

# Day 4. Cancer Genome Analysis - Latin America and the Caribbean

## Assignment of reference mutational signatures



November 30<sup>th</sup>, 2023


Marcos Díaz Gay

Alexandrov lab, University of California San Diego

UC San Diego



# Reference mutational signatures have been extracted from thousands of samples



Catalogue Of Somatic Mutations In Cancer

Projects ▾ Data ▾ Tools ▾ News ▾ Help ▾ About ▾

## 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.



### 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.

### Data downloads

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

### 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:

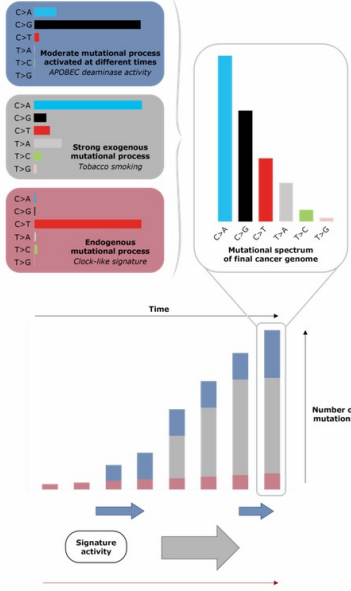
### SigProfiler tools

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.

### 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.



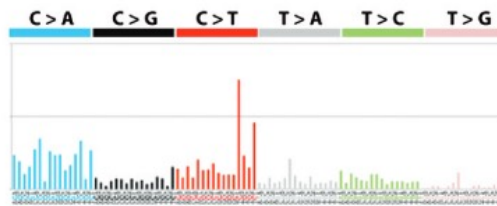


Current set (v3.3)

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

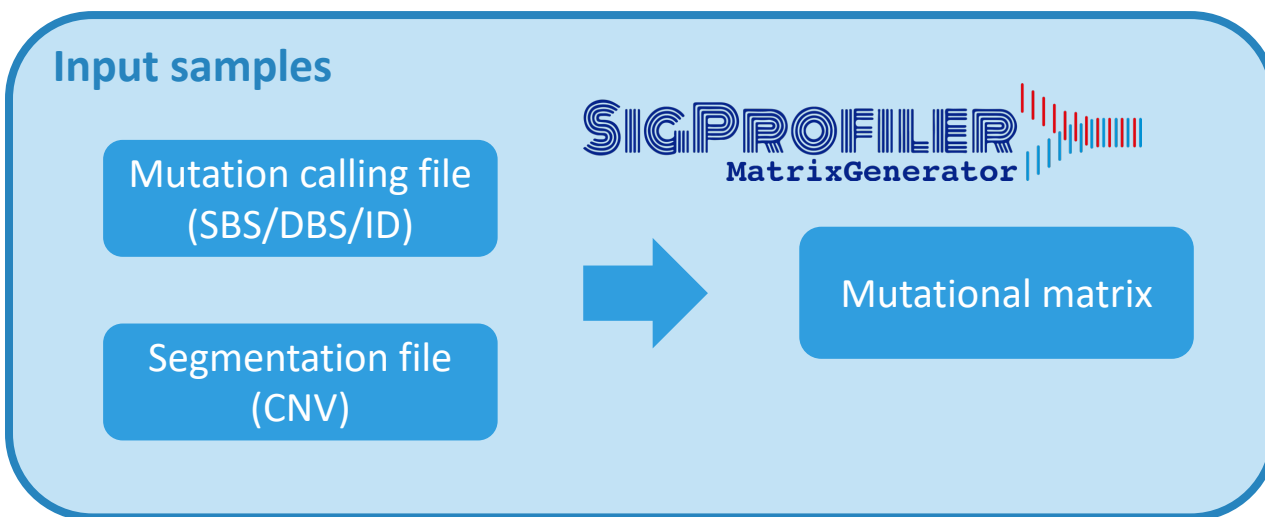
<https://cancer.sanger.ac.uk/signatures/>

