Day 4. Cancer Genome Analysis - Latin America and the Caribbean
Assignment of reference mutational signatures

November 30th, 2023

Marcos Díaz Gay

Alexandrov lab, University of California San Diego

UC San Diego

wellcome
connecting
science

Reference mutational signatures have been extracted from thousands of samples

Current set (v3.3)

- 79 SBS signatures

- 11 DBS signatures

- 18 ID signatures

- 21 CN signatures

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

Tate *et al.* 2019 Nucleic Acids Research | Otlu *et al.* 2023 Cell Reports

Reference mutational signatures allow refitting analysis
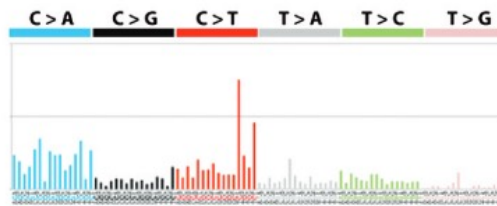
Single sample – clinical setting

Petljak & Alexandrov 2016 Carcinogenesis

Overview of SigProfilerAssignment

Input samples

Mutation calling file (SBS/DBS/ID)

Segmentation file (CNV)

SIGPROFILER MatrixGenerator

Mutational matrix

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 SigProfilerAssignment

SIGPROFILER
Assignment

Input samples

Mutation calling file
(SBS/DBS/ID)

Segmentation file
(CNV)

Mutational matrix

Reference mutational signatures

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

GRCh37/38, mm9/10, rn6

Genome / Exome

COSMIC
Catalogue Of Somatic Mutations In Cancer

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

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

Overview of SigProfilerAssignment

SIGPROFILER *Assignment*

Input samples

| Mutation calling file (SBS/DBS/ID) | → | Mutational matrix |

Segmentation file (CNV)

Reference mutational signatures

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

COSMIC Catalogue Of Somatic Mutations In Cancer

GRCh37/38, mm9/10, rn6

Genome / Exome

$$M = S \times A$$
$$t \times n \quad t \times k \quad k \times n$$

t mutational contexts
n samples
k signatures

Overview of SigProfilerAssignment

SIGPROFILER
Assignment

Input samples

Mutation calling file (SBS/DBS/ID)

Segmentation file (CNV)

→ Mutational matrix

Reference mutational signatures

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

COSMIC Catalogue Of Somatic Mutations In Cancer

GRCh37/38, mm9/10, rn6

Genome / Exome

$$M = S \times A$$

t x n   t x k   k x n

t mutational contexts
n samples
k signatures

Overview of SigProfilerAssignment

SIGPROFILER
Assignment

Input samples

Mutation calling file
(SBS/DBS/ID)

Segmentation file
(CNV)

→ Mutational matrix

Reference mutational signatures

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

COSMIC
Catalogue Of Somatic Mutations In Cancer

GRCh37/38, mm9/10, rn6

Genome / Exome

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

t x n   t x k   k x n

t mutational contexts
n samples
k signatures

Assigning reference signatures to individual samples

Mutational profile

Mutation calling file

Assigning reference signatures to individual samples
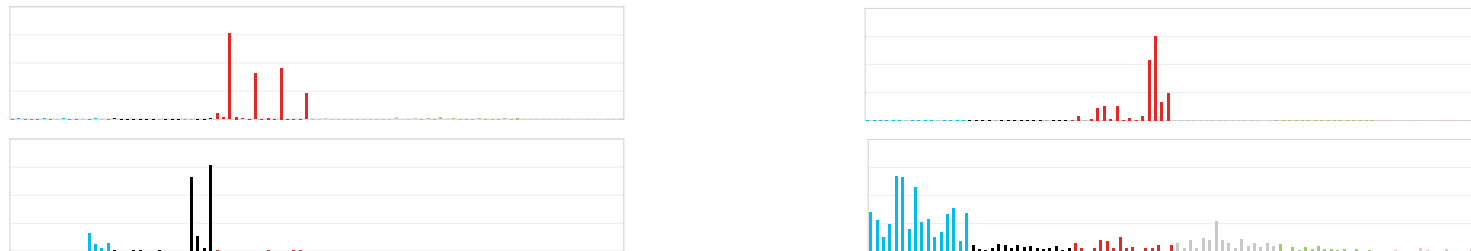
M

