wellcome connecting science

connecting people with science

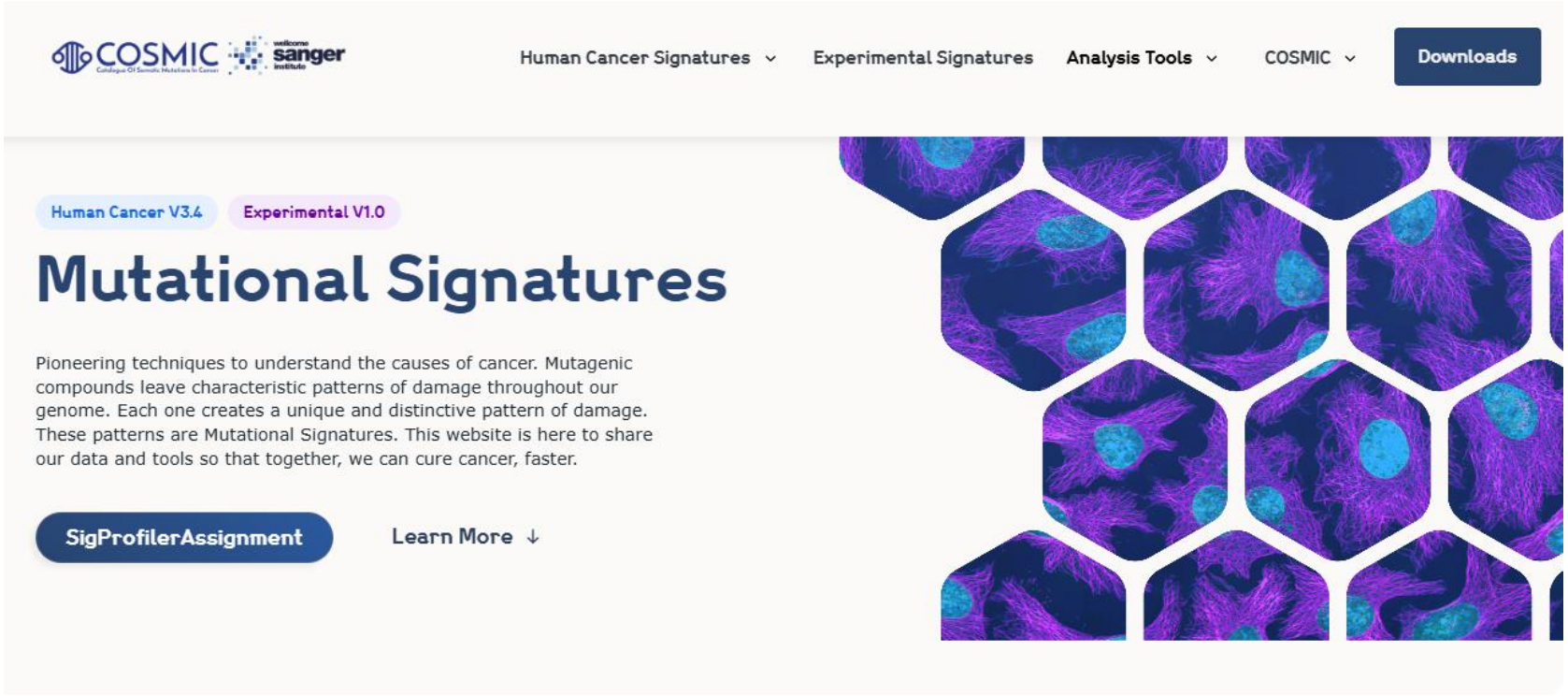
Dr Jia-Wern Pan
Cancer Research Malaysia

Mutational signatures:
Databases, aetiology, and clinical utility

(adapted from Dr Marcos Díaz Gay and
Dr Mariano Golubicki, WCS CGA 2023)

October 2025

COSMIC Mutational Signatures Database



October v3.4 (June 2023)

Genomics
- 99 SBS signatures
- 20 DBS signatures
- 23 ID signatures
- 25 CN signatures
- 10 SV signatures
- 5 RNA-SBS signatures

Experimental
- 140 signatures

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID
- CNV
- SV
- RNA-SBS

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS (78 classes of strand-agnostic doublet base substitutions)
- ID
- CNV
- SV
- RNA-SBS

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID (83 classes based on size, nucleotides affected and presence on repetitive and/or microhomology regions)
- CNV
- SV
- RNA-SBS

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID
- CNV (48 classes based on loss-of-heterozygosity status, total copy number state, and segment length of allele-specific copy number segments)
- SV
- RNA-SBS

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID
- CNV
- SV (32 classes based on deletions, inversions, and tandem duplications, size, and distance between adjacent SVs)
- RNA-SBS

The COSMIC mutational signature database

Mini-practical 1:

Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID
- CNV
- SV
- RNA-SBS (192 stranded trinucleotide classes)

The COSMIC mutational signature database

Mini-practical 1:

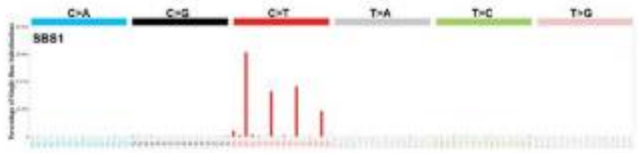
Go to the COSMIC website. The input matrix for SBS consists of 96 classes based on the trinucleotide context of single base substitutions. What are the input matrix classes for:

- DBS
- ID
- CNV
- SV
- RNA-SBS (192 stranded trinucleotide classes)

Linking mutational signatures to specific causes

Many signatures in COSMIC have a specific proposed aetiology, i.e. SBS 4 is linked with tobacco smoking.

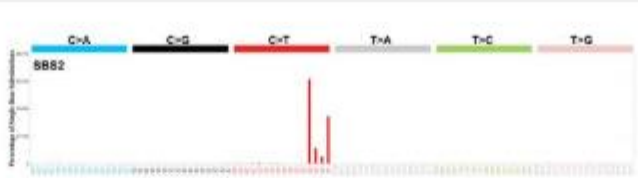
How do we know this? What is the evidence?



SBS1

Proposed Aetiology

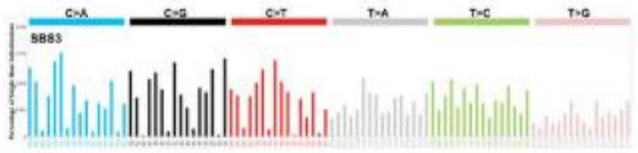
Spontaneous deamination of 5-methylcytosine (clock-like signature)



SBS2

Proposed Aetiology

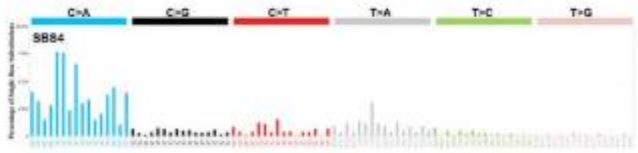
Activity of APOBEC family of cytidine deaminases



SBS3

Proposed Aetiology

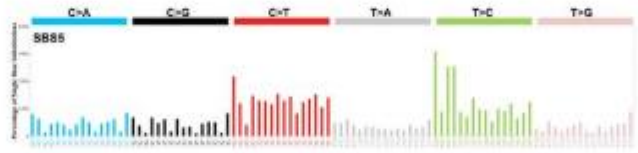
Defective homologous recombination DNA damage repair



SBS4

Proposed Aetiology

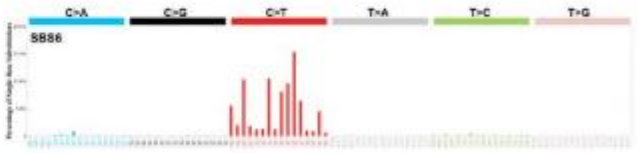
Tobacco smoking



SBS5

Proposed Aetiology

Unknown (clock-like signature)