# Reference mutational signatures allow refitting analysis

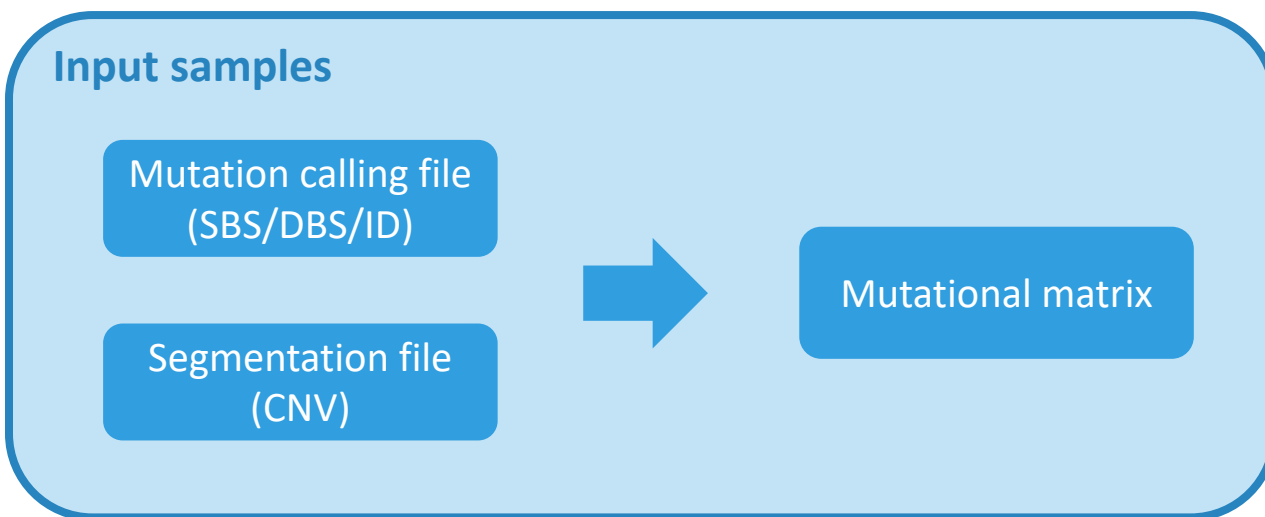


Single sample – clinical setting

# Overview of SigProfilerAssignment



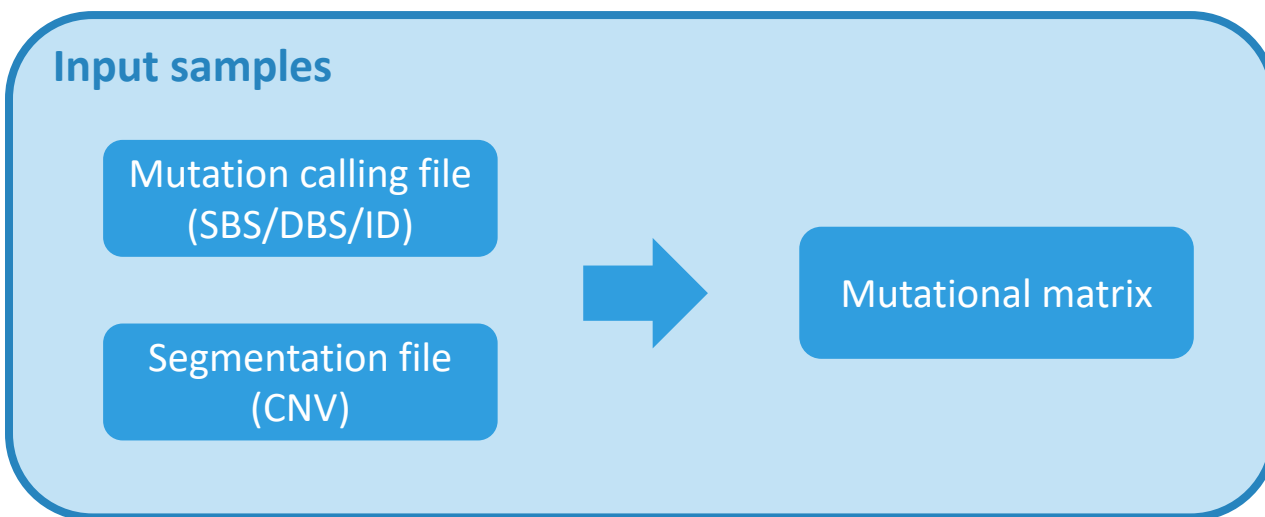
# Overview of SigProfilerAssignment



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



# Overview of SigProfilerAssignment



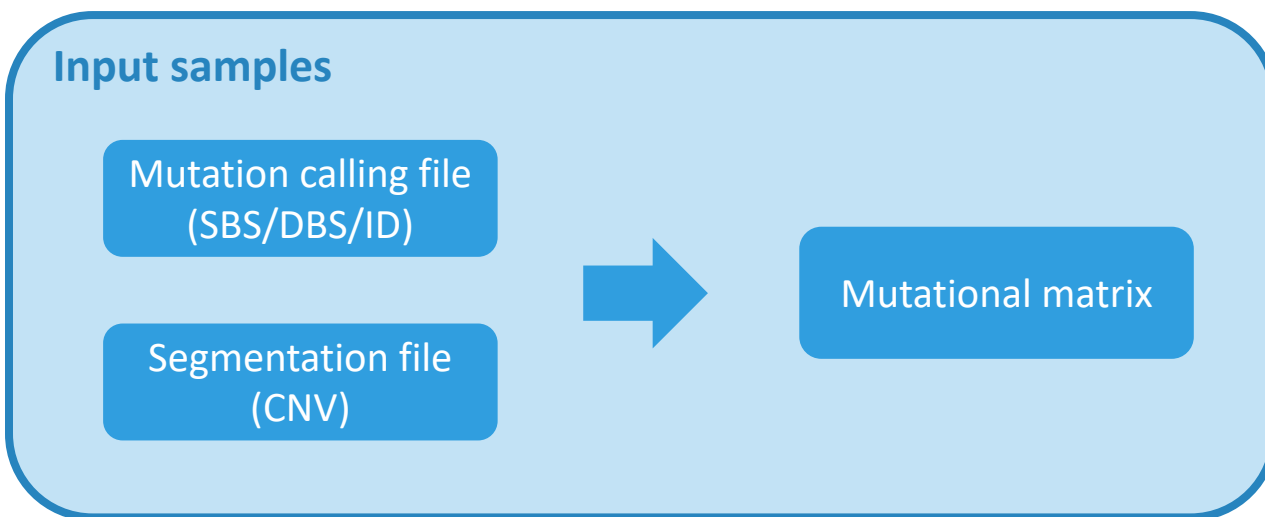
$$\mathbf{M} = \mathbf{S} \times \mathbf{A}$$

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

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



# Overview of SigProfilerAssignment



$$\underset{t \times n}{M} = \underset{t \times k}{S} \times \underset{k \times n}{A}$$

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



# Overview of SigProfilerAssignment



## Input samples

Mutation calling file  
(SBS/DBS/ID)

Segmentation file  
(CNV)



Mutational matrix

$$\underset{t \times n}{\mathbf{M}} = \underset{t \times k}{\mathbf{S}} \times \boxed{\underset{k \times n}{\mathbf{A}}}$$