Sample #1

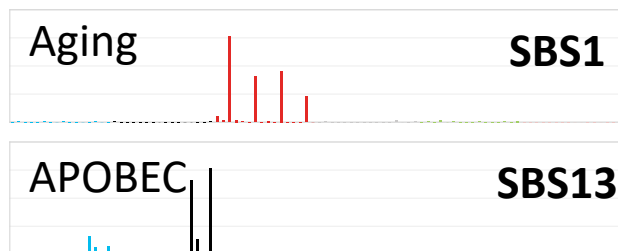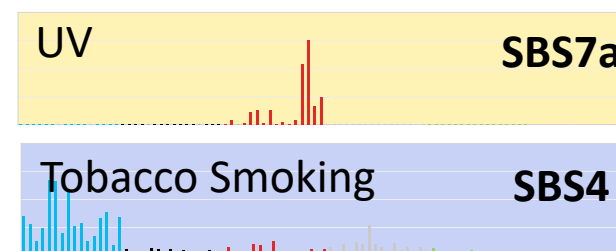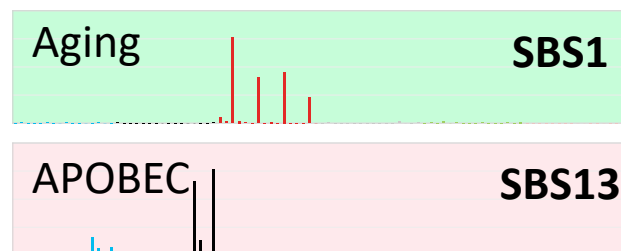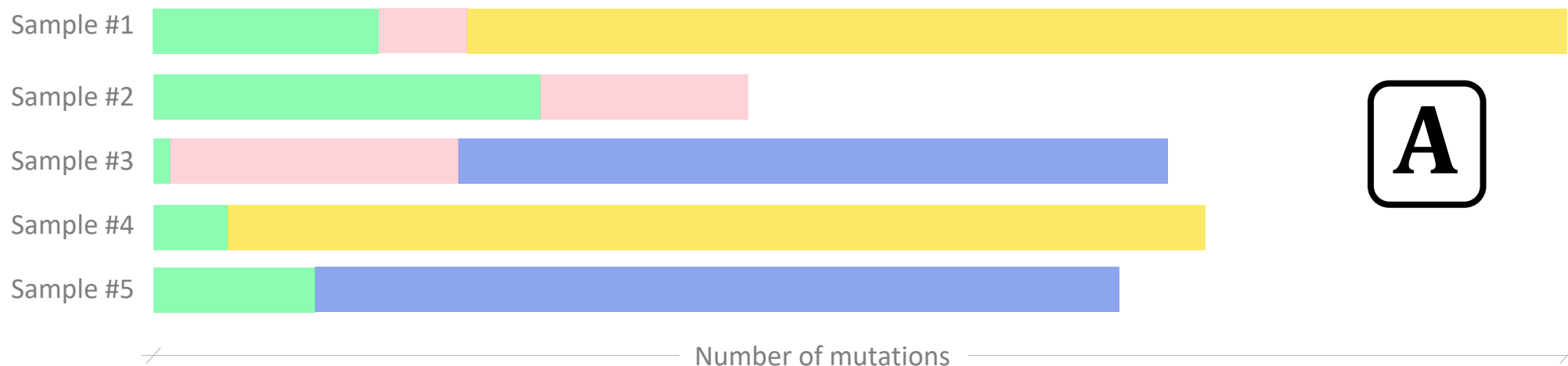Sample #2

