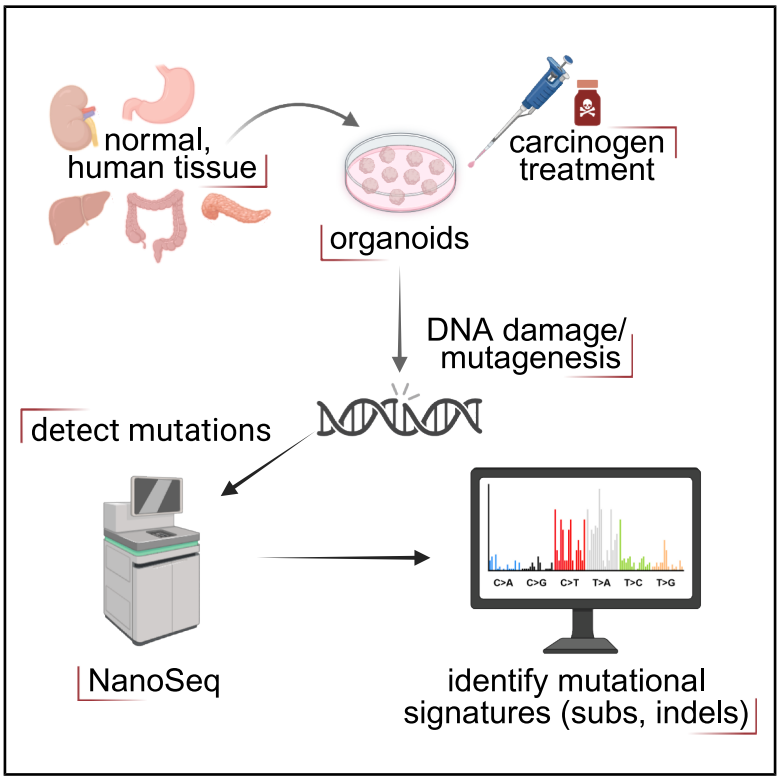
Cell Reports

Article

Mutational signatures of environmental carcinogens in human tissue organoids revealed by duplex sequencing

Graphical abstract



Highlights

- Mutations induced by environmental carcinogens in human tissue organoids
- Direct detection of mutations without clonal expansion
- Five different human tissues investigated—colon, stomach, liver, kidney, and pancreas
- Several incidences of strong concordance with human tumor signatures

Authors

Jill E. Kucab, Shuvro P. Nandi, Halh Al-Serori, ..., Michael R. Stratton, Ludmil B. Alexandrov, David H. Phillips

Correspondence

jill.kucab@kcl.ac.uk (J.E.K.), david.phillips@kcl.ac.uk (D.H.P.)

In brief

Mutational signatures of a panel of environmental carcinogens were determined in organoids derived from five different human tissues through duplex sequencing. Similarities with several signatures in human tumors were observed and the approach provides a basis for further investigating the environmental origins of human cancer.

Kucab et al., 2026, Cell Reports *45*, 117406
June 23, 2026
https://doi.org/10.1016/j.celrep.2026.117406

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Mutational signatures of environmental carcinogens in human tissue organoids revealed by duplex sequencing

Jill E. Kucab,[1,9,*] Shuvro P. Nandi,[2,9] Halh Al-Serori,[1,9] Ellie Dunstone,[3,9,10] Rebekah S.S. Beck,[1,11] Angela L. Caipa Garcia,[1] Eleanor C. Wilde,[1,14] Safa Saeed,[2] Hayley Francies,[3] Mathew J. Garnett,[3] Beiyuan Fu,[3] Fengtang Yang,[3,12] Kourosh Saeb-Parsy,[4] Meritxell Huch,[5] Jarno Drost,[6] Matthias Zilbauer,[7] Laura Humphreys,[3] Glen Kisby,[8] Volker M. Arlt,[1,13] Michael R. Stratton,[3] Ludmil B. Alexandrov,[2] and David H. Phillips[1,15,*]

[1]Department of Analytical, Environmental and Forensic Sciences, School of Cancer and Pharmaceutical Sciences, King's College London, 150 Stamford Street, London SE1 9NH, UK
[2]Department of Cellular and Molecular Medicine, UC San Diego, La Jolla, CA, USA
[3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK
[4]Department of Surgery, University of Cambridge, NIHR Cambridge Biomedical Research Centre, Cambridge, UK
[5]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[6]Princess Máxima Center for Pediatric Oncology, Oncode Institute, Utrecht 3584 CS, the Netherlands
[7]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK
[8]Department of Biomedical Sciences, Western University of Health Sciences, Lebanon, OR 97355, USA
[9]These authors contributed equally
[10]Present address: Quotient Therapeutics, Chesterford Research Park, Little Chesterford, Saffron Walden CB10 1XL, UK
[11]Present address: Pharmaceutical Sciences, pRED Innovation Center Basel, Hoffman-La Roche Ltd, Basel, Switzerland
[12]Present address: School of Life Sciences and Medicine, Shandong University of Technology, Zibo 255000, China
[13]Present address: Toxicology Department, SynTech Regulatory Europe GmbH, 69126 Heidelberg, Germany
[14]This publication is dedicated to the memory of our colleague Eleanor C. Wilde (1990–2019)
[15]Lead contact
*Correspondence: jill.kucab@kcl.ac.uk (J.E.K.), david.phillips@kcl.ac.uk (D.H.P.)
https://doi.org/10.1016/j.celrep.2026.117406

SUMMARY

Environmental exposures play a pivotal role in carcinogenesis, yet their molecular imprints in human tissues remain incompletely understood. Here, we present an extensive catalog of mutational signatures induced by a panel of environmental carcinogens using human tissue-derived organoids coupled with high-fidelity duplex sequencing (NanoSeq). This unique combination enables direct detection of mutations without clonal expansion and reveals consistent carcinogen-specific signatures across multiple organ types (i.e., colon, stomach, liver, kidney, and pancreas). We identify mutational signatures for agents such as benzo[*a*]pyrene, aflatoxin B_1, aristolochic acid I, and alkylating agents, some of which show strong concordance with known tumor signatures (e.g., SBS4, SBS11, SBS22, and SBS24) and previous experimentally-derived signatures. Our findings validate organoid models as physiologically relevant platforms for chemical mutagenesis and provide a foundational resource for decoding the environmental origins of human cancer.

INTRODUCTION

With the advent of massively parallel (whole genome) DNA sequencing, it has become clear that human tumors possess thousands of mutations that have accumulated during their evolution and growth. The Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have characterized more than 85 million mutations in around 23,000 human tumors from multiple cancer types.[1] Some of these mutations are drivers occurring in key genes involved in oncogenic processes, while many others are passengers at sites in the genome that do not result in any phenotypic change, but which, nevertheless, can collectively give clues to the processes that have caused them. These may be the consequence of endogenous processes, such as errors in DNA replication and defects in DNA repair, or they may be caused by exposure to endogenous or exogenous mutagens.[2–4] Using various algorithms, mutational events have been extracted from the raw sequencing data that reveal specific patterns, termed mutational signatures.[5]

The Catalogue of Somatic Mutations in Cancer (COSMIC)[6] and the Signal database[7] currently describe signatures for four

different mutation classes: single base substitutions (SBSs), presented in 96 contexts by considering the mutated base and the bases immediately 5′ and 3′; double base substitutions (DBSs), for which there are 78 strand-agnostic types; signatures comprising insertions and deletions (indels), categorized into 83 subclasses; and copy-number (CN) variation signatures, incorporating loss-of-heterozygosity status, total CN state, and segment length.