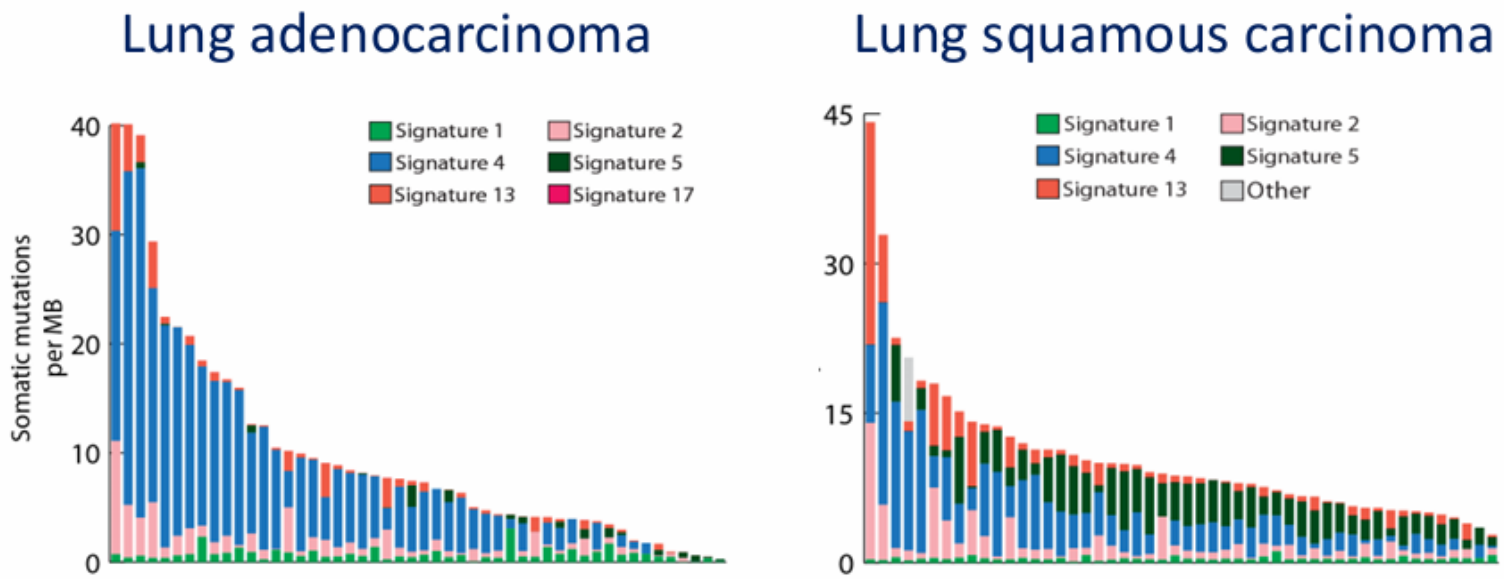
SBS6

Proposed Aetiology

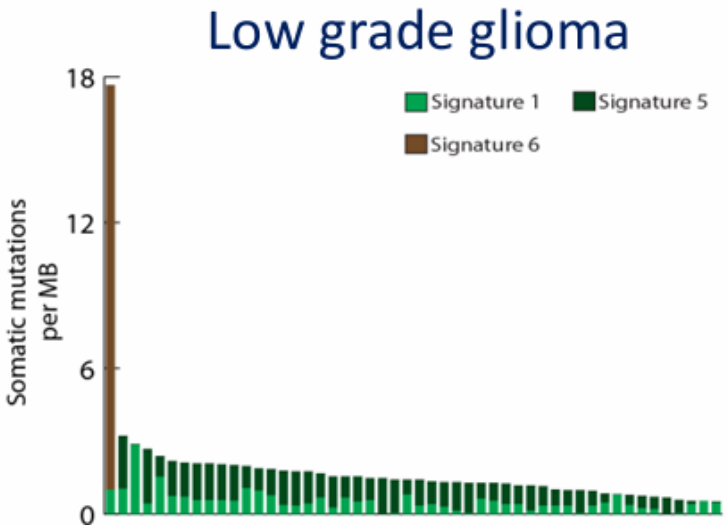
Defective DNA mismatch repair

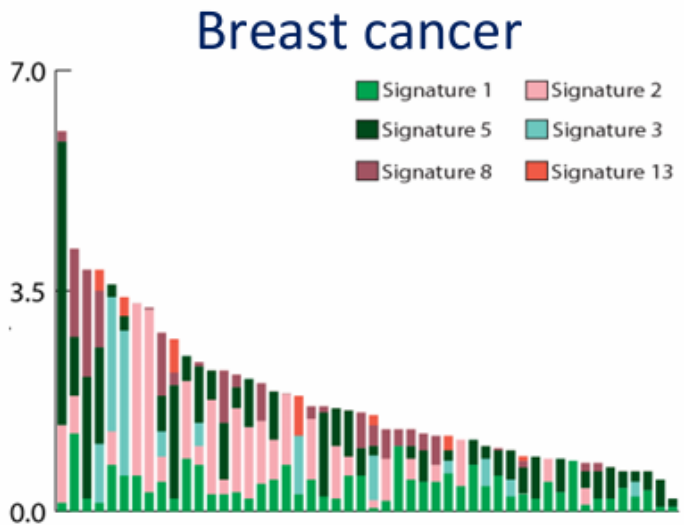
Signature 4 is associated with smoking induced cancer types