Sample #3

Sample #4

Sample #5

Number of mutations

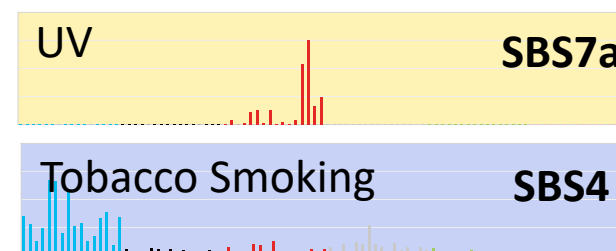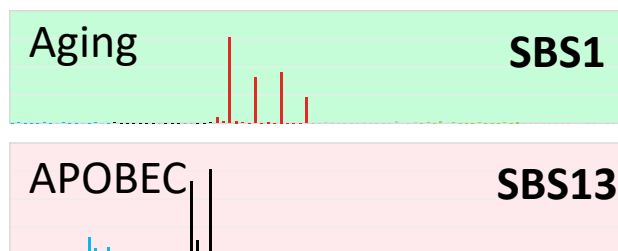**S**

Assigning reference signatures to individual samples

Assigning reference signatures to individual samples

M

Sample #1

Sample #2

Sample #3

Sample #4

Sample #5

Number of mutations

S

Aging **SBS1**

APOBEC **SBS13**

UV **SBS7a**

Tobacco Smoking **SBS4**
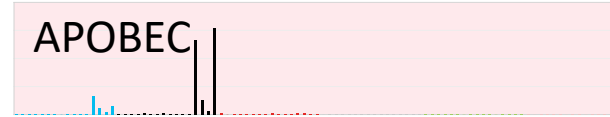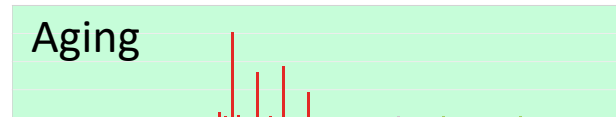
Assigning reference signatures to individual samples

M

Sample #1

Sample #2

Sample #3

Sample #4

Sample #5

Number of mutations

A

S

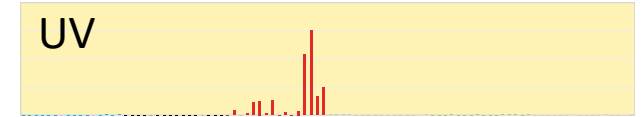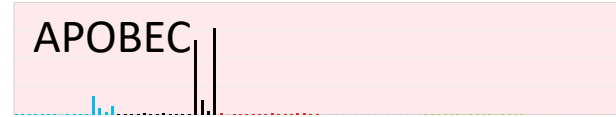| Aging | SBS1 |
| APOBEC | SBS13 |

