

SigProfilerExtractor

Presented by:

Shakuntala Baichoo¹, Marcos Diaz-Gay² & Mariya Kazachkova²

¹University of Mauritius

²Alexandrov Lab, University of California San Diego

Cancer Genome Analysis

12 – 16 September 2022 – Virtual course

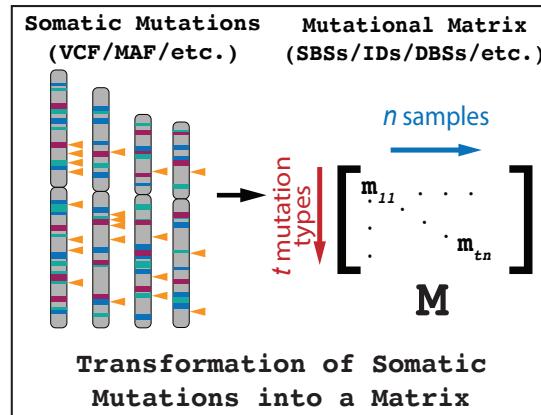


UC San Diego

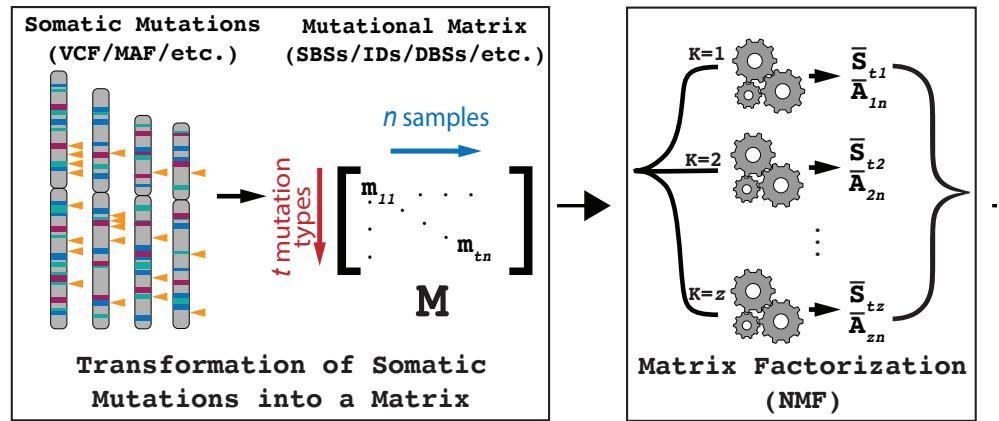
WELLCOME GENOME CAMPUS

CONNECTING
SCIENCE
ADVANCED
COURSES +
SCIENTIFIC
CONFERENCES

SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures



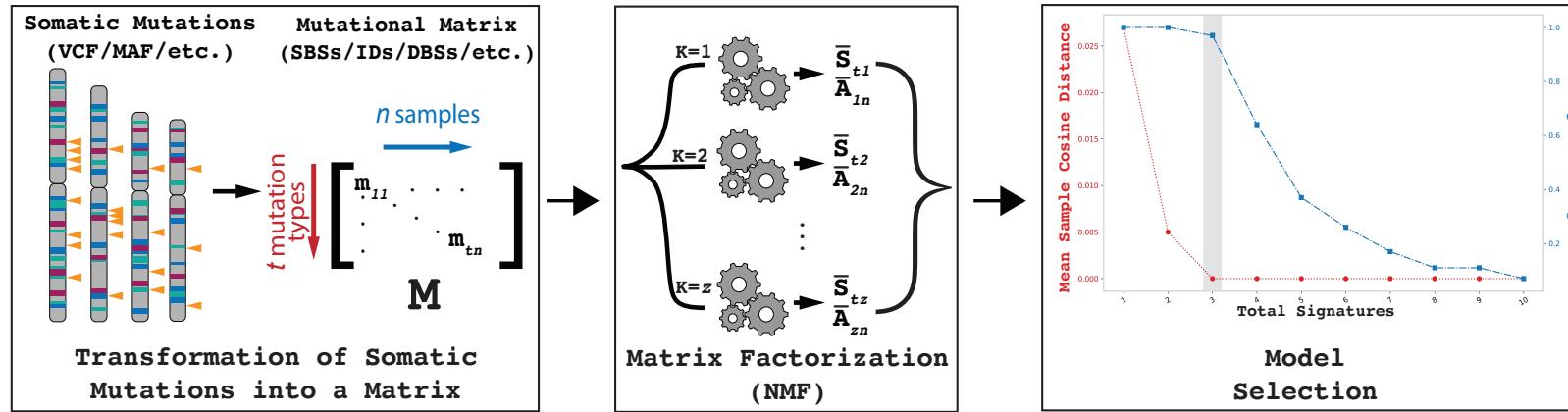
SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures



$$M = S \times A$$

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

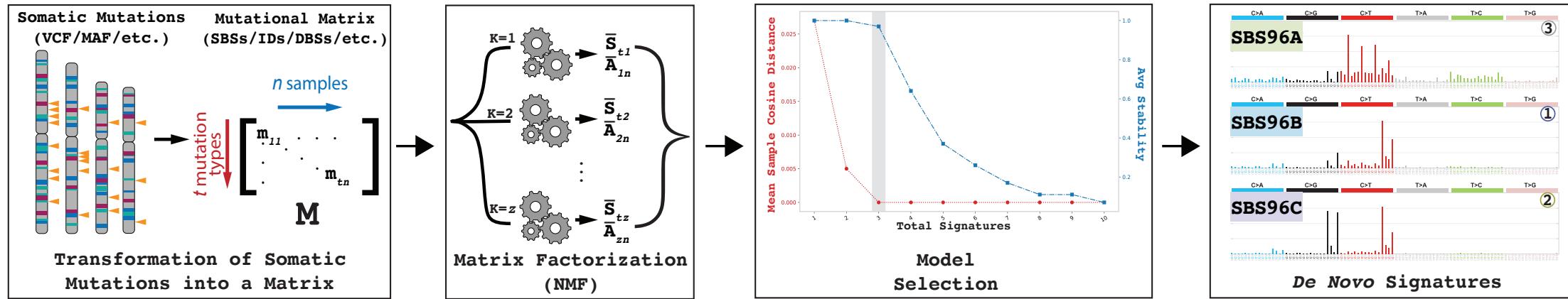
SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures



