

Module 4: Mutational Signature Analysis

Presented by:

Shakuntala Baichoo¹, Marcos Diaz-Gay² & Mariya Kazachkova²

¹University of Mauritius

²Alexandrov Lab, University of California San Diego

Based mostly on slides by:

Ludmil Alexandrov, Marcos Díaz-Gay & Mariya Kazachkova,
from University of California San Diego

 @ShakunBaichoo
 shakunb@uom.ac.mu

 @mdiazgay
 mdiazgay@health.ucsd.edu

 mkazachk@health.ucsd.edu

Cancer Genome Analysis

12 – 16 September 2022 – Virtual course



UNIVERSITY OF
MAURITIUS

UC San Diego

WELLCOME GENOME CAMPUS

CONNECTING SCIENCE
ADVANCED COURSES + SCIENTIFIC CONFERENCES

Outline & learning outcomes

- **Outline of lecture**
 - Refresher on cancer genomics
 - Explanation of mutational signatures and how they're studied
 - Some vocab we will cover:
 - Mutational processes
 - Mutational profiles
 - Mutational signatures
 - Clinical examples using mutational signatures
- **Learning outcomes for the lecture**
 - Participants will know how to interpret a mutational profile
 - Participants will be able to explain the relationship between mutational processes, mutational profiles and mutational signatures
 - Participants will understand the the clinical relevance of mutational signatures

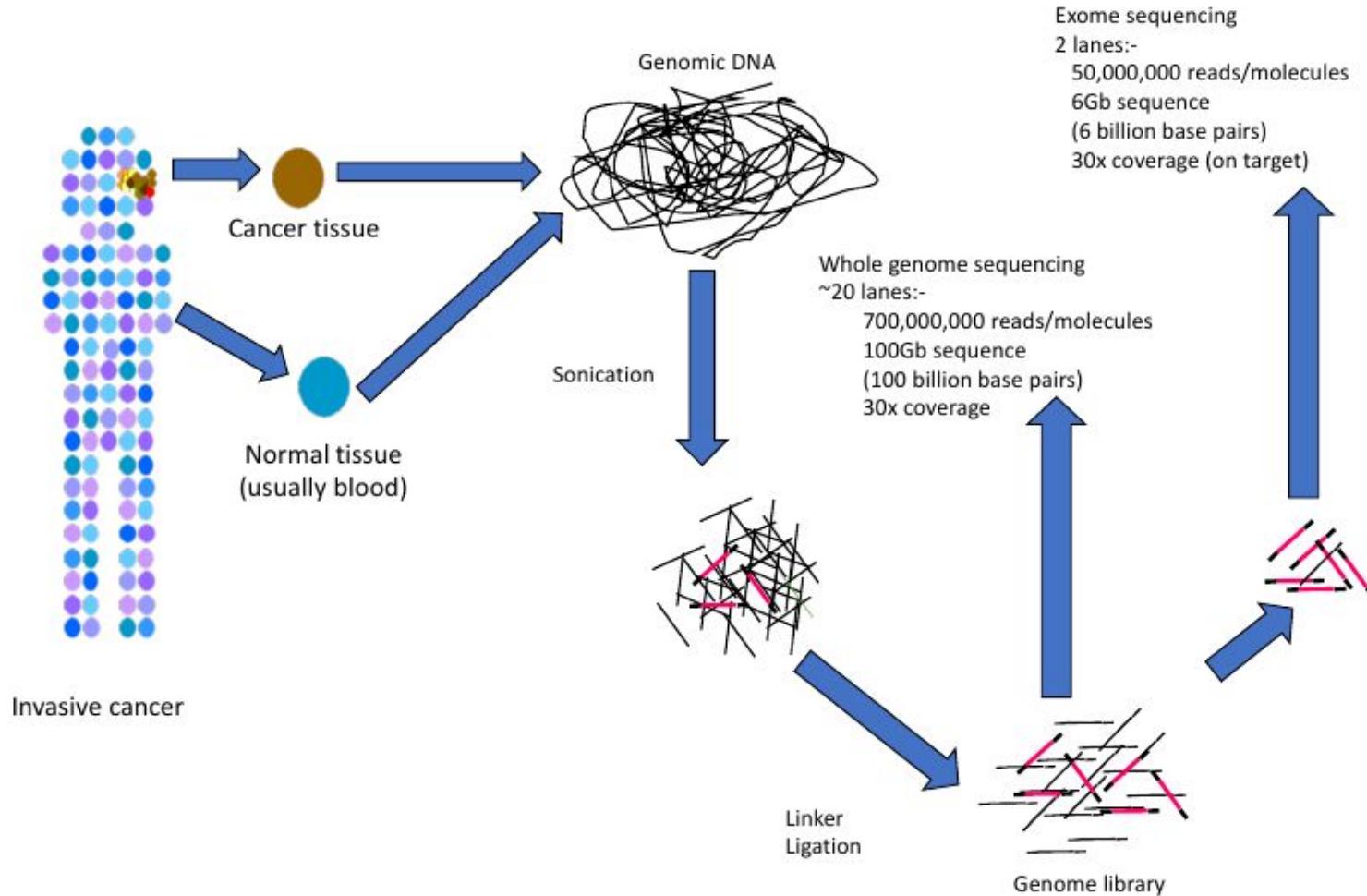
Overview of cancer

- All cancers originate from a single cell that starts to behave abnormally, dividing uncontrollably and invading adjacent tissues
- The reason that this single cell begins to behave abnormally is because of acquired changes in its genome known as somatic mutations
- Cancer is a disease of the genome and the most common human genetics disease

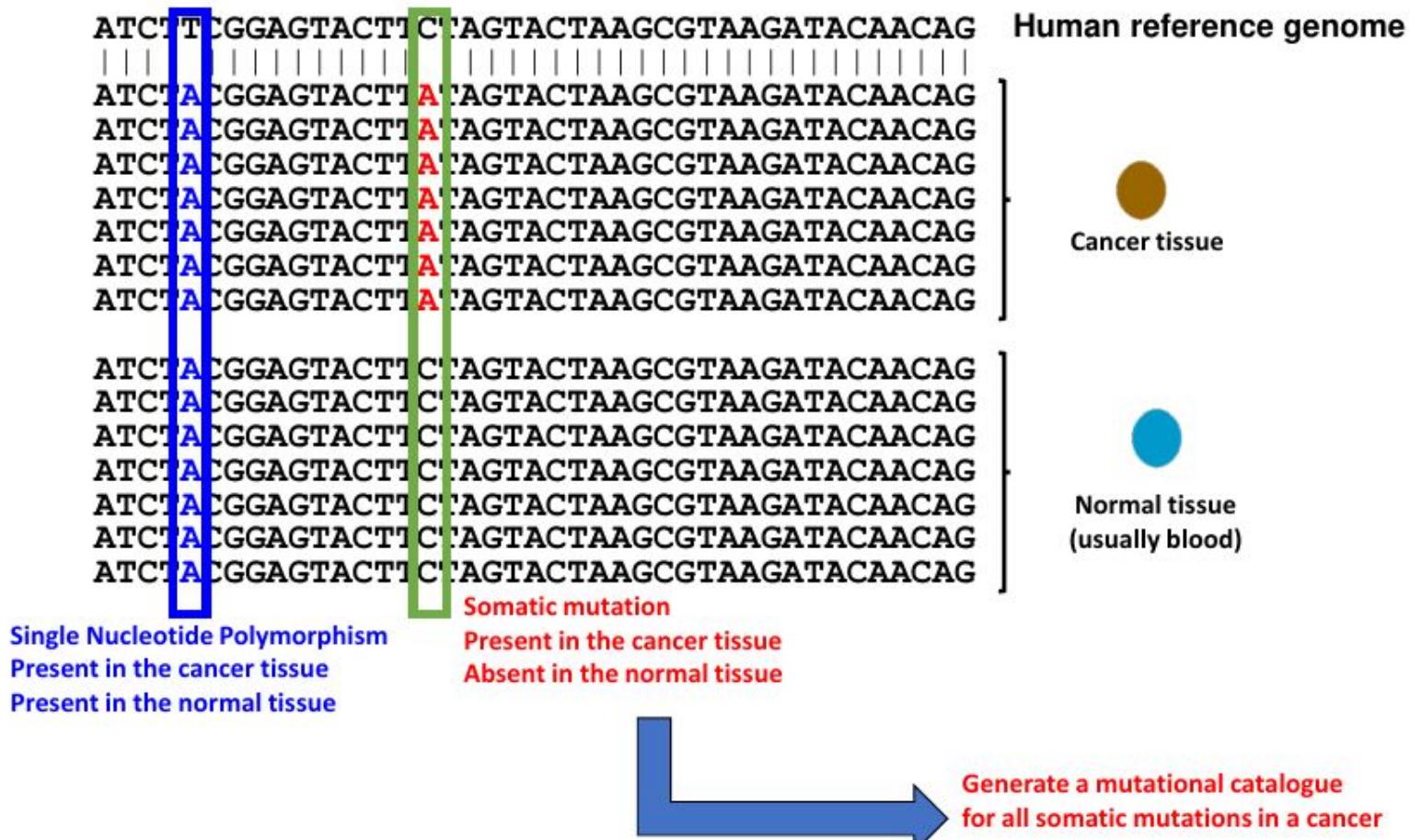
Types of mutations

- DNA molecules in our cells are targeted by diverse mutagenic processes that can occur in:
 - **germ** cells, contributing to species evolution
 - or in **somatic** cells, accumulating with age and contributing to diseases, especially cancer
- Recent mutation rate studies of tumors have focussed on deciphering the **somatically acquired changes** in the DNA of cancer cells to advance our understanding of the relations among mutagenic exposures, DNA damage and repair, and outcomes (such as cancer and uncontrolled cell growth)

Cancer genomics approach



Cancer genomics approach

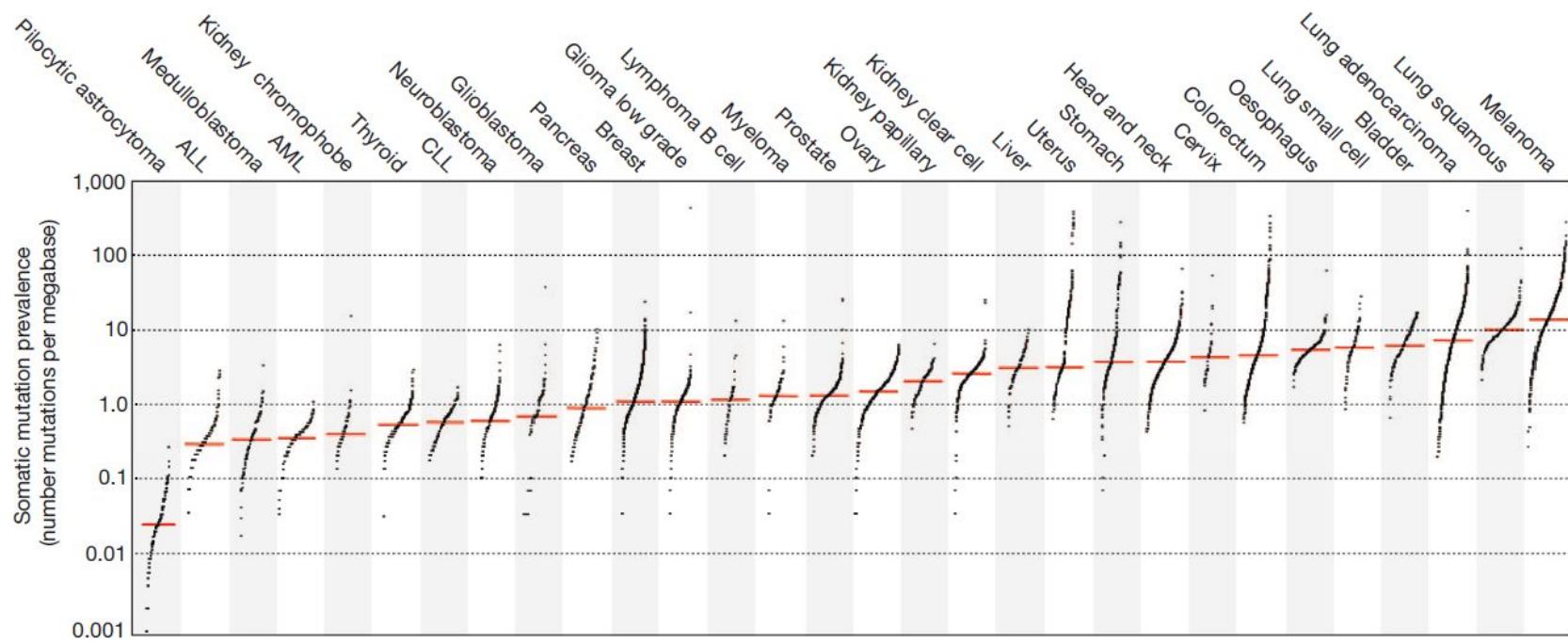


Somatic mutations in cancer

- The burden of somatic mutations is highly variable among different cancer types
- The most mutated cancer types (lung and skin cancers) are associated with well-known environmental mutagens (tobacco smoking and UV light exposure, respectively)

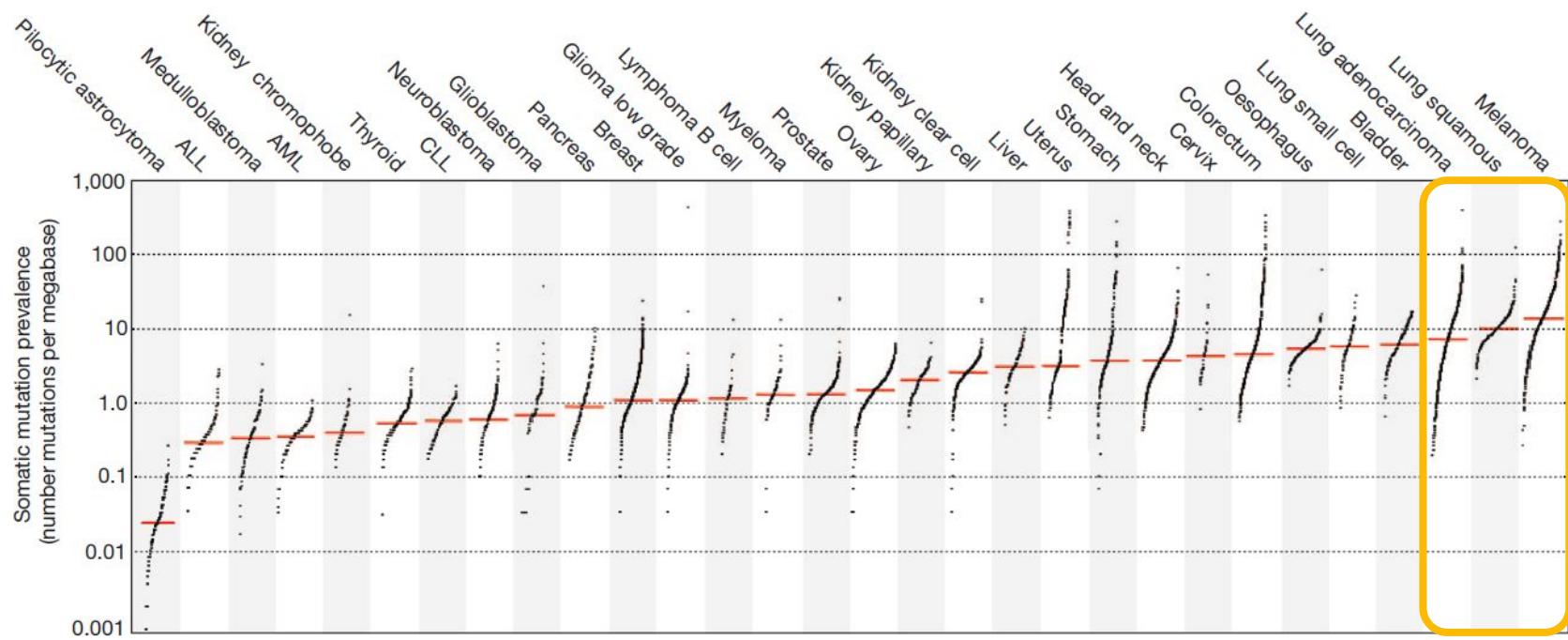
Somatic mutations in cancer

- The burden of somatic mutations is highly variable among different cancer types
- The most mutated cancer types (lung and skin cancers) are associated with well-known environmental mutagens (tobacco smoking and UV light exposure, respectively)



Somatic mutations in cancer

- The burden of somatic mutations is highly variable among different cancer types
- The most mutated cancer types (lung and skin cancers) are associated with well-known environmental mutagens (tobacco smoking and UV light exposure, respectively)



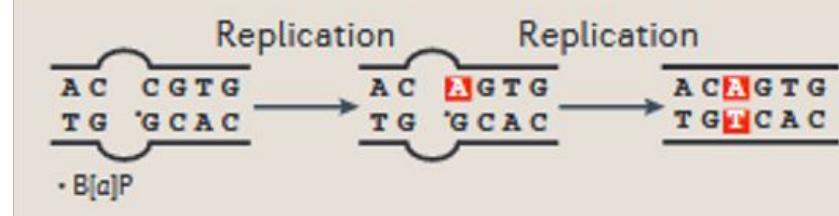
Mutational processes

- Cancer genomes accumulate a large number of somatic mutations resulting from various endogenous and exogenous causes, including normal DNA damage and repair, cancer-related aberrations of the DNA maintenance machinery, and mutations triggered by carcinogenic exposures

Mutational processes

- Cancer genomes accumulate a large number of somatic mutations resulting from various endogenous and exogenous causes, including normal DNA damage and repair, cancer-related aberrations of the DNA maintenance machinery, and mutations triggered by carcinogenic exposures