| UV | SBS7a |
| Tobacco Smoking | SBS4 |

Assigning reference signatures to <u>individual mutations</u>

Assigning reference signatures to individual mutations

Aging

APOBEC

UV
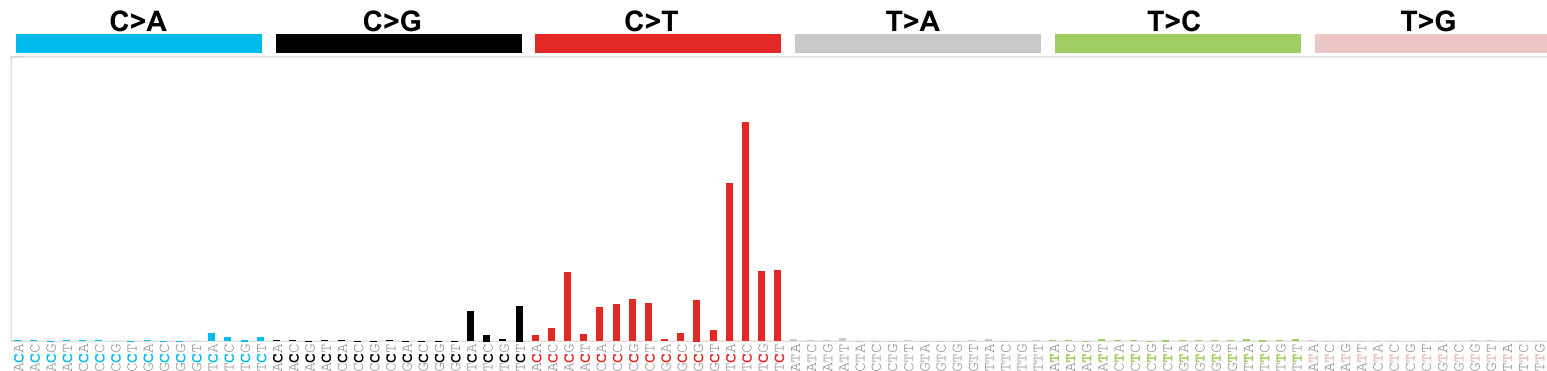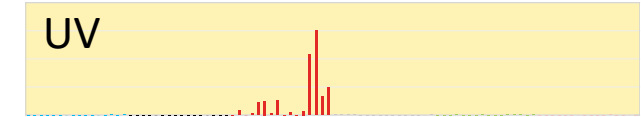
$P(Signature \mid Mut.context) =$

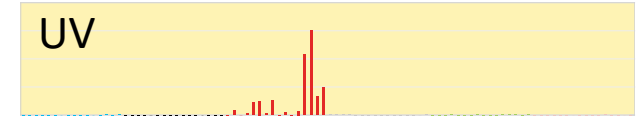# Assigning reference signatures to individual mutations

