

Sensitivity analysis of mutational signature deconstruction

Maths4Life project

Ferran López Guitart

Summer 2021



Mutational signatures

- A person is subjected to multiple mutation-causing processes through their life. They can be exogenous or endogenous.
- Single base substitution (SBS) mutations can be classified into 96 types.
- Each mutational process results in a different pattern of mutation types, called its **signature**.

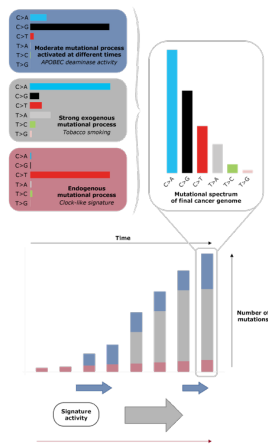


Diagram from COSMIC

signature catalog

Mutational signatures

- In real patients, multiple processes are usually involved. Thus, multiple signatures are present when a DNA sample from a patient is sequenced.
- Computational methods can be used to identify which signatures are present in a sample.
- This can give us information about the processes this cell has been subjected to.

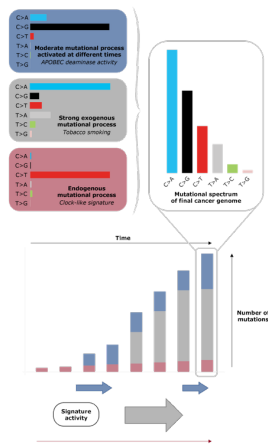
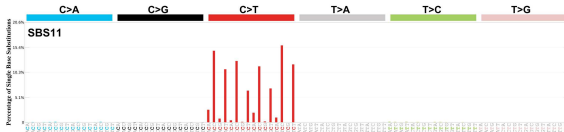


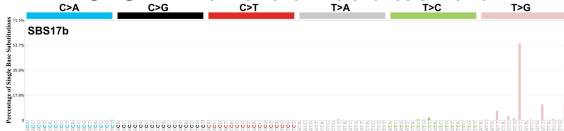
Diagram from COSMIC

signature catalog

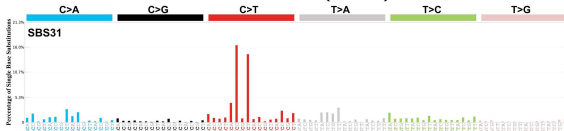
Mutational signatures



SBS11: Temozolomide treatment



SBS17b: Fluorouracil (5FU) treatment



SBS31: Platinum treatment

The signature deconstruction problem

Given a mutational spectrum $G \in \mathbb{R}^n$ (extracted from sequencing a sample), and known mutational signatures $S_1, \dots, S_m \in \mathbb{R}^n$, decompose G into a linear combination

$$G \approx e_1 S_1 + \dots + e_m S_m$$

where $e_1, \dots, e_m > 0$ are exposures to each mutational process.

Warning

This is an ill-posed problem with an infinite number of solutions. We must include additional requirements, such as sparsity.

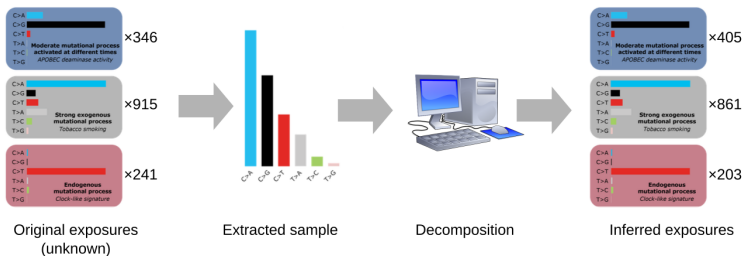
Common algorithms: `deconstructSigs`¹ (exposure inference), `mSigAct`² (exposure inference + presence test)

¹[doi:10.1186/s13059-016-0893-4](https://doi.org/10.1186/s13059-016-0893-4)