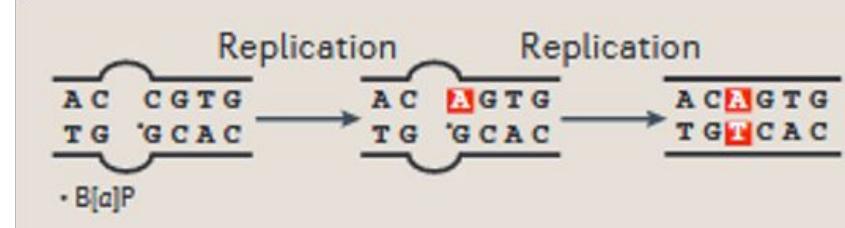
Environmental exposures Tobacco smoking or chewing



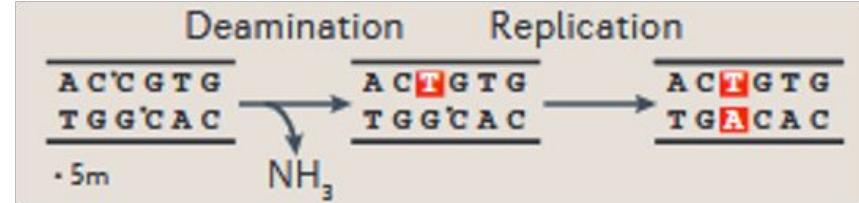
Mutational processes

- Cancer genomes accumulate a large number of somatic mutations resulting from various endogenous and exogenous causes, including normal DNA damage and repair, cancer-related aberrations of the DNA maintenance machinery, and mutations triggered by carcinogenic exposures

Environmental exposures
Tobacco smoking or chewing



Normal cellular activities
Spontaneous deamination of methylated cytosines

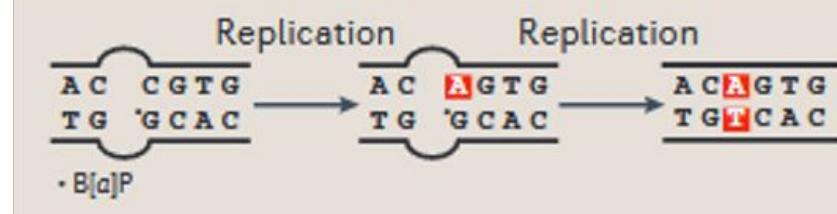


Mutational processes

- Cancer genomes accumulate a large number of somatic mutations resulting from various endogenous and exogenous causes, including normal DNA damage and repair, cancer-related aberrations of the DNA maintenance machinery, and mutations triggered by carcinogenic exposures

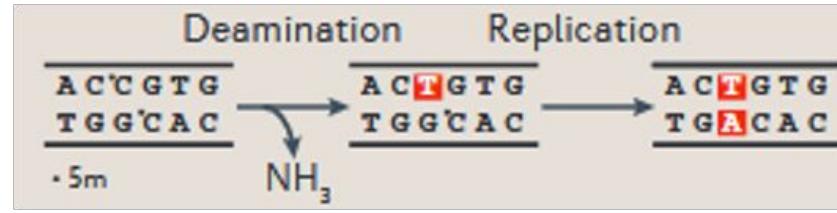
Environmental exposures

Tobacco smoking or chewing



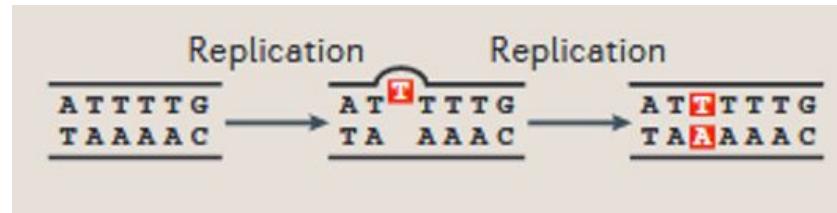
Normal cellular activities

Spontaneous deamination of methylated cytosines



Failure in DNA replication or repair

Aberrant mismatch repair pathway



Classification of base substitution mutations

C>T

C>A

C>G

T>A

T>C

T>G

6 mutation classes

Classification of base substitution mutations

.....ATCGGGAAAT**C**GGACCCGATG.....

.....ATCGGGAAAT**T**GGACCCGATG.....

Classification of base substitution mutations

.....ATCGGGAA**TCG**GACCCGATG.....

.....ATCGGGAA**TTG**GACCCGATG.....

Classification of base substitution mutations

.....ATCGGGAA**TCG**GACCCGATG.....

.....ATCGGGAA**TTG**GACCCGATG.....

.....ATCGGGAA**ACG**GACCCGATG.....

.....ATCGGGAA**ATG**GACCCGATG.....

Classification of base substitution mutations

.....ATCGGGAA**TCG**GACCCGATG.....

.....ATCGGGAA**TTG**GACCCGATG.....

.....ATCGGGAA**ACG**GACCCGATG.....

.....ATCGGGAA**ATG**GACCCGATG.....

.....ATCGGGAA**ACC**GACCCGATG.....

.....ATCGGGAA**ATC**GACCCGATG.....

Classification of base substitution mutations

6 mutation classes

C>T
C>A
C>G
T>A
T>C
T>G

ACA>ATA
ACC>ATC
ACG>ATG
ACT>ATT
CCA>CTA
CCC>CTC
CCG>CTG
CCT>CTT
GCA>GTA
GCC>GTC
GCG>GTG
GCT>GTT
TCA>TTA
TCC>TTC
TCG>TTG
TCT>TTT

Classification of base substitution mutations

6 mutation classes

C>T

C>A

C>G

T>A

T>C

T>G

A CA > A TA	ATA > AAA
A CC > A TC	ATC > AAC
A CG > A TG	ATG > AAG
A CT > A TT	ATT > AAT
C CA > C TA	CTA > CAA
C CC > C TC	CTC > CAC
C CG > C TG	CTG > CAG
C CT > C TT	CTT > CAT
G CA > G TA	GTA > GAA
G CC > G TC	GTC > GAC
G CG > G TG	GTC > GAG
G CT > G TT	GTT > GAT
T CA > T TA	TTA > TAA
T CC > T TC	TTC > TAC
T CG > T TG	TTC > TAG
T CT > T TT	TTT > TAT
A CA > A AA	ATA > ACA
A CC > A AC	ATC > ACC
A CG > A AG	ATG > ACG
A CT > A AT	ATT > ACT
C CA > C AA	CTA > CCA
C CC > C AC	CTC > CCC
C CG > C AG	CTG > CCG
C CT > C AT	CTT > CCT
G CA > G AA	GTA > GCA
G CC > G AC	GTC > GCC
G CG > G AG	GTC > GCG
G CT > G AT	GTT > GCT
T CA > T AA	TTA > TCA
T CC > T AC	TTC > TCC
T CG > T AG	TTC > TCG
T CT > T AT	TTT > TCT
A CA > A GA	ATA > AGA
A CC > A GC	ATC > AGC
A CG > A GG	ATG > AGG
A CT > A GT	ATT > AGT
C CA > C GA	CTA > CGA
C CC > C GC	CTC > CGC
C CG > C GG	CTG > CGG
C CT > C GT	CTT > CGT
G CA > G GA	GTA > GGA
G CC > G GC	GTC > GGC
G CG > G GG	GTC > GGG
G CT > G GT	GTT > GGT
T CA > T GA	TTA > TGA
T CC > T GC	TTC > TGC
T CG > T GG	TTC > TGG
T CT > T GT	TTT > TGT

96 mutation classes

Classification of base substitution mutations

- Six classes of single-base mutations
 - Reported by pyrimidine

Classification of base substitution mutations

- Six classes of single-base mutations
 - Reported by pyrimidine



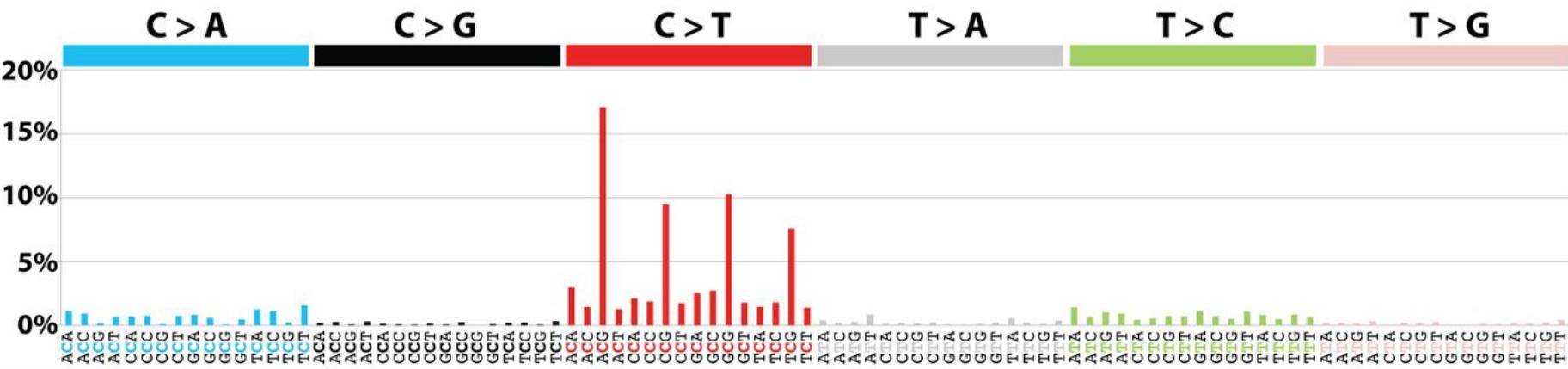
Classification of base substitution mutations

- Six classes of single-base mutations
 - Reported by pyrimidine
- Adding 5' and 3' adjacent bases
 - 96 possibilities considering context



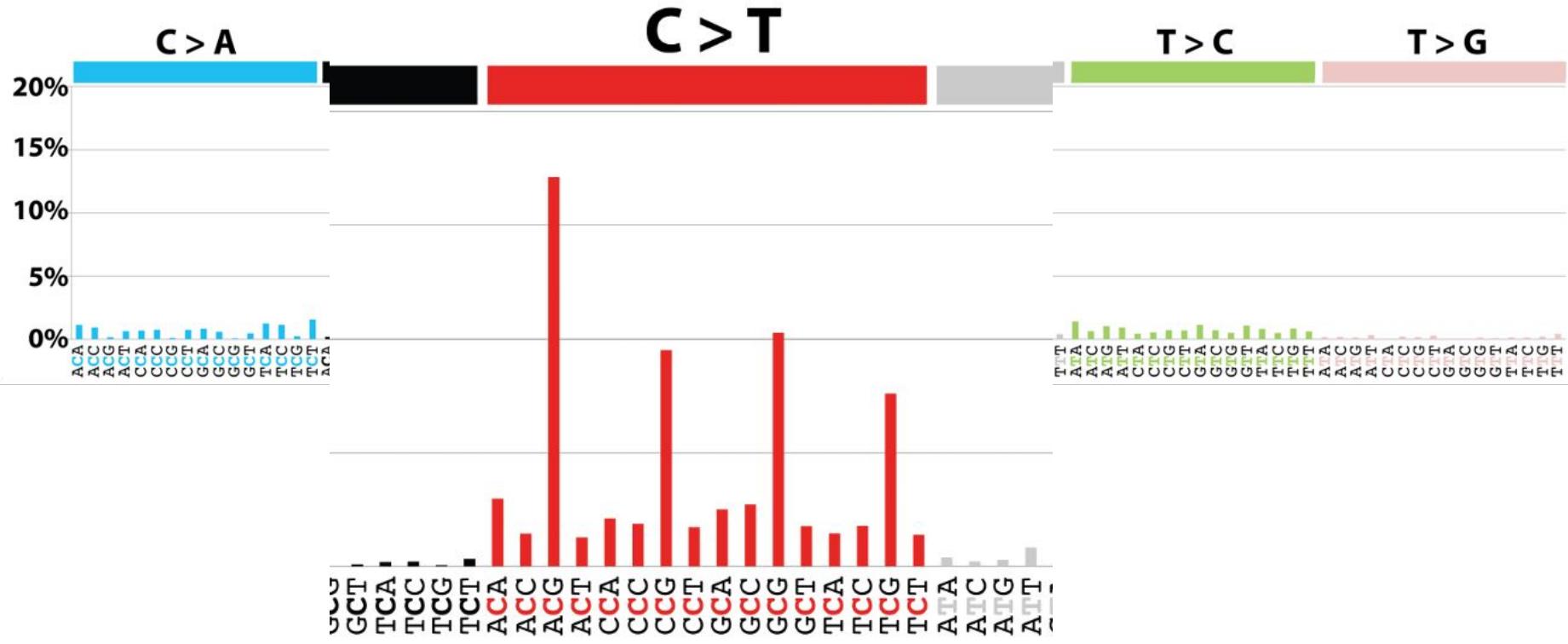
Classification of base substitution mutations

- Six classes of single-base mutations
 - Reported by pyrimidine
- Adding 5' and 3' adjacent bases
 - 96 possibilities considering context



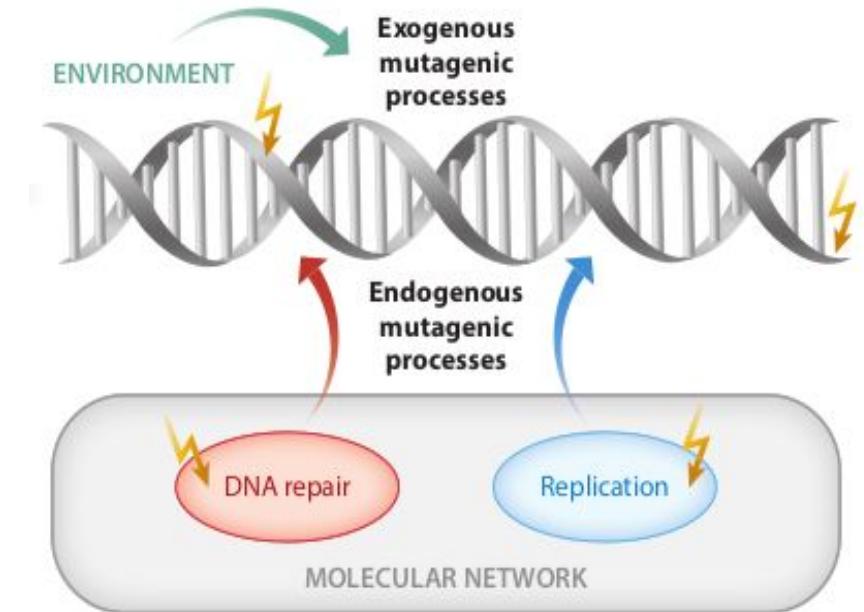
Classification of base substitution mutations

- Six classes of single-base mutations
 - Reported by pyrimidine
- Adding 5' and 3' adjacent bases
 - 96 possibilities considering context



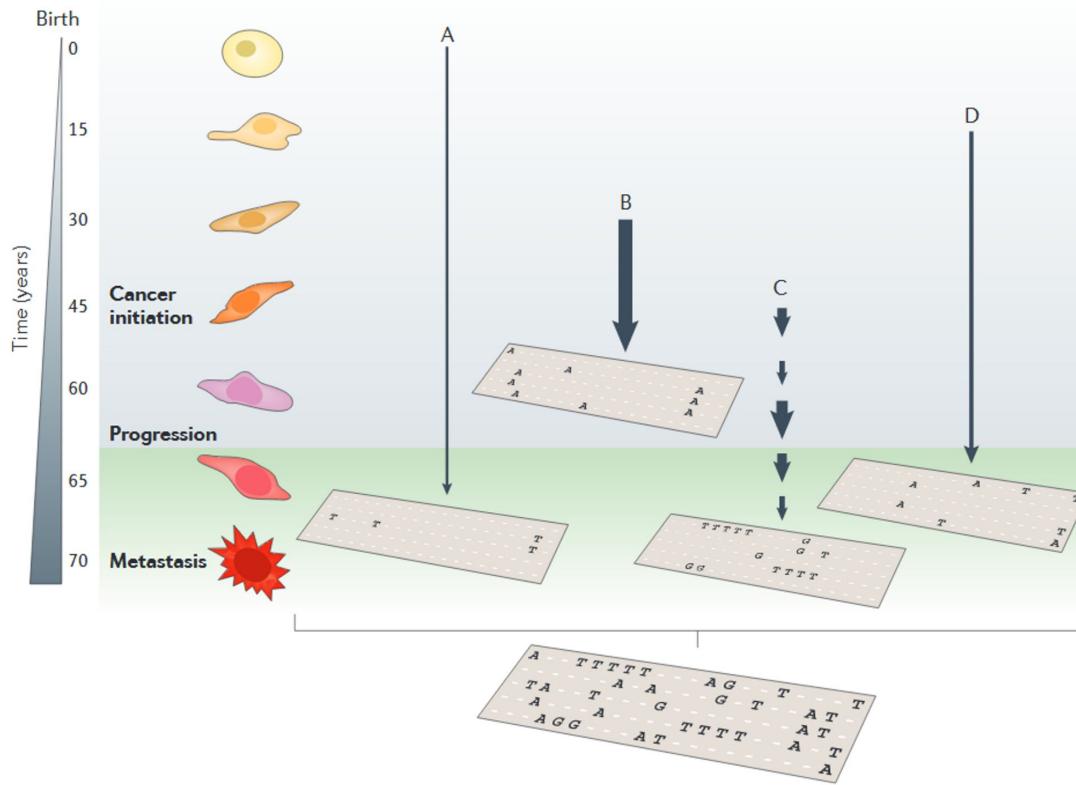
Mutational signatures

- DNA mutations can arise due to exogenous (environmental) or endogenous (caused by the natural stochasticity of biological processes and the dysregulation of specific molecular pathways) factors
- The imprint of a particular mutational process in the genome is a **mutational signature**