Of the 67 COSMIC SBS signatures from human cancer, around one-third are attributed to endogenous processes, another one-third are attributed to exogenous influences, and the remaining one-third are currently of unknown etiology. Similarly, of the 19 extracted DBS signatures, four are attributed to endogenous processes, four to environmental agents, and 11 have unidentified origins. Of the 23 indel signatures, six are attributed to endogenous processes, four to exogenous agents, and the remaining nine are of unknown etiology.[6]

It is important to generate mutational signatures in experimental systems to verify the assignments of signatures observed in human tumors and to investigate the origins of the many signatures not yet identified.[3,8] Signatures derived from systems that include *C. elegans* nematodes, *S. cerevisiae* yeast, mammalian primary cells, and immortalized cell lines[4] are also found in the COSMIC[6] and Signal[7] databases. We have previously treated mouse embryo fibroblasts[2] and human induced pluripotent stem cells (hiPSCs)[9] with environmental carcinogens and chemotherapeutic agents and have had some success in replicating COSMIC cancer signatures. We have now sought to use more physiologically relevant cell culture systems by using human tissue organoids.[10] These three-dimensional structures can be derived from normal tissues and retain some of the architecture and functions of the organs from which they are derived, thereby affording the opportunity to observe the effects of carcinogens in both their target and non-target organs. Organoids enable long-term propagation of normal cells in culture, making them an attractive model system to examine mutational signatures in human cells experimentally.

In the present study, we determined the mutational signatures of four environmental carcinogens, aristolochic acid I (AAI), benzo[*a*]pyrene (BaP), 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP), and aflatoxin B_1 (AFB_1) in organoids derived from five different human tissues, namely stomach, colon, kidney, pancreas, and liver tissues. We investigated an additional 13 environmental carcinogens in two gastric organoid lines. These additional carcinogens can damage DNA in different ways and all but one were formerly tested in hiPSCs; some of these agents induced mutational signatures in hiPSCs (e.g., ethylnitrosourea [ENU] and potassium bromate [$KBrO_3$]) and some did not (e.g., acetaldehyde and o-toluidine [*o*-TOL]).[9] Only methylazoxymethanol acetate (MAM) was not previously tested. Previously, we used whole genome sequencing to determine carcinogen-induced signatures in cell-based models, which required subcloning after treatment to identify mutations.[2,9] Here, we have bypassed this restriction by utilizing NanoSeq,[11] an error-corrected next-generation sequencing (ecNGS) method, enabling the detection of mutations in nonclonal samples and facilitating the extraction of mutational signatures more rapidly and efficiently than hitherto.

RESULTS

Organoid growth and treatment

Characterization of organoid models

To examine the mutational signatures of carcinogens in normal human cells, we used organoids derived from several different tissues that are susceptible to carcinogenesis: stomach, colon, kidney, pancreas, and liver (Figure 1A). Organoids from these tissues have been shown to contain a self-renewing stem cell population as well as tissue-specific cell types, as characterized previously (Table S1). Two donor lines each for stomach, colon, and kidney and one donor line each for pancreas and liver were included in the study. We confirmed that each organoid model was karyotypically normal using multiplex fluorescence *in situ* hybridization (M-FISH) (Table S5; Figure S1), except for the kidney line JD021, which carried a small translocation (1;9) in 1 out of 20 cells, and the colon line SC311, which exhibited tetraploidy in 1 out of 50 cells.

Organoid treatment for mutagenesis

To standardize the treatment regimen for mutation induction, two concentrations of each carcinogen of interest were selected for each organoid model, one close to the IC_{50} and one close to the IC_{75}, where possible (Figure 1B). Organoids were also treated with a range of solvent controls that matched the conditions used in the carcinogen treatments (including 0.1%–1% DMSO, 0.5 mM acetic acid, 2% ultrapure water, and culture media alone). Organoids were treated with carcinogens or solvent controls for up to 48 h and then expanded for 7–10 days to allow mutation fixation prior to NanoSeq (Figure 1D).

Mutagenesis by four carcinogens in different tissue organoids

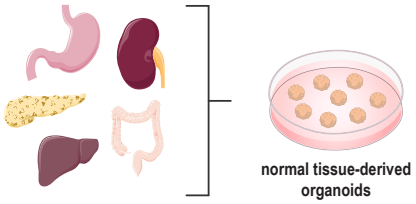
To assess and compare mutagenesis in the eight different organoid models, we treated them with four well-studied environmental carcinogens: AAI; BaP; PhIP or its reactive metabolite *N*-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (*N*-OH-PhIP); and AFB_1 (Figure 1B).

Cytotoxicity of carcinogens across the organoid models

To identify suitable concentrations for mutagenesis, organoid cell viability was measured after carcinogen treatment (Figure S2A; Table S6). Each of these chemicals requires metabolic activation to reactive intermediates that form promutagenic DNA adducts. These adducts can be cytotoxic if, for example, they block DNA replication or transcription or generate strand breaks during DNA repair.

Sensitivity to AAI and AFB_1 varied widely between tissue types. Colon was the most sensitive to both AAI and AFB_1 (AAI IC_{50} = 0.2–0.3 μM; AFB_1 IC_{50} = 0.7–2.9 μM), while kidney was the least sensitive (AAI IC_{50} = 147.7–148.0 μM; AFB_1 IC_{50} = 55.8–64.7 μM). Kidney was also the least sensitive to BaP treatment (BaP IC_{50} > 50.0 μM). When gastric and colon organoids were treated with PhIP (up to 250 μM), limited toxicity was observed (48%–85% viability at 250 μM). We, therefore, tested PhIP again (in gastric and colon organoids only) with rat liver S9 mix. The addition of S9 mix increased the toxicity of PhIP (gastric IC_{50} = 18.7–54.3 μM; colon IC_{50} = 12.9 μM) and led to a concomitant induction of DNA damage response markers (phospho-CHK2, phospho-p53, γH2A.X, and p21), suggesting

A



Tissue	Donor Line	Age (years)	Gender
Stomach/ Gastric	D88	76	Male
	D95	69	Male
Sigmoid colon	SC311	6	Male
	SC351	11	Female
Kidney; cortex	JD021	3	Female
	JD050	4	Female
Pancreas; ductal	D39	58	Male
Liver; ductal, undifferentiated	D4	44	Female

B

For mutation induction: [μM]