Smoking induced cancer types

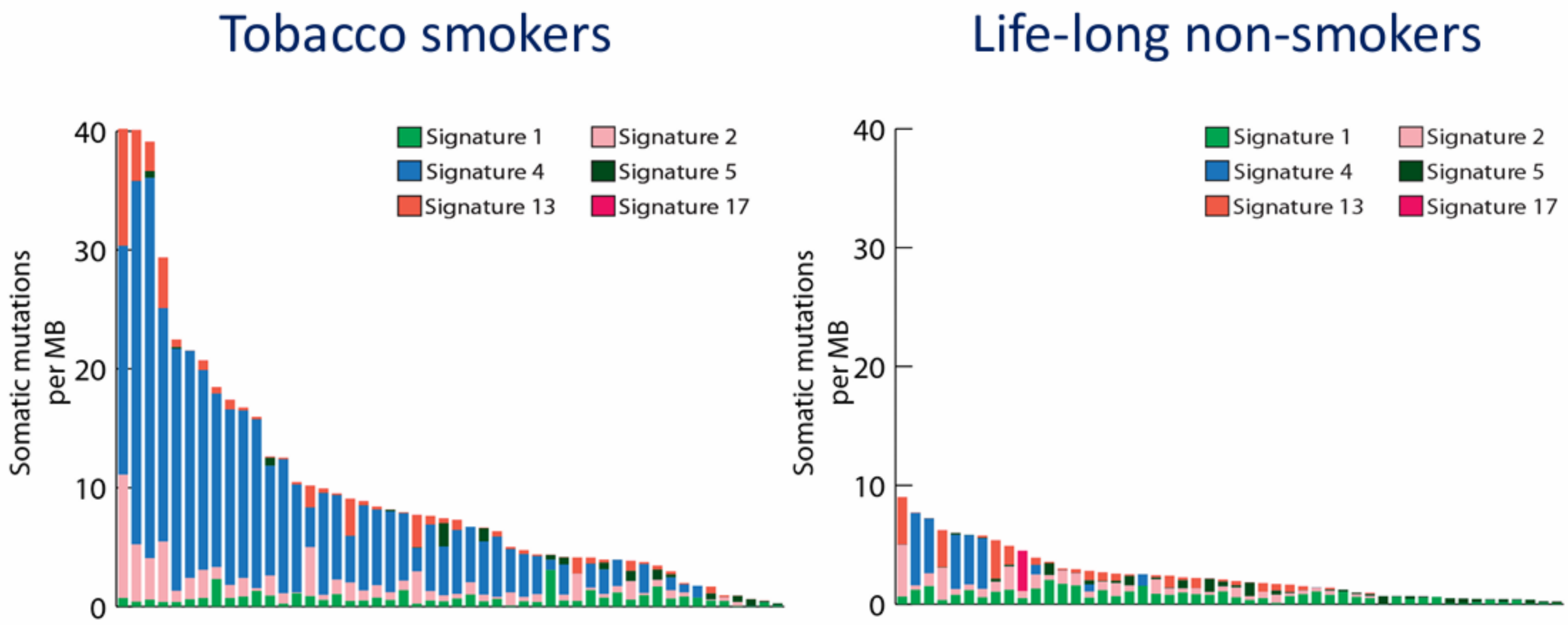


Non-smoking induced cancer types





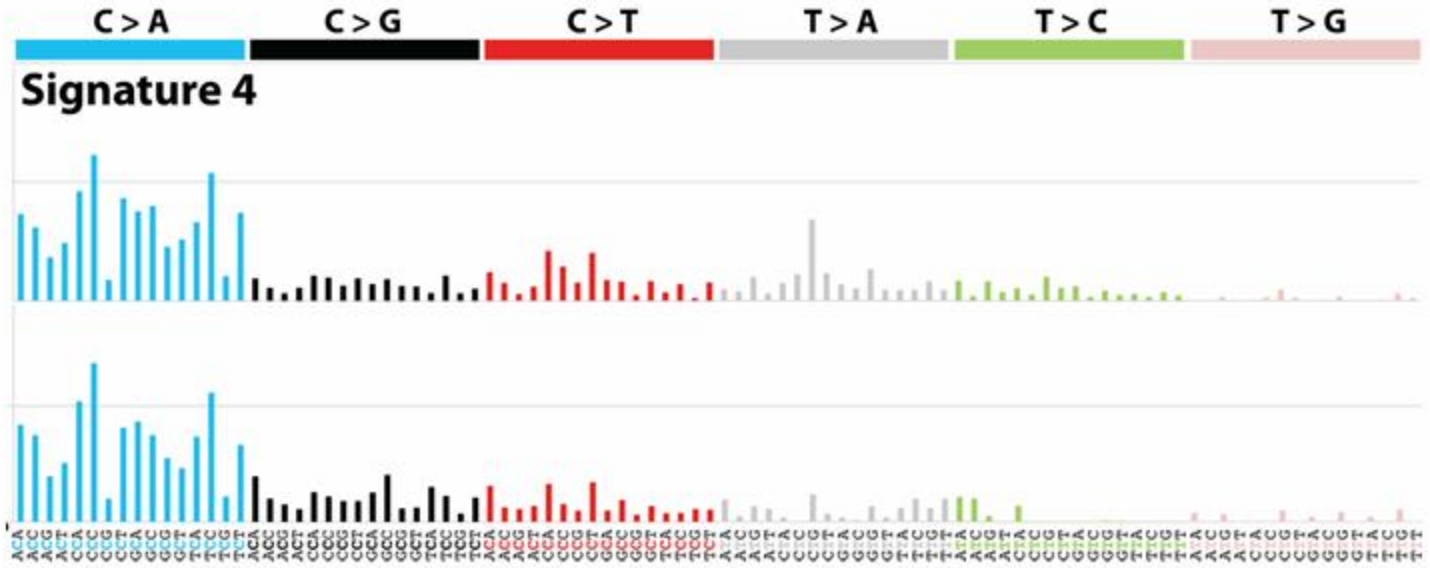
Within lung cancer, SBS 4 is much more prevalent in tobacco smokers compared to non-smokers



SBS 4 matches the signature of benzo[a]pyrene exposure, a common chemical component of cigarette smoke, in cell lines and mouse models

Signature 4 extracted from human cancers

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



Evidence for the aetiology of SBS 4

- Identified only in cancer types epidemiologically known to be caused by tobacco smoking
- Highly enriched in tobacco smokers compared to non-smokers
- The pattern of SBS 4 matches *in vitro* experiments exposing cell lines to known tobacco carcinogens

Summary of evidence for SBS 4

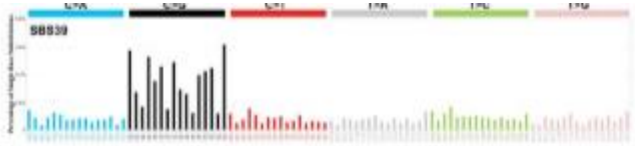
Background	**Identification study**		**First included in COSMIC**
	Alexandrov *et al.* 2013 Nature		v1
Identification	**NGS technique**	**Different variant callers**	**Multiple sequencing centres**
	WES & WGS	Yes	Yes
Technical validation	**Validated in orthogonal techniques**	**Replicated in additional studies**	**Extended context enrichment**
	Yes	Yes	-
Proposed aetiology	**Mutational process**		**Support**
	Tobacco smoking		Experimental confirmation
Experimental validation	**Experimental study**		**Species**
	Nik-Zainal *et al.* 2015 Mutagenesis		Mouse

Summary of evidence for SBS 92

Background	**Identification study**		**First included in COSMIC**
	Lawson *et al.* 2020 Science		v3.2
Identification	**NGS technique**	**Different variant callers**	**Multiple sequencing centres**
	WGS	Yes	Yes
Technical validation	**Validated in orthogonal techniques**	**Replicated in additional studies**	**Extended context enrichment**
	Yes	Yes	-
Proposed aetiology	**Mutational process**		**Support**
	Tobacco smoking		Statistical association
Experimental validation	**Experimental study**		**Species**
	-		-

Also associated with tobacco smoking, but only in bladder cancer and no experimental validation

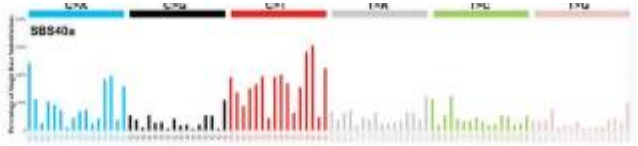
Many signatures still have unknown aetiology



SBS39

Proposed Aetiology

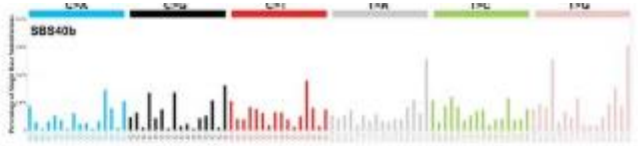
Unknown



SBS40a

Proposed Aetiology

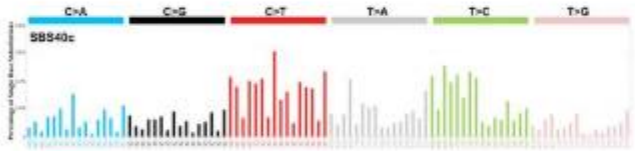
Unknown



SBS40b

Proposed Aetiology

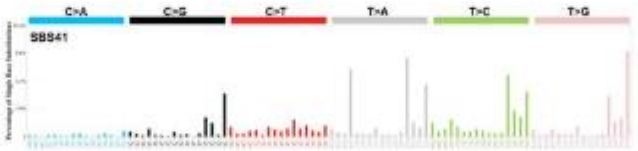
Unknown



SBS40c

Proposed Aetiology

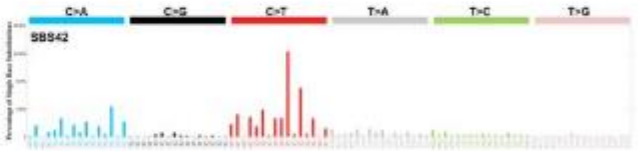
Unknown



SBS41

Proposed Aetiology

Unknown



SBS42

Proposed Aetiology

Haloalkane exposure

Caveats to consider

Discuss: How might the COSMIC database be biased in terms of mutational signatures included and also their proposed aetiologies?

Caveats to consider

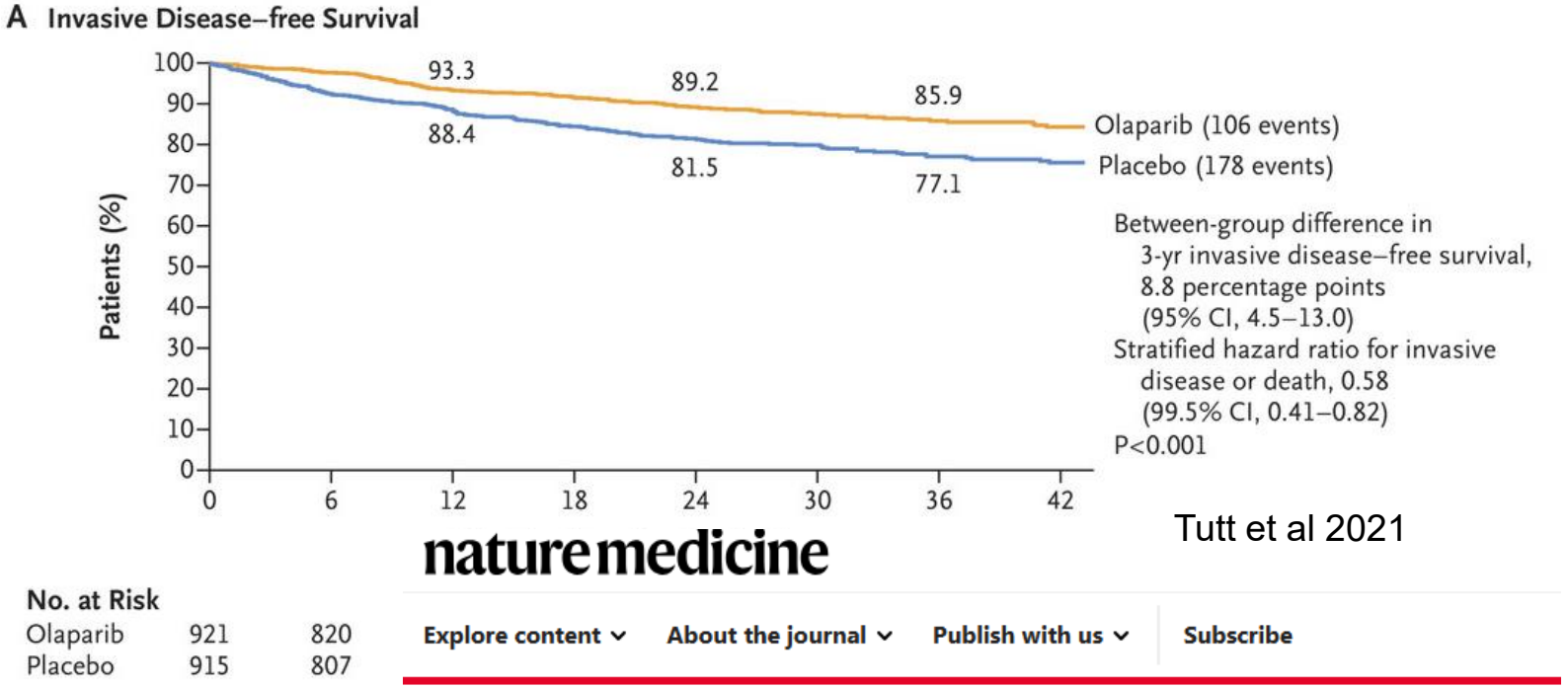
Discuss: How might the COSMIC database be biased in terms of mutational signatures included and also their proposed aetiologies?