Mutational signatures

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



Mutational signatures

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

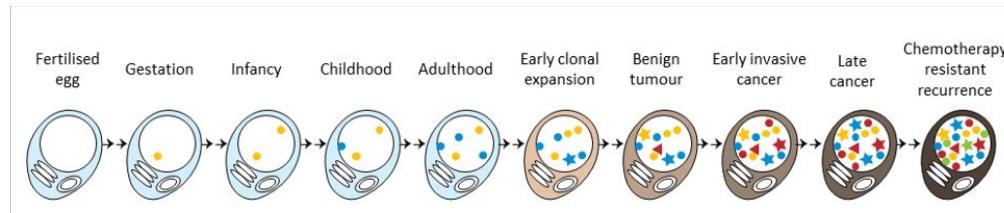
Mutational signatures

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



Mutational signatures

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



Mutational signatures

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



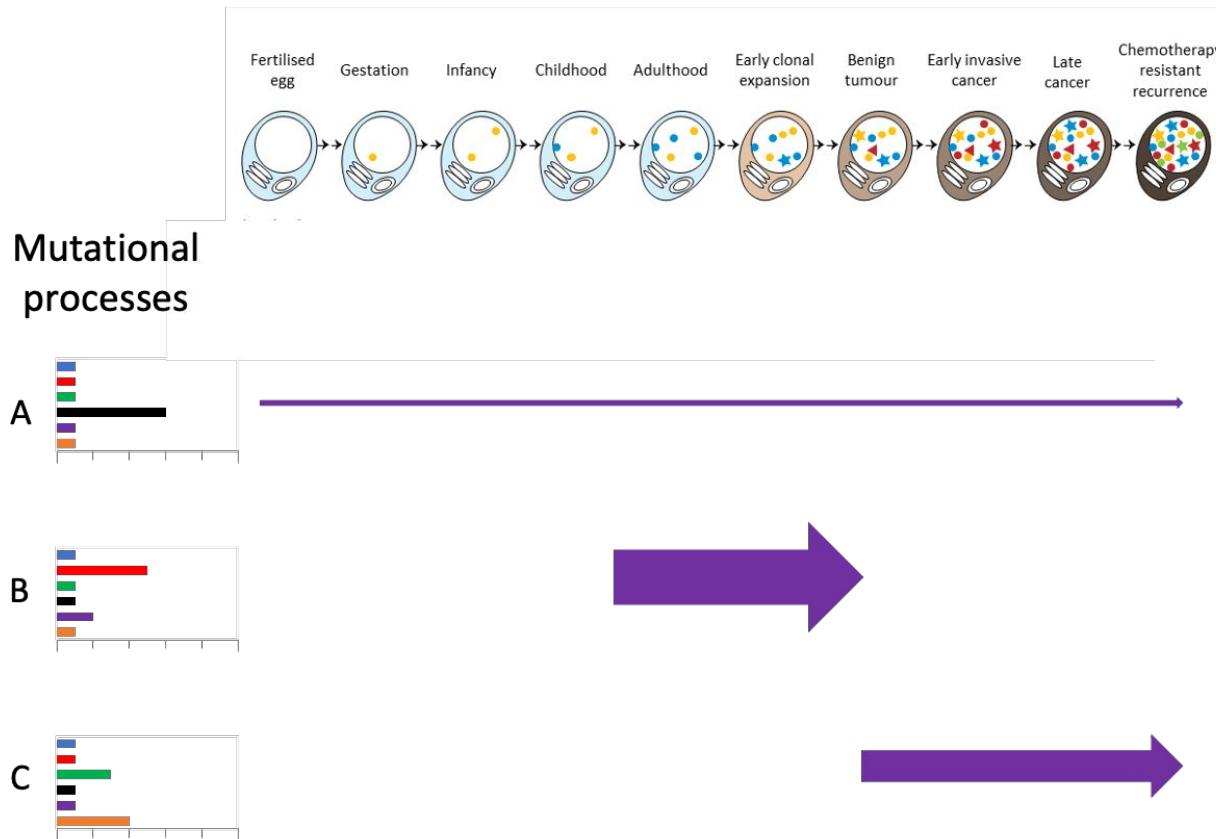
Mutational signatures

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



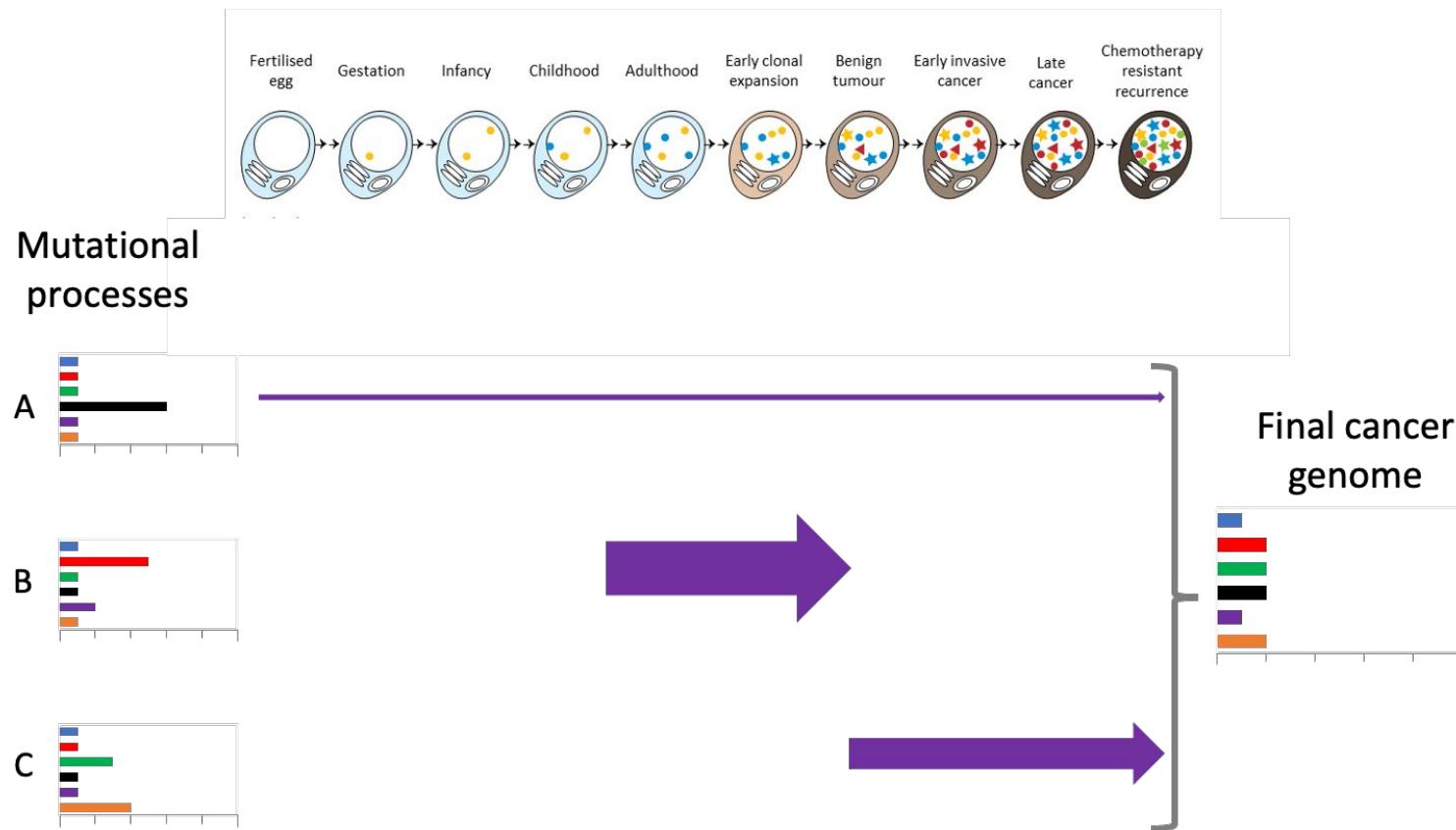
Mutational signatures

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



Mutational signatures

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

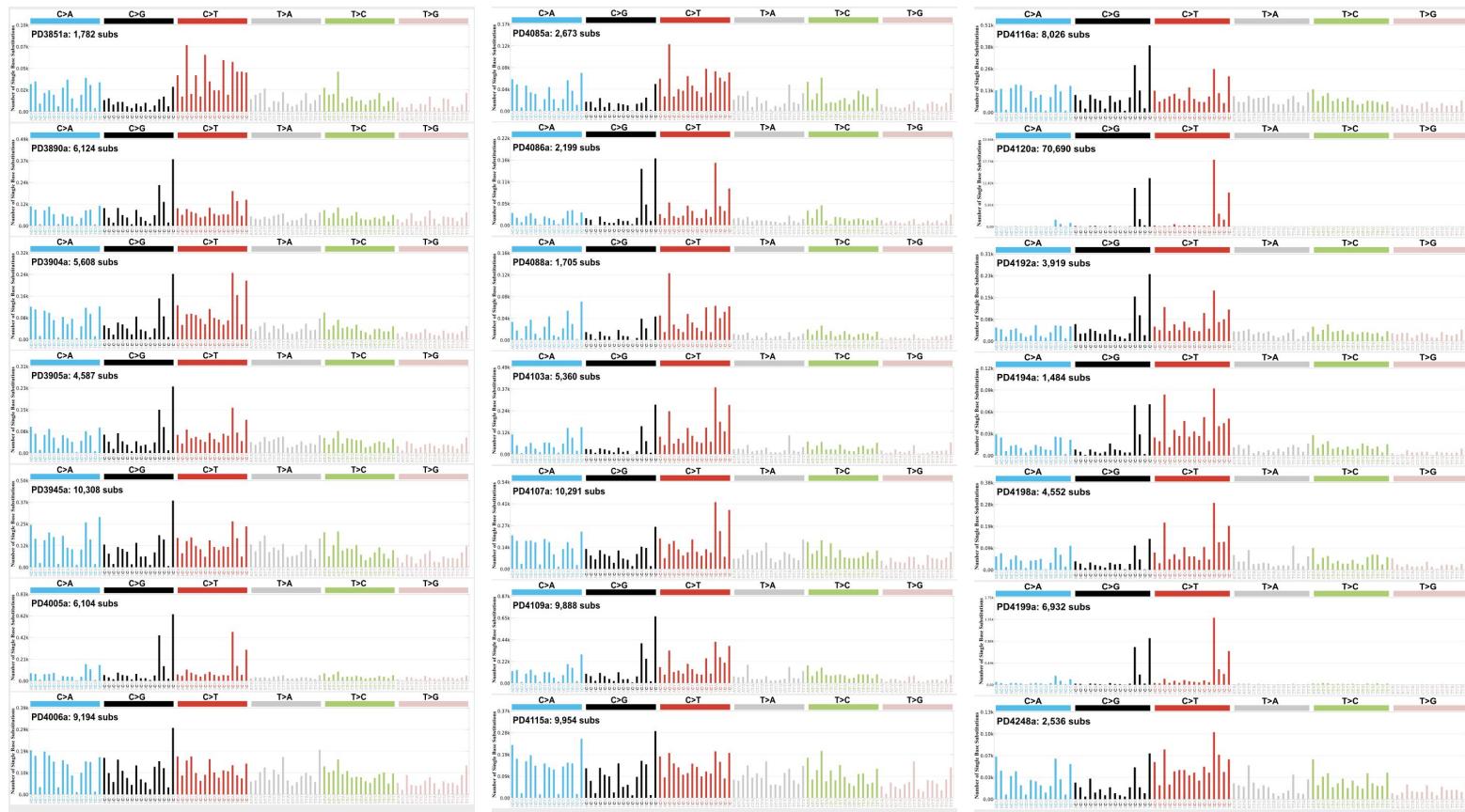


Identifying mutational signatures - visual example

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

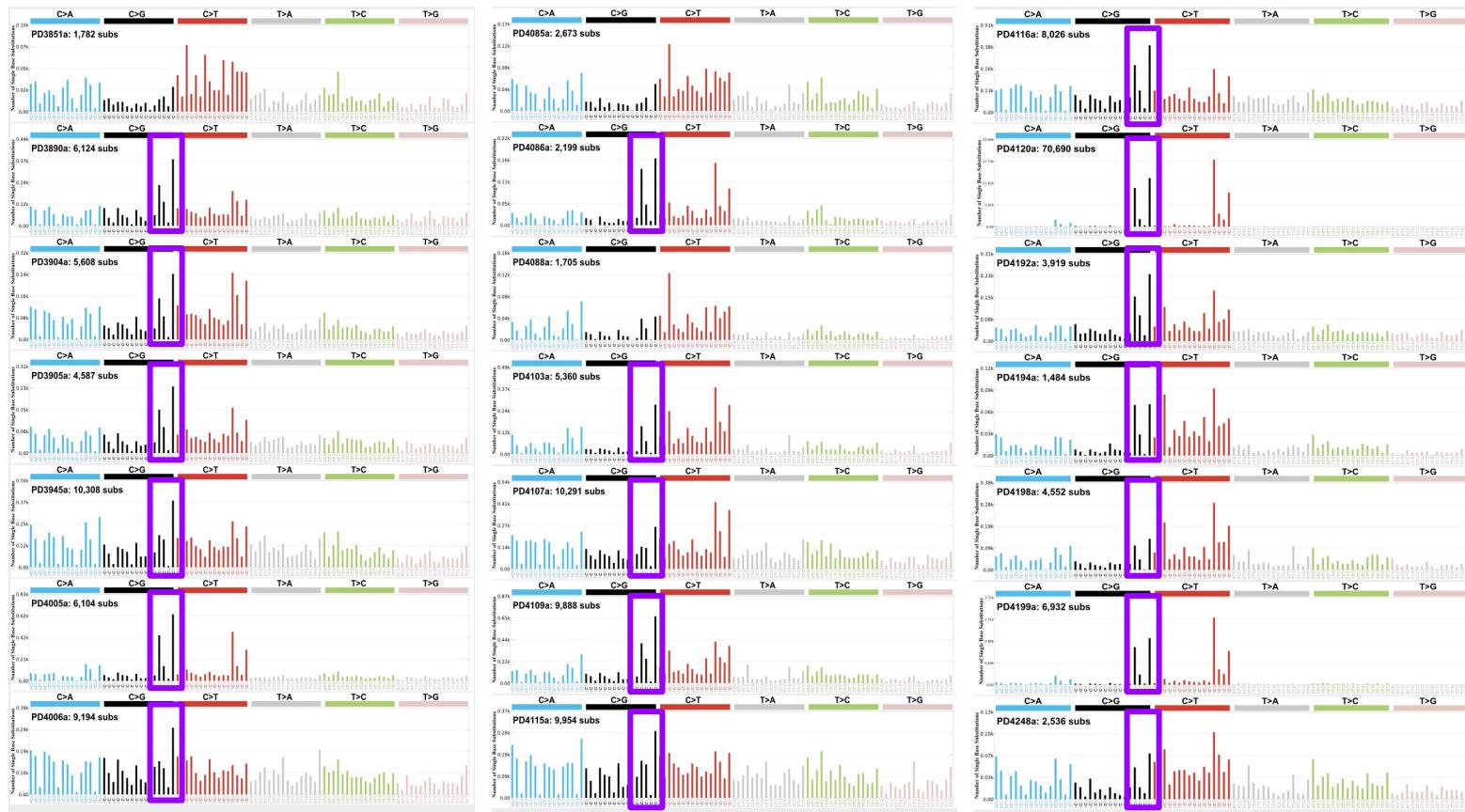
Identifying mutational signatures - visual example