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

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

SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures

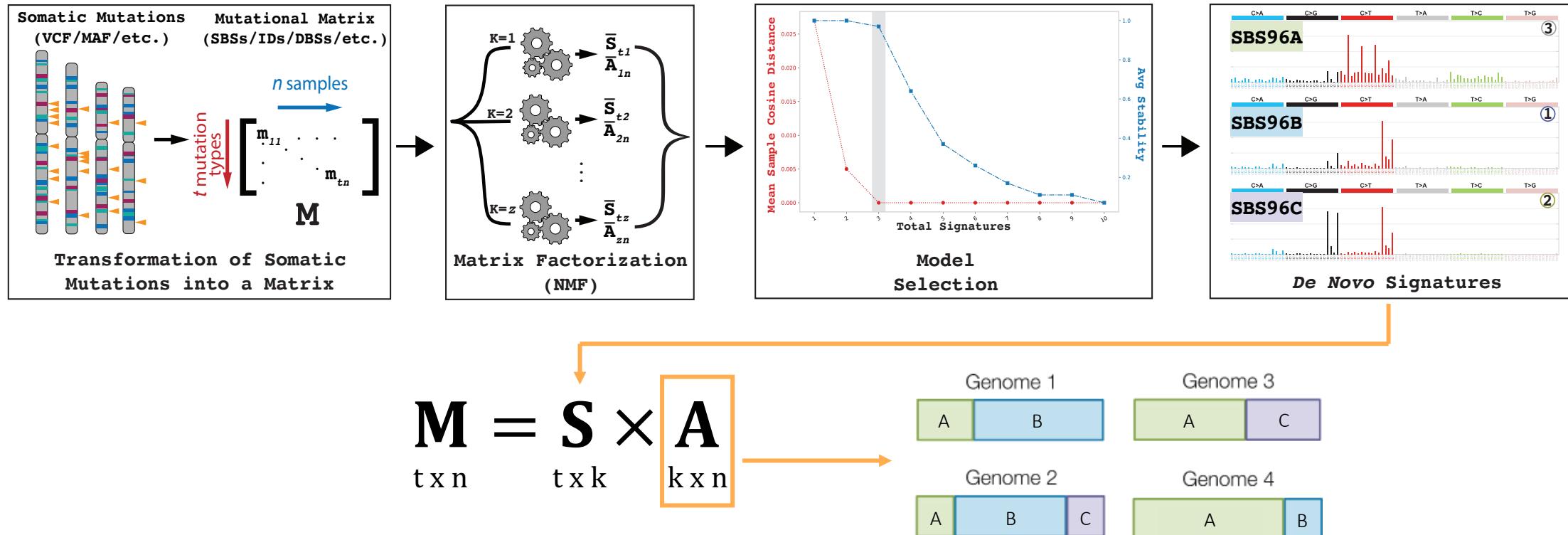


$$M = S \times A$$

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

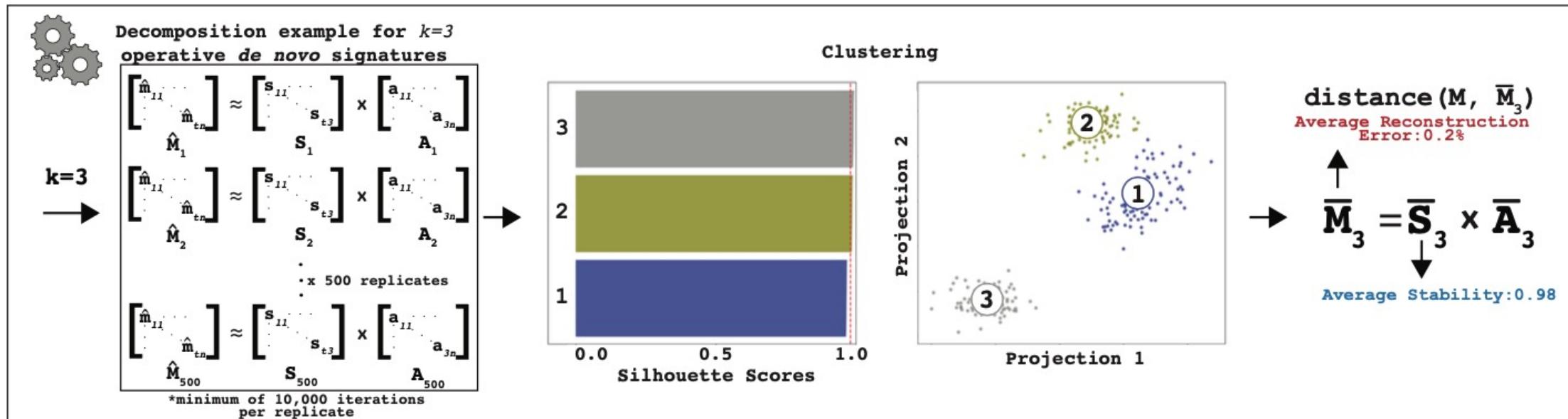
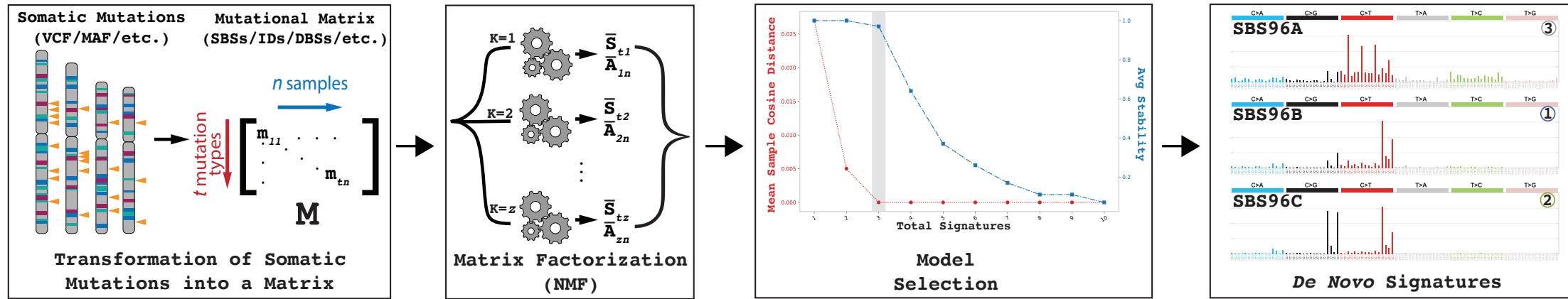
SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures

SIGPROFILER
Extractor

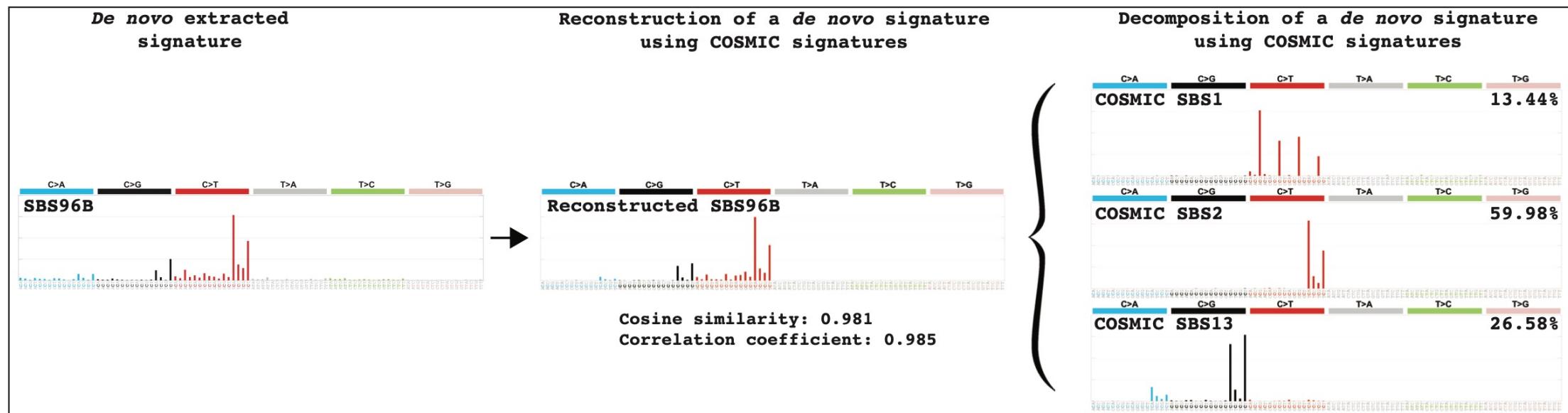


SIGPROFILER
Assignment

SigProfilerExtractor uses NMF for identifying *de novo* mutational signatures



SigProfilerExtractor decomposes *de novo* mutational signatures using COSMIC reference signatures



Different tools are currently available to perform *de novo* signature extraction

| Tool | Platform | Factorization Approach | | Selection Approach | | Reference |
|----------------------|--------------------|------------------------|---------------------------|--------------------|--------------|-------------------------------|
| | | Method | Computational Engine | Type | Algorithm | |
| EMu | C++ | EM | Original implementation | M/A | BIC | Fischer <i>et al.</i> 2013 |
| Maftools | R-Bioconductor | NMF | NMF R package | M | - | Mayakonda <i>et al.</i> 2018 |
| MutationalPatterns | R-Bioconductor | NMF | NMF R package | M | - | Blokzijl <i>et al.</i> 2018 |
| MutSignatures | R | NMF | Brunet <i>et al.</i> 2004 | - | - | Fantini <i>et al.</i> 2020 |
| MutSpec | R/Galaxy | NMF | NMF R package | M | - | Ardin <i>et al.</i> 2016 |
| SigFit | R | Bayesian inference | Stan R package | M/A | Elbow method | Gori <i>et al.</i> 2020 |
| SigMiner | R | NMF/Bay. NMF | NMF R package/SA | M/A | ARD | Wang <i>et al.</i> 2021 |
| SignatureAnalyzer | R/Python | Bayesian NMF | Original implementation | A | ARD | Kasar <i>et al.</i> 2015 |
| SignatureToolsLib | R | NMF | NMF R package | M | - | Degasperi <i>et al.</i> 2020 |
| SigneR | C++/R-Bioconductor | Bayesian NMF | Original implementation | M/A | BIC | Rosales <i>et al.</i> 2017 |
| SigProfilerExtractor | Python/R | NMF | Original implementation | M/A | NMFk | Islam <i>et al.</i> 2021 |
| SigProfiler_PCAWG | Python/MATLAB | NMF | Brunet <i>et al.</i> 2004 | M | - | Alexandrov <i>et al.</i> 2013 |
| SomaticSignatures | R-Bioconductor | NMF | NMF R package | M | - | Gehring <i>et al.</i> 2015 |
| TensorSignatures | Python | NTF | TensorFlow | M/A | BIC | Vöhringer <i>et al.</i> 2021 |

Different tools are currently available to perform *de novo* signature extraction