- Underlying cohorts primarily Western
- Relies mostly on short-read sequencing
- Biased towards common cancer types and common risk factors

wellcome connecting science

Clinical and research applications of mutational signatures

Mutational signatures as predictive biomarkers for targeted therapy



Tutt et al 2021

Analysis | Published: 13 March 2017

HRDetect is a predictor of *BRCA1* and *BRCA2* deficiency based on mutational signatures

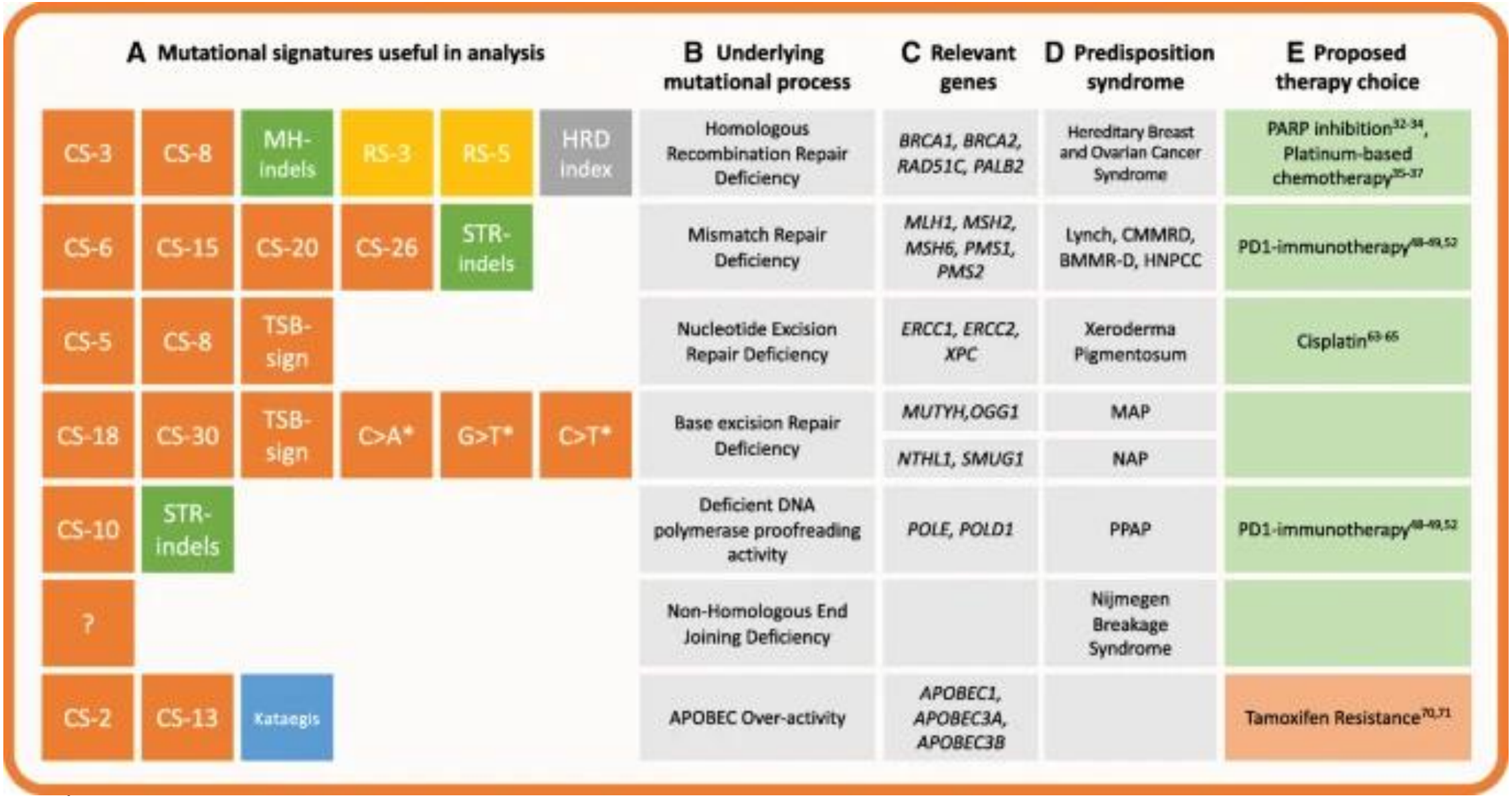
Helen Davies, Dominik Glodzik, Sandro Morganella, Lucy R Yates, Johan Staaf, Xueqing Zou, Manasa Ramakrishna, Sancha Martin, Sandrine Boyault, Anieta M Sieuwerts, Peter T Simpson, Tari A King, Keiran Raine, Jorunn E Eyfjord, Gu Kong, Åke Borg, Ewan Birney, Hendrik G Stunnenberg, Marc J van de Vijver,

Homologous recombination repair deficiency (HRD) is a common many cancer types

Patients with HRD may be more likely to respond to platinum chemotherapy and PARP inhibitor targeted therapy

Mutational signatures associated with HRD may be able to identify patients who should receive PARP inhibitors, particularly in breast and ovarian cancer

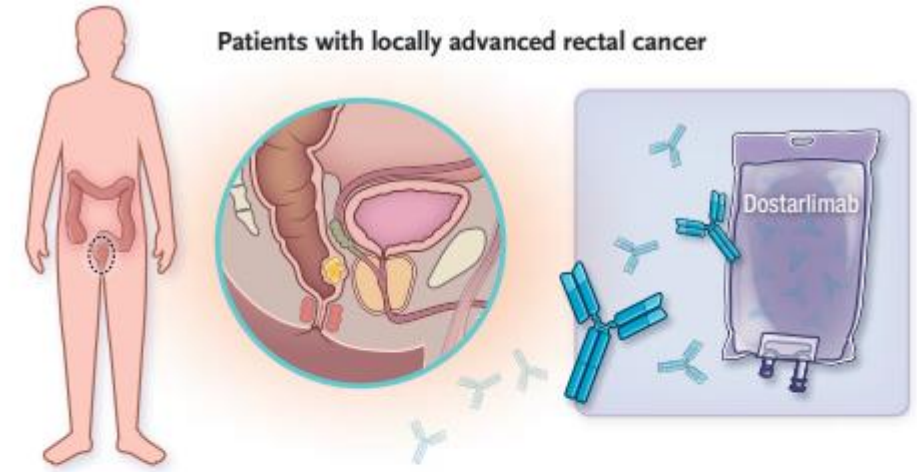
Mutational signatures as predictive biomarkers for targeted therapy