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



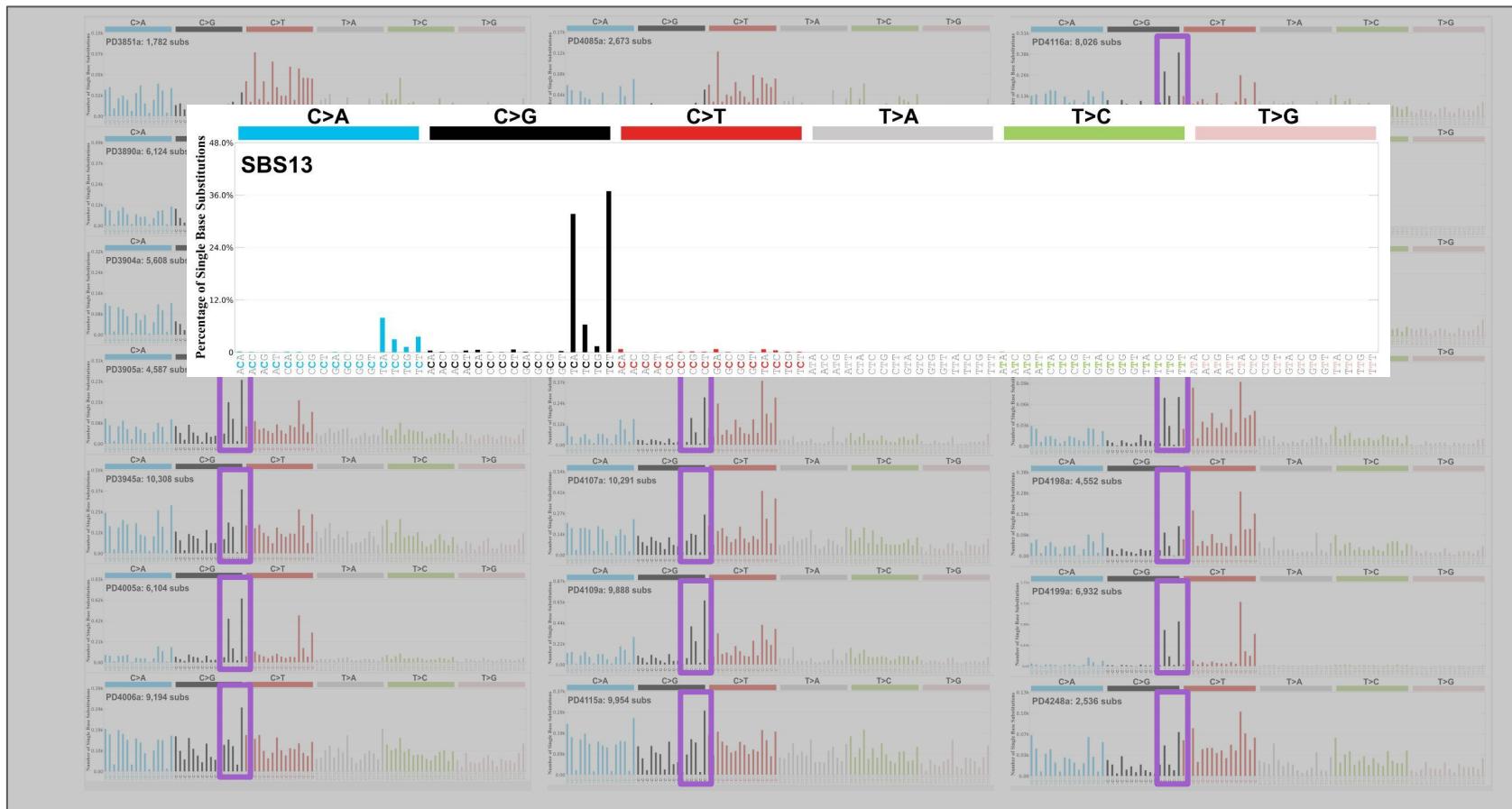
Identifying mutational signatures - visual example

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



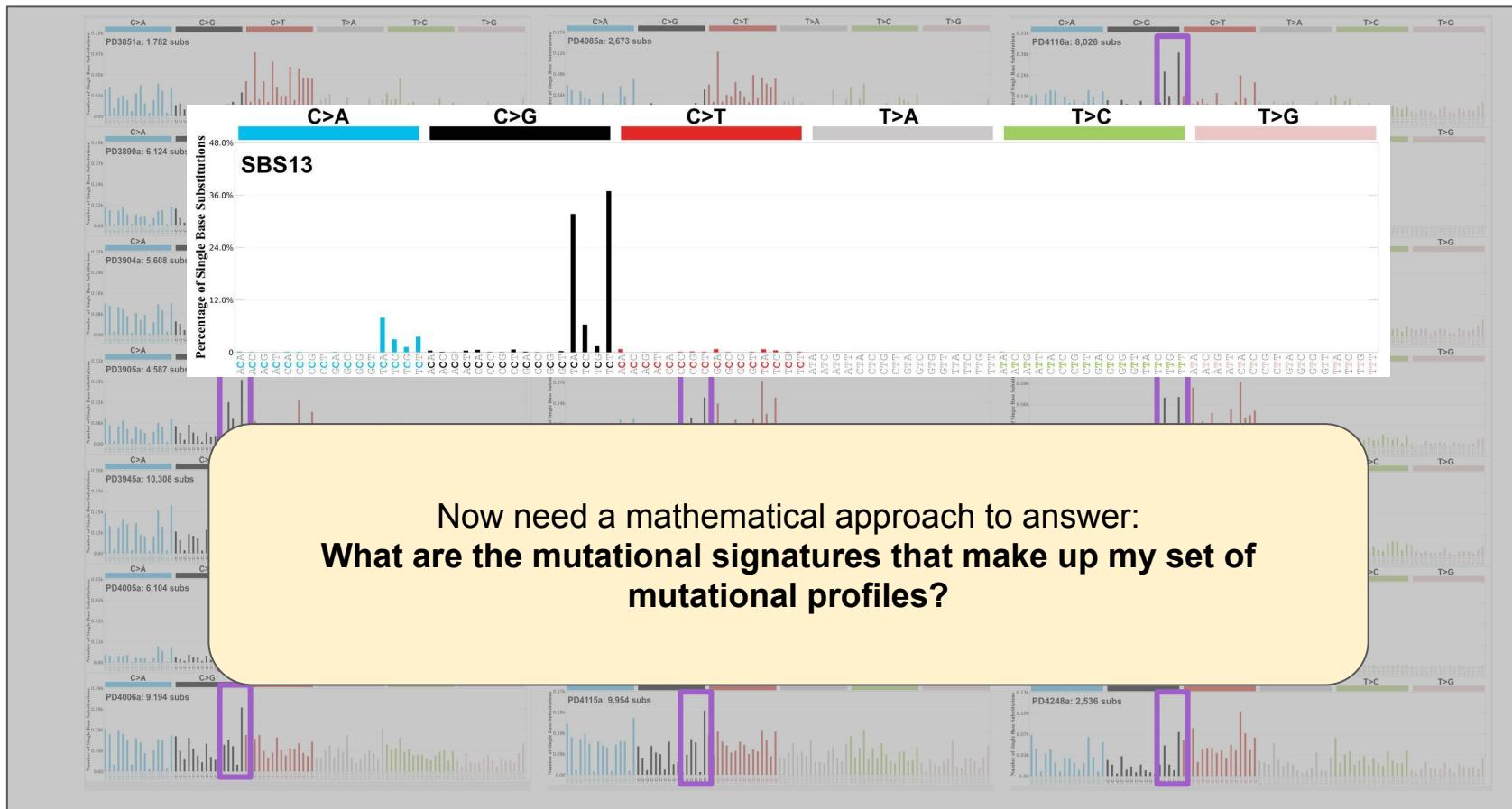
Identifying mutational signatures - visual example

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



Identifying mutational signatures - visual example

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



Computational identification of mutational signatures

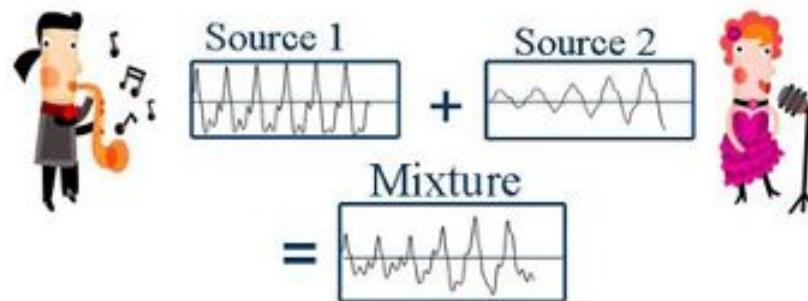
- Mutational signatures can be determined based on mutational profiles across a set of individuals

Computational identification of mutational signatures

- Mutational signatures can be determined based on mutational profiles across a set of individuals
- Mathematical models allows the *un-mixing* and extraction of mutational signatures by solving a **blind source separation problem**

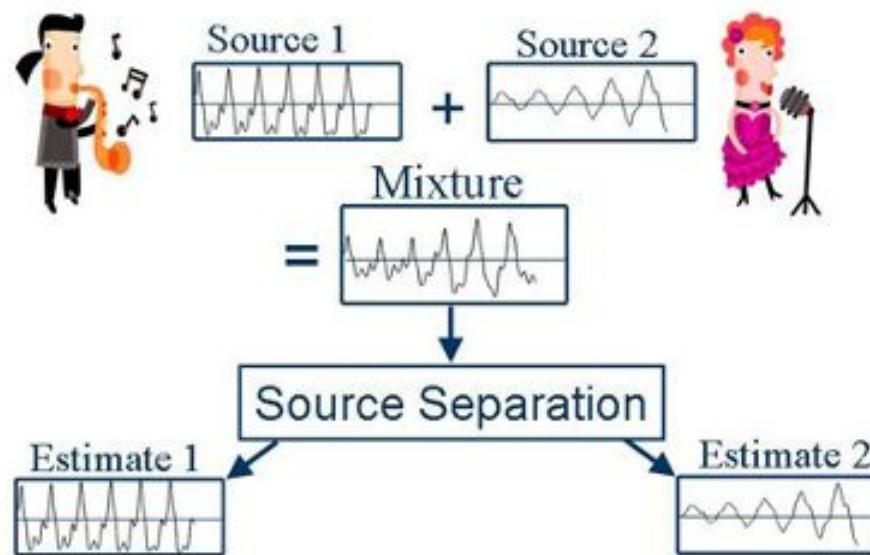
Computational identification of mutational signatures

- Mutational signatures can be determined based on mutational profiles across a set of individuals
- Mathematical models allows the *un-mixing* and extraction of mutational signatures by solving a **blind source separation problem**



Computational identification of mutational signatures

- Mutational signatures can be determined based on mutational profiles across a set of individuals
- Mathematical models allows the *un-mixing* and extraction of mutational signatures by solving a **blind source separation problem**



Computational identification of mutational signatures

- **Non-negative matrix factorization (NMF)** for solving the blind source separation problem

Computational identification of mutational signatures

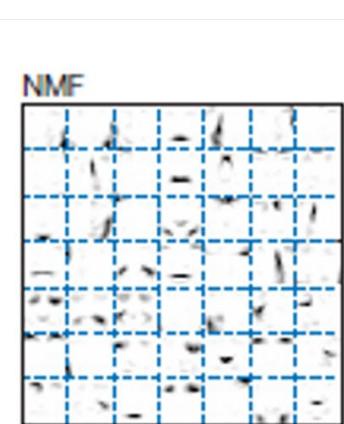
- Non-negative matrix factorization (NMF) for solving the blind source separation problem

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



Computational identification of mutational signatures

- **Non-negative matrix factorization (NMF)** for solving the blind source separation problem
 - Infinite solutions as a matrix can be approximately decompose in two matrices in infinite number of ways
 - BSS problem is usually solved by constraining the solutions
 - Intrinsic nonnegative constraints from our theoretical modal
 - One main hyperparameter, the rank k of the latent matrices S and A , which corresponds to the number of mutational signatures present in the input data (matrix M)

$$M \approx S \times A$$

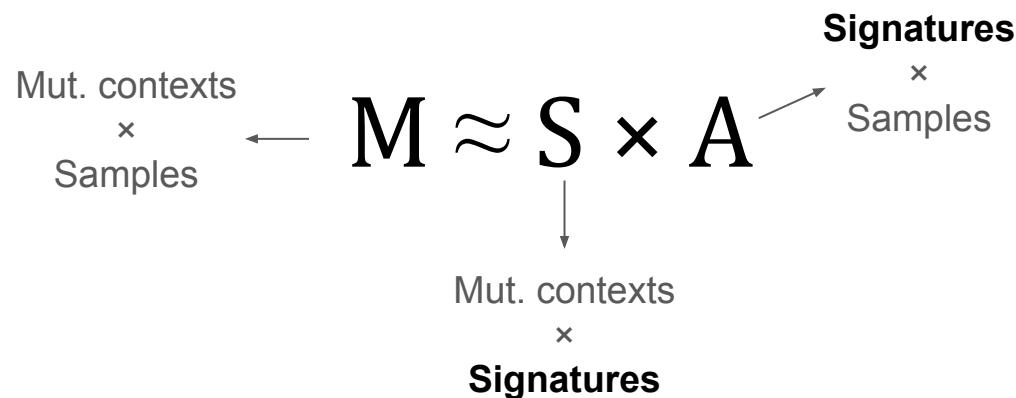
Computational identification of mutational signatures

- **Non-negative matrix factorization (NMF)** for solving the blind source separation problem
 - Infinite solutions as a matrix can be approximately decompose in two matrices in infinite number of ways
 - BSS problem is usually solved by constraining the solutions
 - Intrinsic nonnegative constraints from our theoretical modal
 - One main hyperparameter, the rank k of the latent matrices S and A , which corresponds to the number of mutational signatures present in the input data (matrix M)

$$\begin{array}{c} \text{Mut. contexts} \\ \times \\ \text{Samples} \end{array} \leftarrow M \approx S \times A$$

Computational identification of mutational signatures

- **Non-negative matrix factorization (NMF)** for solving the blind source separation problem
 - Infinite solutions as a matrix can be approximately decompose in two matrices in infinite number of ways
 - BSS problem is usually solved by constraining the solutions
 - Intrinsic nonnegative constraints from our theoretical modal
 - One main hyperparameter, the rank k of the latent matrices S and A , which corresponds to the number of mutational signatures present in the input data (matrix M)



Computational identification of mutational signatures

- There exist a wide variety of NMF methods for the specific application to mutational signatures
- The original and one of the most commonly used methods is SigProfilerExtractor from the Alexandrov lab, which solves the NMF problem by using the stochastic multiplicative update method based on the Frobenius norm
- There are other methods using a Bayesian form of NMF and relying on Automatic Relevance Determination to automatically infer the rank k of the latent matrices

Computational identification of mutational signatures

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

Computational identification of mutational signatures

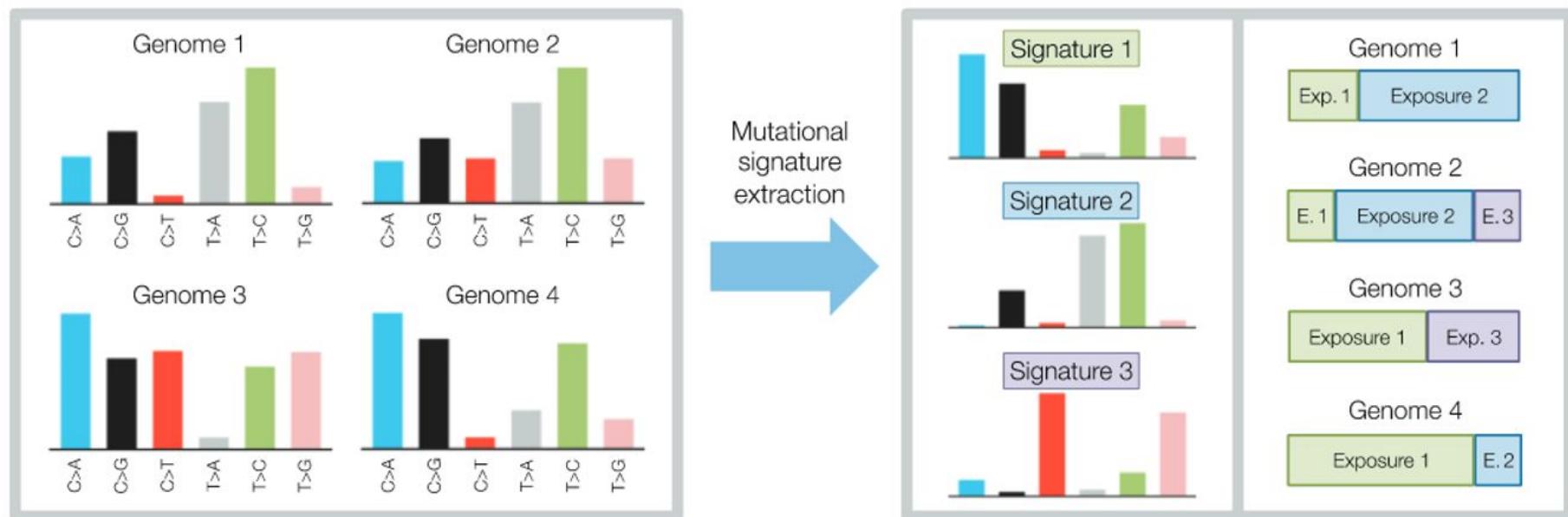
- Mutational signature extraction relies on a large number of samples (and mutations) to get accurate results

Computational identification of mutational signatures

- Mutational signature extraction relies on a large number of samples (and mutations) to get accurate results
- Large international consortia have analyzed thousands of whole genome and whole exome sequenced samples to generate a consensus set of reference mutational signatures (deposited in the COSMIC database)

Computational identification of mutational signatures

- Mutational signature extraction relies on a large number of samples (and mutations) to get accurate results
- Large international consortia have analyzed thousands of whole genome and whole exome sequenced samples to generate a consensus set of reference mutational signatures (deposited in the COSMIC database)

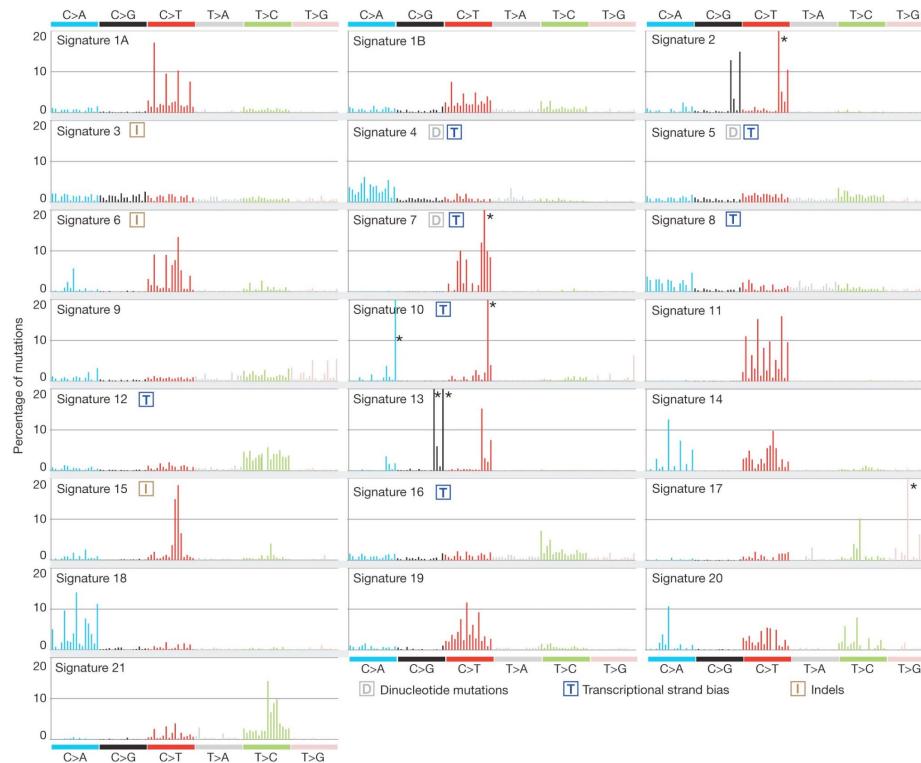


Reference mutational signatures

- The COSMIC database has been growing over the years with the addition of novel samples, and considering different variant classes

Reference mutational signatures

- The COSMIC database has been growing over the years with the addition of novel samples, and considering different variant classes



v1 (August 2013)

- 21 SBS signatures

Reference mutational signatures

- The COSMIC database has been growing over the years with the addition of novel samples, and considering different variant classes



v1 (August 2013)

- 21 SBS signatures

v2 (March 2015)

- 30 SBS signatures

Reference mutational signatures

- The COSMIC database has been growing over the years with the addition of novel samples, and considering different variant classes



v3 (May 2019)

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

Reference mutational signatures

The current set of COSMIC reference signatures is available at <https://cancer.sanger.ac.uk/signatures/>, and encompasses:

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

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 Catalogue Of Somatic Mutations In Cancer**

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.