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

## Reference mutational signatures



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

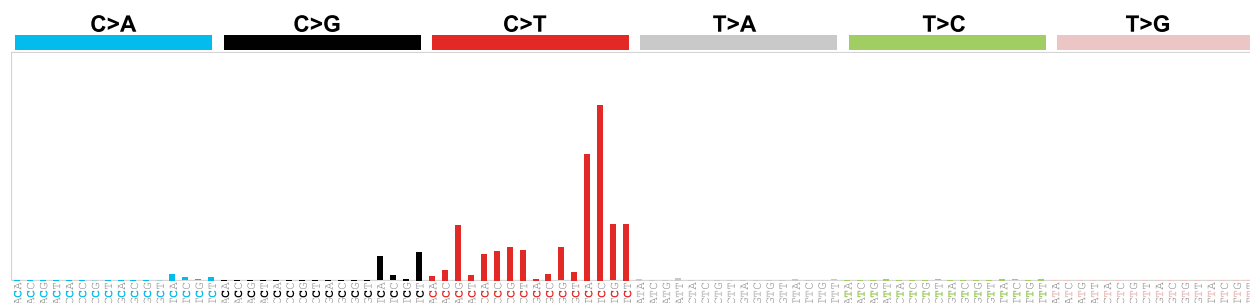
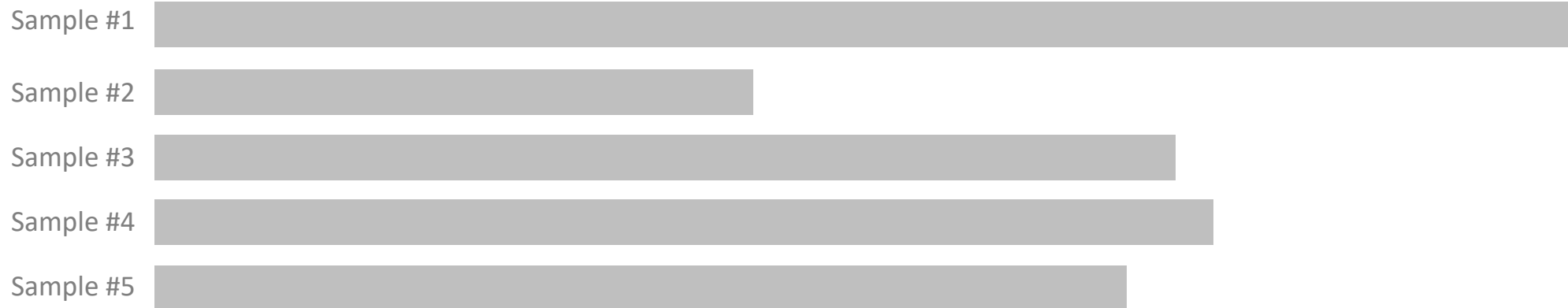
GRCh37/38, mm9/10, rn6