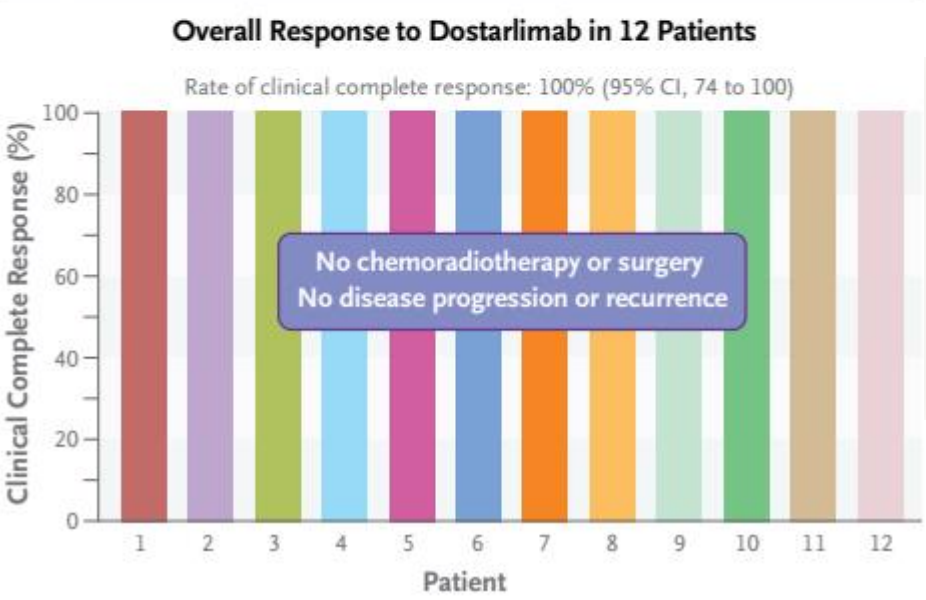


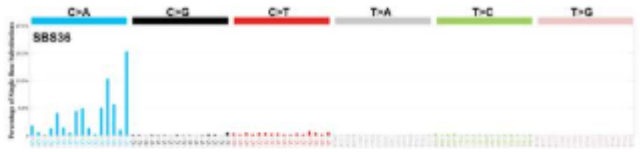
A Mutational signatures useful in analysis						B Underlying mutational process	C Relevant genes	D Predisposition syndrome	E Proposed therapy choice
CS-3	CS-8	MH-indels	RS-3	RS-5	HRD index	Homologous Recombination Repair Deficiency	*BRCA1, BRCA2, RAD51C, PALB2*	Hereditary Breast and Ovarian Cancer Syndrome	PARP inhibition[32-34], Platinum-based chemotherapy[35-37]
CS-6	CS-15	CS-20	CS-26	STR-indels		Mismatch Repair Deficiency	*MLH1, MSH2, MSH6, PMS1, PMS2*	Lynch, CMMRD, BMMR-D, HNPCC	PD1-immunotherapy[48-49,52]
CS-5	CS-8	TSB-sign				Nucleotide Excision Repair Deficiency	*ERCC1, ERCC2, XPC*	Xeroderma Pigmentosum	Cisplatin[63-65]
CS-18	CS-30	TSB-sign	C>A*	G>T*	C>T*	Base excision Repair Deficiency	*MUTYH,OGG1*	MAP	
							NTHL1, SMUG1	NAP	
CS-10	STR-indels					Deficient DNA polymerase proofreading activity	*POLE, POLD1*	PPAP	PD1-immunotherapy[48-49,52]
?						Non-Homologous End Joining Deficiency		Nijmegen Breakage Syndrome	
CS-2	CS-13	Kataegis				APOBEC Over-activity	*APOBEC1, APOBEC3A, APOBEC3B*		Tamoxifen Resistance[70,71]

Van Hoeck et al 2019

Mutational signatures as predictive biomarkers for immunotherapy



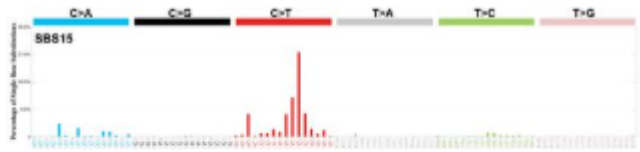




SBS36

Proposed Aetiology

Defective DNA base excision repair due to MUTYH mutations



SBS15

Proposed Aetiology

Defective DNA mismatch repair

- Microsatellite instability (MSI) due to deficiencies in mismatch repair (dMMR) is a common in colorectal cancer
- Patients with MSI are more likely to respond to checkpoint immunotherapy
- Mutational signatures associated with MSI (SBS15, SBS36, etc.) may be able to identify colorectal patients who should be treated with immunotherapy

Cercek et al 2022

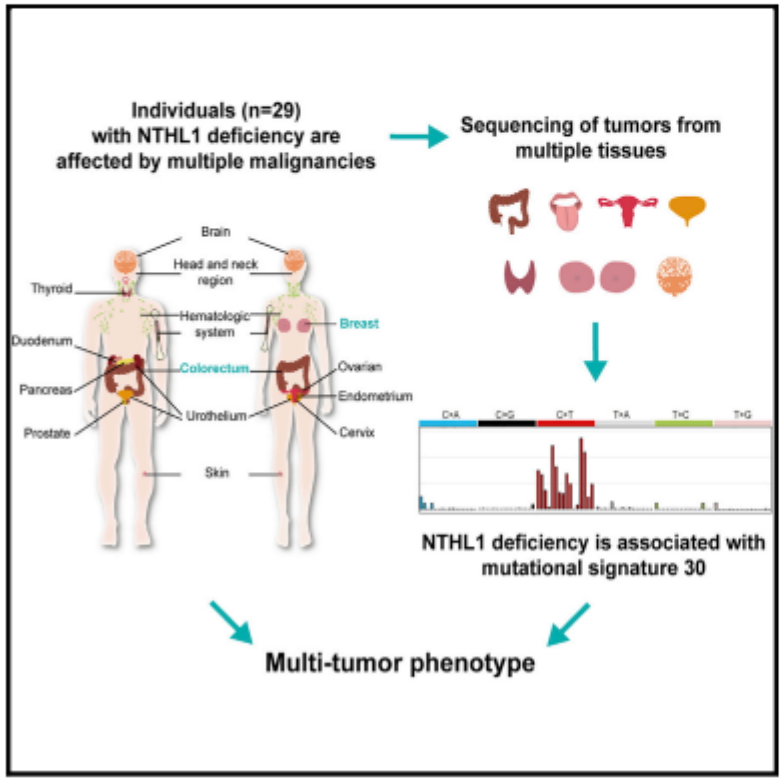
Mutational signatures as a tool to quantify cancer risk in individuals with *NTHL1* mutations

Article

Cancer Cell

Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype

Graphical Abstract



Authors

Judith E. Grolleman, Richarda M. de Voer, Fadwa A. Elsayed, ..., Tom van Wezel, Nicoline Hoogerbrugge, Roland P. Kuiper

Correspondence

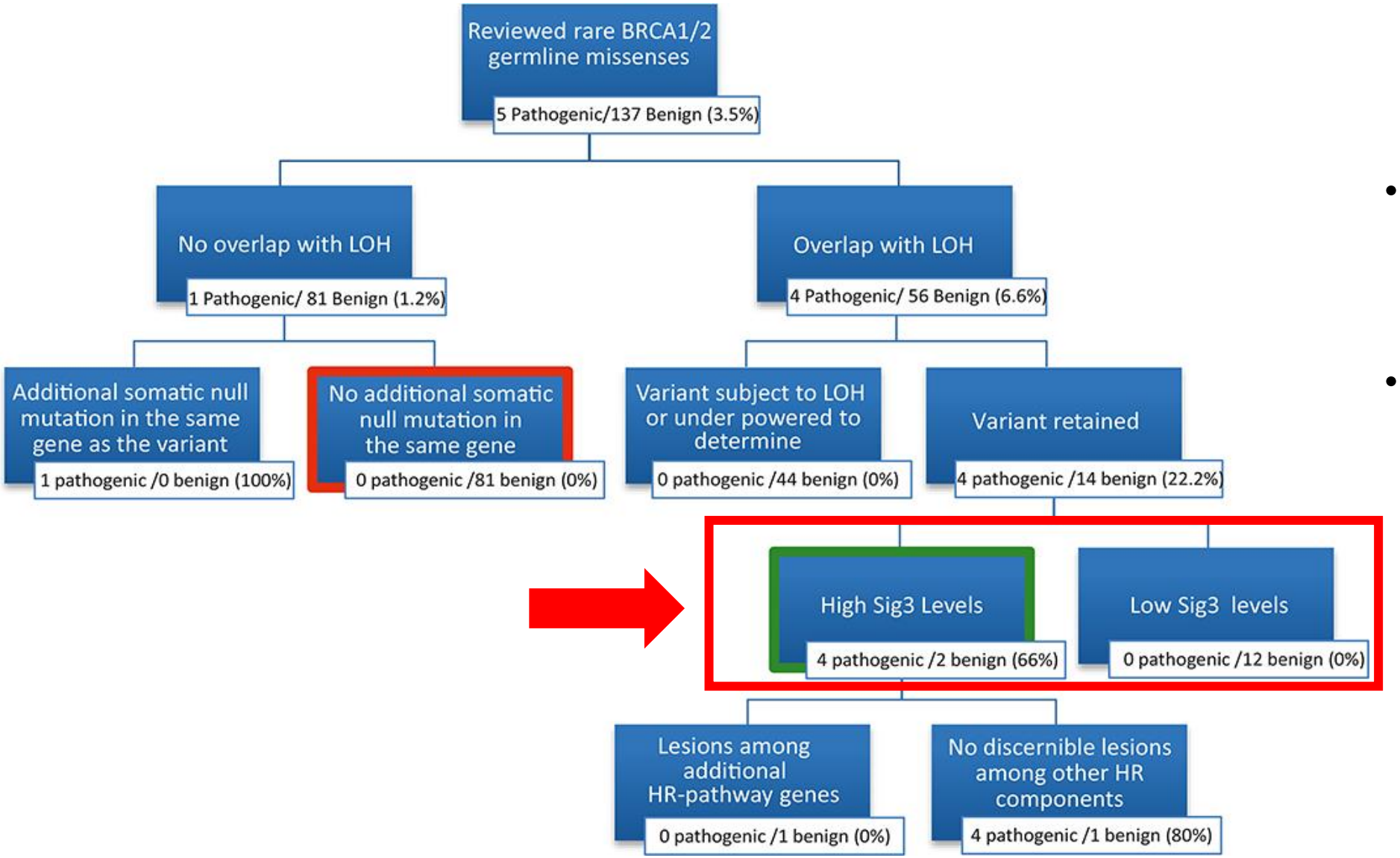
richarda.devoer@radboudumc.nl (R.M.d.V.), r.kuiper@prinsesmaximacentrum.nl (R.P.K.)