| Tool | Platform | Factorization Approach | | Selection Approach | | Reference |
|----------------------|--------------------|------------------------|---------------------------|--------------------|--------------|-------------------------------|
| | | Method | Computational Engine | Type | Algorithm | |
| EMu | C++ | EM | Original implementation | M/A | BIC | Fischer <i>et al.</i> 2013 |
| Maftools | R-Bioconductor | NMF | NMF R package | M | - | Mayakonda <i>et al.</i> 2018 |
| MutationalPatterns | R-Bioconductor | NMF | NMF R package | M | - | Blokzijl <i>et al.</i> 2018 |
| MutSignatures | R | NMF | Brunet <i>et al.</i> 2004 | - | - | Fantini <i>et al.</i> 2020 |
| MutSpec | R/Galaxy | NMF | NMF R package | M | - | Ardin <i>et al.</i> 2016 |
| SigFit | R | Bayesian inference | Stan R package | M/A | Elbow method | Gori <i>et al.</i> 2020 |
| SigMiner | R | NMF/Bay. NMF | NMF R package/SA | M/A | ARD | Wang <i>et al.</i> 2021 |
| SignatureAnalyzer | R/Python | Bayesian NMF | Original implementation | A | ARD | Kasar <i>et al.</i> 2015 |
| SignatureToolsLib | R | NMF | NMF R package | M | - | Degasperi <i>et al.</i> 2020 |
| SigneR | C++/R-Bioconductor | Bayesian NMF | Original implementation | M/A | BIC | Rosales <i>et al.</i> 2017 |
| SigProfilerExtractor | Python/R | NMF | Original implementation | M/A | NMFk | Islam <i>et al.</i> 2021 |
| SigProfiler_PCAWG | Python/MATLAB | NMF | Brunet <i>et al.</i> 2004 | M | - | Alexandrov <i>et al.</i> 2013 |
| SomaticSignatures | R-Bioconductor | NMF | NMF R package | M | - | Gehring <i>et al.</i> 2015 |
| TensorSignatures | Python | NTF | TensorFlow | M/A | BIC | Vöhringer <i>et al.</i> 2021 |

Different tools are currently available to perform *de novo* signature extraction

| Tool | Platform | Factorization Approach | | Selection Approach | | Reference |
|----------------------|--------------------|------------------------|-----------------------------------|--------------------|--------------|-------------------------------|
| | | Method | Computational Engine | Type | Algorithm | |
| EMu | C++ | EM | Original implementation | M/A | BIC | Fischer <i>et al.</i> 2013 |
| Maftools | R-Bioconductor | NMF | NMF R package | M | - | Mayakonda <i>et al.</i> 2018 |
| MutationalPatterns | R-Bioconductor | NMF | NMF R package | M | - | Blokzijl <i>et al.</i> 2018 |
| MutSignatures | R | NMF | Brunet <i>et al.</i> 2004 | - | - | Fantini <i>et al.</i> 2020 |
| MutSpec | R/Galaxy | NMF | NMF R package | M | - | Ardin <i>et al.</i> 2016 |
| SigFit | R | Bayesian inference | Stan R package | M/A | Elbow method | Gori <i>et al.</i> 2020 |
| SigMiner | R | NMF/Bay. NMF | NMF R package /SA | M/A | ARD | Wang <i>et al.</i> 2021 |
| SignatureAnalyzer | R/Python | Bayesian NMF | Original implementation | A | ARD | Kasar <i>et al.</i> 2015 |
| SignatureToolsLib | R | NMF | NMF R package | M | - | Degasperi <i>et al.</i> 2020 |
| SigneR | C++/R-Bioconductor | Bayesian NMF | Original implementation | M/A | BIC | Rosales <i>et al.</i> 2017 |
| SigProfilerExtractor | Python/R | NMF | Original implementation | M/A | NMFk | Islam <i>et al.</i> 2021 |
| SigProfiler_PCAWG | Python/MATLAB | NMF | Brunet <i>et al.</i> 2004 | M | - | Alexandrov <i>et al.</i> 2013 |
| SomaticSignatures | R-Bioconductor | NMF | NMF R package | M | - | Gehring <i>et al.</i> 2015 |
| TensorSignatures | Python | NTF | TensorFlow | M/A | BIC | Vöhringer <i>et al.</i> 2021 |

Ground truth signatures allow evaluation of tool performance on synthetic data

| | Extracted Signature A | Extracted Signature B | Extracted Signature C | Extracted Signature D |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Ground Truth Signature 1 | 0.14 | 0.98 | 0.56 | 0.36 |
| Ground Truth Signature 2 | 0.35 | 0.29 | 0.93 | 0.46 |
| Ground Truth Signature 3 | 0.31 | 0.56 | 0.78 | 0.66 |
| Ground Truth Signature 4 | 0.34 | 0.08 | 0.57 | 0.67 |
| Ground Truth Signature 5 | 0.95 | 0.15 | 0.81 | 0.39 |
| Ground Truth Signature 6 | 0.23 | 0.74 | 0.48 | 0.26 |

Ground truth signatures allow evaluation of tool performance on synthetic data

| | Extracted Signature A | Extracted Signature B | Extracted Signature C | Extracted Signature D |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Ground Truth Signature 1 | 0.14 | 0.98 | 0.56 | 0.36 |
| Ground Truth Signature 2 | 0.35 | 0.29 | 0.93 | 0.46 |
| Ground Truth Signature 3 | 0.31 | 0.56 | 0.78 | 0.66 |
| Ground Truth Signature 4 | 0.34 | 0.08 | 0.57 | 0.67 |
| Ground Truth Signature 5 | 0.95 | 0.15 | 0.81 | 0.39 |
| Ground Truth Signature 6 | 0.23 | 0.74 | 0.48 | 0.26 |

Cosine similarity between Extracted Signature C and Ground Truth Signature 6



Ground truth signatures allow evaluation of tool performance on synthetic data

| | Extracted Signature A | Extracted Signature B | Extracted Signature C | Extracted Signature D |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Ground Truth Signature 1 | 0.14 | 0.98 | 0.56 | 0.36 |
| Ground Truth Signature 2 | 0.35 | 0.29 | 0.93 | 0.46 |
| Ground Truth Signature 3 | 0.31 | 0.56 | 0.78 | 0.66 |
| Ground Truth Signature 4 | 0.34 | 0.08 | 0.57 | 0.67 |
| Ground Truth Signature 5 | 0.95 | 0.15 | 0.81 | 0.39 |
| Ground Truth Signature 6 | 0.23 | 0.74 | 0.48 | 0.26 |

True Positives (TP; ≥ 0.90)

Extracted Signature A
Extracted Signature B
Extracted Signature C

Signatures correctly extracted from the *dataset*

False Positives (FP)