SSS 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, other processes get activated and the mutational spectrum of the cancer genome continues to change.

Time

Number of mutations

Signature activity

Mutational spectrum of final cancer genome

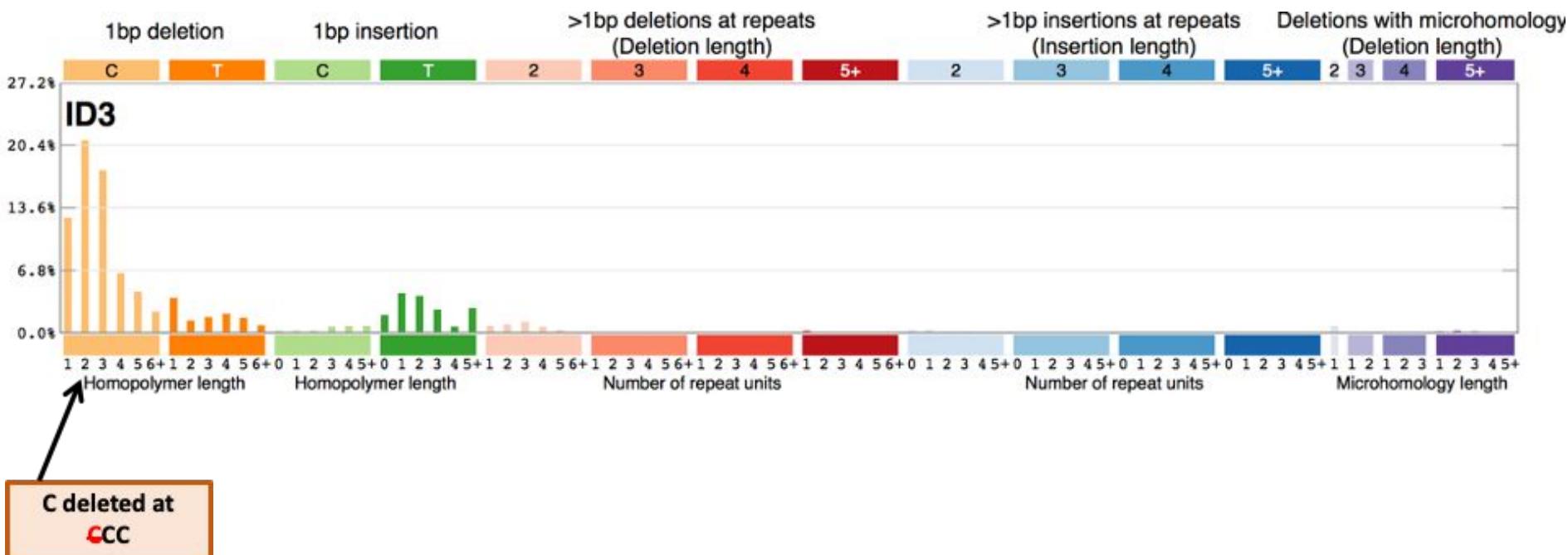
ID mutational signatures

- Also known as indels, ID are defined as the incorporation or loss of small fragments of DNA (usually between 1 and 50 base pairs) in a specific genomic location
- Although there is no single intuitive and naturally constrained set of ID mutation types (as there arguably are for single base substitutions and doublet base substitutions), a compilation of **83** different types considering size, nucleotides affected and presence on repetitive and/or microhomology regions was used to extract mutational signatures
- More details can be found here:
https://cancer.sanger.ac.uk/signatures/documents/4/PCAWG7_indel_classification_2021_08_31.xlsx

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

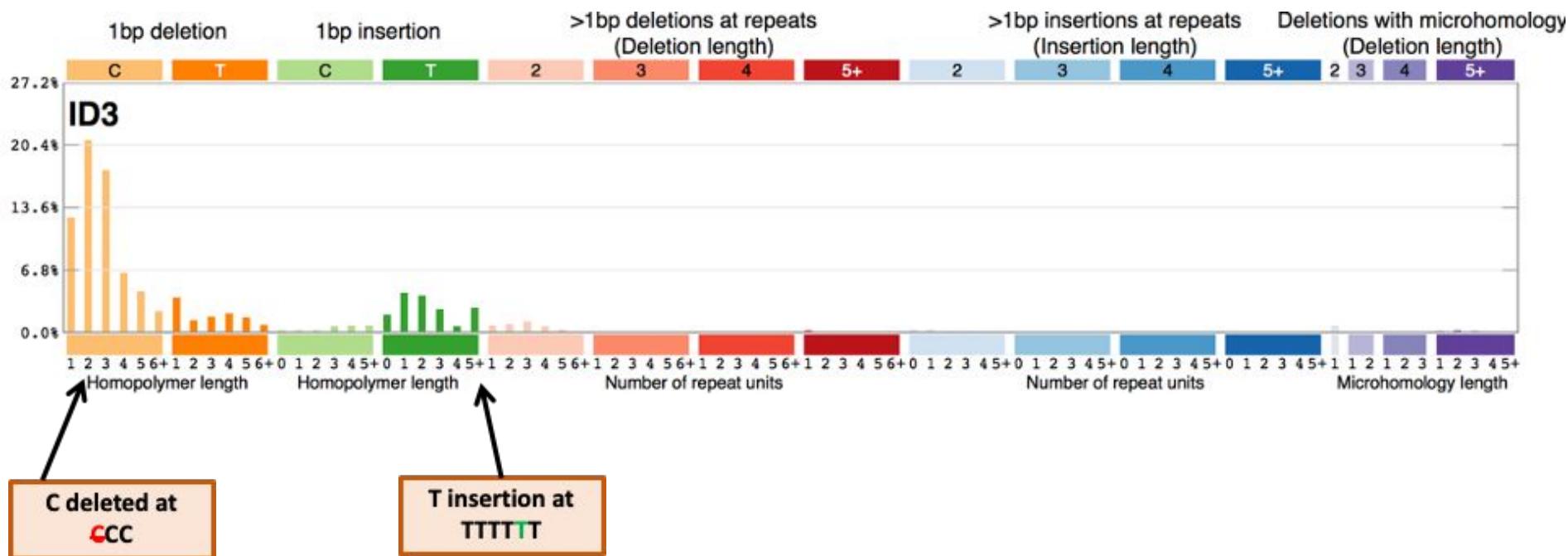
ID mutational signatures

- Example pattern of an indel (ID) mutational signature



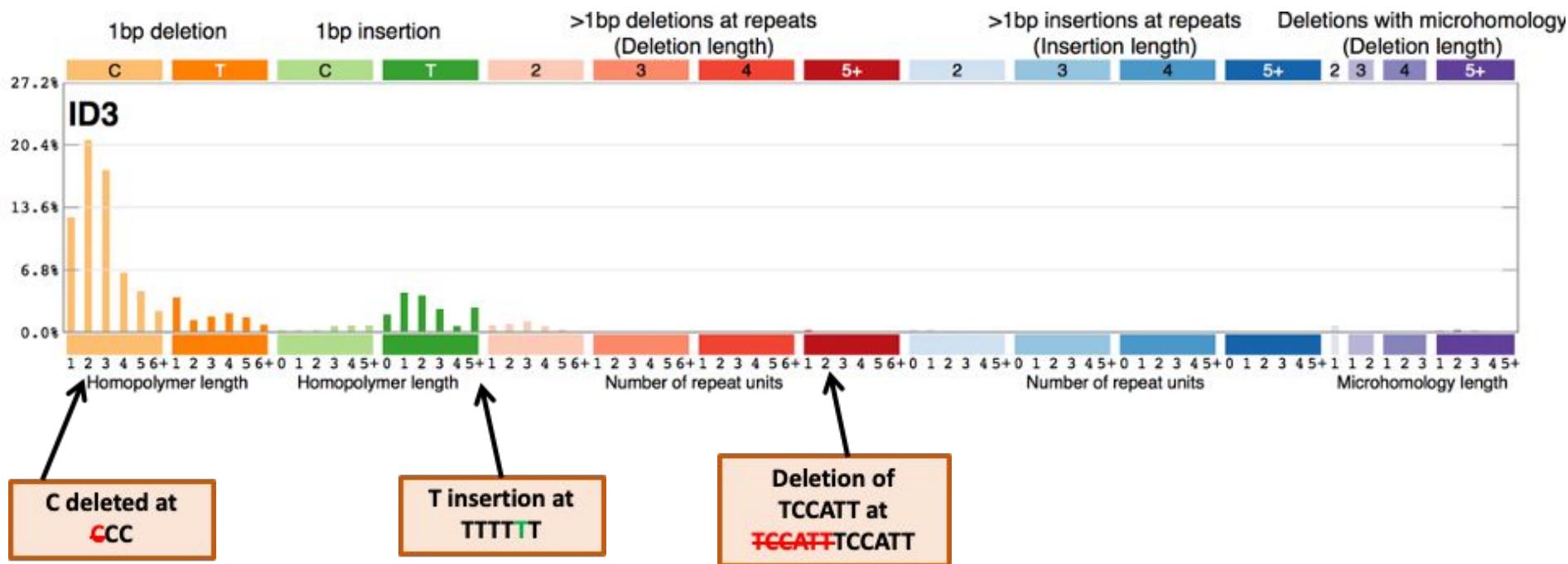
ID mutational signatures

- Example pattern of an indel (ID) mutational signature



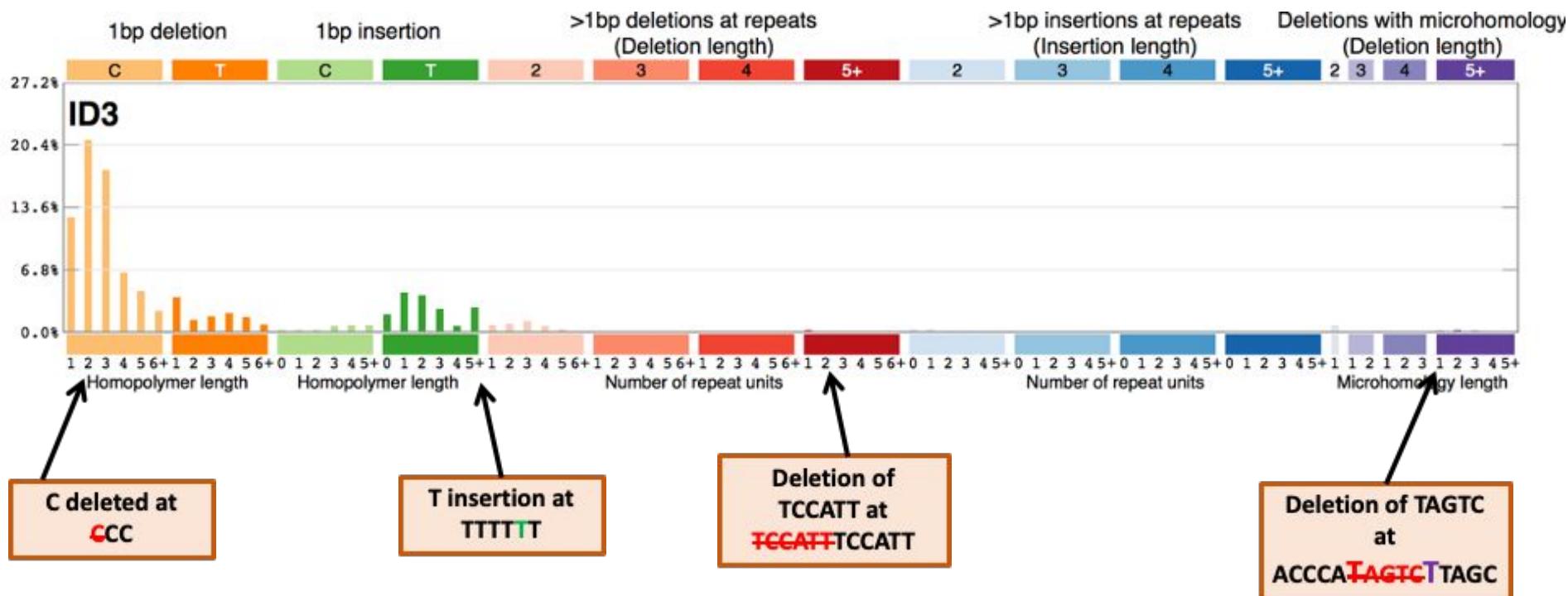
ID mutational signatures

- Example pattern of an indel (ID) mutational signature



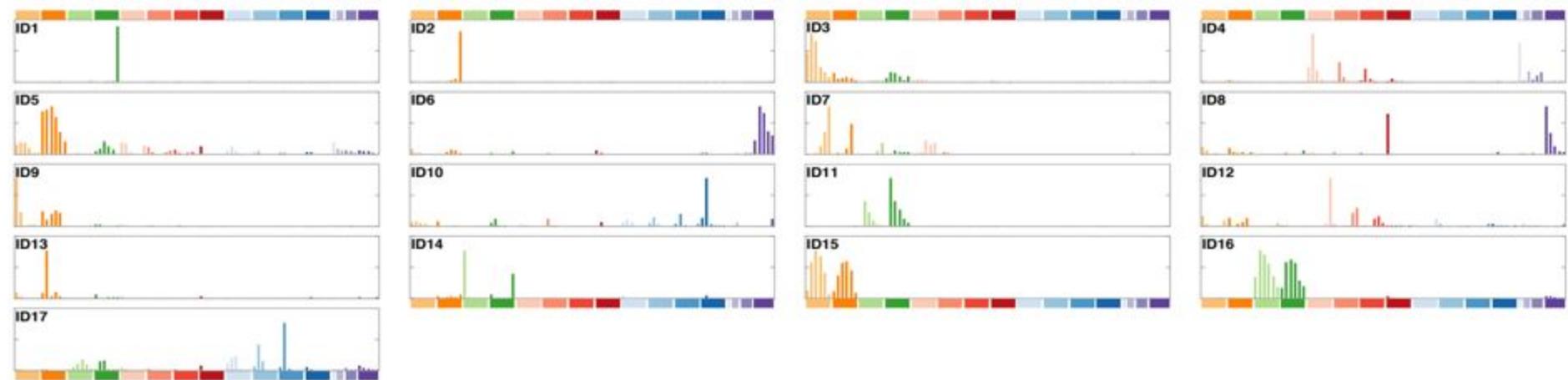
ID mutational signatures

- Example pattern of an indel (ID) mutational signature



ID mutational signatures

- Reference ID mutational signatures from COSMIC



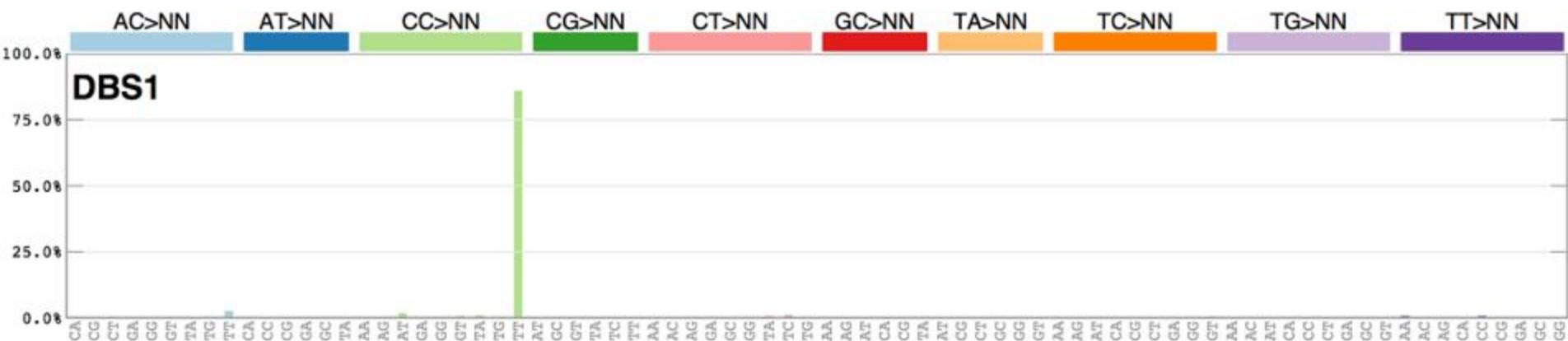
DBS mutational signatures

- DBS are generated after the concurrent modification of two consecutive nucleotide bases.
- There are 78 strand-agnostic DBS mutation types
- More specifically, there are 16 possible source doublet bases (4×4)
- Of these, AT, TA, CG, and GC are their own reverse complement
- The remaining 12 can be represented as 6 possible strand-agnostic doublets
- Thus, there are $4+6=10$ source doublet bases
- Because they are their own reverse complements, AT, TA, CG, and GC can each be substituted by only 6 doublets
- For the remaining doublets, there are 9 possible DBS mutation types (3×3)
- Therefore, in total there are $4 \times 6 + 6 \times 9 = 78$ strand-agnostic DBS mutation types.