In Brief

In addition to the know colorectal tumors, Grolleman et al. find tumors in 13 tissue types, including a high breast cancer incidence, among 29 carriers of biallelic germline *NTHL1* mutations and identify a mutation signature across tumor types, which may facilitate the identification and management of new cases.

Mutational signature analyses of tumours in people with a rare germline mutation in NTHL1 reveals that they are at risk to develop multiple different types of tumours

Mutational signatures as a tool to classify VUS



- Genes such as *BRCA1*/2 are associated with specific mutational signatures such as HRD-associated SBS3
- Rare variants in these genes are often classified as VUS
- The presence of HRD mutational signatures can be used as evidence for whether a VUS is benign or deleterious

Polak et al. 2017

Mutational signatures can identify individuals with pathogenic germline variants

nature genetics

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › nature genetics › articles › article

Article | Published: 21 August 2017

A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer

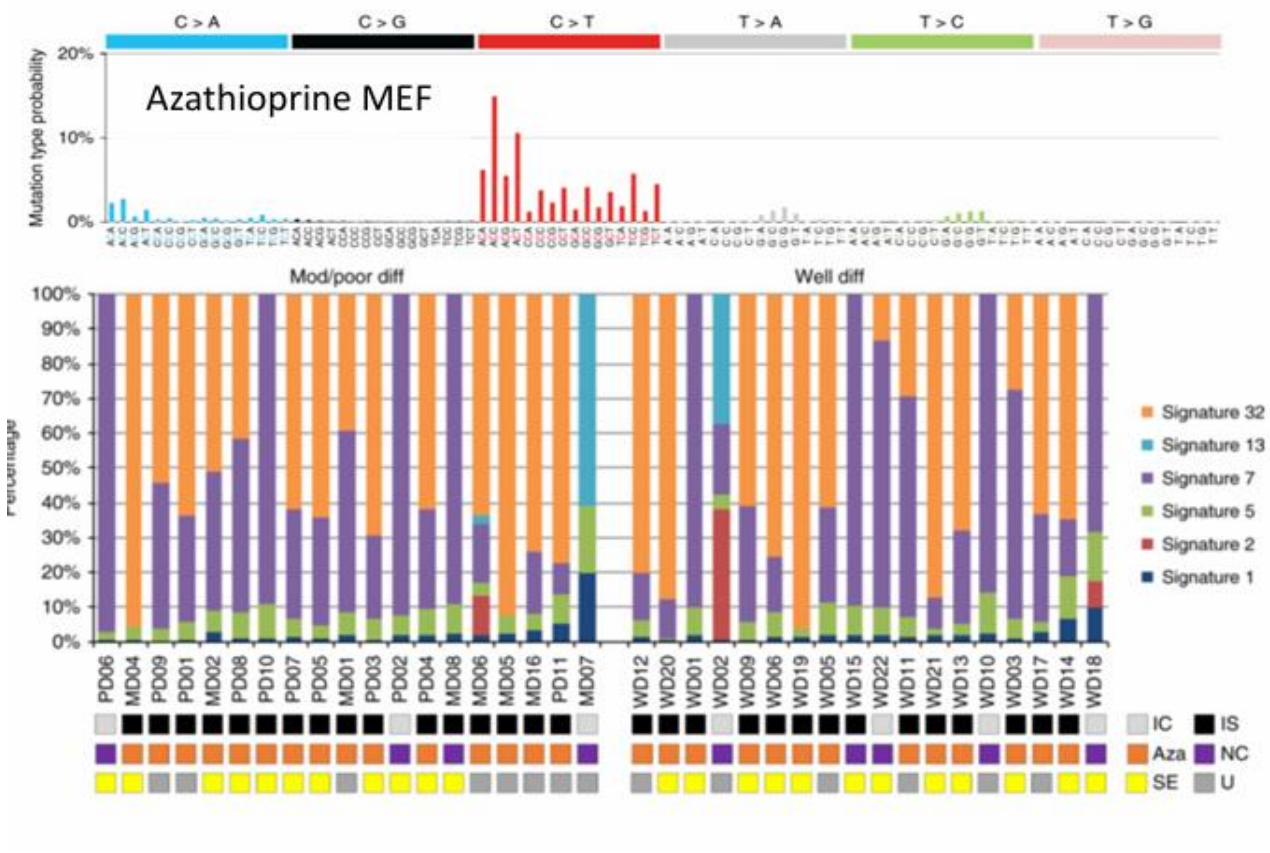
Paz Polak, Jaegil Kim, Lior Z Braunstein, Rosa Karlic, Nicholas J Haradhavala, Grace Tiao, Daniel Rosebrock, Dimitri Livitz, Kirsten Kübler, Kent W Mouw, Atanas Kamburov, Yosef E Maruvka, Ignaty Leshchiner, Eric S Lander, Todd R Golub, Aviad Zick, Alexandre Orthwein, Michael S Lawrence, Rajbir N Batra, Carlos Caldas, Daniel A Haber, Peter W Laird, Hui Shen, Leif W Ellisen, ... Gad Getz + Show authors

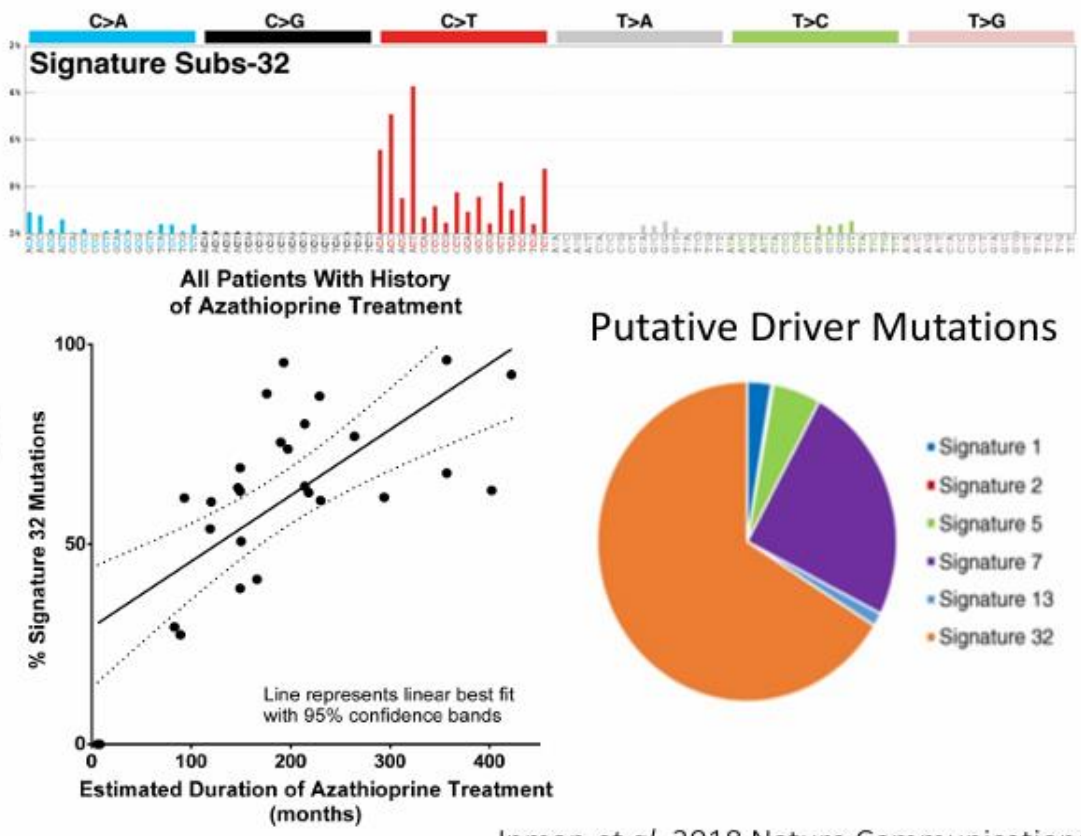
Nature Genetics **49**, 1476–1486 (2017) | Cite this article

Quantification of HRD-associated mutational signatures in a patient's tumour can help doctors and patients decide if the patient needs genetic testing and/or genetic counselling.