$$P(Signature \mid Mut.\,context) = \frac{\overbrace{P(Mut.\,context \mid Signature)}^{S} \times \overbrace{P(Signature)}^{A}}{\underbrace{P(Mut.\,context)}_{M}}$$
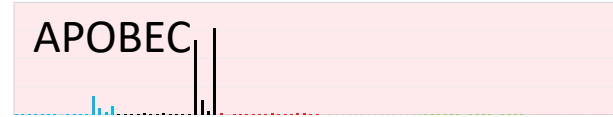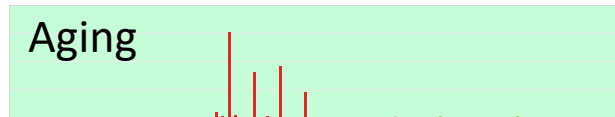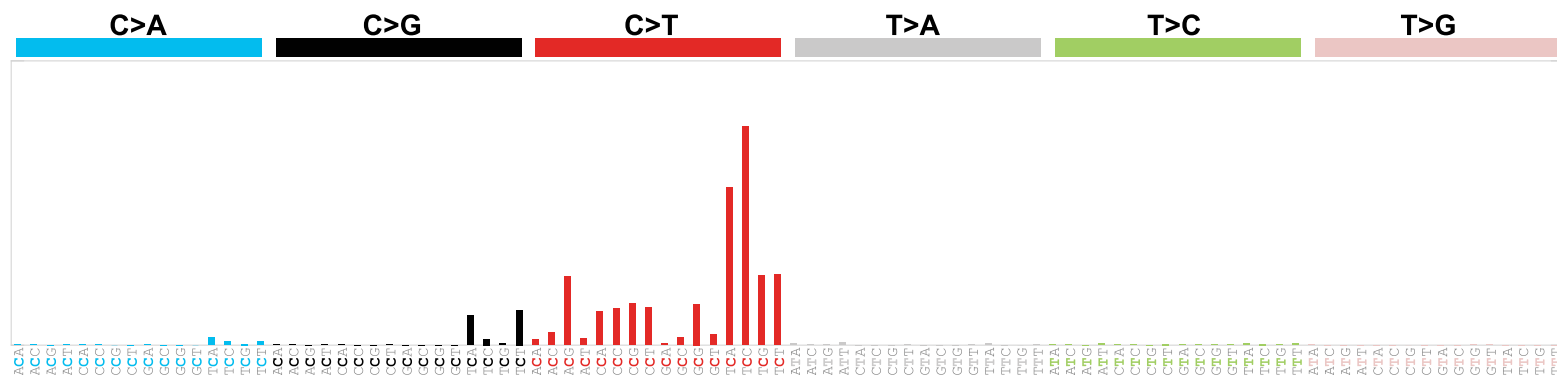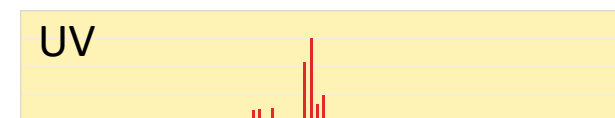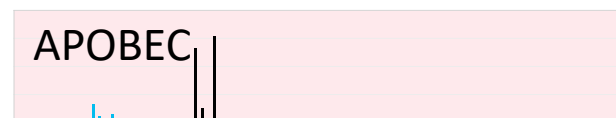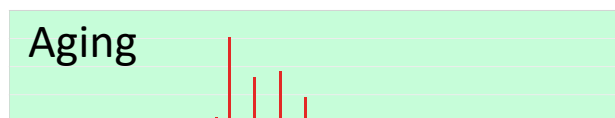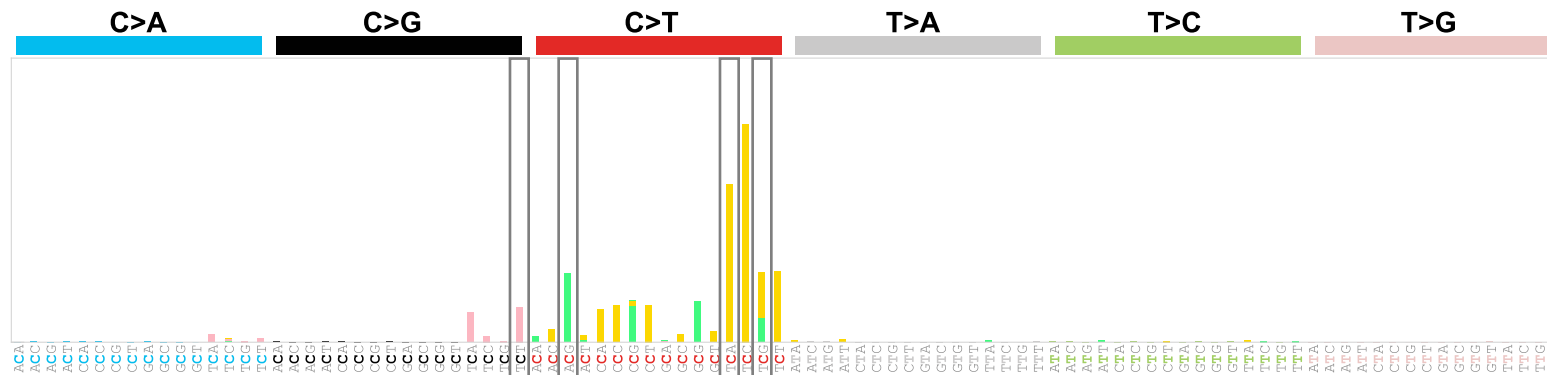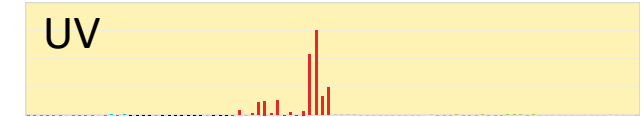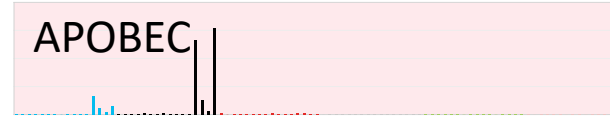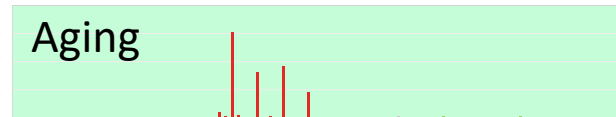
Assigning reference signatures to individual mutations