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

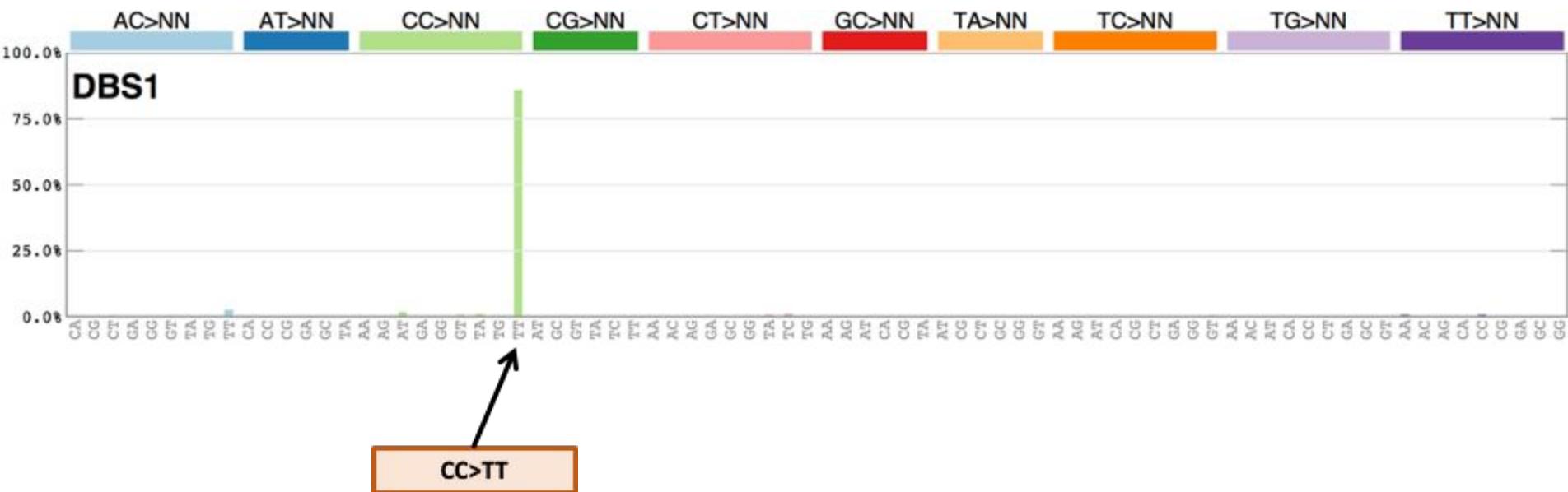
DBS mutational signatures

- Example pattern of a DBS mutational signature



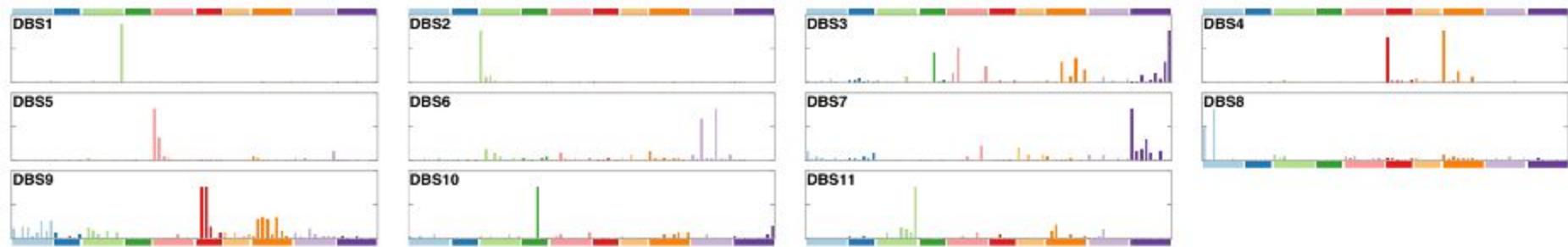
DBS mutational signatures

- Example pattern of a DBS mutational signature



DBS mutational signatures

- Reference DBS mutational signatures from COSMIC



CN mutational signatures

- Copy number variants are defined using the 48-channel copy number classification scheme
- To categorise segments from allele-specific copy number profiles (as major copy number and minor copy number respectively i.e. non-phased profiles) the scheme incorporates:
 - loss-of-heterozygosity status
 - total copy number state
 - segment length
- The signatures displayed in COSMIC were identified from 9,873 tumour copy number profiles obtained from The Cancer Genome Atlas (TCGA) SNP6 array data spanning 33 cancer types and called using ASCAT.

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

Reference mutational signatures

- Other reference databases exist for different variant classes

Reference mutational signatures

- Other reference databases exist for different variant classes

MUTAGENE

MUTAGENE [Home](#) [Explore](#) [Compare](#) [Identify](#) [Analyze gene](#) [Help](#) [Contact](#)

Explore context-dependent mutational profiles and signatures

[Mutational profiles](#) [Mutational signatures](#)

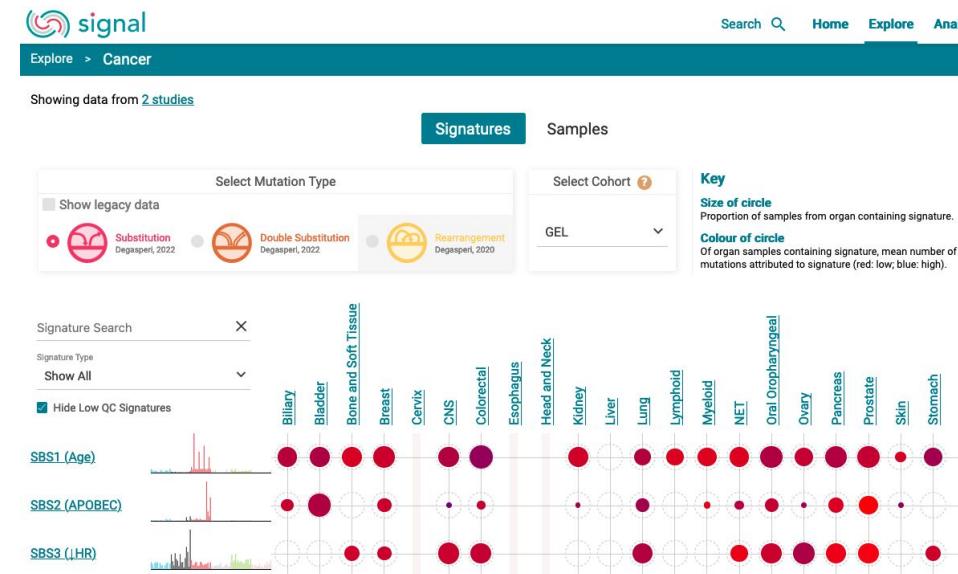
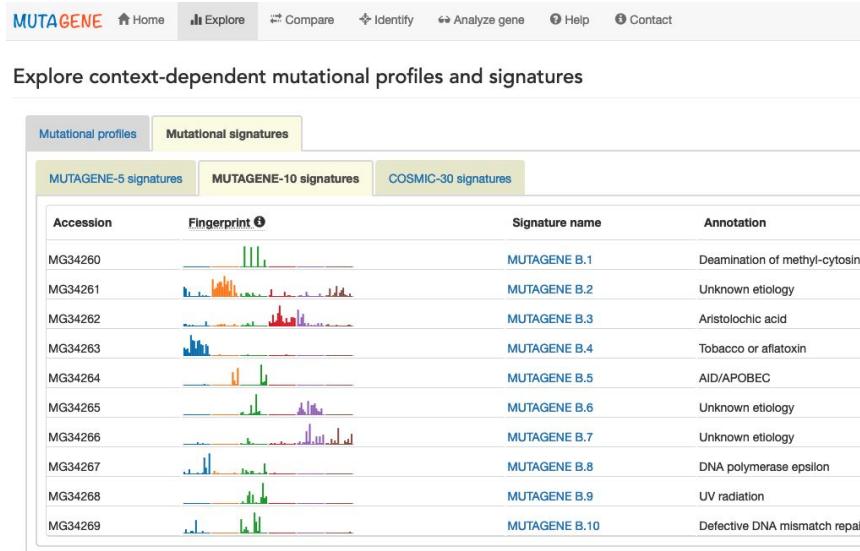
[MUTAGENE-5 signatures](#) [MUTAGENE-10 signatures](#) [COSMIC-30 signatures](#)

Accession	Fingerprint	Signature name	Annotation
MG34260		MUTAGENE B.1	Deamination of methyl-cytosine
MG34261		MUTAGENE B.2	Unknown etiology
MG34262		MUTAGENE B.3	Aristolochic acid
MG34263		MUTAGENE B.4	Tobacco or aflatoxin
MG34264		MUTAGENE B.5	AID/APOBEC
MG34265		MUTAGENE B.6	Unknown etiology
MG34266		MUTAGENE B.7	Unknown etiology
MG34267		MUTAGENE B.8	DNA polymerase epsilon
MG34268		MUTAGENE B.9	UV radiation
MG34269		MUTAGENE B.10	Defective DNA mismatch repair

Reference mutational signatures

- Other reference databases exist for different variant classes

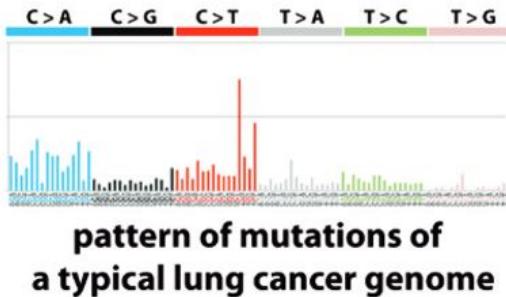
MUTAGENE



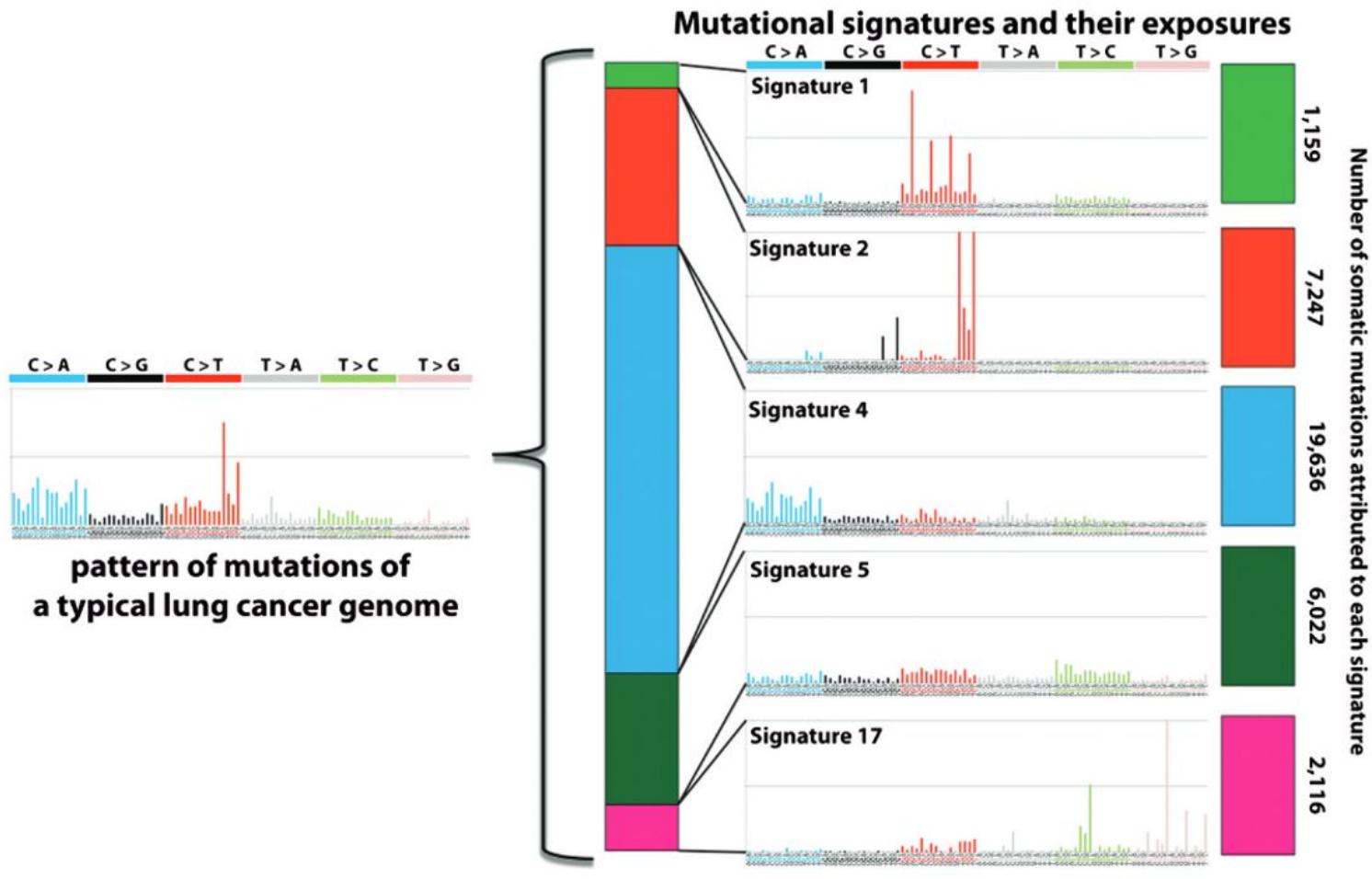
Refitting mutational signatures

- For mutational signature refitting, the set of mutational signatures is given (matrix S) apart from the input mutational matrix (matrix M), and the goal is to infer the activities or exposures of each signature in each sample (matrix A)
- Most methods are based on the non-negative least squares algorithm
- The signature matrix can consist of either the full set of COSMIC signatures, a subset thereof, or signatures extracted from a specific cancer cohort using a *de novo* method
- The refitting methods are especially useful when the analyzed set of mutations is too small for *de novo* signature extraction, for example, in the case of small sample size, targeted sequencing panels, or samples with few mutations such as in healthy tissues or in slowly growing tumors
- Also, refitting allows extending the applicability of validated mutational signatures in small targeted studies and even in clinical settings for individual patients

Refitting mutational signatures

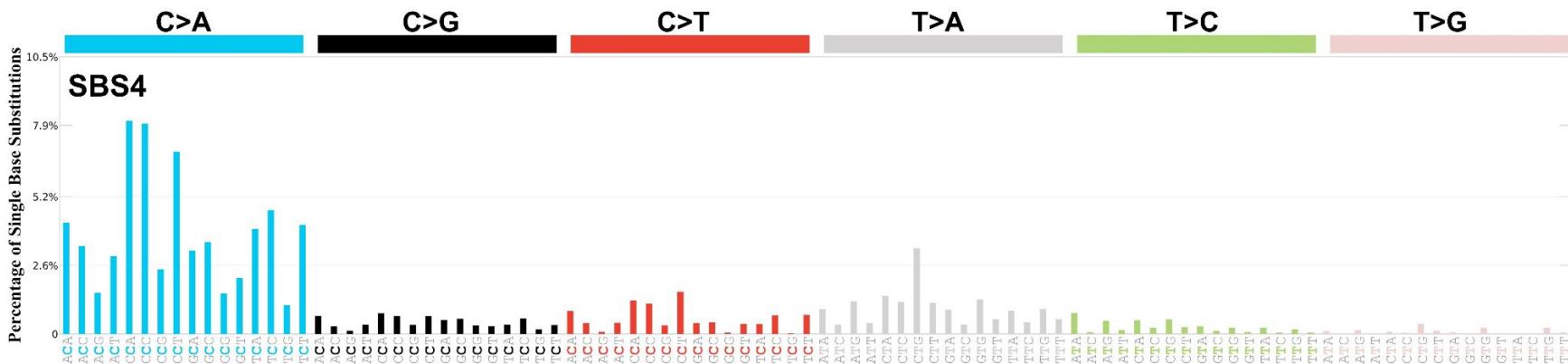


Refitting mutational signatures



Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking



Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?

