caused by tobacco smoking

# Evidence for the aetiology of signature 4

Identified only in cancer types epidemiologically known to be caused by tobacco smoking

Highly enriched in tobacco smokers when compared to tobacco non-smokers



# Evidence for the aetiology of signature 4

Identified only in cancer types epidemiologically known to be caused by tobacco smoking

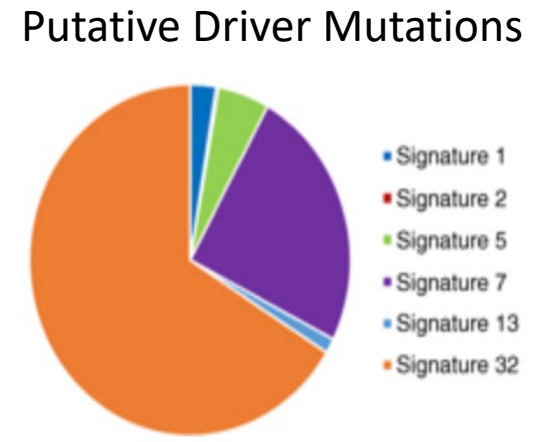
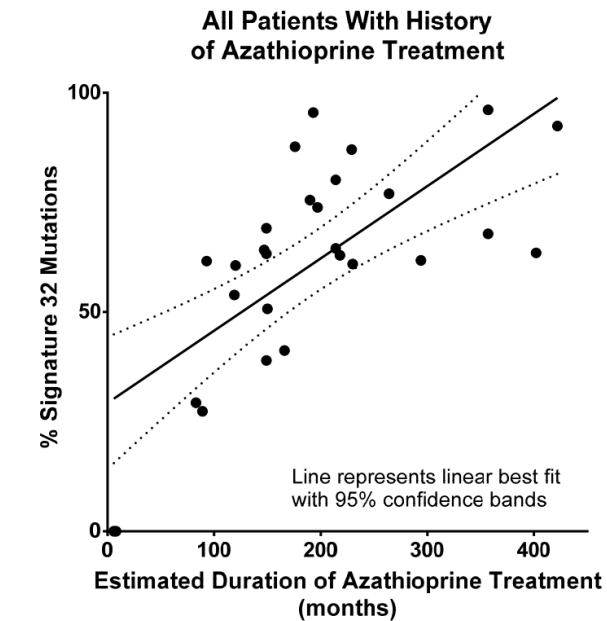
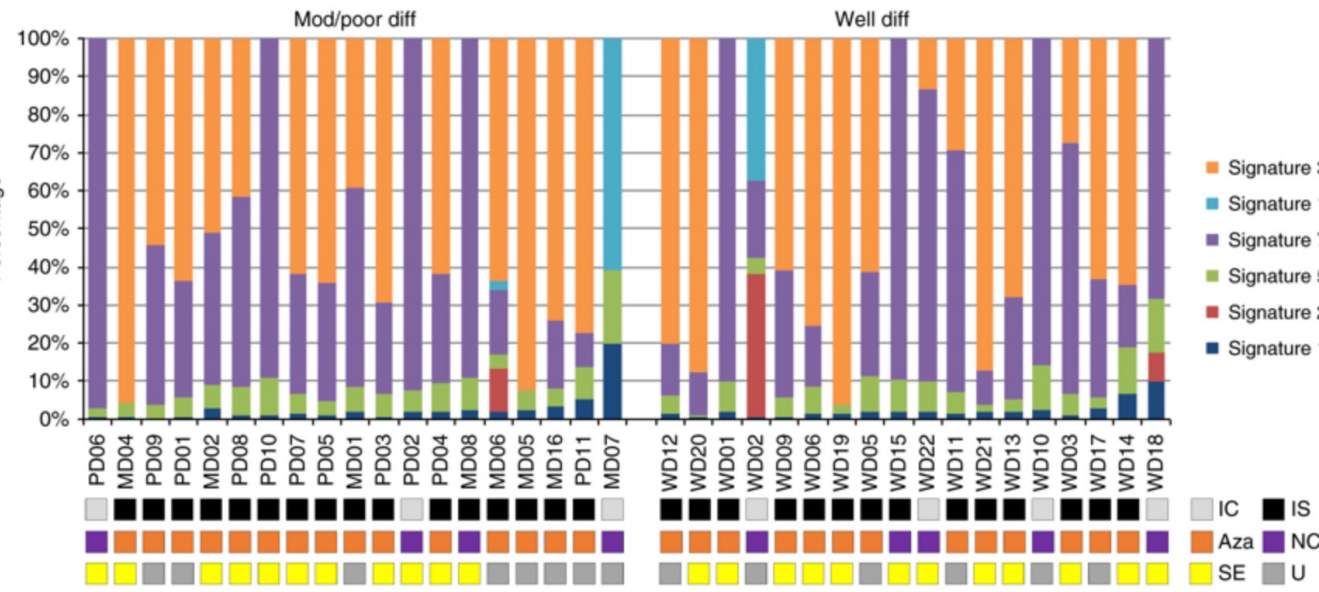
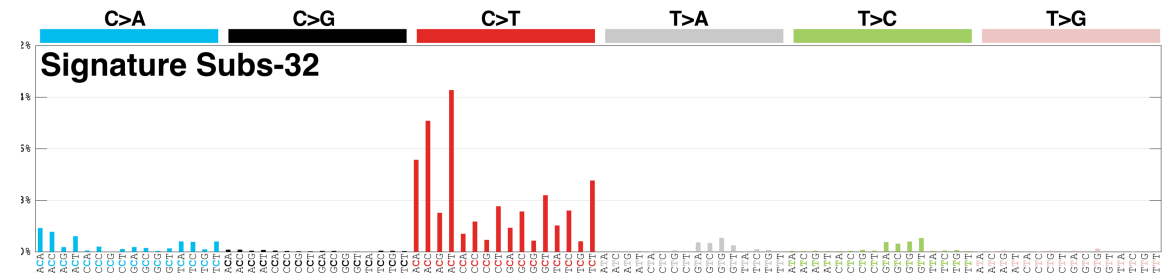
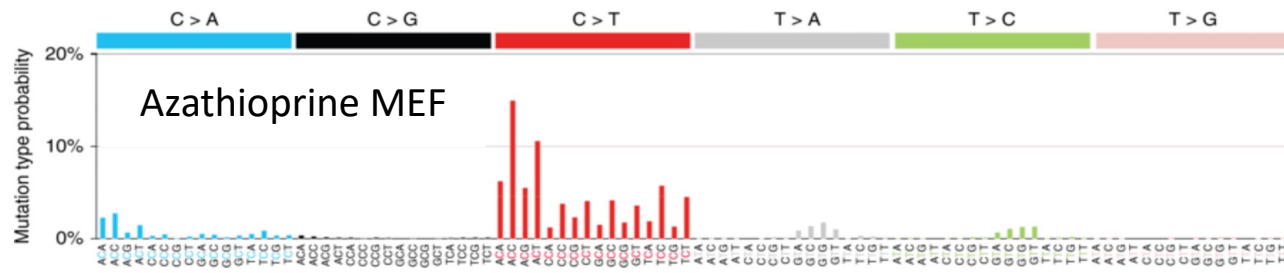
Highly enriched in tobacco smokers when compared to tobacco non-smokers