Group	Carcinogen	Gastric D88	Gastric D95	Colon SC311	Colon SC351	Kidney JD021	Kidney JD050	Liver D4	Pancreas D39
Mycotoxin	Aflatoxin B1 (AFB_1)	37.5, 75	6.25, 12.5	1.25, 2.5	0.5, 1	12.5, 25	12.5, 25	15, 20	2.25, 4.5
Phytochemical	Aristolochic acid I (AAI)	3.13, 6.25	25, 37.5	0.25, 1	0.25, 0.5	50, 100	50, 100	50, 75	50, 75
Polycyclic aromatic hydrocarbon	Benzo(a)pyrene (BaP)	25, 50	12.5, 25	12.5, 25	12.5, 20	25, 50	12.5, 25	25, 50	6.25, 12.5
Hetercyclic aromatic amine (metabolite)	2-Hydroxyamino-1-methyl-6-phenyl-imidazo(4,5-b)pyridine (*N*-OH-PhIP)	15, 17.5	3.75, 5	20, 30	15, 20	5, 6	3.75, 5	N/S	2.5, 5
Hetercyclic aromatic amine	2-Amino-1-methyl-6-phenyl-imidazo(4,5-b)pyridine (PhIP)	N/S	250.00	187.5, 250	187.5, 250	N/S	N/S	N/S	N/S
Hetercyclic aromatic amine	PhIP +S9	37.5, 50	17.5, 20	12.5, 50	12.5, 50	N/S	N/S	N/S	N/S

C

For mutation induction: [μM]

Group	Agent	Gastric D88	Gastric D95
Alkylating agent	1,2-Dimethylhydrazine (12DMH)	20000.00	10000, 20000
Alkylating agent	12DMH +S9	15000, 17500	12500, 15000
Alkylating agent	Ethylnitrosourea (ENU)	2500, 2750	1500, 2000
Alkylating agent	Methylazoxymethanol acetate (MAM)	750, 1000	250, 375
Alkylating agent	1-Methyl-3-nitro-1-nitrosoguanidine (MNNG)	50, 150, 200	37.5, 50, 100
Heterocyclic aromatic amine	2-Amino-3,8-dimethyl-imidazo[4,5-f] quinoxaline (MeIQx)	400, 500	300, 400
Heterocyclic aromatic amine	MeIQx +S9	400, 500	400, 500
Other (Epoxide)	Propylene oxide (PO)	40000, 60000	40000, 60000
Other (Epoxide)	Styrene oxide (SO)	500, 1000	500, 1000
Other (Aldehyde)	Acetaldehyde (ACE)	25000, 50000	50000, 75000
Other (Epoxide metabolite from acrylamide)	Glycidamide (GLY)	1500, 2000	1250, 1500
Other (Chlorination disinfection byproduct)	Mutagen X; 3-Chloro-4-(dichloro-methyl)-5-hydroxy-5H-furan-2-one (MX)	350	400
Other (Chlorination disinfection byproduct)	MX +S9	300, 350	300, 350
Other (Aromatic amine)	o-Toluidine HCl (o-TOL)	10000	7500, 10000
Other (Aromatic amine)	o-TOL +S9	7500, 10000	7500, 10000
Other (Oxidising agent)	Potassium bromate ($KBrO_3$)	3000, 4000	2000, 3000
Other (Oxidising agent)	Potassium chromate (K_2CrO_4)	20, 30	15, 20

For mutation induction: [μM]

Group	Agent	Colon SC311	Colon SC351
Alkylating agent	12DMH +S9	1800, 3000	1800, 3000
Alkylating agent	MAM	375, 500	375, 500

D

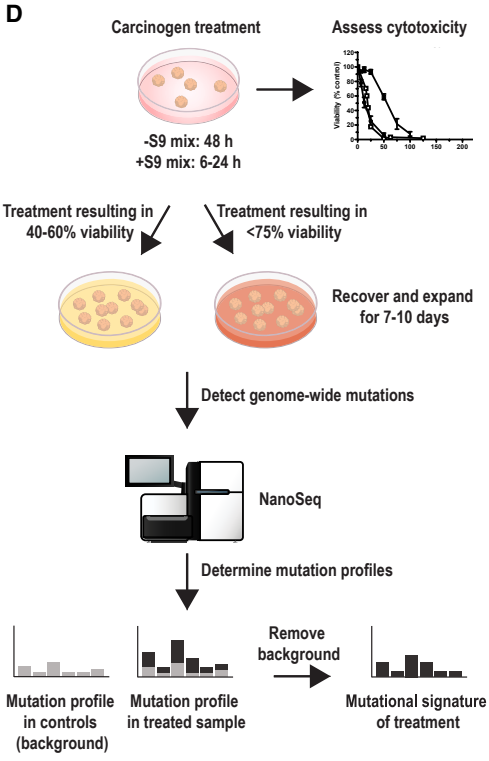


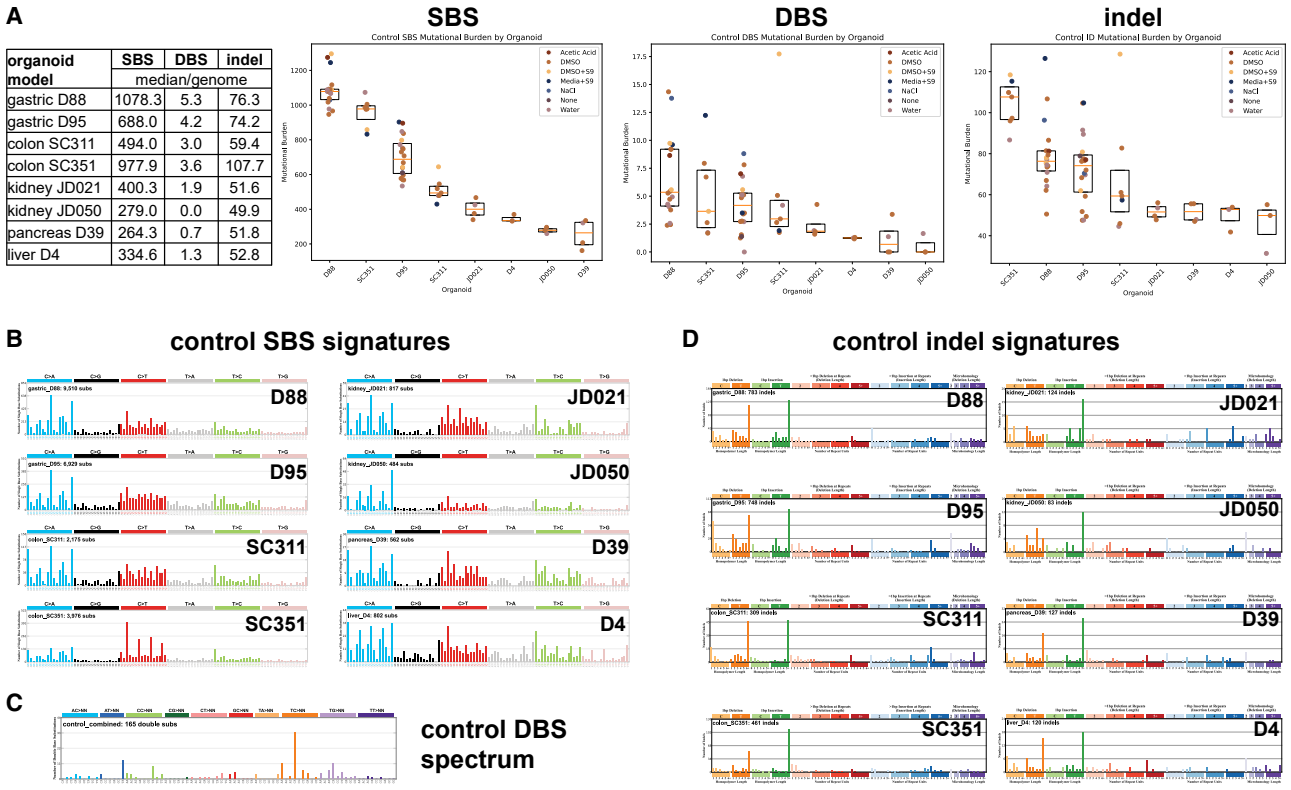
Figure 1. Summary of the organoid models and carcinogens assessed and the experimental workflow

(A) Information on the human tissue organoid models.