²[doi:10.1126/scitranslmed.aan6446](https://doi.org/10.1126/scitranslmed.aan6446)

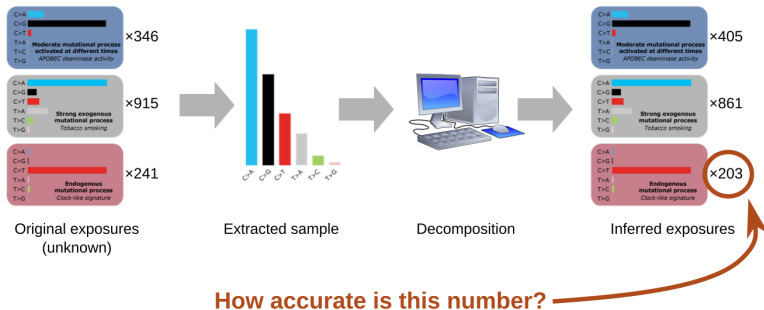
Uncertainty of deconstruction algorithms

Goal



Uncertainty of deconstruction algorithms

Goal



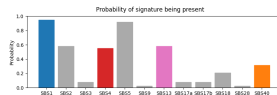
Goal

Estimate the original exposure, using the output of the deconstruction method.

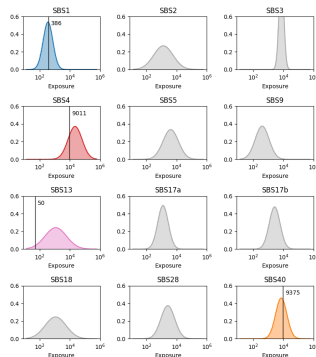
Synthetic sample generation

From a cancer type cohort

- 1 Deconstruct the samples into signatures
- 2 Calculate signature frequency and exposure distribution from cohort
- 3 Select signatures and their exposures according to the distribution
- 4 Generate realistic mutational profiles from each signature with the selected exposure
- 5 Add all the mutations



Distribution of signature exposures

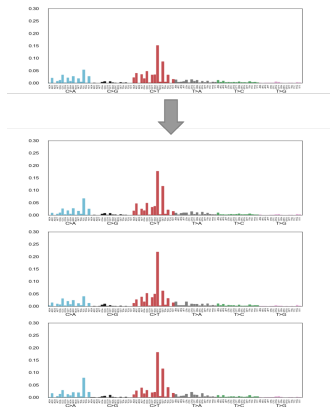


Mutational signatures in

Synthetic sample generation

From a single sample

We apply the negative binomial distortion process we applied to the signatures to the sample itself:



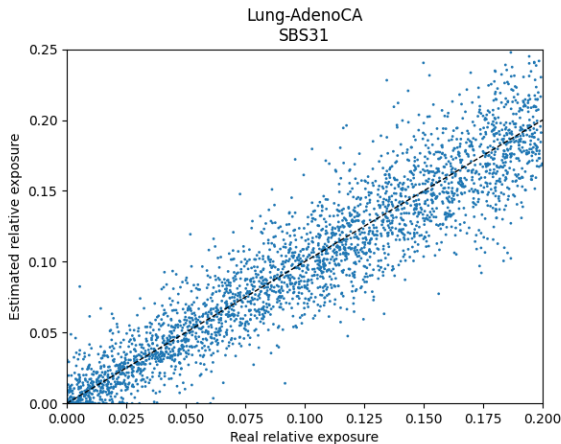
Uncertainty of deconstruction algorithms

Method

- 1 From a cohort of the same cancer type, generate a large number of synthetic samples (not containing the target signature)
- 2 Inject a certain exposure to the target signature, different for each sample
- 3 Use the algorithm to decompose the samples
- 4 Plot the exposure inferred by the algorithm against the actual injected exposure