Extracted Signature D
Signatures extracted but absent in the dataset

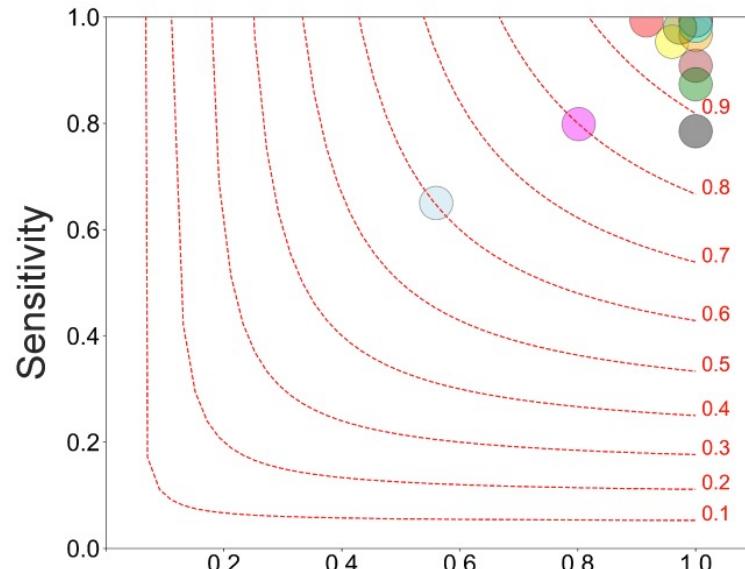
False Negatives (FN)

Ground Truth Signature 3
Ground Truth Signature 4
Ground Truth Signature 6

Signatures not extracted but used in simulating the *dataset*

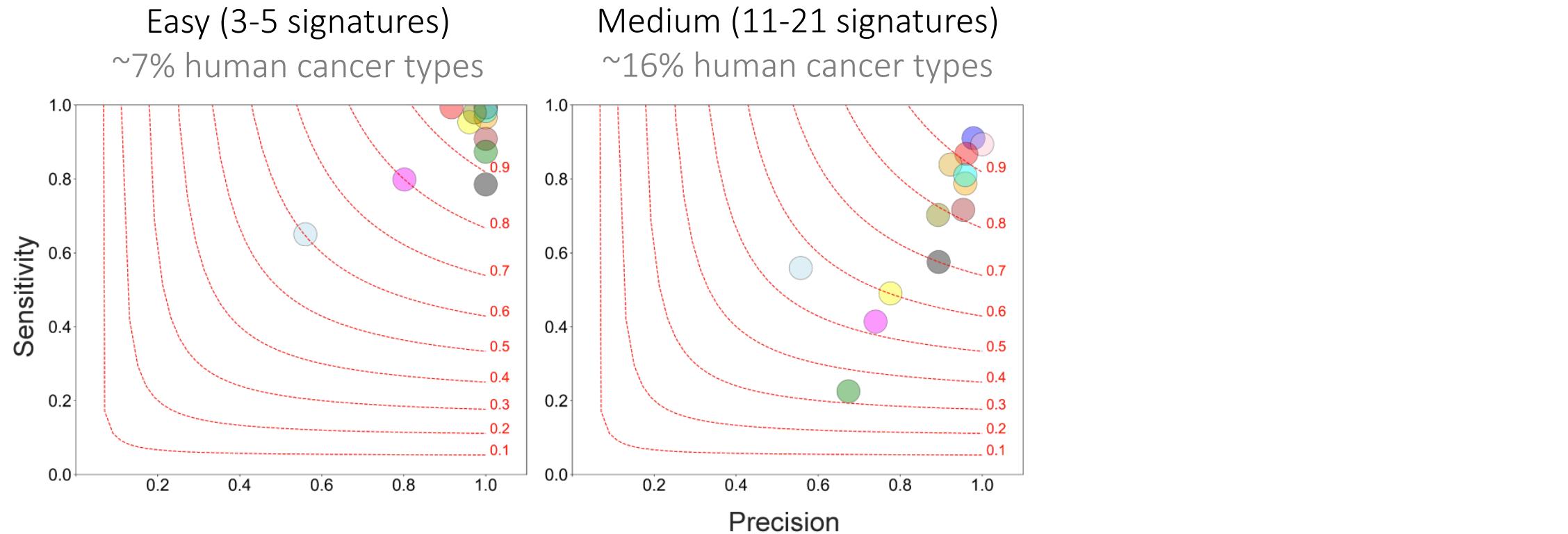
SigProfilerExtractor outperformed other approaches in WGS noiseless data, specially in hard scenarios

Easy (3-5 signatures)
~7% human cancer types



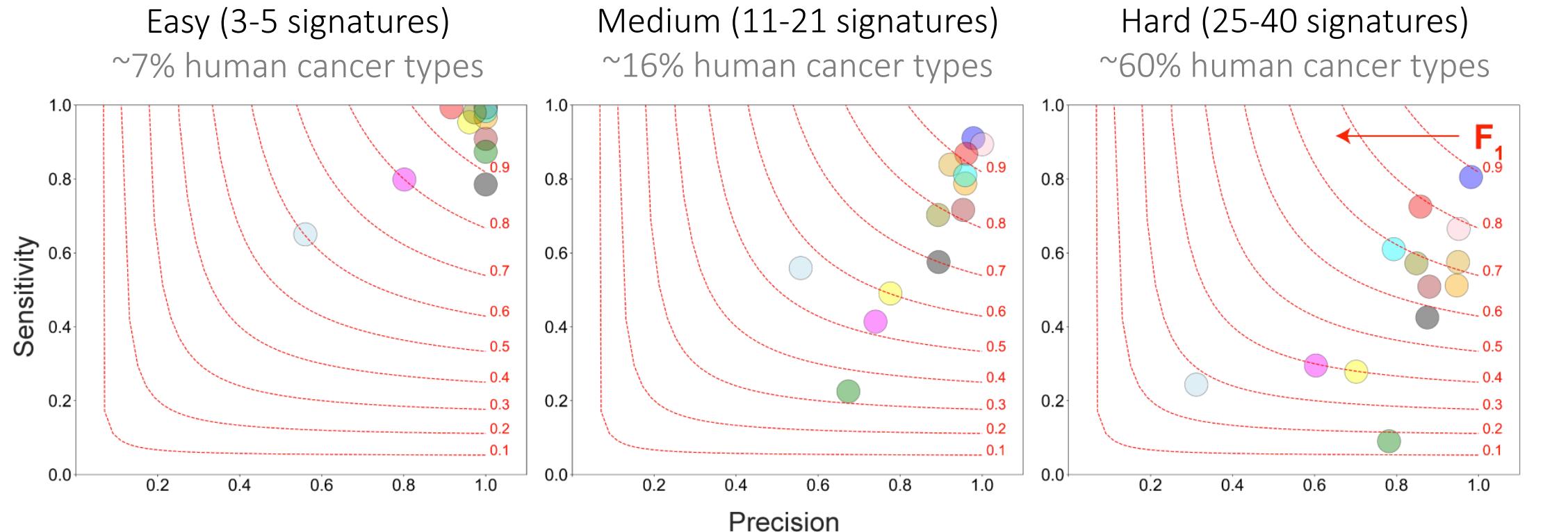
- █ SigProfilerExtractor
- █ SignatureAnalyzer
- █ SigProfiler_PCAWG
- █ SigneR
- █ MutationalPatterns
- █ SomaticSignatures
- █ MutSpec
- █ SignatureToolsLib
- █ SigMiner
- █ MutSignatures
- █ Maftools
- █ EMu
- █ SigFit
- █ TensorSignatures