(B) Carcinogens tested in gastric, colon, kidney, pancreas, and liver organoids and the concentrations used to assess mutagenesis.

(C) Additional carcinogens tested in gastric or colon organoids and the concentrations used to assess mutagenesis.

(D) The experimental workflow. Organoid models were treated with carcinogens for 48 h without (−) S9 mix or 6–24 h with (+) S9 mix to assess cytotoxicity and determine effective concentrations. To induce mutations, organoids were treated with carcinogens at concentrations inducing 40%–60% or >75% cytotoxicity and then expanded for 7–10 days. DNA from carcinogen-treated or solvent control-treated organoids was sequenced by NanoSeq. Mutational signatures of carcinogens were extracted after background mutation subtraction using ExpSigFinder. Genome sequencer icon: Bioicons DBCLS https://togotv.dbcls.jp/en/pics.html, licensed under CC-BY 4.0; digestive system icons by Servier https://smart.servier.com/, licensed under CC-BY 3.0; and organoid culture icons by Marcel Tisch https://twitter.com/MarcelTisch licensed under CC0.



organoid model	SBS	DBS	indel
	median/genome		
gastric D88	1078.3	5.3	76.3
gastric D95	688.0	4.2	74.2
colon SC311	494.0	3.0	59.4
colon SC351	977.9	3.6	107.7
kidney JD021	400.3	1.9	51.6
kidney JD050	279.0	0.0	49.9
pancreas D39	264.3	0.7	51.8
liver D4	334.6	1.3	52.8

Figure 2. Background mutation burden and mutational signatures present in control samples of organoids derived from gastric, colon, kidney, pancreas, and liver tissues

(A) SBS, DBS, and indel burdens of control samples treated with solvents and/or rat liver S9 mix at equivalent concentrations to those used for carcinogen treatments. Points show the estimated number of mutations per genome, calculated by correcting raw SBS, DBS, or indel calls for total duplex sequencing coverage, colored according to solvent control treatment. Boxes indicate medians and 95% confidence intervals for each organoid model. Median values are summarized in the table on the left.

(B) SBS background signatures were estimated by aggregating SBS reads from all control samples of each organoid model.

(C) The DBS background spectrum is aggregated from control samples of all organoid lines.

(D) Indel background signatures were estimated by aggregating indels from all untreated control samples of each organoid model.

that S9 was able to activate PhIP (Figure S2B). To bypass the need for metabolic activation, gastric, colon, kidney, and pancreas organoids were treated with *N*-OH-PhIP, the reactive intermediate of PhIP, which was cytotoxic to all tissue types. Kidney, pancreas, and one gastric line (D95) were the most sensitive to *N*-OH-PhIP (IC_{50} = 3.7–6.9 μM).

Overall, while we observed similar levels of cytotoxicity between the two different colon and kidney organoid donor lines following carcinogen treatment, the two gastric donor lines generally responded differently to each other; gastric donor D88 was less sensitive to the carcinogens than gastric donor D95 (e.g., AFB1: D95 IC_{50} = 7.8 μM vs. D88 IC_{50} = 52.5 μM), except for AAI where the reverse was observed (i.e., AAI: D95 IC_{50} = 21.9 μM vs. D88 IC_{50} = 2.5 μM).

Mutations in control cultures

To establish the background mutational burden and spectra in each model, we examined SBS, DBS, and indels in DNA from solvent control-treated organoids (Figure 2A; Table S7). In total, we produced 66 control samples (gastric D88 = 18, gastric D95 = 20, colon SC311 = 7, colon SC351 = 7, kidney JD021 = 4, kidney JD050 = 3, pancreas D39 = 4, and liver D4 = 3). The median number of SBS/genome in the control samples varied 4-fold between the organoid models, ranging from 264 in kidney JD050 to 1078 in gastric D88. Very few DBS were found in the controls of any organoid model (median burden from 0.68 to 5.3 DBS/genome). Median control indel burdens varied about 2-fold, from 49.9 indel/genome in kidney JD050 to 107.7 indel/genome in colon SC351. SBS, DBS, and indel burdens in control samples showed a moderate positive correlation (SBS-DBS: 0.56, SBS-indel: 0.59, and DBS-indel: 0.47; $p < 0.001$, Pearson's product moment correlation) (Figure S3).

SBS and indel signatures of background mutagenesis were estimated for each organoid model using the mean duplex coverage-corrected SBS or indel mutation catalogs of all control samples for each model, similar to our previous work[9] (Figures 2B and 2D). Due to very low DBS burdens in control samples, the spectra for individual organoid models were not informative for estimation of background DBS signatures. DBS were, therefore, aggregated across all control organoid samples to compile a background spectrum (Figure 2C).

The background SBS and indel mutation profiles between different models were compared by calculating pairwise cosine

similarities (cossim) (Figure S3). The SBS control profiles of most of the organoid lines were highly similar (cossim = 0.83–0.97), consisting mostly of C>A, C>T, and T>C mutations (throughout this article, we have followed the convention of stating mutations with reference to the pyrimidine bases, C and T; however, it should be noted that for many of the agents described, covalent modification occurs to the purine bases, G and A, from which mutations result). However, the colon organoid line SC351 was less similar to the other lines (cossim = 0.69–0.83) due to a greater proportion of C>T mutations. The background indel signatures from the gastric, colon, pancreas, and liver organoid lines showed high-to-very-high similarity (cossim = 0.83–0.94), consisting predominantly of 1-bp T insertions and deletions at >6-bp homopolymer sequences, while the kidney lines JD021 and JD050 were less similar to the other organoids (cossim = 0.71–0.81).

Due to the polyclonal nature of each organoid sample, the background signatures extracted from control samples following NanoSeq reflect mutations accrued during a human donor's lifetime in addition to mutations that arose during organoid culture. To investigate sources of background mutagenesis, a subset of reference signatures from COSMIC were fitted to the control sample mutation catalogs (see STAR Methods and Figure S4).[1,6] We found that the background SBS signatures were composed of COSMIC signatures SBS1, SBS5, SBS18, and SBS40 (Figure S4A). SBS1 is a clock-like signature associated with spontaneous deamination of 5-methylcytosine.[12] SBS5 and SBS40 are also considered clock-like, but their origins are unknown.[5] SBS18, reportedly caused by reactive oxygen species (ROS), has been observed previously in cultured cells.[5,9] The background indel signatures (IDs) were predominantly composed of COSMIC ID1, ID2, ID3, and ID5 (Figure S4B). ID1 and ID2 are clock-like signatures associated with DNA slippage during replication, ID3 is associated with tobacco smoking[1], and the etiology of ID5 is unknown, although it also shows clock-like accumulation with age.[1]

Mutation burden and putative mutational signature analysis of four carcinogens across the organoid models