The pattern of signature 4 matches *in vitro* experimental results in which cells were exposed to known tobacco carcinogens

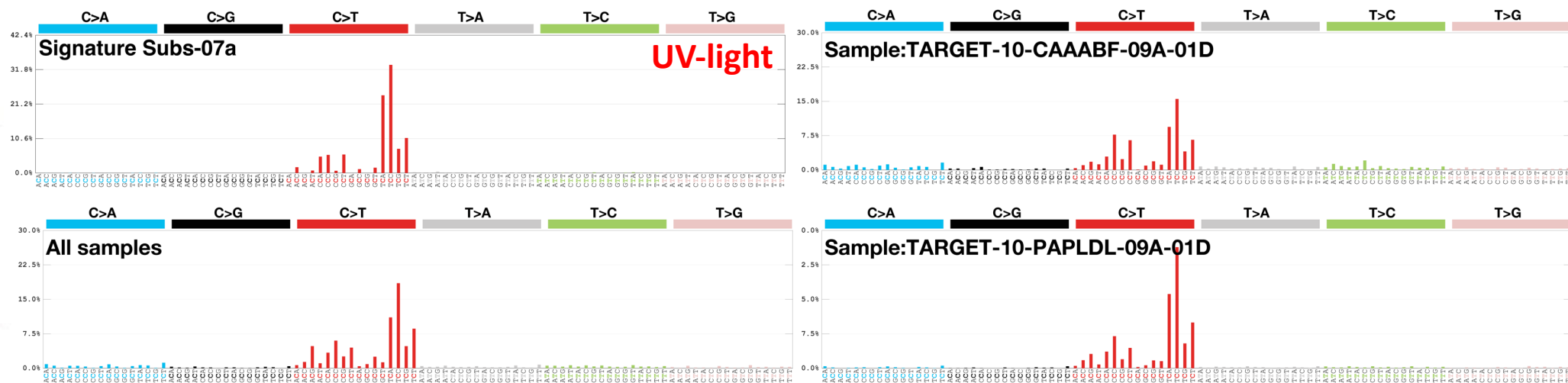
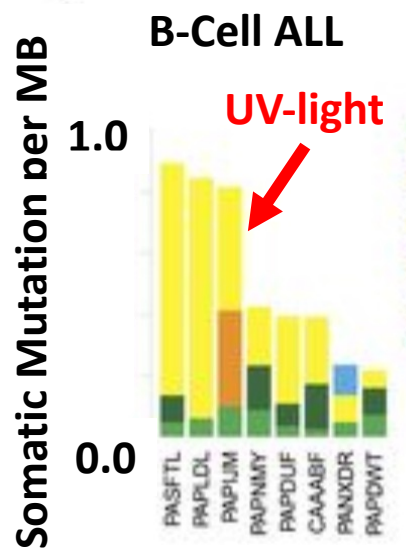
## Clinical examples

# (Somewhat) unexpected carcinogens: Azathioprine

Azathioprine, sold under the brand name Imuran among others, is an immunosuppressive medication. Azathioprine is on the World Health Organization's List of Essential Medicines, the most effective and safe medicines needed in a health system. **Epidemiological studies by International Agency for Research on Cancer have provided "sufficient" evidence of azathioprine carcinogenicity in humans (Group 1), although the methodology of past studies and the possible underlying mechanisms are questioned.**



# Known carcinogen in unexpected cancer types: **UV-light**



**Similarity extends to strand bias, dinucleotide, and indel patterns. Confirmed in three other cohorts. Signature found only in white Caucasian children. Much lower mutation burden compared to skin cancer.**




[Cancer Causes & Control](#)

October 2017, Volume 28, [Issue 10](#), pp 1075–1083 | [Cite as](#)

Residential exposure to ultraviolet light and risk of precursor B-cell acute lymphoblastic leukemia: assessing the role of individual risk factors, the ESCALE and ESTELLE studies

Authors

[Authors and affiliations](#)

Astrid Coste , Denis Hémon, Laurent Orsi, Mathieu Boniol, Jean-François Doré, Laure Faure, Jacqueline Clavel,

Stéphanie Goujon

**UV-light high confidence cancer types:**

- Basal Cell Carcinoma
- Squamous cell carcinoma
- Cutaneous melanoma (NOT in uveal melanoma)
- Lip cancer (H&N)
- B-cell ALL (childhood)
- Sarcomas (adulthood)
- Squamous cell lung carcinoma (all melanoma metastasis)







UC San Diego  
SCHOOL OF MEDICINE