Genome / Exome



# Assigning reference signatures to individual samples

## M



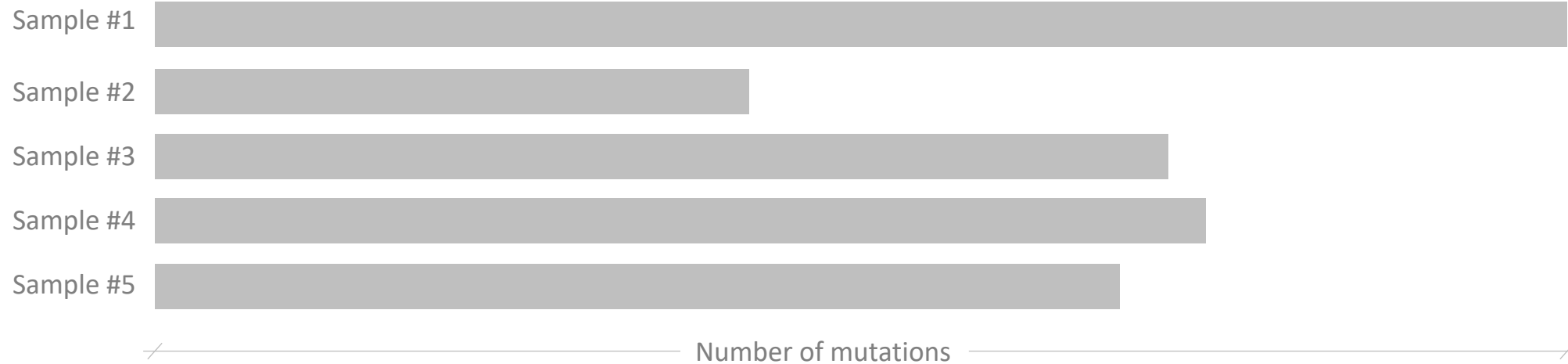
Mutational profile

#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

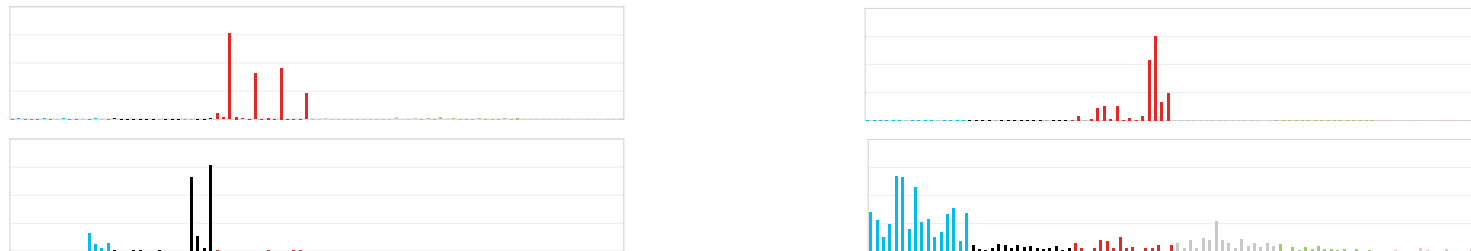
Mutation calling file

# Assigning reference signatures to individual samples

**M**



**S**

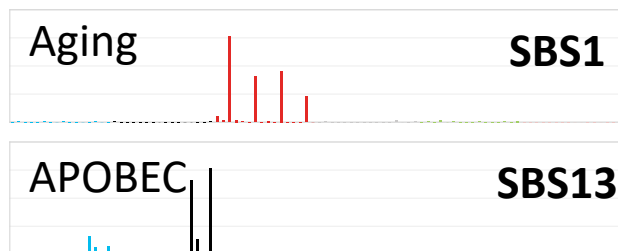


# Assigning reference signatures to individual samples

## M



## S

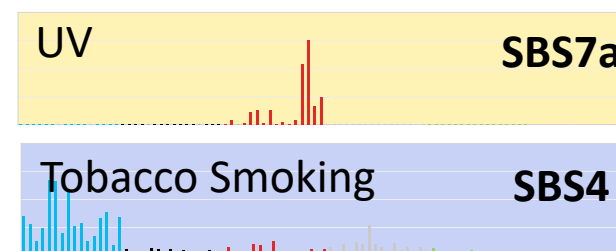
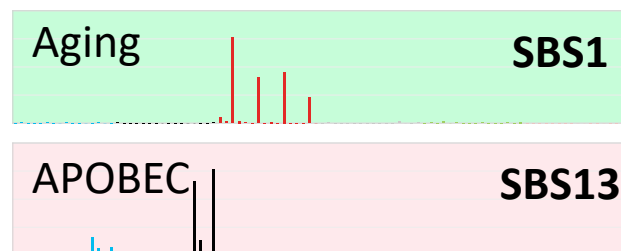


# Assigning reference signatures to individual samples

## M

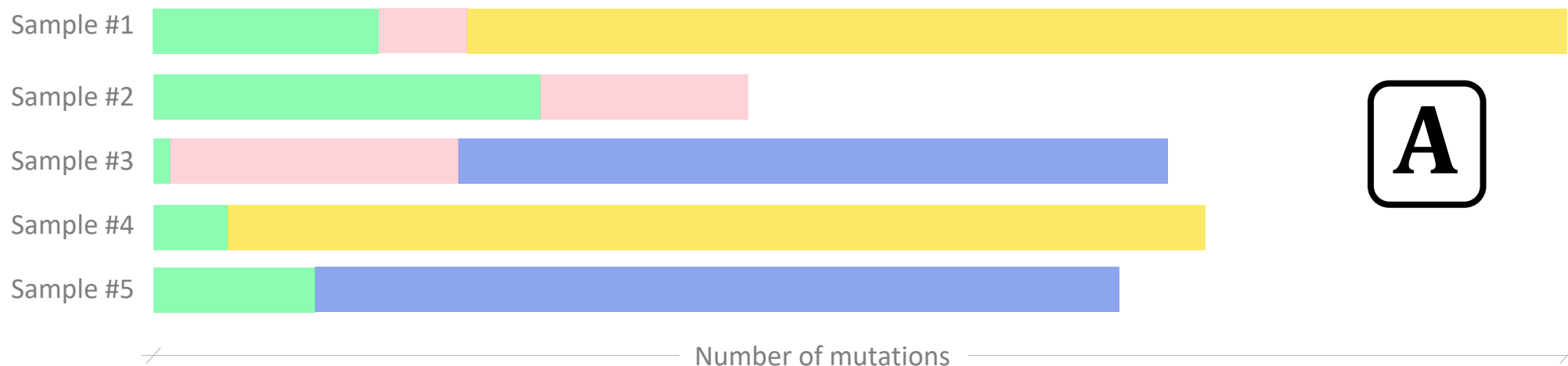


## S



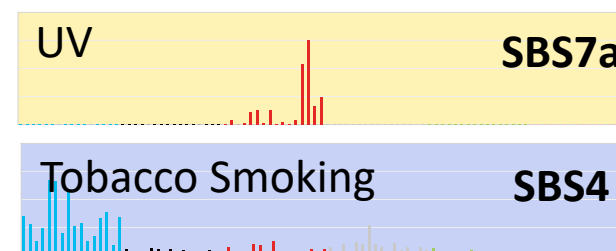
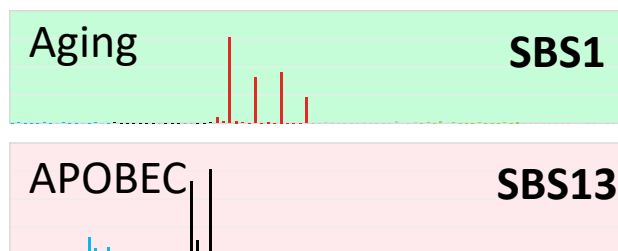
# Assigning reference signatures to individual samples

**M**

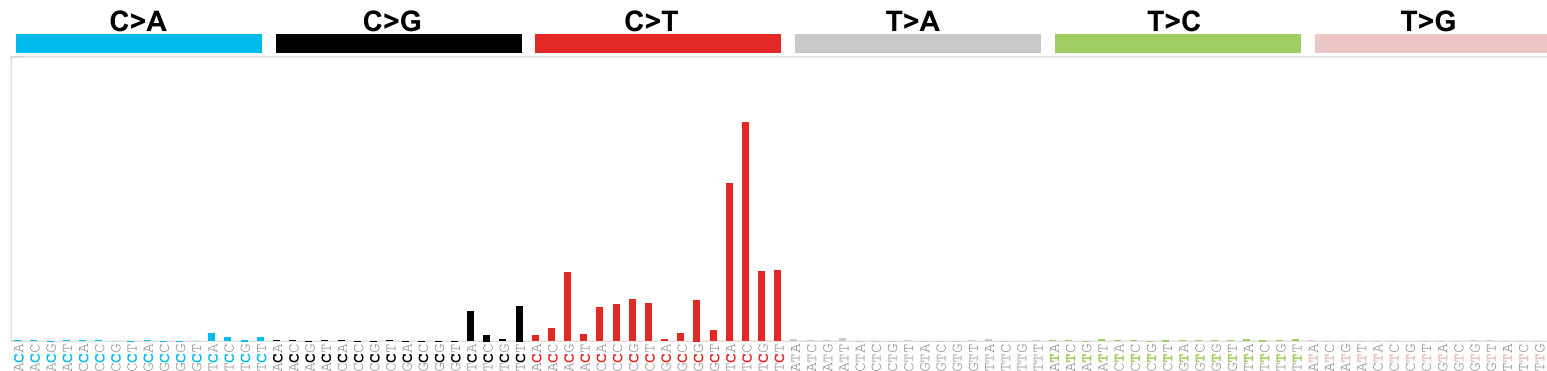
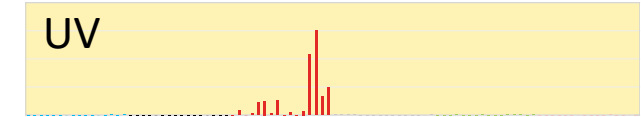
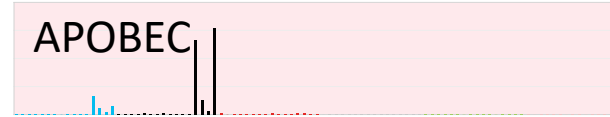
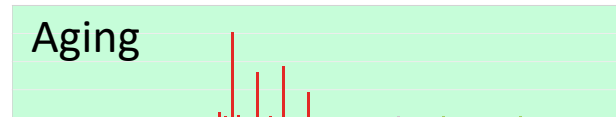