To assess carcinogen-induced mutagenesis in the organoids, we first determined the mutational burden of SBS, DBS, and indels in DNA from treated samples (Figure 3A; Table S8). Samples with mutation calls lower than our set thresholds of 100 for SBS and 20 for DBS and indels were excluded (Table S11; also see data and code availability). In terms of SBS burden, AAI was the most mutagenic agent in this sample set (mutagenicity index [MI]: 4.1–54.1), inducing 1,226–34,231 SBS/genome. Interestingly, the colon organoid models accumulated many more AAI-induced mutations than the kidney models, despite the kidney being a target organ of AAI; the relationship between mutation burden and dose or organoid model is explored in greater detail later. BaP was also highly mutagenic to each organoid model (6,335–16,579 SBS/genome; MI: 10.5–42.3), followed by AFB_1 (1,133–4,892 SBS/genome; MI: 3.0–11.3), *N*-OH-PhIP (1,091–3,886 SBS/genome; MI: 3.0–7.9), and PhIP+S9 (252–2,327 SBS/genome; MI: 1.3–4.3). For DBS, only AAI (34–145 DBS/genome, with three samples excluded due to insufficient DBS) and BaP (27–122 DBS/genome) induced enough mutations for signature analysis. We also observed a significantly increased indel burden in all organoid lines treated with AAI (55–498 indels/genome, MI: 2.2–7.3; one sample excluded), BaP (188–833 indels/genome, MI 4.1–17.1), *N*-OH-PhIP (52–290 indels/genome, MI: 2.0–4.7), and PhIP+S9 (50–109 indels/genome, MI: 1.5–2.5). AFB_1 induced a significant increase in indels in five models (28–113 indels/genome, MI: 1.6–3.2), while the indel burden induced by PhIP without S9 was significant only in one sample from the colon model SC351 (38 indels/genome, MI 1.4).

Putative signatures were then extracted from the treated samples that harbored mutation burdens significantly higher than controls, and signatures that did not show a significant difference to the respective background signature (or random noise, in the case of DBS) were excluded (Table S11; also see data and code availability). We thus identified putative SBS signatures in all organoid models treated with AAI, BaP, PhIP+S9, *N*-OH-PhIP, and AFB_1, as well as DBS signatures for AAI (13/16 samples) and BaP (14/15 samples) and indel signatures for AAI (15/16 samples), BaP (15/15 samples), PhIP+S9 (8/8 samples), and *N*-OH-PhIP (14/14 samples) (refer to "Putative Signatures" files at Zenodo: https://doi.org/10.5281/zenodo.15911686). The remaining putative signatures were inspected for similarities between samples treated with the same carcinogen (Figures S6 and S7). Samples that exhibited poor similarity to other samples from the same treatment group (i.e., did not share cossim $\geq$0.8 with at least one other sample) were excluded from final signatures (Table S11).

Although mutagenesis for each carcinogen differed quantitatively across the models, the SBS signatures were highly similar in each tissue type and donor line, as shown by cosine similarity-based clustering (AAI: cossim = 0.95–0.99, AFB_1: cossim = 0.93–0.99, BaP: cossim >0.99, *N*-OH-PhIP: cossim = 0.96–0.99, and PhIP+S9: cossim = 0.78–0.99 for 7/8 samples [Figure S7A]). Extracted DBS signatures were also highly similar in the samples treated with AAI (cossim = 0.66–0.92) or BaP (cossim = 0.68–0.98), as were the putative indel signatures for AAI (cossim = 0.69–0.98), BaP (cossim = 0.90–0.99), and *N*-OH-PhIP (cossim = 0.65–0.98, excluding one outlier: D88 + *N*-OH-PhIP [15 μM]). For PhIP+S9 treatment, only samples from gastric organoids harbored comparable indel signatures (cossim = 0.80–0.92 for 4/8 samples). Compound-associated mutations from all concordant samples treated with each individual carcinogen were thereafter combined to generate our final, refined signatures described in the following sections (Figure 3B; Table S11).

Final mutational signatures of AAI, BaP, activated PhIP, and AFB_1 in organoids

AAI undergoes metabolic activation to *N*-hydroxyaristolactam I, which forms DNA adducts, mainly 7-(deoxyadenosin-N^6-yl)-aristolactam I.[13] Such modifications on adenine tend to mispair with adenine when DNA replicates, resulting in A>T (or T>A) transversions. Treatment of organoids with the phytochemical AAI, found in the *Aristolochia* species used in traditional herbal medicine, generated the expected SBS signature of T>A mutations, with a dominant sequence context of GTC. A DBS signature for AAI composed of several peaks (including TC>AA, TG>AT, TT>AA, and TA>AT/CT/GT) and an indel signature predominantly consisting of 1-bp deletions of T or C were also detected.

BaP, a polycyclic aromatic hydrocarbon found in urban air, tobacco smoke, and cooked food, requires metabolic activation