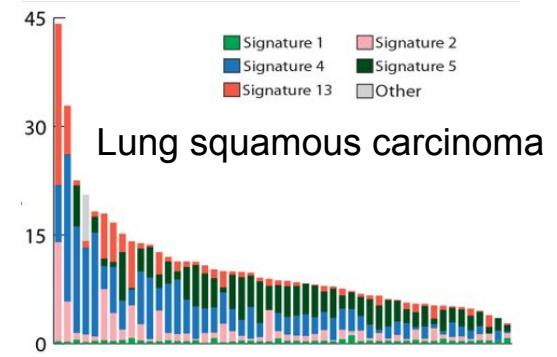
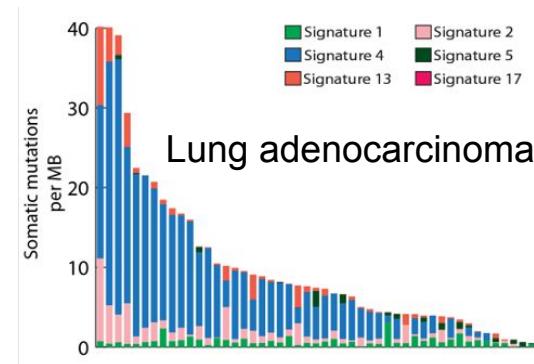
Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types

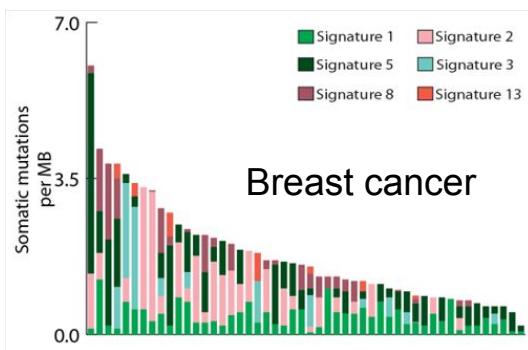
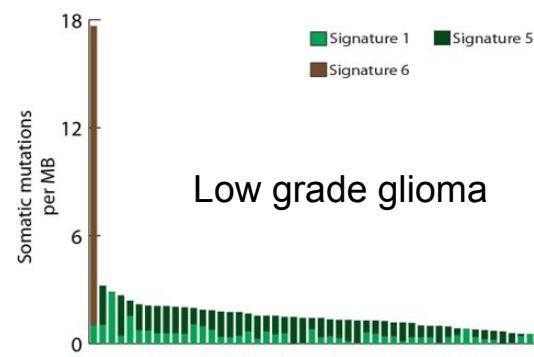
Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types

Smoking induced
cancer types



Non-smoking induced
cancer types

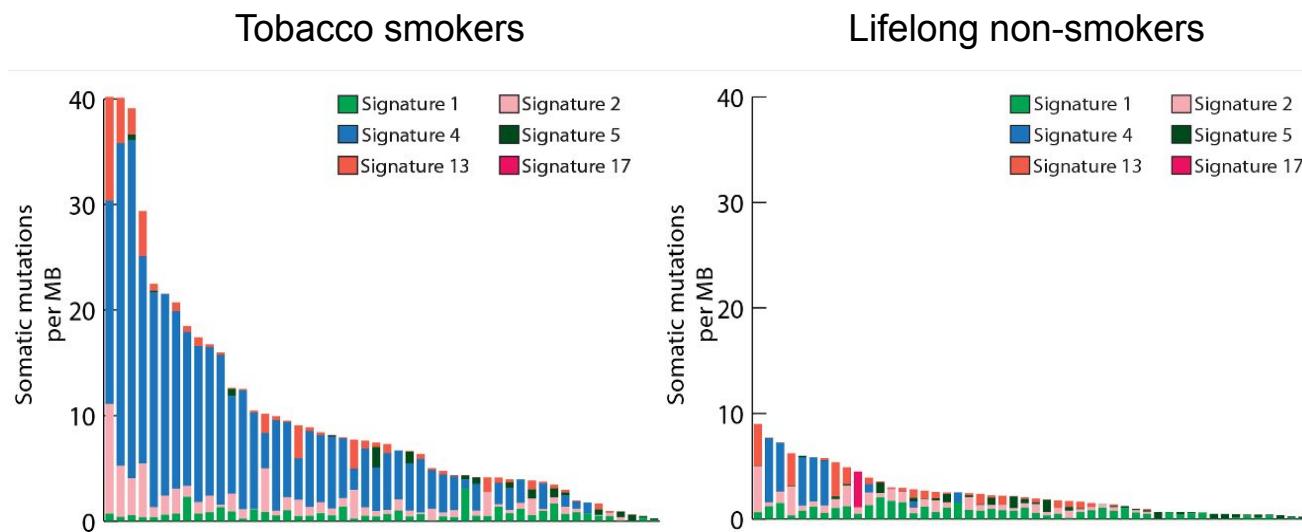


Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types
 - Contributions of signatures to lung adenocarcinoma in smokers vs. non-smokers

Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types
 - Contributions of signatures to lung adenocarcinoma in smokers vs. non-smokers



Validating the aetiologies of mutational signatures

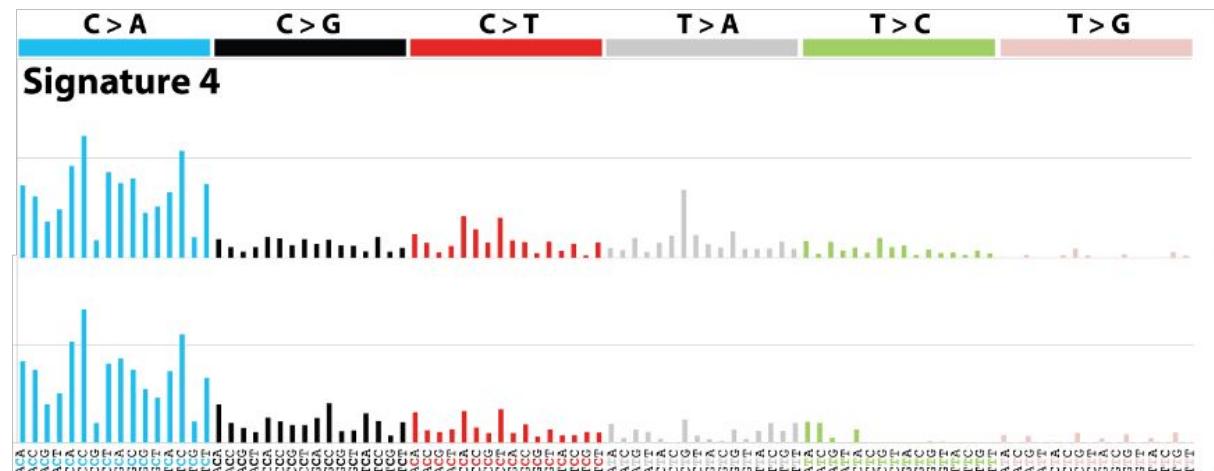
- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types
 - Contributions of signatures to lung adenocarcinoma in smokers vs. non-smokers
 - Similarity with experimental benzo[a]pyrene (known tobacco carcinogen) signature

Validating the aetiologies of mutational signatures

- Signature SBS4 is likely related to tobacco smoking
- How this is validated?
 - Contributions of signatures to smoking induced and non-smoking induced cancer types
 - Contributions of signatures to lung adenocarcinoma in smokers vs. non-smokers
 - Similarity with experimental benzo[a]pyrene (known tobacco carcinogen) signature

Signature SBS4 extracted from human cancers

Signature of benzo[a]pyrene exposure *in vitro*



Clinical Example #1 - unexpected carcinogen: 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.

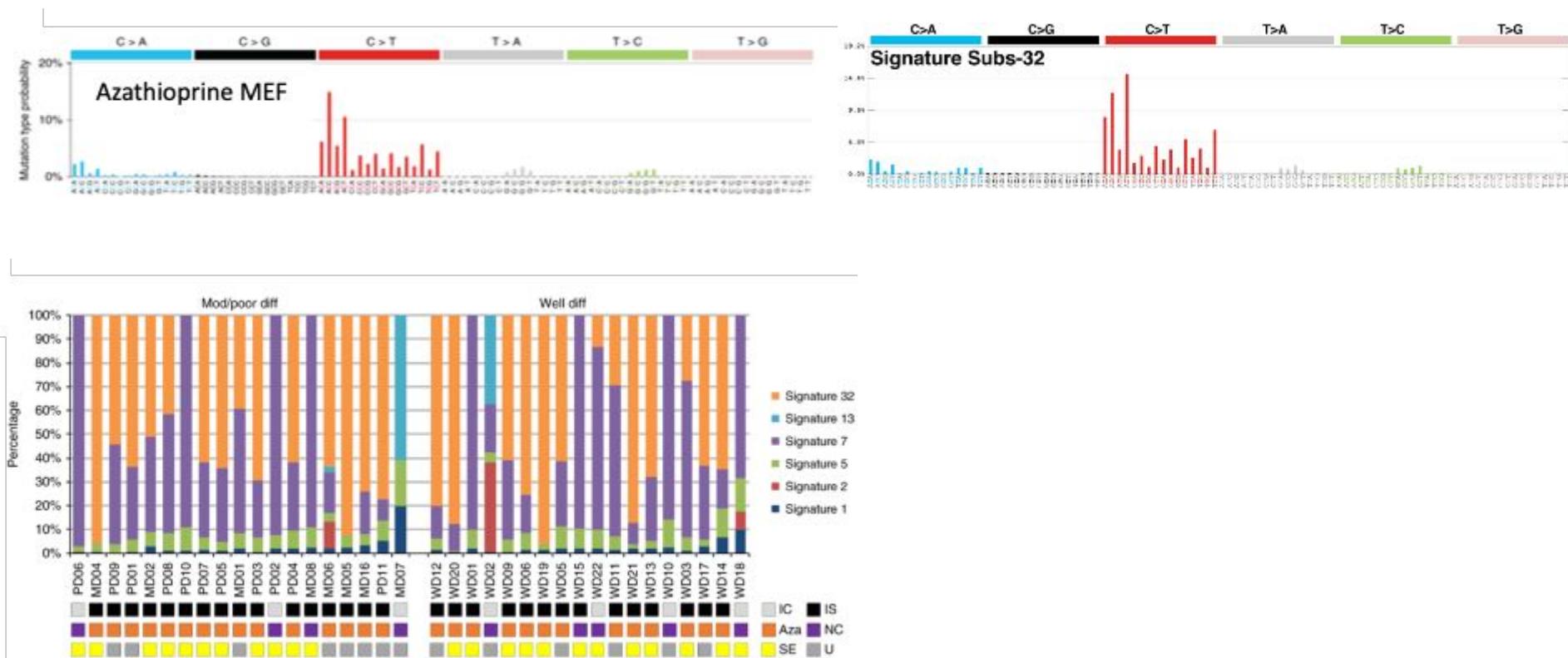
Clinical Example #1 - unexpected carcinogen: Azathioprine

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.



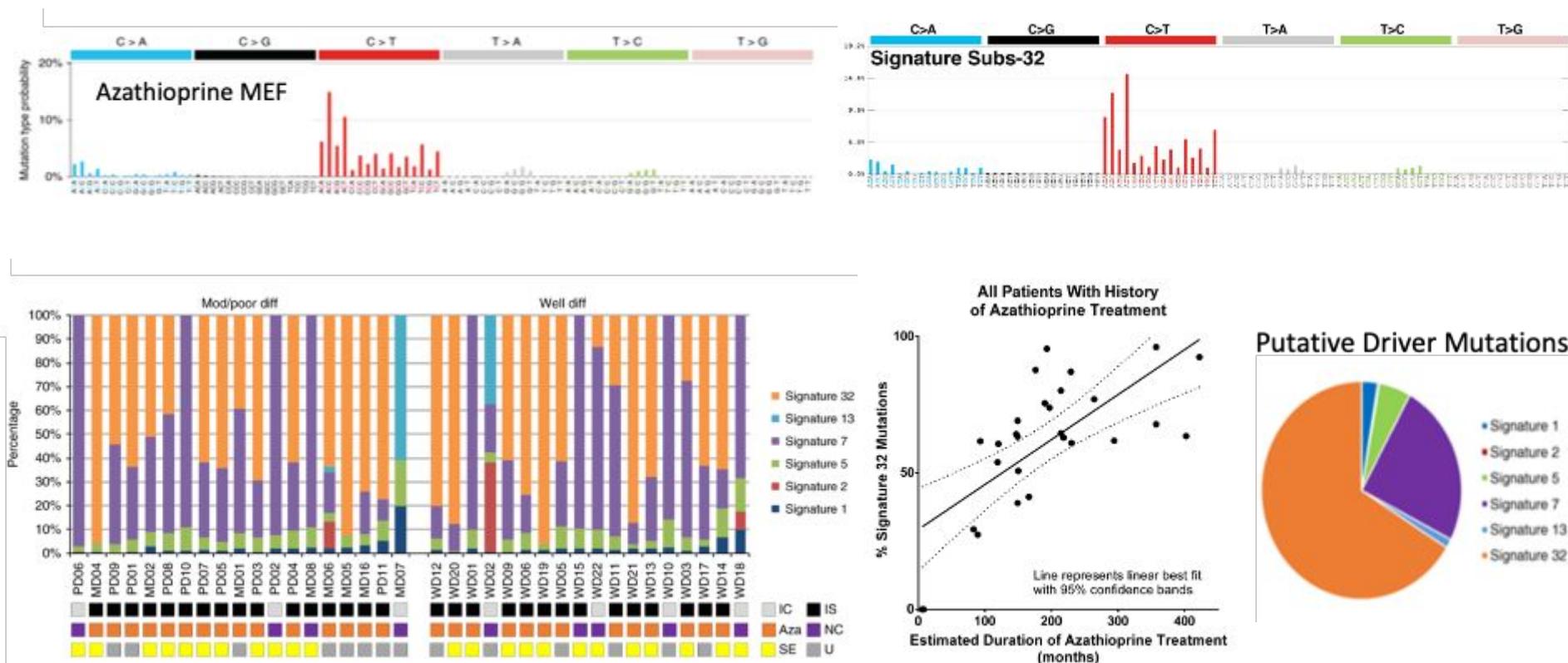
Clinical Example #1 - unexpected carcinogen: Azathioprine

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.

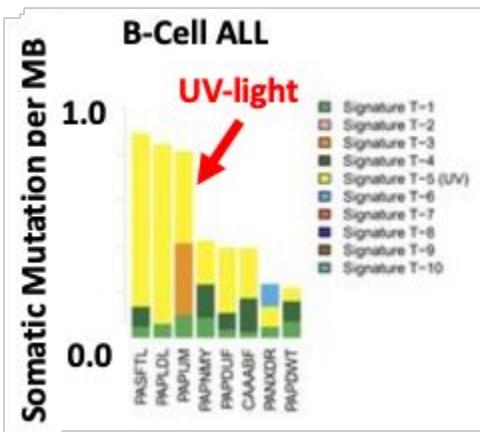


Clinical Example #1 - unexpected carcinogen: Azathioprine

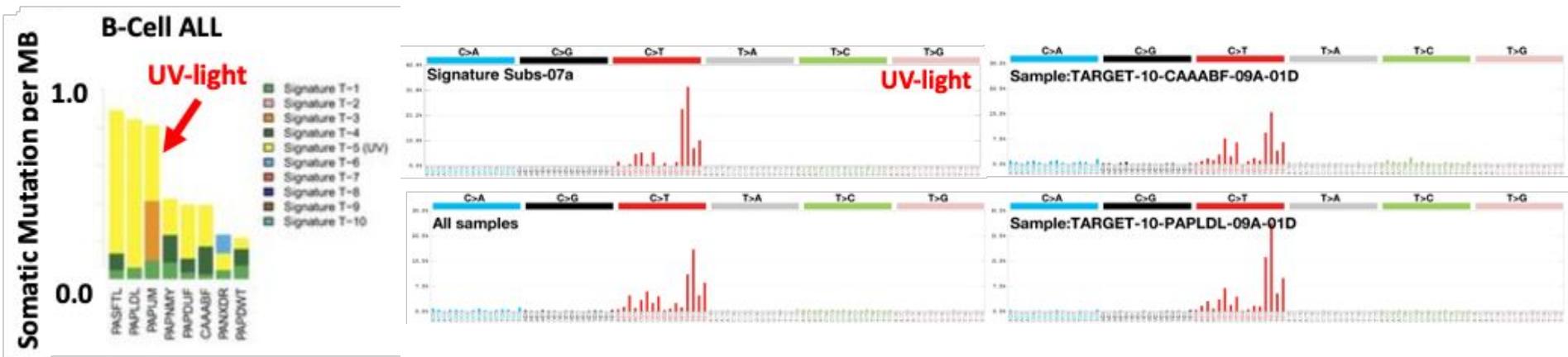
Azathioprine, sold under the brand name Imuran among others, is an immunosuppressive medication. Azathioprine is on the World Health Organization's List of Essential Medicines, the most effective and safe medicines needed in a health system. Epidemiological studies by International Agency for Research 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.