**A**

**S**

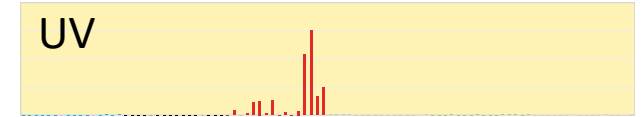
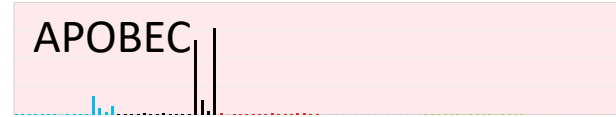
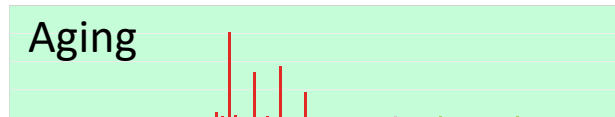


# Assigning reference signatures to individual mutations



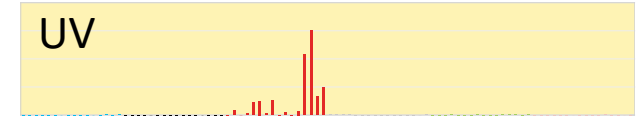
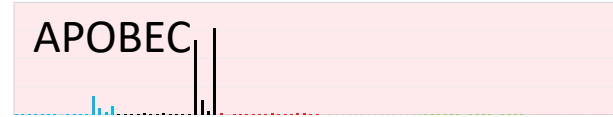
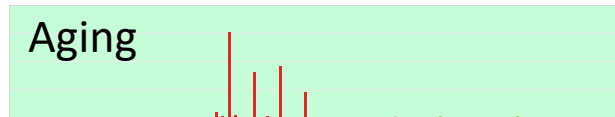
CHROM	POS	REF	ALT	Mut. Context
1	221436661	G	A	A[C>T]G
3	178936091	G	A	T[C>T]A
4	119785075	C	G	T[C>G]T
6	162294115	C	T	T[C>T]G
...	...	...	...	...

# Assigning reference signatures to individual mutations



$$P(\textit{Signature} \mid \textit{Mut.context}) =$$

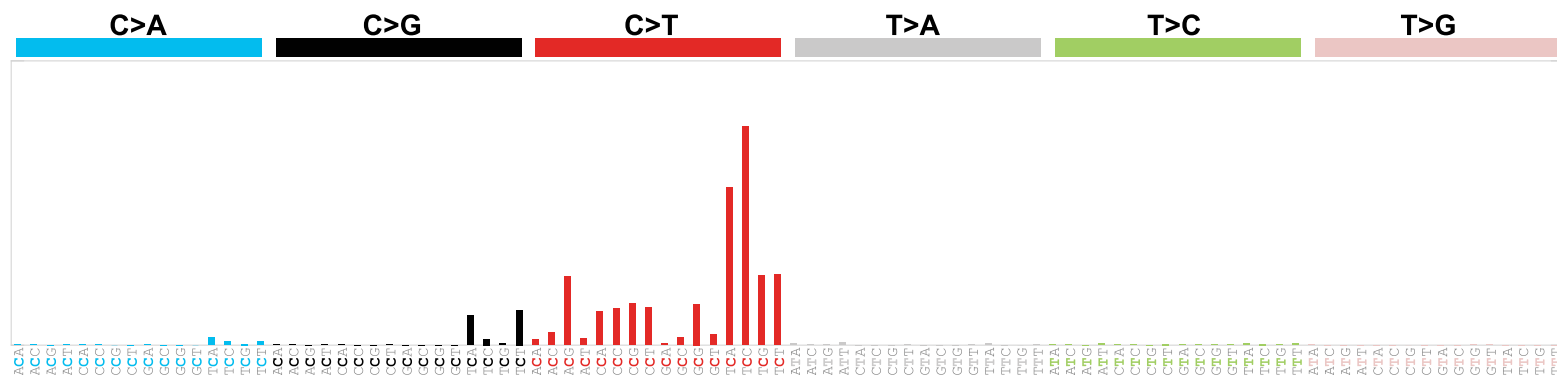
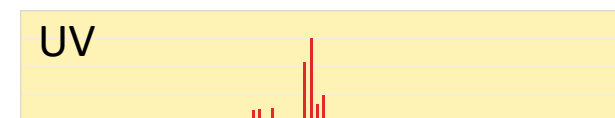
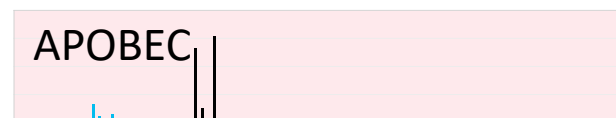
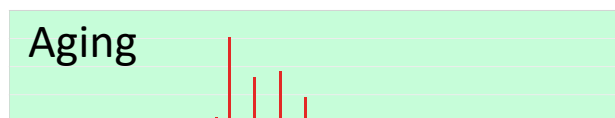
# Assigning reference signatures to individual mutations



$$P(\text{Signature} \mid \text{Mut.context}) = \frac{\overbrace{P(\text{Mut.context} \mid \text{Signature})}^{\text{S}} \times \overbrace{P(\text{Signature})}^{\text{A}}}{\underbrace{P(\text{Mut.context})}_{\text{M}}}$$