Assigning reference signatures to individual mutations

Assigning reference signatures to individual mutations

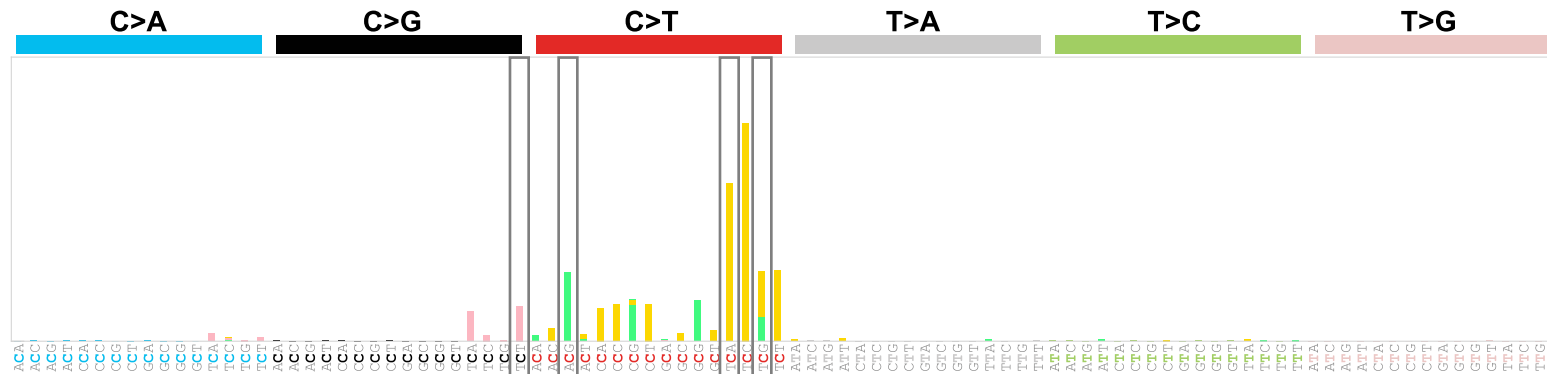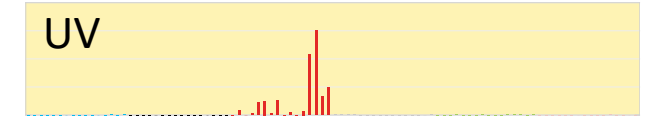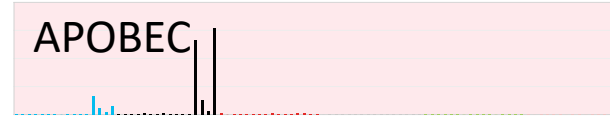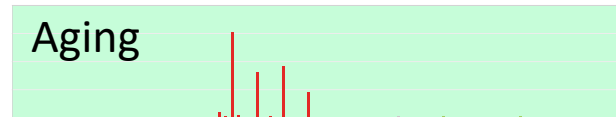# Different tools are currently available to perform mutational signatures refitting