SigProfilerExtractor outperformed other approaches in WGS noiseless data, specially in hard scenarios



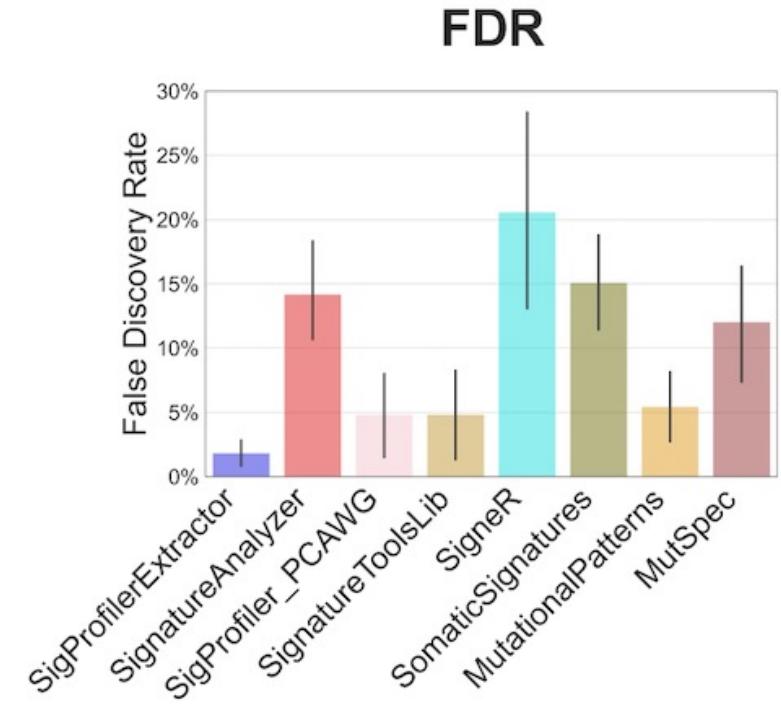
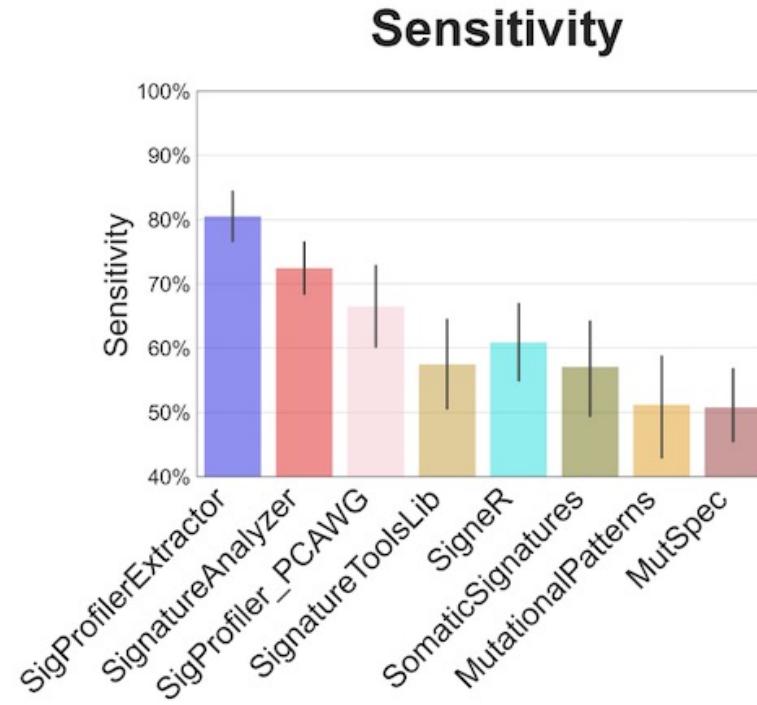
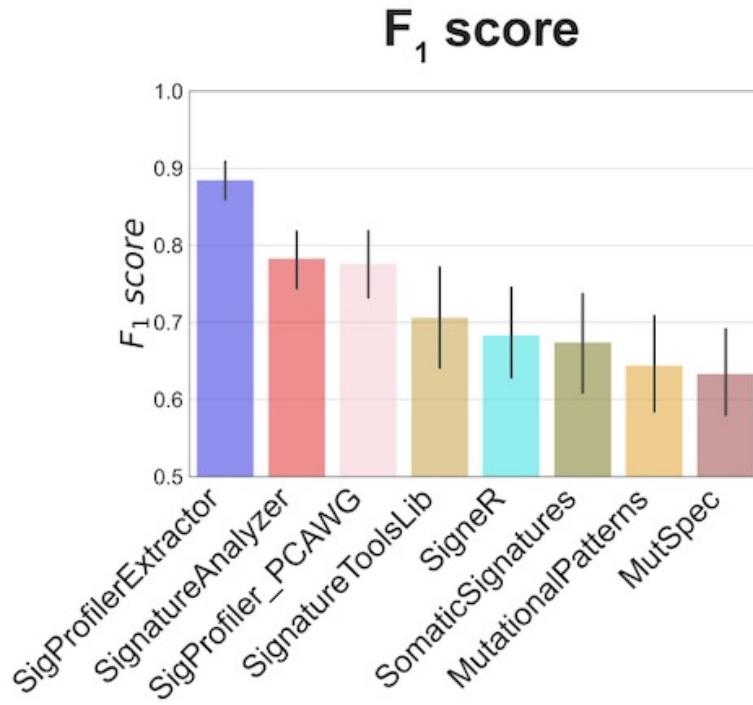
- | | | | | |
|--|--|---|---|---|
| █ SigProfilerExtractor | █ SigneR | █ MutSpec | █ MutSignatures | █ SigFit |
| █ SignatureAnalyzer | █ MutationalPatterns | █ SignatureToolsLib | █ Maftools | █ TensorSignatures |
| █ SigProfiler_PCAWG | █ SomaticSignatures | █ SigMiner | █ EMu | |