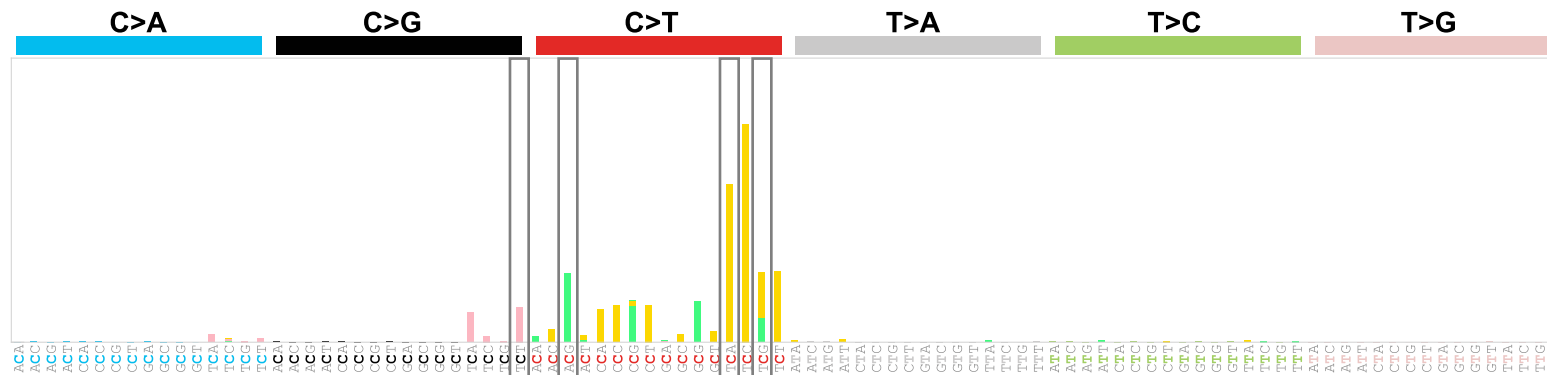
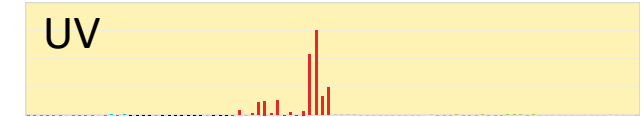
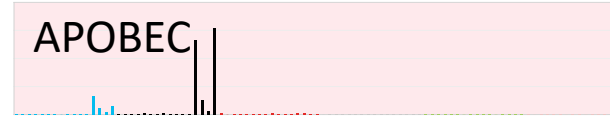
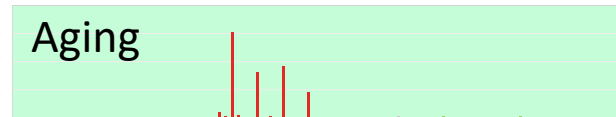


## Assigning reference signatures to individual mutations



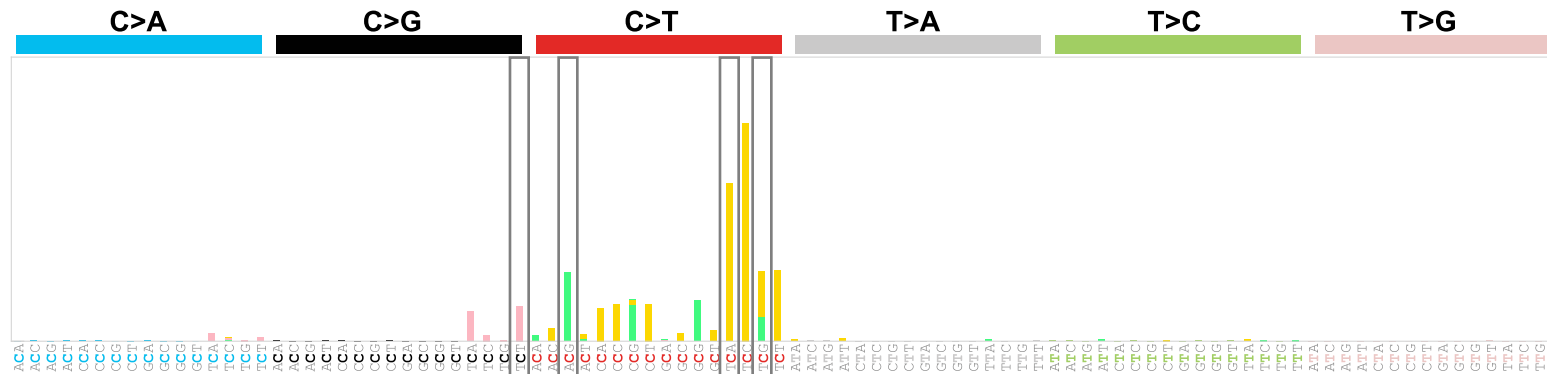
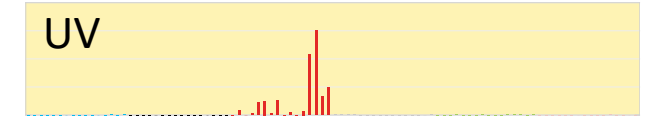
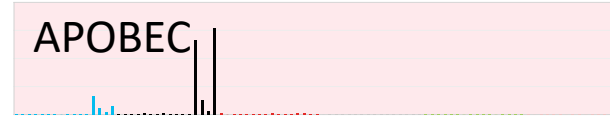
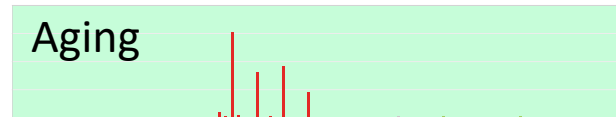
CHROM	POS	REF	ALT	Mut. Context
1	221436661	G	A	A[C>T]G
3	178936091	G	A	T[C>T]A
4	119785075	C	G	T[C>G]T
6	162294115	C	T	T[C>T]G
...	...	...	...	...