Uncertainty of deconstruction algorithms

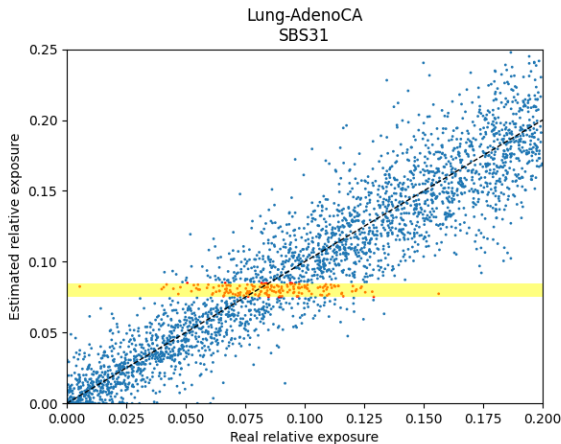
Result



Real and inferred exposure to signature SBS31 in a synthetic cohort of Lung-AdenoCA

Uncertainty of deconstruction algorithms

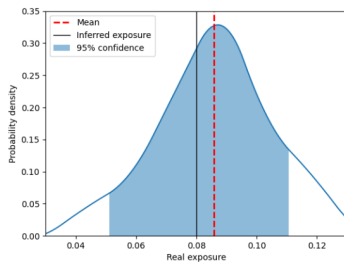
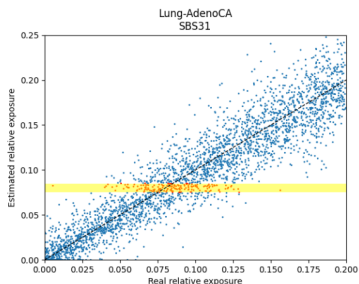
Result



Real and inferred exposure to signature SBS31 in a synthetic cohort of Lung-AdenoCA

Uncertainty of deconstruction algorithms

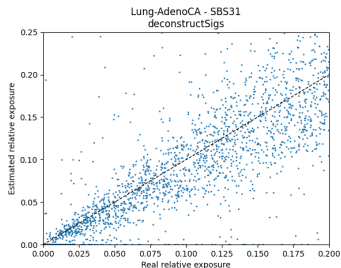
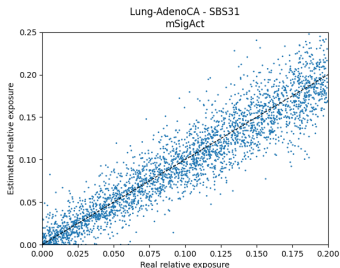
Result



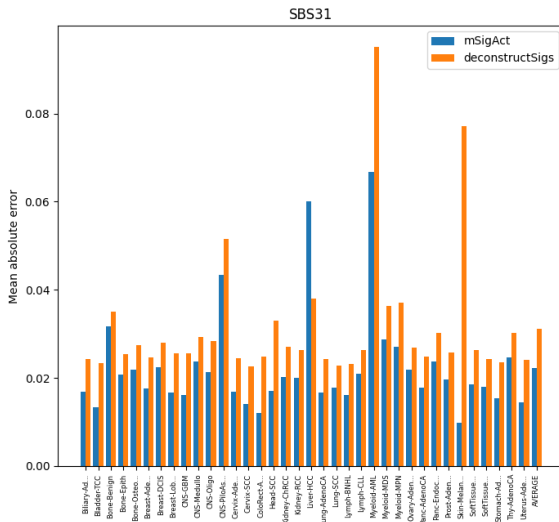
Distribution of real exposures if the inferred exposure is 0.08

Comparison of deconstruction algorithms

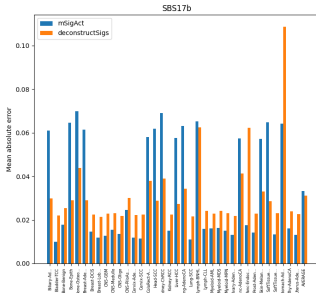
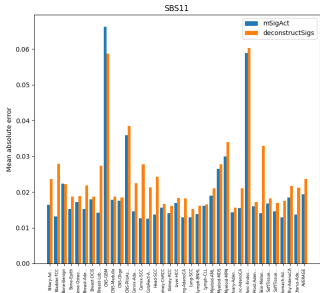
We can use the same method with different algorithms to compare their accuracy:



Comparison of deconstruction algorithms



Comparison of deconstruction algorithms

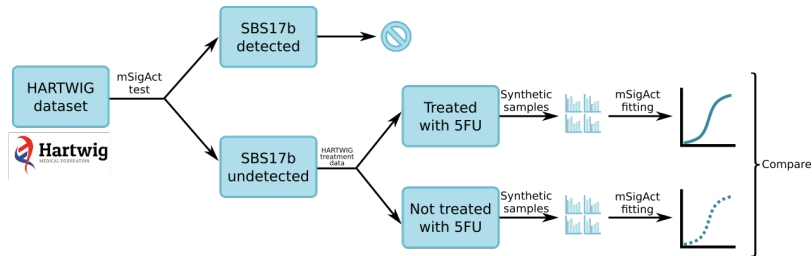


Does undetected mean unexposed?

Question

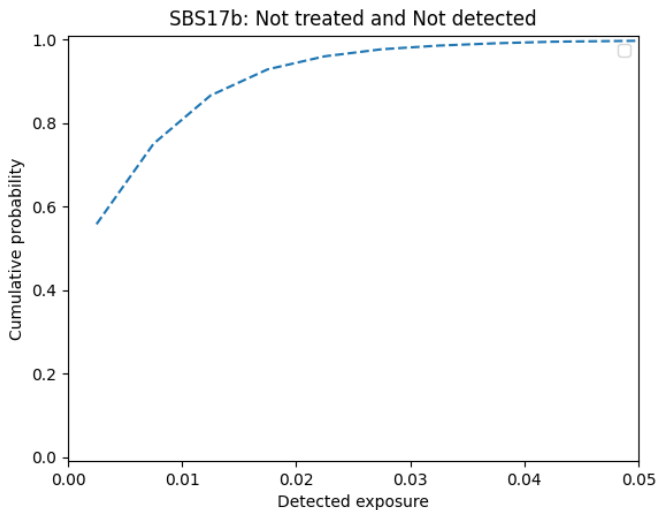
Question

Is it possible that a signature is present in a sample but is not detected by mSigAct?



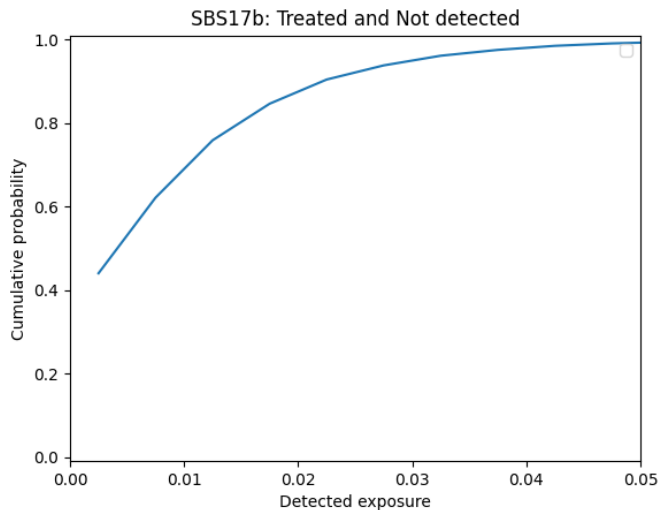
Does undetected mean unexposed?