Mutational signatures as a tool to discover/validate new carcinogens

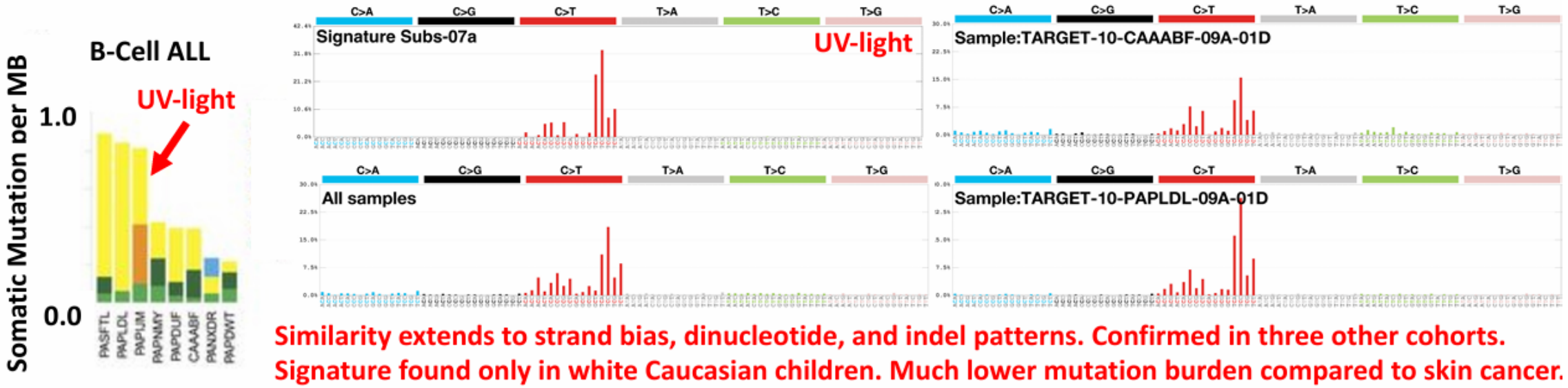
Mutational signature analysis provides strong evidence that azathioprine, an immunosuppressive medication, is the carcinogen responsible for some cutaneous squamous cell carcinomas





Inman *et al.* 2018 Nature Communications

Mutational signatures as a tool to discover known carcinogens in unexpected settings



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

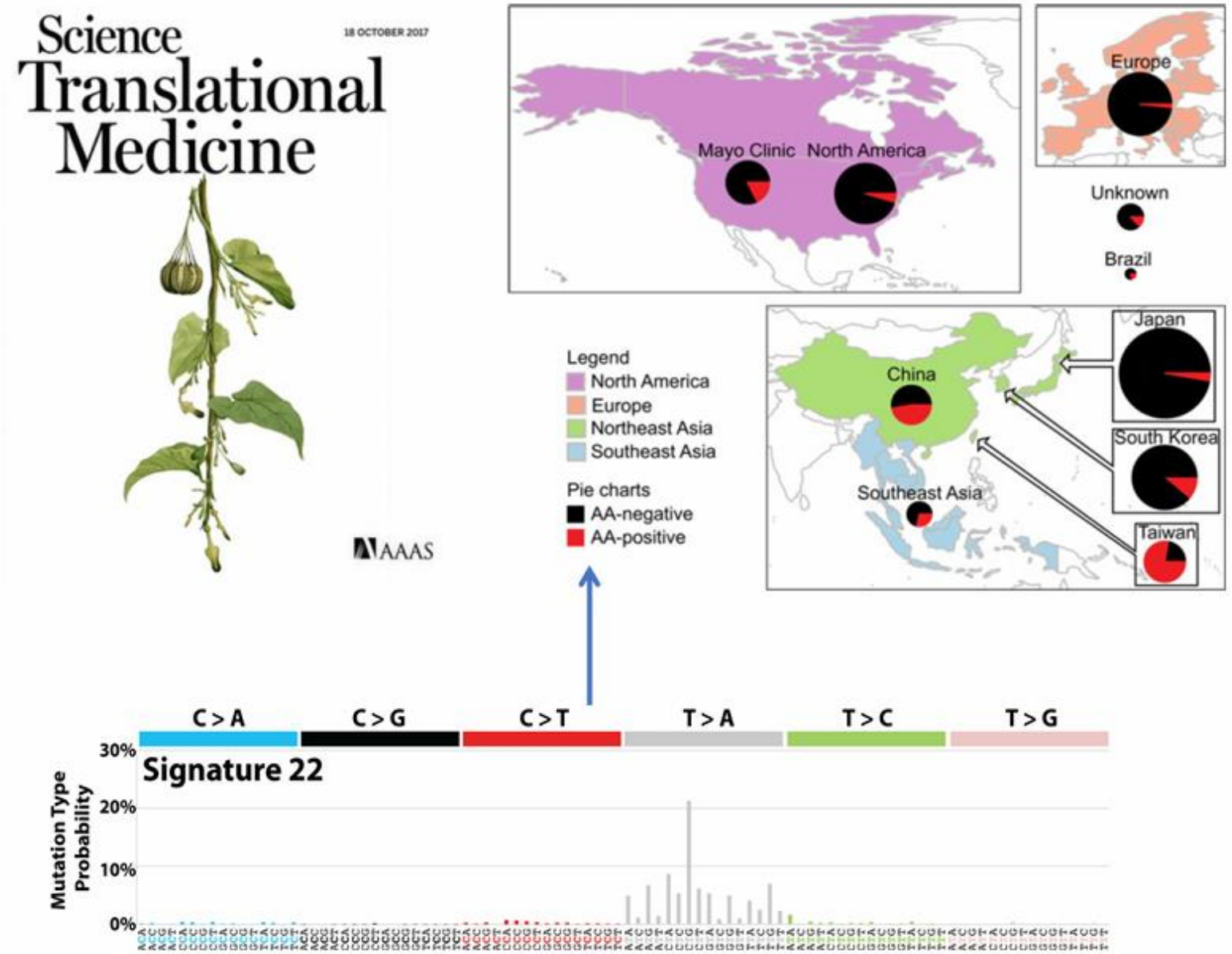
Residential exposure to ultraviolet light and risk of precursor B-cell acute lymphoblastic leukemia: assessing the role of individual risk factors, the ESCALE and ESTELLE studies

Authors | Authors and affiliations

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

Mutational signature analysis revealed the surprising association between UV light and some childhood leukemias (B-cell ALL), identifying UV light exoposure as a new potential risk factor for B-cell ALL.

Mutational signatures as a tool to quantify the relative importance of different carcinogens in specific cancer types



Quantification of the aristolochic acid mutational signature 22 in Asian liver tumours provide strong evidence that some traditional medicines (birthwort - 马兜铃科) are strong carcinogens and important drivers of liver cancer in East Asia

Ng *et al.* 2017 Science Translational Medicine

Mutational signatures as a tool to quantify the relative importance of different carcinogens in specific cancer types

nature

Explore content | About the journal | Publish with us | Subscribe

nature > articles > article

Article | Published: 02 July 2025

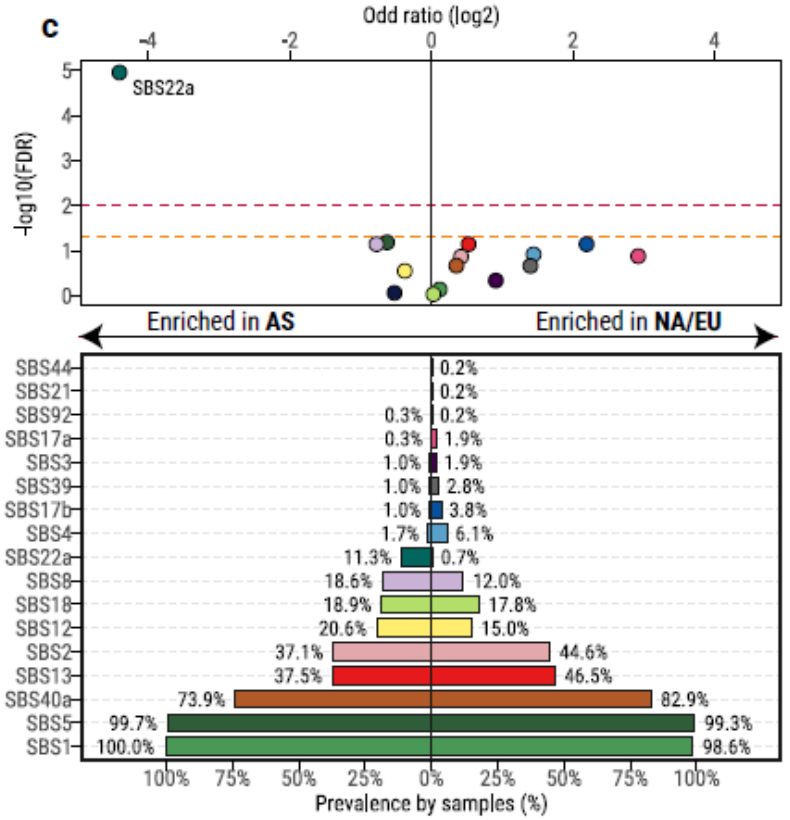
The mutagenic forces shaping the genomes of lung cancer in never smokers