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

NNLS: non-negative least squares

SigProfilerAssignment outperformed other approaches while maintaining similar computational performance
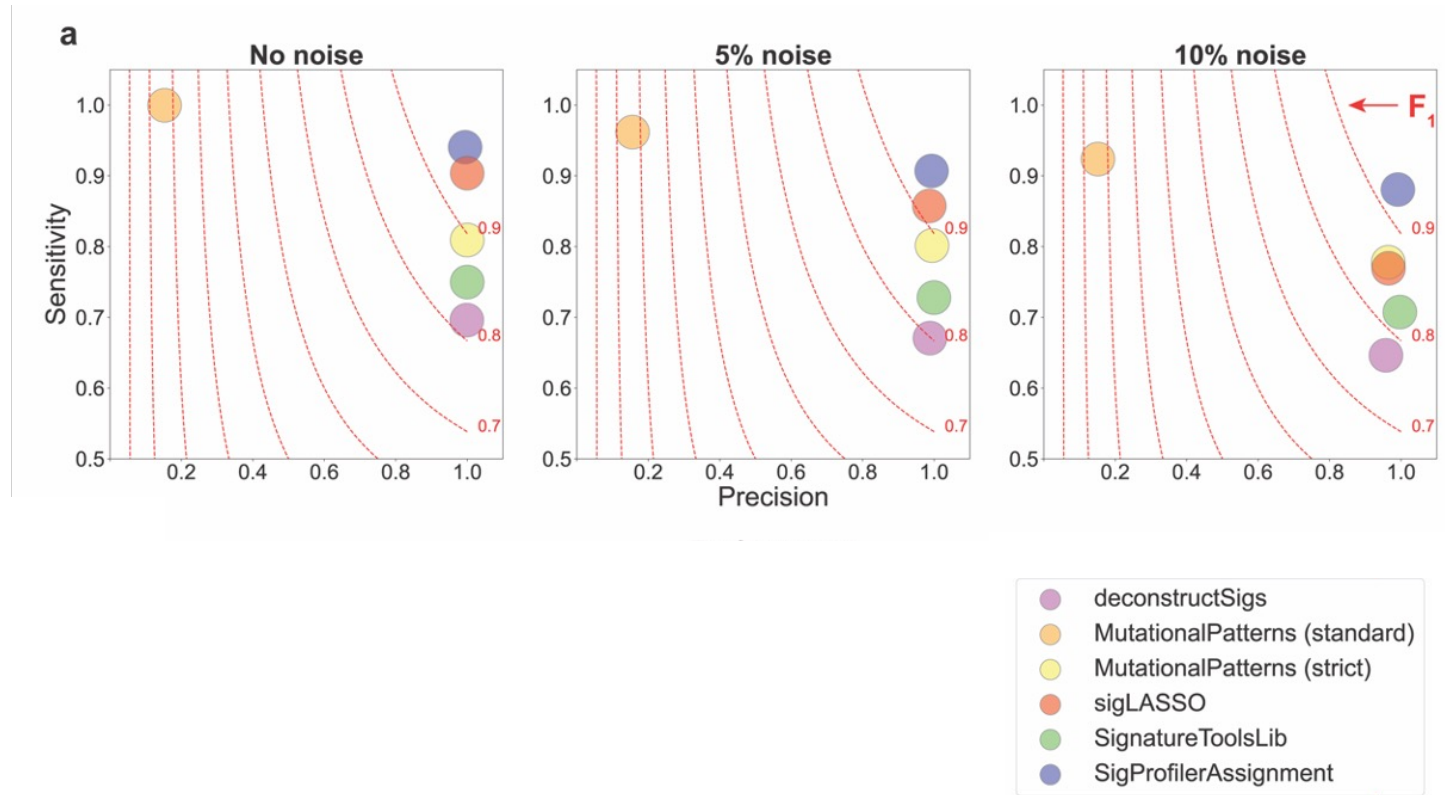