# Assigning reference signatures to individual mutations



CHROM	POS	REF	ALT	Mut. Context
1	221436661	G	A	A[C>T]G
3	178936091	G	A	T[C>T]A
4	119785075	C	G	T[C>G]T
6	162294115	C	T	T[C>T]G
...	...	...	...	...

# Assigning reference signatures to individual mutations



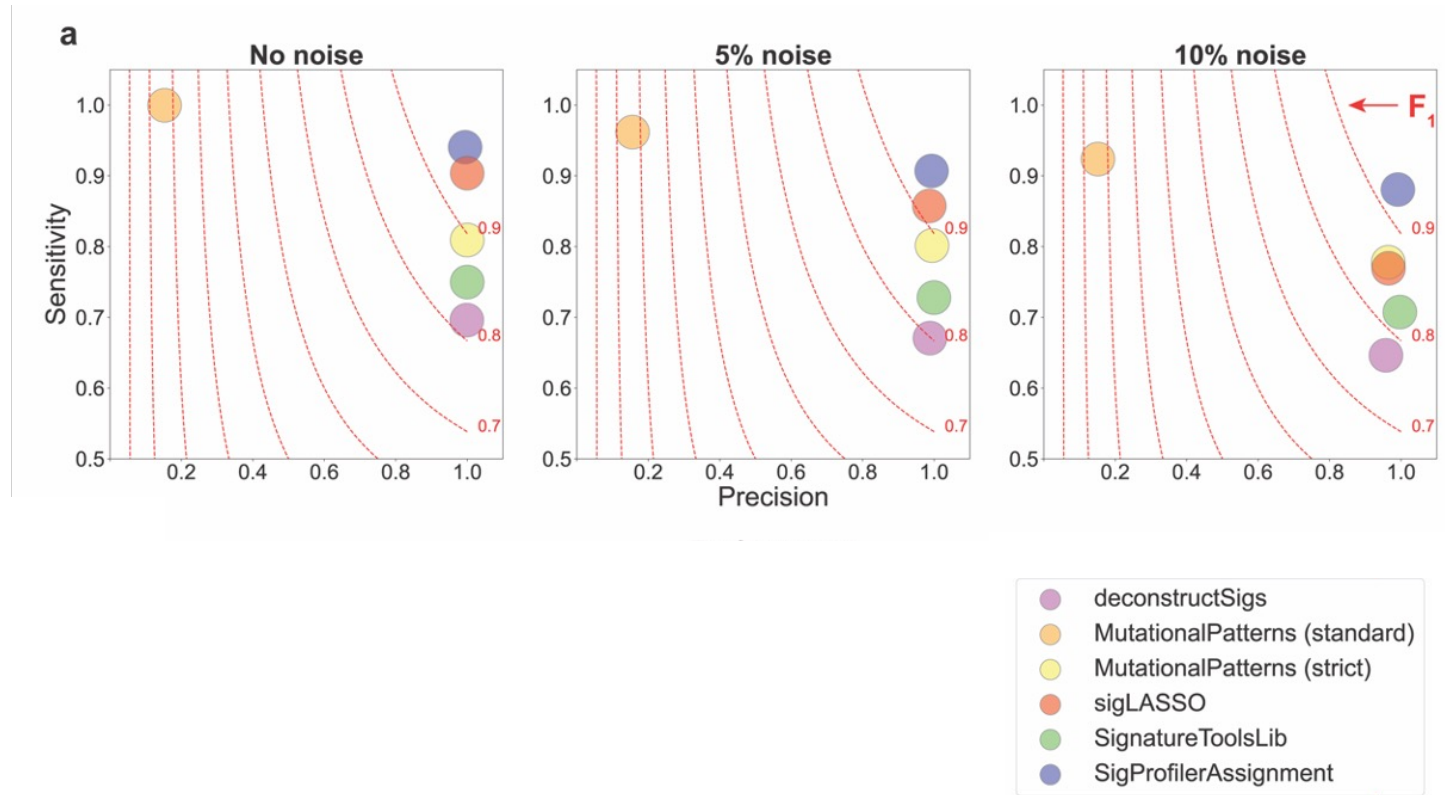
CHROM	POS	REF	ALT	Most Prob. Signature	Prob.
1	221436661	G	A	SBS1	99.9%
3	178936091	G	A	SBS7a	99.5%
4	119785075	C	G	SBS13	98.6%
6	162294115	C	T	SBS7a	70.1%
...	...	...	...	...	...