SigProfilerExtractor outperformed other approaches in WGS noiseless data, specially in hard scenarios

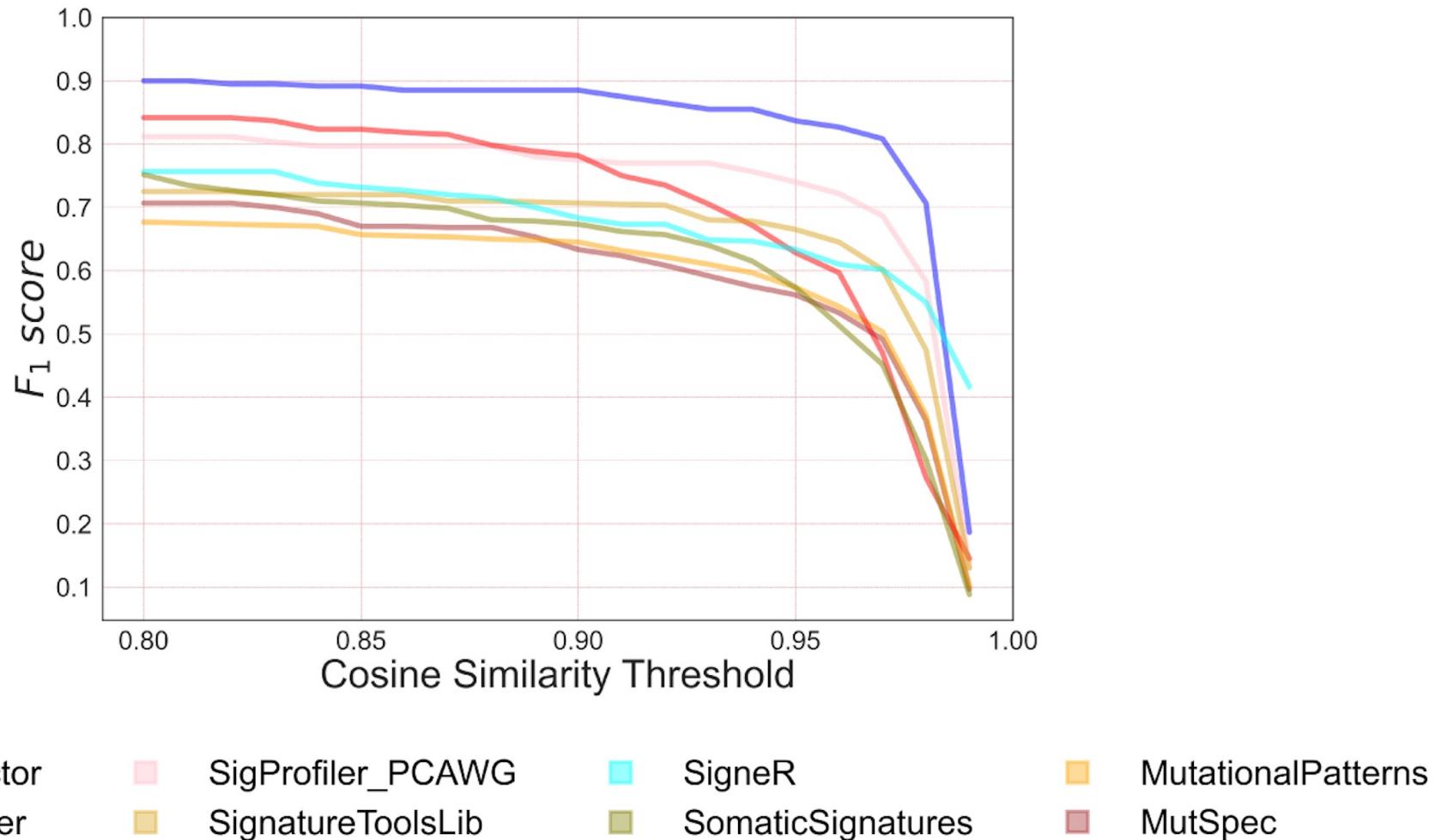


- | | | | | |
|--|--|---|---|---|
| █ SigProfilerExtractor | █ SigneR | █ MutSpec | █ MutSignatures | █ SigFit |
| █ SignatureAnalyzer | █ MutationalPatterns | █ SignatureToolsLib | █ Maftools | █ TensorSignatures |
| █ SigProfiler_PCAWG | █ SomaticSignatures | █ SigMiner | █ EMu | |

SigProfilerExtractor outperformed other approaches in WGS noiseless data, specially in hard scenarios

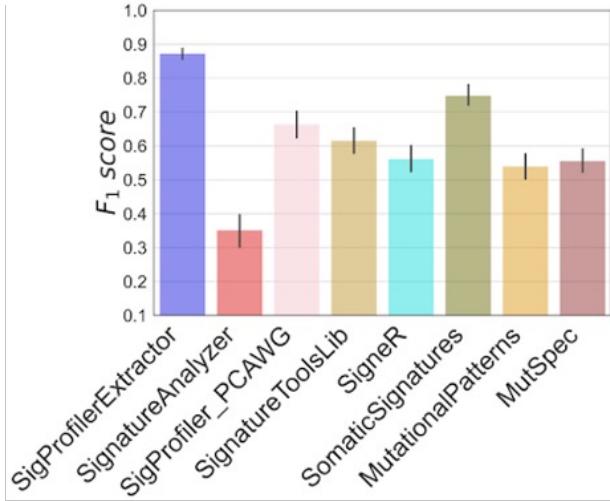


SigProfilerExtractor outperformed other approaches in WGS noiseless data, specially in hard scenarios