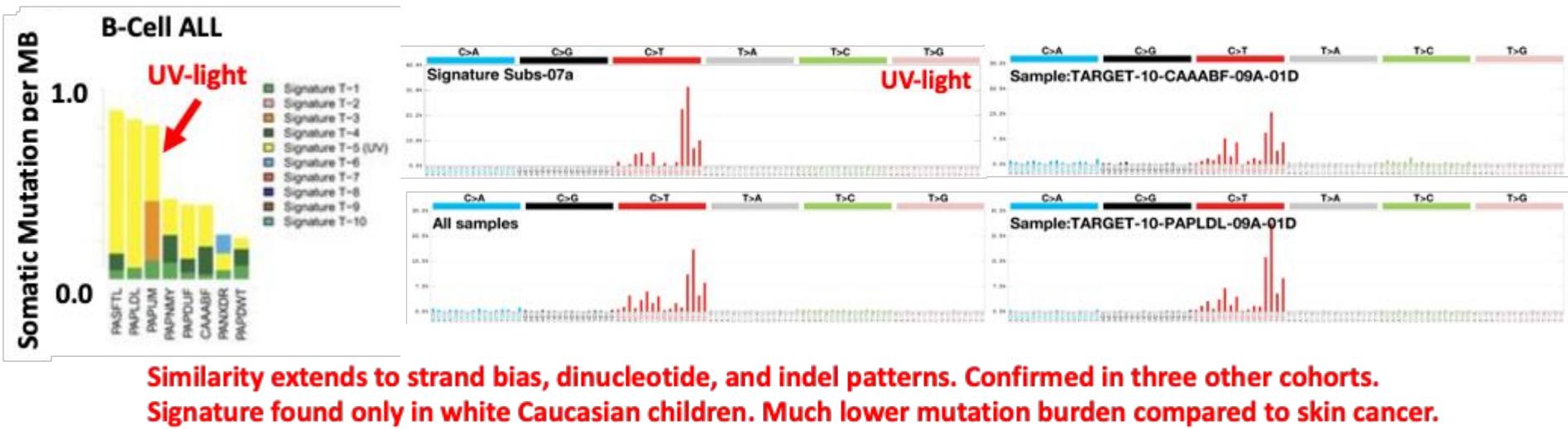
Clinical Example #2 - Known carcinogen in unexpected cancer types: UV-light



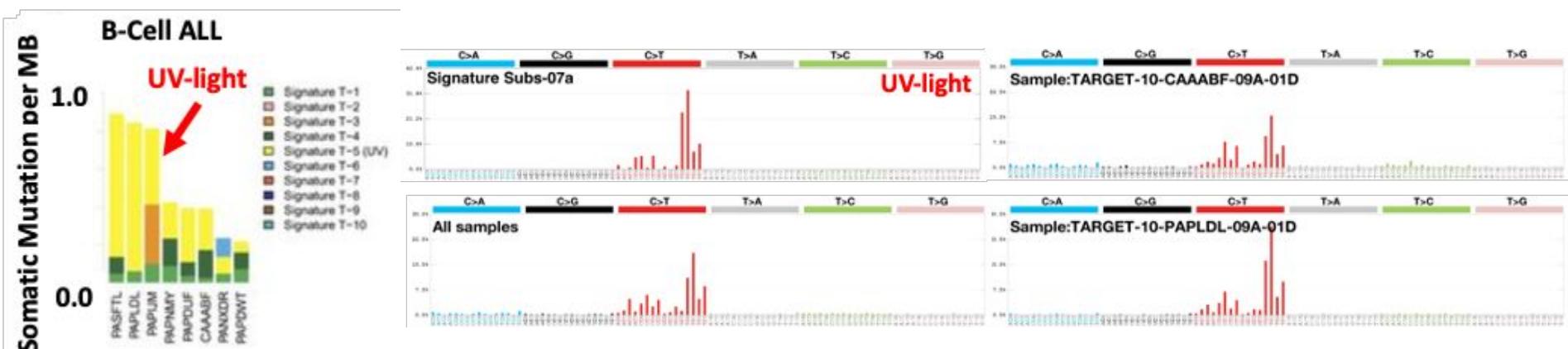
Clinical Example #2 - Known carcinogen in unexpected cancer types: UV-light



Clinical Example #2 - Known carcinogen in unexpected cancer types: UV-light



Clinical Example #2 - 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

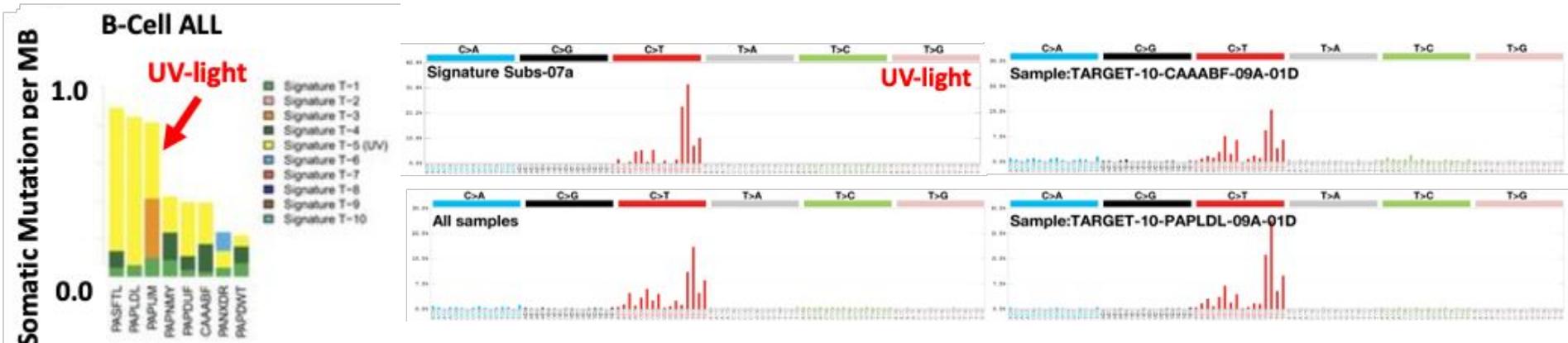
Residential exposure to ultraviolet light and risk of precursor B-cell acute lymphoblastic leukemia: assessing the role of individual risk factors, the ESCALE and ESTELLE studies

Authors

Authors and affiliations

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

Clinical Example #2 - 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 Hénon, 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)**

Clinical example #3 - Quantification of known carcinogens in suspected cancer types

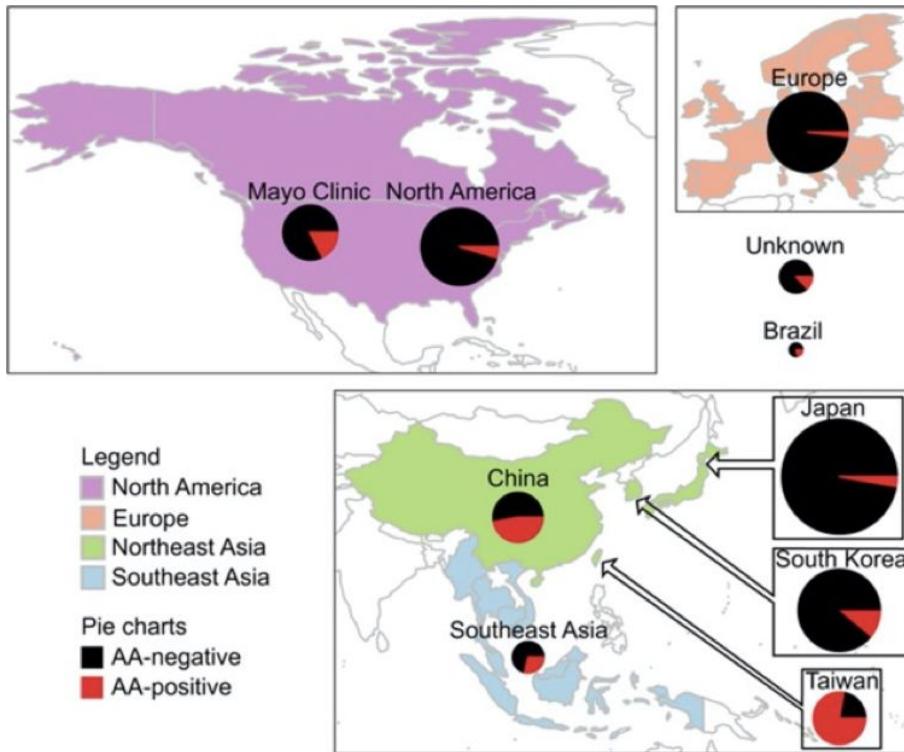
Science
Translational
Medicine

18 OCTOBER 2017



AAAS

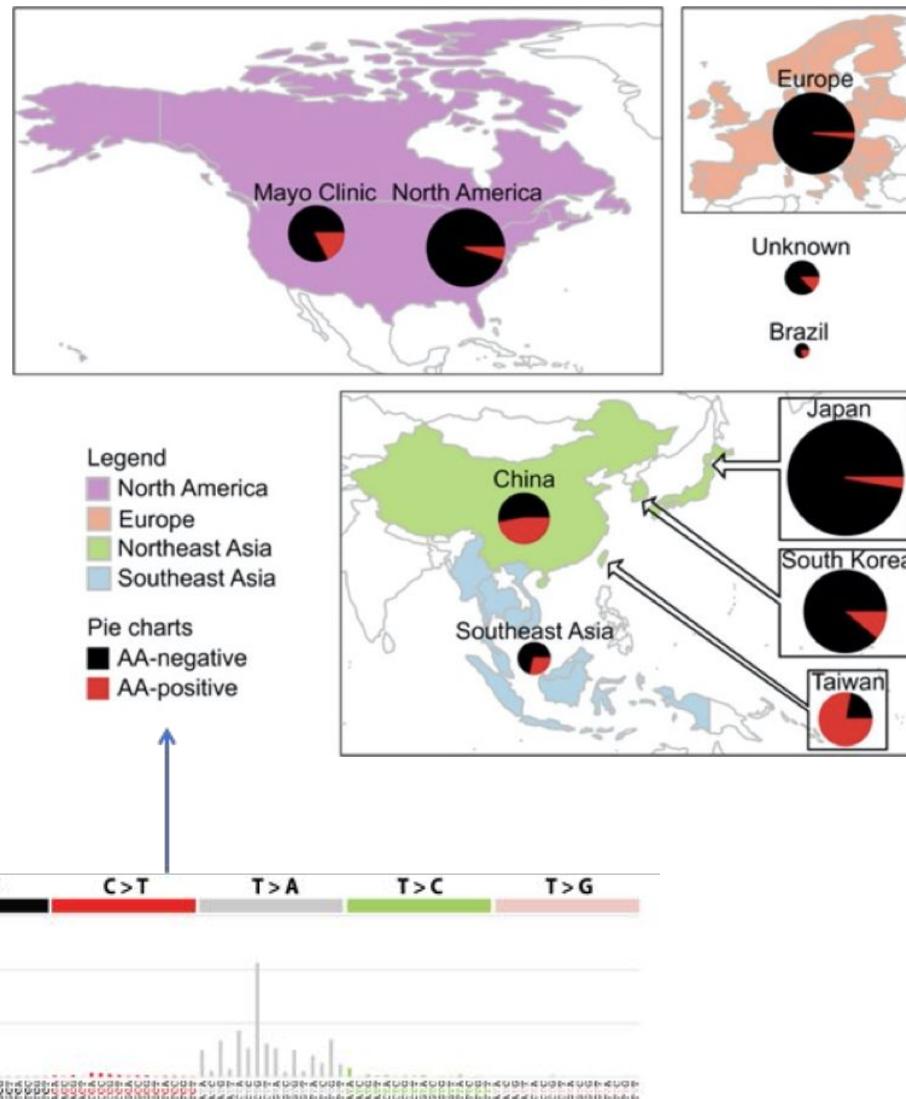
Clinical example #3 - Quantification of known carcinogens in suspected cancer types



Clinical example #3 - Quantification of known carcinogens in suspected cancer types



18 OCTOBER 2017

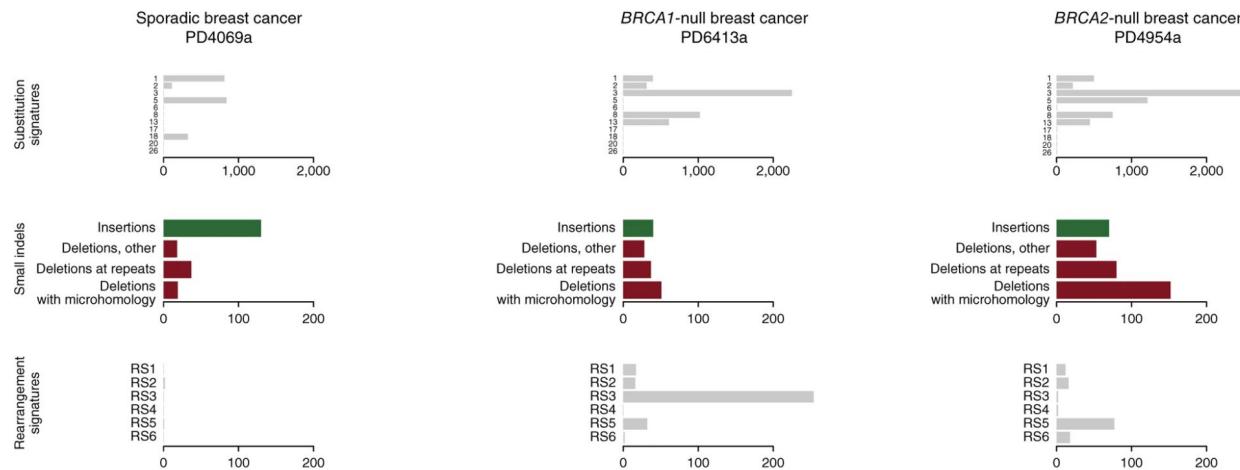


Clinical example #4 - HRDetect

Dysfunction of *BRCA1* or *BRCA2* is associated with selective sensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors, making it an important clinical biomarker. Six mutational signatures have been found to be predictive of *BRCA1/BRCA2* deficiency. **HRDetect** is a tool used to detect *BRCA1/BRCA2*-deficient samples.

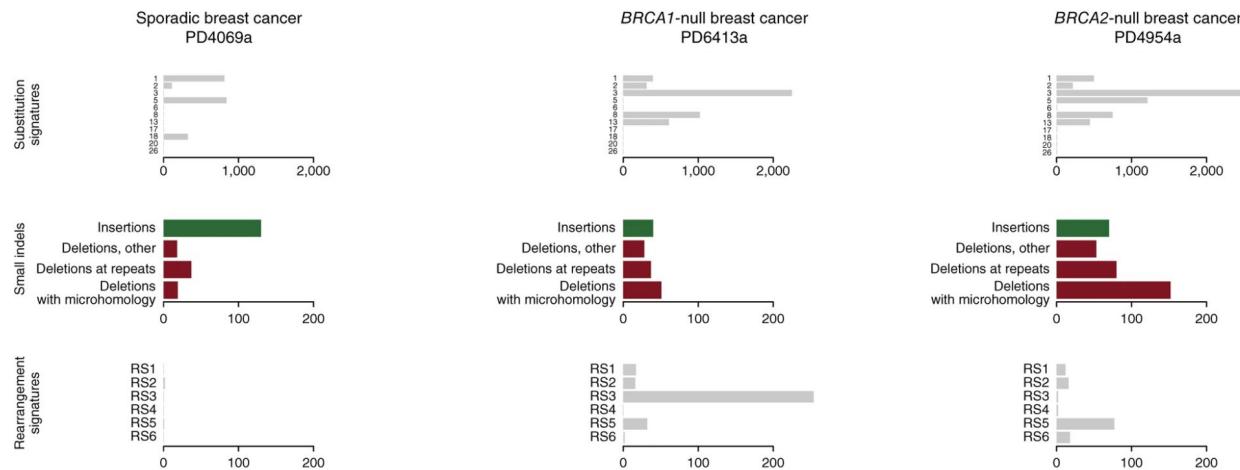
Clinical example #4 - HRDetect

Dysfunction of *BRCA1* or *BRCA2* is associated with selective sensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors, making it an important clinical biomarker. Six mutational signatures have been found to be predictive of *BRCA1/BRCA2*-deficient samples. **HRDetect** is a tool used to detect *BRCA1/BRCA2*-deficient samples.

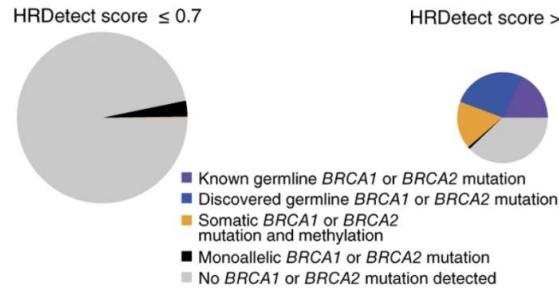


Clinical example #4 - HRDetect

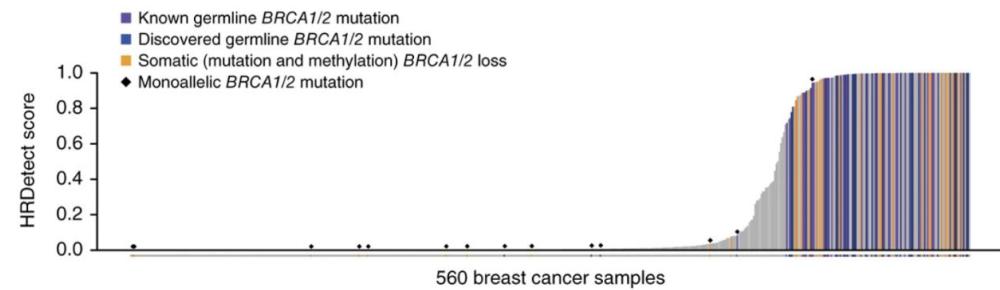
Dysfunction of *BRCA1* or *BRCA2* is associated with selective sensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors, making it an important clinical biomarker. Six mutational signatures have been found to be predictive of *BRCA1*/*BRCA2* deficiency. **HRDetect** is a tool used to detect *BRCA1*/*BRCA2*-deficient samples.



a



b



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
- Mutational signatures can be identified computationally by using NMF
- Reference mutational signatures have been identified and deposited in COSMIC after the analysis of thousands of cancer samples
- Mutational signatures can be used clinically as biomarkers for cancer prevention, prognosis and treatment

Module 4: Mutational Signature Analysis

Presented by:

Shakuntala Baichoo¹, Marcos Diaz-Gay² & Mariya Kazachkova²

¹University of Mauritius

²Alexandrov Lab, University of California San Diego

Based mostly on slides by:

Ludmil Alexandrov, Marcos Díaz-Gay & Mariya Kazachkova,
from University of California San Diego

 @ShakunBaichoo
 shakunb@uom.ac.mu

 @mdiazgay
 mdiazgay@health.ucsd.edu

 mkazachk@health.ucsd.edu

Cancer Genome Analysis

12 – 16 September 2022 – Virtual course



UNIVERSITY OF
MAURITIUS

UC San Diego

WELLCOME GENOME CAMPUS

CONNECTING SCIENCE
ADVANCED COURSES + SCIENTIFIC CONFERENCES