SigProfilerAssignment outperformed other approaches while maintaining similar computational performance
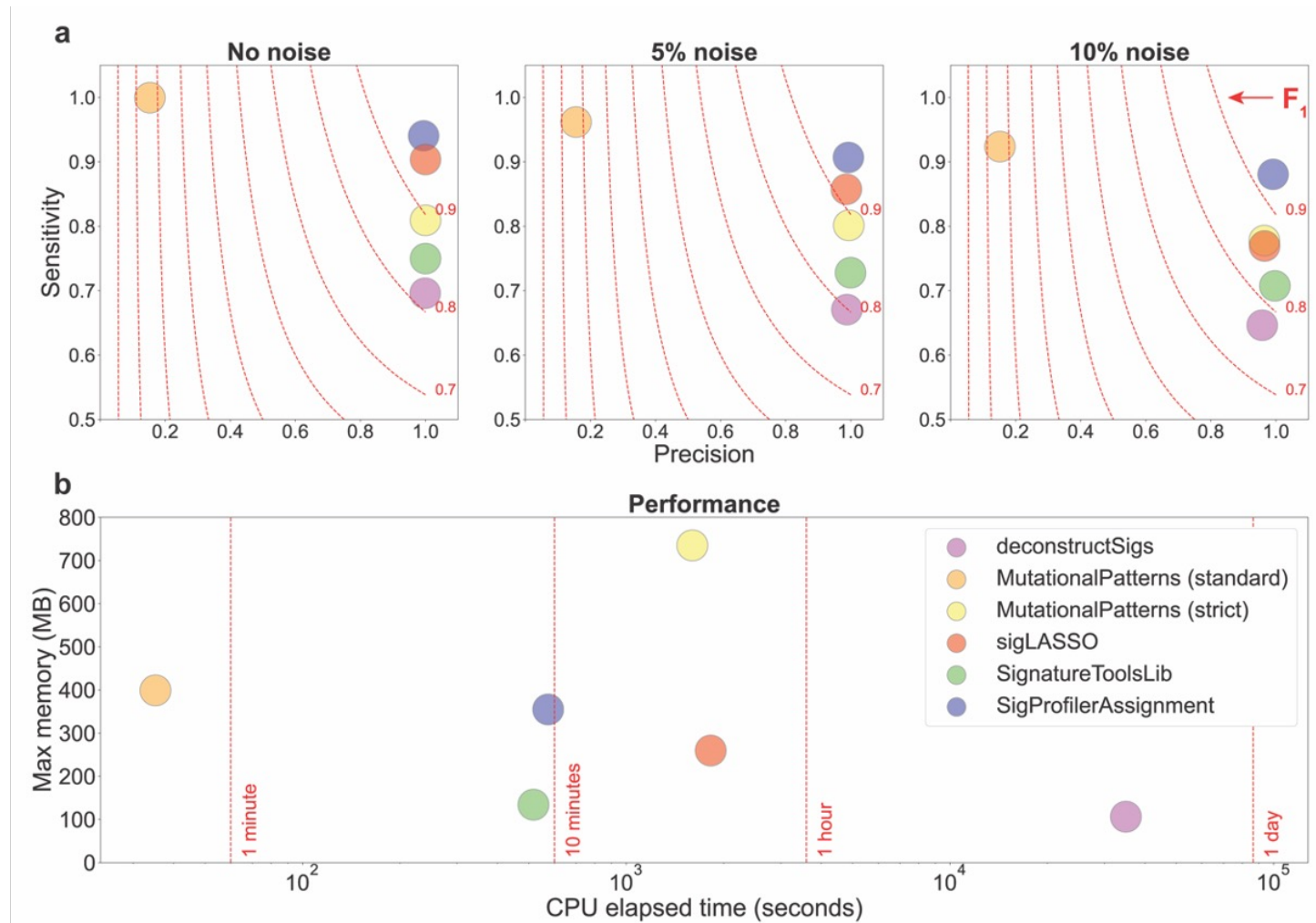
SigProfilerAssignment – ONLINE TOOL
https://cancer.sanger.ac.uk/signatures/assignment/

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/