Result



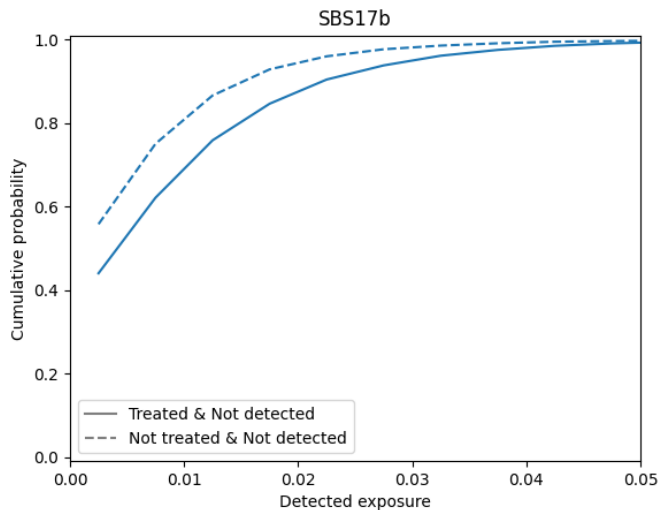
Does undetected mean unexposed?

Result



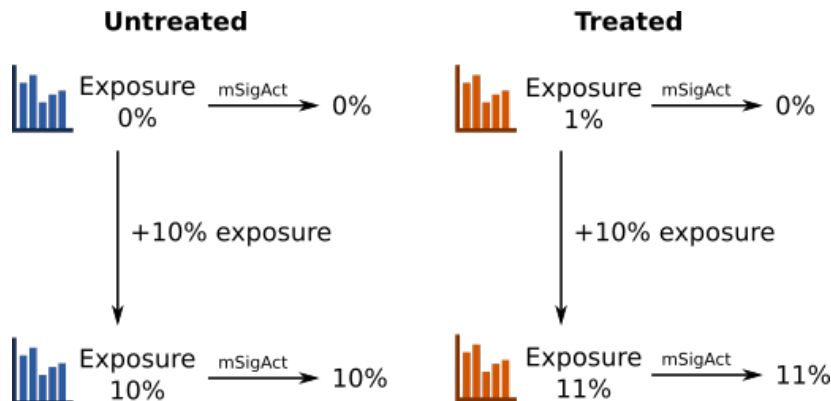
Does undetected mean unexposed?

Result



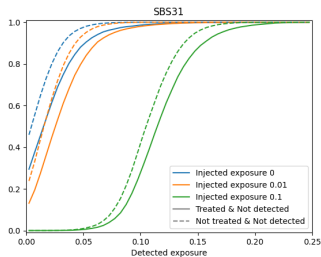
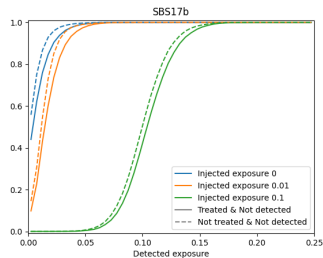
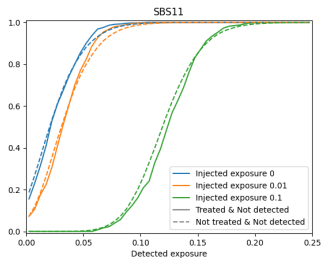
Does undetected mean unexposed?

Adding extra exposure



Does undetected mean unexposed?

Adding extra exposure



Added exposure	0	0.01	0.1
SBS11	0.49	0.044	$2.4 \cdot 10^{-5}$
SBS17b	0.067	$< 10^{-20}$	$< 10^{-20}$
SBS31	$< 10^{-20}$	$< 10^{-20}$	$< 10^{-20}$

p-value of
Kolmogorov-Smirnov test

Conclusions

- 1 We propose a stochastic model to randomly generate biologically plausible mutational catalogues.
- 2 We study the distribution of exposures to a specific mutational process based on the exposure estimated by a fitting method.
- 3 We propose a method to enhance the statistical power of the mSigAct by pooling synthetic samples.
- 4 We provide a ready-to-use tool to reproduce our analyses.³

³<https://github.com/flopgui/signature-deconstruction-sensitivity> 