# Different tools are currently available to perform mutational signatures refitting

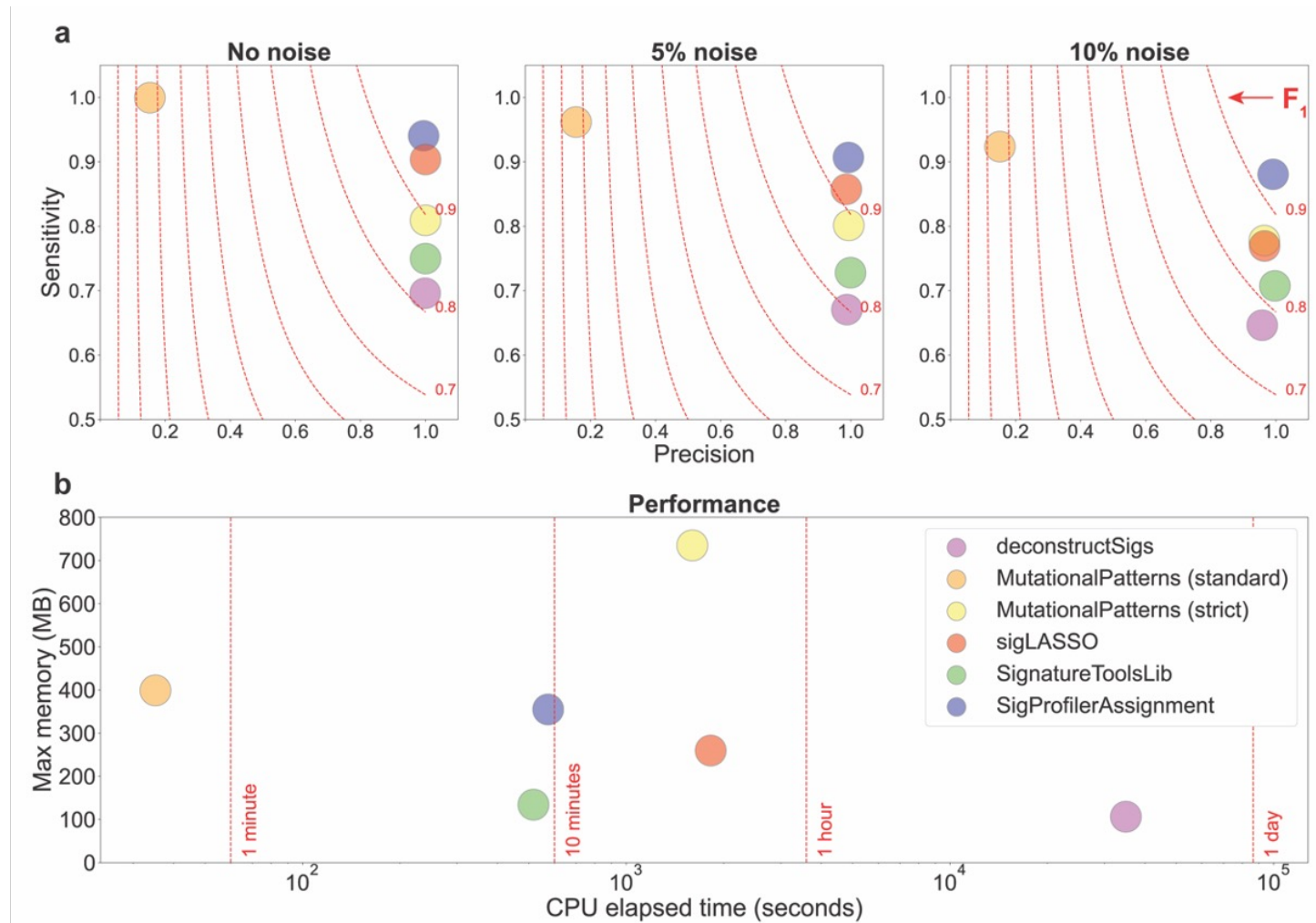
Tool	Platform	Refitting Approach		Reference
		Method	Computational Engine	
<b>deconstructSigs</b>	R	Non-negative linear regression	Original implementation	Rosenthal <i>et al.</i> 2016 Genome Biology
<b>MSA</b>	Python / Nextflow	NNLS	Original implementation / Scipy python package	Senkin 2021 BMC Bioinformatics
<b>MutationalPatterns (standard)</b>	R	NNLS	Pracma R package	Blokzijl <i>et al.</i> 2018 Genome Medicine
<b>MutationalPatterns (strict)</b>	R	NNLS	Original implementation / Pracma R package	Manders <i>et al.</i> 2022 BMC Genomics
<b>sigLASSO</b>	R	Lasso regression	Original implementation / glmnet R package	Li <i>et al.</i> 2020 Nature Communications
<b>SignatureToolsLib</b>	R / Web app	Non-negative linear regression	NNLM R package	Degasperi <i>et al.</i> 2022 Science
<b>SigProfilerAssignment</b>	Python / R / Web app	NNLS	Original implementation / Scipy python package	Díaz-Gay <i>et al.</i> 2023 bioRxiv

NNLS: non-negative least squares

SigProfilerAssignment outperformed other approaches while maintaining similar computational performance



SigProfilerAssignment outperformed other approaches while maintaining similar computational performance



# SigProfilerAssignment – ONLINE TOOL

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



[Projects](#) [Data](#) [Tools](#) [News](#) [Help](#) [About](#)  [SEARCH](#)

[Home](#) [SBS Signatures](#) [DBS Signatures](#) [ID Signatures](#) [CN Signatures](#) [Version 2](#) [Downloads](#) [Assignment](#) [SigProfiler Tools](#)

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

### SigProfiler Assignment

#### Use the SigProfilerAssignment web tool



Assign COSMIC mutational signatures to your samples with SigProfilerAssignment

[Get started >](#)

[Help and instructions](#) [Contact us](#) [How to cite](#) [View source code](#)

#### How do I use SigProfilerAssignment?

Our guided tour of SigProfilerAssignment demonstrates how to assign reference mutational signatures to individual samples.



SigProfiler Assignment Live Demo

#### About the SigProfilerAssignment tool

SigProfilerAssignment is a novel tool developed by the [Alexandrov lab](#) at UC San Diego. The tool is used to assign previously known mutational signatures to individual sequenced samples, utilising a custom implementation of the forward stagewise algorithm and nonnegative least squares.

#### How are mutational signatures assigned to individual samples?

Assigning previously known mutational signatures to individual samples, also known as *refitting* of known signatures, enables identifying the set of operative mutational signatures in a particular sample and quantifying the number of mutations attributed to each signature found in that sample.

and exogenous processes. Leveraging the vast amount of DNA sequencing data accumulated in recent years, these mutational processes have been characterised, resulting in reference signatures which are recorded in [COSMIC Mutational Signatures](#).

# Useful links – SigProfilerAssignment

- Pre-print manuscript  
<https://doi.org/10.1101/2023.07.10.548264>
- GitHub repository (python package)  
<https://github.com/AlexandrovLab/SigProfilerAssignment>
- GitHub repository (R wrapper)  
<https://github.com/AlexandrovLab/SigProfilerAssignmentR>
- Wiki page (usage instructions)  
<https://osf.io/mz79v/wiki/home/>
- COSMIC Mutational Signatures website  
<https://cancer.sanger.ac.uk/signatures/>