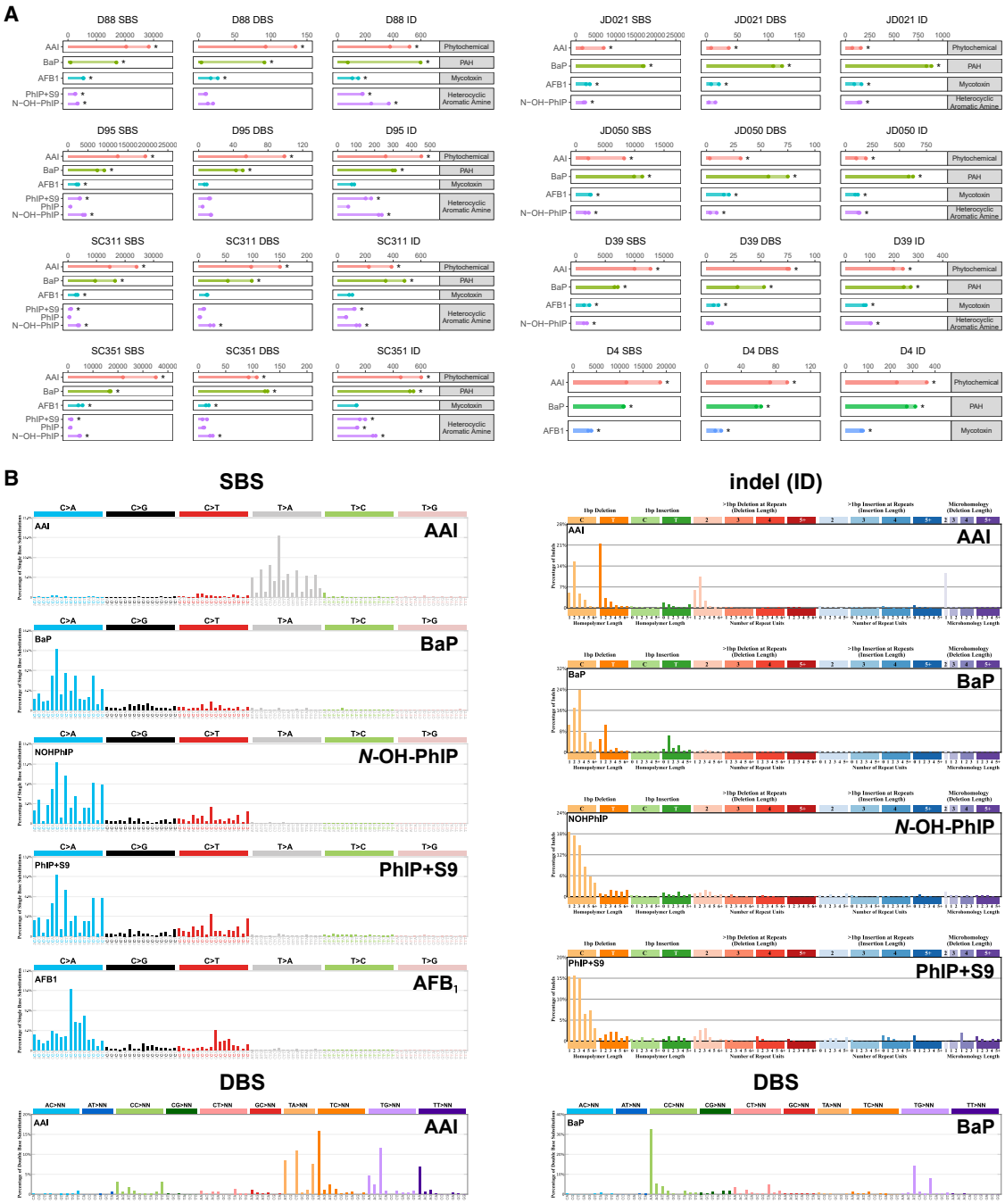


Figure 3. Mutational signatures of AAI, BaP, PhIP, and AFB$_1$ in human tissue organoids

(A) Genome-wide burdens of SBS, DBS, and indels in eight organoid models treated with AAI, BaP, PhIP, PhIP+S9, *N*-OH-PhIP, or AFB$_1$. Asterisks indicate a significant increase in mutation burden in one or more treated samples compared to the controls. Colored points show the estimated number of mutations per genome for each sample, calculated by correcting raw mutation calls according to total duplex sequencing coverage. Error bars in black show 95% confidence intervals of individual sample burdens for all samples. Full data on mutation burdens for all samples are given in Table S8.

(B) Final mutational signatures of SBS, DBS, or indels present in carcinogen-treated organoids. Final signatures were extracted from a weighted mean of the concordant, per-sample putative signatures of each treatment group.

by cytochrome P450 (CYP) 1A1, CYP1B1, and epoxide hydrolase to exert its genotoxicity.[14] The ultimate reactive species BaP-7,8-diol-9,10-epoxide (BPDE) forms bulky DNA adducts preferentially at the N^2 position of guanines.[15] Bulky adducts formed on guanine tend to mispair mainly with adenine during DNA replication, resulting predominantly in G>T (or C>A)

transversion mutations. All organoid models tested here were able to activate BaP, unlike in previous studies (e.g., hiPSCs[9]) where activation by S9 mix was required. The SBS signature for BaP in the organoids featured mainly mutations at guanines, predominantly C>A transversions at CC dinucleotides, in various contexts. BaP also generated a DBS signature consisting mostly of CC>AA, with some TG>AT/CT mutations, and an indel signature consisting mostly of 1-bp C deletions but with some 1-bp T deletions or insertions.

PhIP, a carcinogenic heterocyclic aromatic amine produced during the grilling of meat and fish, was not toxic or mutagenic to the organoids on its own. As PhIP undergoes metabolic activation to create the reactive intermediate *N*-OH-PhIP,[16] which forms adducts principally at C8 of guanine, these organoid cultures likely do not express sufficiently the required metabolic enzymes (e.g., CYP1A2). However, treatment with the reactive metabolite *N*-OH-PhIP, or with PhIP and rat liver S9 mix, was cytotoxic and mutagenic to the organoids. Interestingly, the SBS signatures of PhIP+S9 and *N*-OH-PhIP (collectively referred to as "activated PhIP") were similar to that of BaP (cossim = 0.94–0.96), suggesting a similar sequence context for DNA adduct formation and/or translesion bypass of their adducts, despite the different structures of the compounds and sites of modification of guanine in DNA. However, while BaP caused a distinct DBS signature in organoids, a DBS signature was not extracted for activated PhIP. Treatment with PhIP+S9 or *N*-OH-PhIP generated an indel signature that was somewhat similar to the indel signature of BaP (cossim = 0.86–0.88) but exhibited more 1-bp C deletions at homopolymer lengths of 1 or 2, and more T deletions at homopolymers of 6 or more.

AFB_1, a mycotoxin and potent liver carcinogen produced by *Aspergillus flavus* and *Aspergillus parasiticus*, is activated by CYP3A4 and CYP1A2 to produce the highly reactive metabolite AFB_1-8,9-epoxide, which can form DNA adducts at N7-Gua and cause primarily G>T (or C>A) mutations.[17] As expected, the SBS signature for AFB_1 in organoids featured mostly C>A transversions with some C>T. While AFB_1, BaP, and PhIP all form adducts at guanines, the sequence context for AFB_1 C>A mutations was quite different from those of BaP and activated PhIP, with most C>A mutations induced by AFB_1 occurring in the context of GCN and a dominant peak of GCA. AFB_1 treatment did not generate a DBS or indel signature.

This initial analysis demonstrated that organoid lines from all five tissue types were able to activate several procarcinogens, rat liver S9 mix supplementation was useful for activating compounds that require additional metabolic enzymes (e.g., PhIP), NanoSeq could be used to detect carcinogen-induced mutations in polyclonal organoid samples, and the tested carcinogens induced the same mutational signature in each tissue type.

Mutagenesis by additional carcinogens in organoids

Because of signature consistency in different tissues for the four carcinogens tested above, we further tested 13 carcinogens mostly in gastric organoids only (Figure 1C; Table S3). Testing focused on gastric organoids because they exhibited a consistent growth rate; were easier to culture than colon, liver, or pancreatic organoids; and were generally more sensitive to treatment than kidney organoids. The panel of carcinogens included alkylating and oxidizing agents, several chemicals used in manufacturing, and chemicals produced during food cooking. Two agents, 1,2-dimethylhydrazine (1,2-DMH) and MAM, were also examined in colon organoids, as these chemicals are potent colon carcinogens in animal models.[18,19]

Cytotoxicity of the carcinogens in gastric or colon organoids

As before, we first examined the effect of carcinogen treatment on viability (Table S9; Figure S5). Four chemicals that require activation by enzymes enriched in the liver were tested with (+S9) and without rat liver S9 mix (1,2-DMH, 3-chloro-4-(dichloromethyl)-5-hydroxy-5H-furan-2-one [Mutagen X, MX]; 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline [MeIQx]; and *o*-TOL) and the remainder were tested only without rat liver S9 mix. We again found that the gastric line D95 was more sensitive than D88 to most agents in the panel. *o*-TOL without S9 mix did not reach an IC_{50} in either gastric model when tested at maximum solubility (i.e., 10 mM). Addition of S9 mix somewhat lowered the IC_{50} in gastric organoids for 1,2-DMH and MX (e.g., 1,2-DMH-S9 D88 IC_{50} = 15.2 mM vs. 1,2-DMH+S9 D88 IC_{50} = 12.1 mM) but had little to no effect on the toxicity of MeIQx or *o*-TOL. The colon organoids were more sensitive to 1,2-DMH+S9 treatment than the gastric models (e.g., SC311 IC_{50} = 4.0 mM vs. D88 IC_{50} = 16.1 mM) and more sensitive to MAM than the gastric line D88 (e.g., SC311 IC_{50} = 381.5 μM vs. D88 IC_{50} = 1166.0 μM).