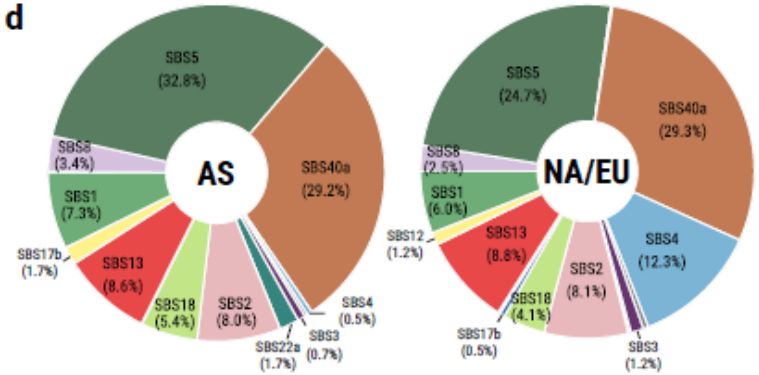
Marcos Díaz-Gay, Tongwu Zhang, Phuc H. Hoang, Charles Leduc, Marina K. Baine, William D
M. Sholl, Philippe Joubert, Azhar Khandekar, Wei Zhao, Christopher D. Steele, Burçak Otlu, S
Raviteja Vangara, Erik N. Bergstrom, Mariya Kazachkova, Oriol Pich, Charles Swanton, Chao
I-Shou Chang, Maria Pik Wong, Kin Chung Leung, Jian Sang, John P. McElderry, ... Maria Ter

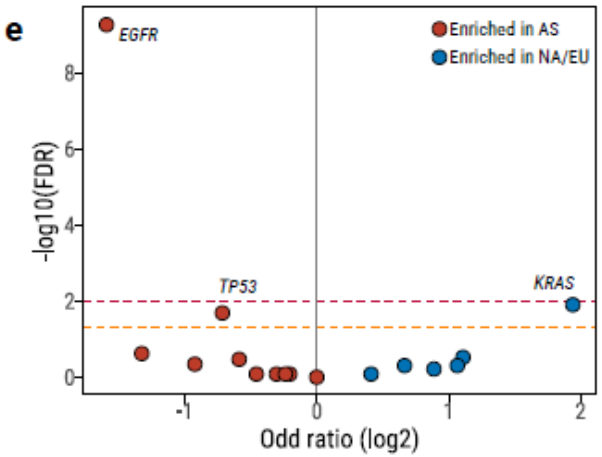
+ Show authors

Nature **644**, 133–144 (2025) | Cite this article

Quantification of mutational signatures in lung cancers from never smokers highlighted the role of **air pollution** (associated with SBS4 and SBS5), and also **aristolochic acid** (SBS22a) in Taiwanese patients.







Mutational signatures as a tool to quantify the relative importance of different carcinogens in specific cancer types

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

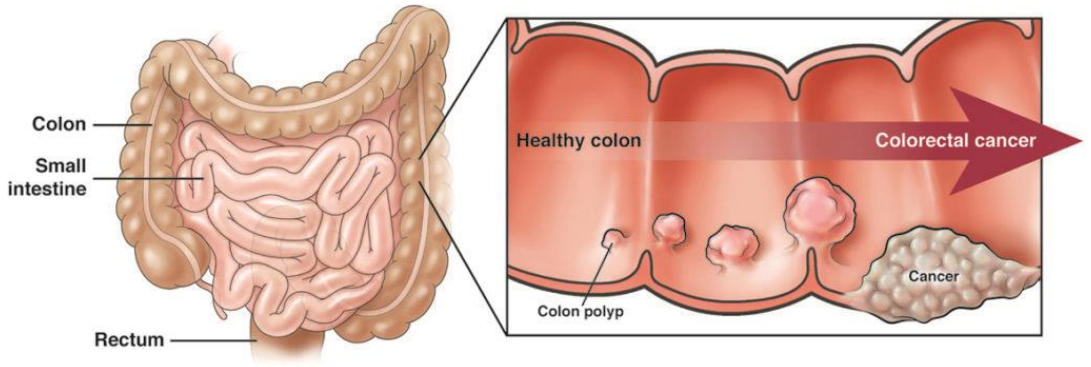
Article | Open access | Published: 23 April 2025

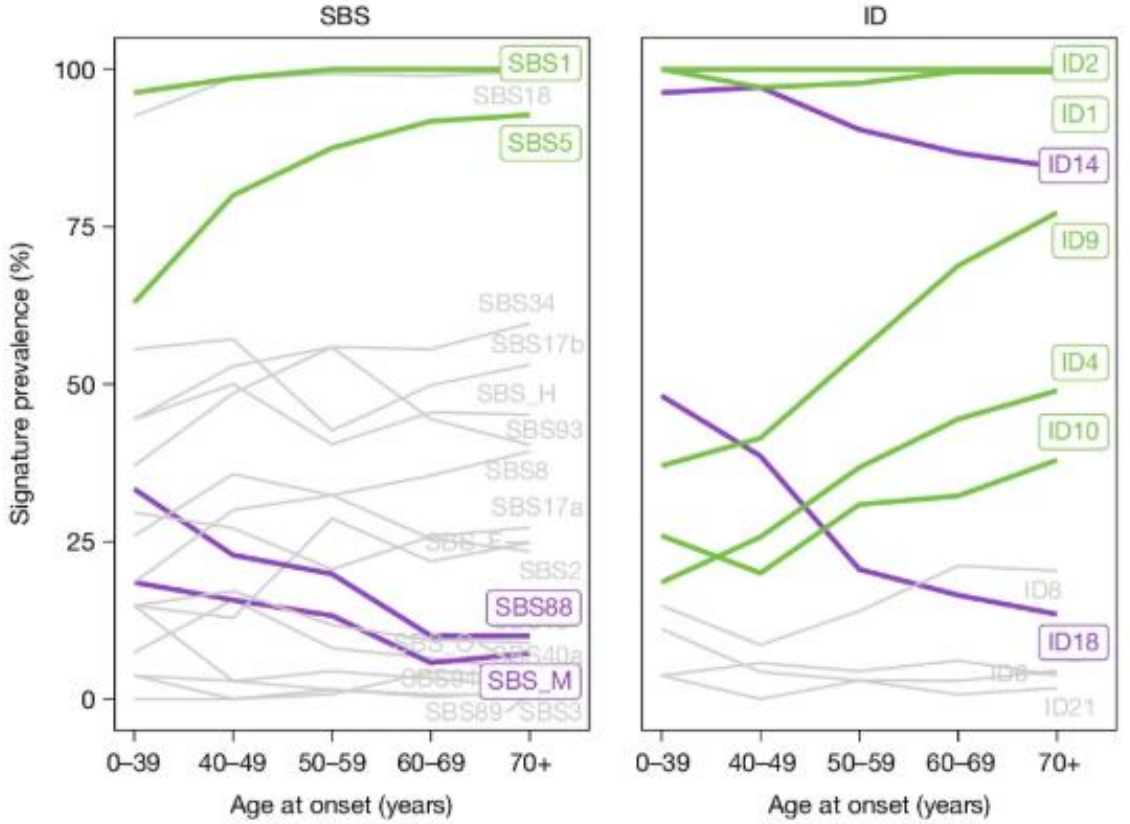
Geographic and age variations in mutational processes in colorectal cancer

Marcos Díaz-Gay, Wellington dos Santos, Sarah Moody, Mariya Kazachkova, Ammal Abbasi, Christopher D. Steele, Raviteja Vangara, Sergey Senkin, Jingwei Wang, Stephen Fitzgerald, Erik N. Bergstrom, Azhar Khandekar, Burçak Otlu, Behnoush Abedi-Ardekani, Ana Carolina de Carvalho, Thomas Cattiaux, Ricardo Cortez Cardoso Penha, Valérie Gaborieau, Priscilia Chopard, Christine Carreira, Saamin Cheema, Calli Latimer, Jon W. Teague, Anush Mukeriya, ... Ludmil B. Alexandrov ✉ + Show authors

Nature **643**, 230–240 (2025) | Cite this article

61k Accesses | **18** Citations | **2419** Altmetric | Metrics





Quantification of mutational signatures in colorectal cancer suggests that **colibactin** exposure (SBS88 and ID18) produced by gut microbiota is an important driver of **early-onset colorectal cancer** and is associated with *APC* mutations

Clinical and research applications

Discuss: How else might you use mutational signatures in your research?

Clinical and research applications

Discuss: How else might you use mutational signatures in your research?

- Track carcinogenic exposure over time
- Compare carcinogenic exposure between different populations
- Orthogonal validation of self-reported data

Summary and takeaways

- The COSMIC database compiles validated mutational signatures across several classes of mutations

- Mutational signatures in COSMIC are linked with specific aetiologies through association studies and experimental validation, but many signatures still have unknown aetiologies

- Mutational signatures do not yet have widespread clinical utility, but they are valuable research tools in cancer biology and cancer epidemiology

wellcome connecting science

thanks!

Please contact wellcomeconnectingscience.org for more information.