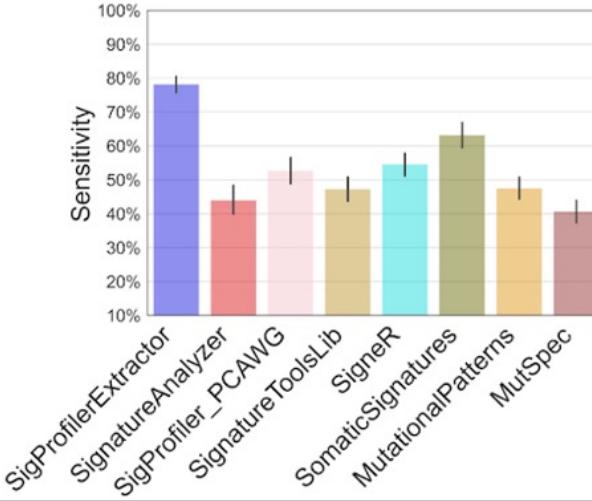


SigProfilerExtractor outperformed other tools in both WGS and WES realistic scenarios (5% noise)

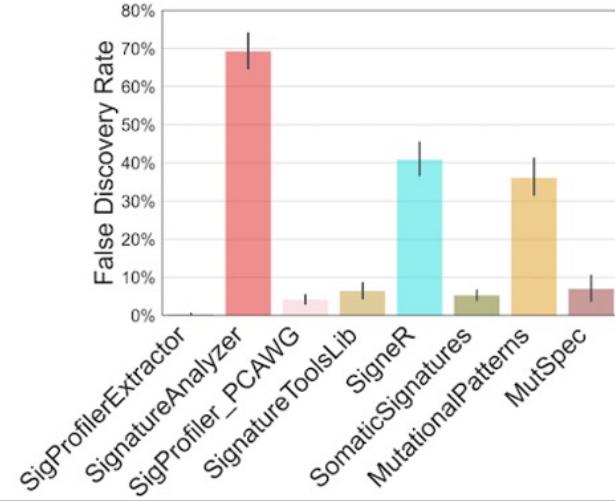
F_1 score WGS



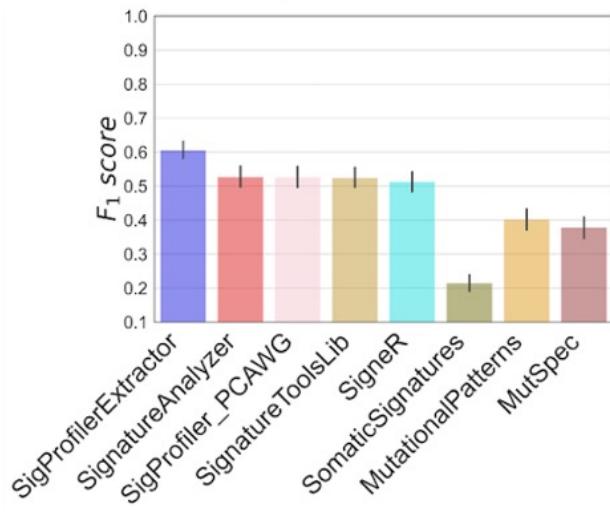
Sensitivity WGS



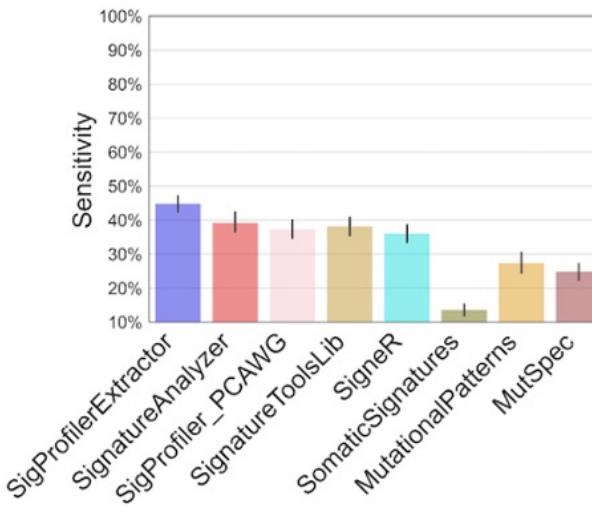
FDR WGS



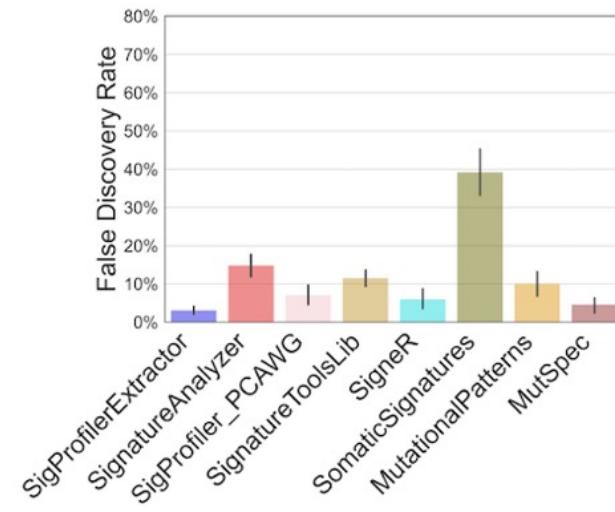
F_1 score WES



Sensitivity WES



FDR WES



Useful links – SigProfilerExtractor

- Pre-print manuscript
<https://www.biorxiv.org/content/10.1101/2020.12.13.422570v3>
- GitHub repository (python package)
<https://github.com/AlexandrovLab/SigProfilerExtractor>
- GitHub repository (R wrapper)
<https://github.com/AlexandrovLab/SigProfilerExtractorR>
- Wiki page (usage instructions)
<https://osf.io/t6j7u/wiki/home/>