Mutation burden and putative mutational signature analysis for the additional carcinogens

A significant increase in SBS burden was observed after treatment with all 13 carcinogens in one or more samples (Figure 4A; Table S10). Three alkylating agents were the most mutagenic compounds tested in this group: ENU (MI: 6.1–10.8), *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG, MI: 2.4–19.2) and MAM (MI: 11.2–64.0). While D88 and D95 gastric organoids treated with ENU harbored a similar number of SBS/genome (5,521–6,012 and 5,541–6,846; respectively), MNNG was much more mutagenic to the D88 gastric organoids than it was to D95, inducing 5,452–19,745 SBS/genome in D88 and 946–1,910 SBS/genome in D95. MAM was also most mutagenic in the D88 gastric model (SBS/genome: D88 max = 68,442; D95 max = 38,849; SC351 max = 18,371; and SC311 max = 10,178). Only 2 of the 13 additional carcinogens induced a significant increase in DBS burden: potassium chromate (KChr, max = 37 DBS/genome) and MAM (max = 86 DBS/genome). Indel burdens were significantly increased after treatment with 7 of the 13 carcinogens in more than one sample.

Putative signatures were extracted from the treated samples that harbored significantly increased mutation burdens; those that did not exhibit significant difference from the respective background signature (or random noise, in the case of DBS) were excluded from final signatures (Table S11). The remaining putative signatures were inspected for similarities between samples treated with the same carcinogen; although in some cases, only a single sample remained from a treatment group (Figure S7B). Final refined signatures were extracted from a weighted mean of the concordant putative signatures of each treatment group and had to consist of at least 3 samples to be considered stable (Figure 4B). Single samples and samples

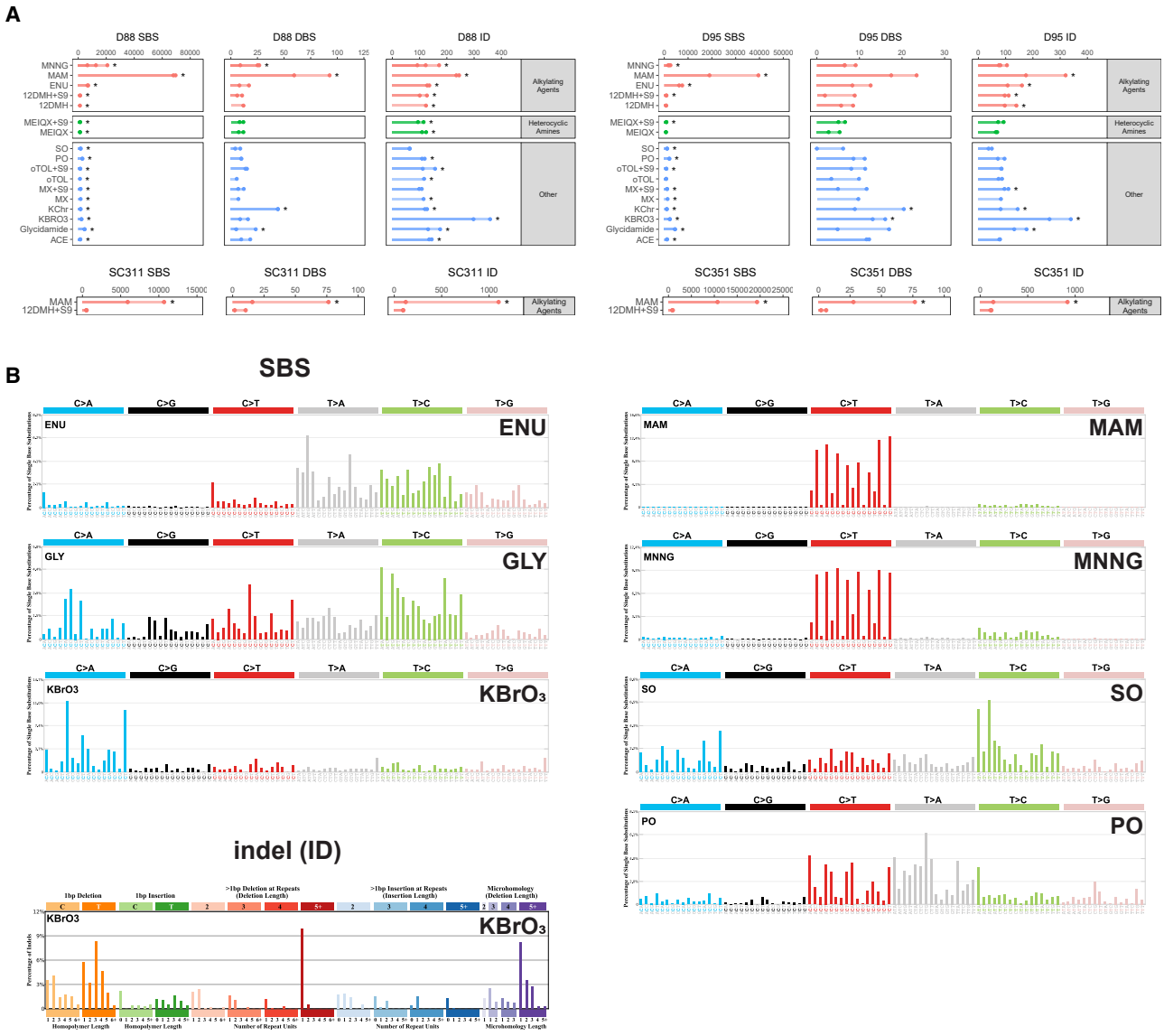


Figure 4. Mutational signatures induced by alkylating agents, glycidamide and $KBrO_3$ in human tissue organoids

(A) Genome-wide burdens of SBS, DBS, and indels in gastric or colon organoid models treated with up to 13 carcinogens, without or with S9 mix (+S9). Asterisks indicate a significant increase in mutation burden in one or more treated samples compared to the controls. Colored points show the estimated number of mutations per genome for each sample, calculated by correcting raw mutation calls according to total duplex sequencing coverage. Error bars in black show 95% confidence intervals of individual sample burdens for all samples. Full data on mutation burdens for all samples are given in Table S10.

(B) Final mutational signatures of SBS or indels present in carcinogen-treated organoids. Final signatures were extracted from a weighted mean of the concordant, per-sample putative signatures of each treatment group.

with putative signatures that exhibited poor similarity to other samples from the same treatment group (i.e., did not share cosine similarity $\geq$0.8 with at least one other sample) were excluded from final refined signatures (Table S11; also see data and code availability). On this basis, refined SBS signatures were identified for ENU, MNNG, MAM, propylene oxide (PO), styrene oxide (SO), glycidamide, and $KBrO_3$. No refined DBS signatures were extracted for these treatments, with the majority of samples excluded due to low mutation burden. Refined indel signatures were identified for $KBrO_3$, acetaldehyde, and MAM (however, these were based on only two samples each for acetaldehyde and MAM, and are not considered stable signatures). All putative signatures can be found at Zenodo: https://doi.org/10.5281/zenodo.15911686.

Mutational signatures induced by alkylating agents in organoids

Alkylation damage is common in DNA and can originate from endogenous or exogenous sources, including the diet, tobacco smoke, occupational exposures, and chemotherapy.[20] We examined the mutagenic effects of DNA alkylation in organoids

using ENU, MNNG, 1,2-DMH, MAM, PO, and SO, and extracted a mutagen-specific mutational signature for all except 1,2-DMH.

ENU and MNNG are nitrosamides that spontaneously decompose in aqueous medium into alkylating species.[21,22] Most mutations induced in organoids by ENU were either T>A (with dominant peaks at the sequence contexts of ATG and GTG) or T>C, along with some T>G and C>T mutations. Ethylation by ENU occurs at nucleophilic oxygen sites, primarily the O^2-and O^4-position of thymine, which have been reported to cause T>A transversions and T>C transitions, respectively.[23,24] ENU can also form O^6-ethylguanine to a lesser extent, which causes C>T transitions.[25] Most MNNG-induced mutations in organoids were C>T transitions, followed by T>C. MNNG decomposes in media into methyldiazonium ions that methylate DNA, primarily forming N7-methylguanine (N7-meG), N3-methyladenine (N3-meA), and O^6-methylguanine (O6-meG).[22] While N7-meG itself is not promutagenic, O6-meG will readily mispair with thymine and cause C>T mutations.[26] The T>C mutations induced by MNNG were likely caused by the mispairing of N3-meA with cytosine. Interestingly, the T>C component of the MNNG signature was more prominent in D95 samples treated with lower concentrations of the compound, particularly 37.5–50 μM (Figure S8).

1,2-DMH is a colon carcinogen that is chemically related to the naturally occurring carcinogen cycasin, an azoxyglycoside produced by palm-like cycad plants.[27] 1,2-DMH, like cycasin, requires metabolic conversion to the proximate carcinogen, MAM, which is then converted to a methyldiazonium ion—the same ultimate carcinogen described above for MNNG.[28] MAM was highly mutagenic to both gastric and colon organoids, with an SBS signature composed mainly of C>T transitions, followed by a smaller number of T>C mutations. As observed for MNNG, the C>T mutations induced by MAM generally occurred with a 3′ neighboring C or T. The MAM signature was essentially identical to that of MNNG (cossim = 0.98), reflecting the same mechanism of action. 1,2-DMH (±S9), however, induced a much smaller number of SBS than MAM. Putative signatures from 1,2-DMH+S9 treatment were dissimilar to MAM or MNNG signatures (Figure S7B). This suggests that 1,2-DMH was not metabolically converted into methyldiazonium ions in our organoid models even with the addition of rat liver S9 mix.

PO, used in the production of polyurethane and propylene glycols, can also alkylate DNA and forms several 2-hydroxyalkyl (HA) adducts (N7-HA-Gua > N3-HA-Ade > N3-HA-Cyt > N1-/N6-HA-Ade).[29] PO was toxic and mutagenic at very high concentrations to both gastric organoid models (40–60 mM), and generated an SBS signature of predominantly T>A and C>T, with some T>C and T>G. SO is a reactive metabolite of styrene, used in the manufacturing of plastics, resins, and polyester.[30] In addition to occupational exposure, styrene can be inhaled from cigarette smoke.[31] SO induced an SBS signature of T>C (dominant peaks at contexts of ATA and ATG), along with C>A, C>T, and T>A mutations. This reflects the ability of SO to alkylate both guanine (N7-, N^2-, and O^6-positions) and adenine (N3-, N1-, and N^6-positions).[32]

No refined DBS signatures for the alkylating agents were identified. Putative DBS signatures were extracted from three MAM-treated samples (D88 + 1000 μM, SC311 + 500 μM, and SC351 + 500 μM), but they were not similar to each other. MAM also induced putative indel signatures in 7 of 8 samples, but only two of these samples (SC311 and SC351 + 500 μM MAM) exhibited cossim >0.8 and were included in a final signature (not considered stable and presented in Figure S9). This MAM indel signature consisted mainly of T insertions and deletions in homopolymeric regions of different lengths as well as C deletions.

Mutations induced by other environmental carcinogens

Of the remaining carcinogens tested, glycidamide (SBS/genome: 3,347–3,877; MI: 4.1–6.5) and $KBrO_3$ (SBS/genome: 1,118–1,695; MI: 2.0–3.4) were the most mutagenic and generated similar putative SBS signatures within all samples of each treatment group, enabling the extraction of refined signatures for both compounds (Figure 4B; Table S11). Putative SBS signatures were extracted from samples treated with MeIQx ± S9 (SBS/genome: 2–273, MI: 1.0–1.3), MX (SBS/genome: 472–692, MI: 1.4–2.0), *o*-TOL±S9 (SBS/genome: 99–414, MI: 1.1–1.4), KChr (SBS/genome: 352–599, MI: 1.5–1.6), and acetaldehyde (SBS/genome: 273–526, MI: 1.3–1.8); however, they did not exhibit sufficient similarity within treatment groups (–S9 and +S9 treatments were analyzed separately). Putative indel signatures were identified in all $KBrO_3$-treated samples (ID/genome: 188–280, MI: 3.6–4.7) and in some acetaldehyde-treated samples (ID/genome: 4–65, MI: 1.1–1.8).

Glycidamide is an epoxide generated from the CYP2E1-catalyzed metabolism of acrylamide, a suspected dietary carcinogen formed in baked or fried starch-rich food.[33] Glycidamide-DNA adducts occur at N7-Gua and N3- or N1-Ade.[34,35] The refined glycidamide SBS signature extracted from treated organoids contained diverse mutational patterns and sequence contexts. It was dominated by T>C, with peaks at several contexts, but also consisted of T>A, C>T, and C>A. The T>C mutations were likely caused by miscoding adducts at N1-Ade, which are the most commonly identified adenine adducts induced by glycidamide *in vitro*.[36] The T>A and C>A mutations likely resulted from depurinating adducts at N3-Ade and N7-Gua, respectively, also known to be induced by glycidamide.[36] Refined DBS and indel signatures could not be extracted from the glycidamide-treated organoids.

$KBrO_3$ is an oxidizing agent used as a flour improver in the United States and formed as a byproduct of water disinfection by ozonation.[37,38] It is a nephrotoxin and a known rodent carcinogen that induces the formation of oxidized DNA lesions such as 8-hydroxydeoxyguanosine (8-OHdG), which can mispair with adenine to cause G>T transversions.[39,40] Here, we found that $KBrO_3$ caused an SBS signature consisting mainly of C>A mutations, with dominant peaks at CCA, CCT, GCA, and TCT. $KBrO_3$ also induced an indel signature in the organoids, which included 1-bp deletions of T and some C, as well as larger deletions. Previous studies have also observed deletions and C>A transversions induced by $KBrO_3$.[41,42]

As mentioned above, we did not observe similarity in the putative SBS signatures extracted from 1,2-DMH-, MeIQx-, or MX-treated D88 and D95 samples when separated into –S9 or +S9 treatment groups. However, the putative SBS signatures in D88 organoids alone treated with 1,2-DMH (+S9), MeIQx (±S9), and MX (±S9) were similar to each other (cossim >0.8; Figure S9). As this similarity was only observed for one organoid model and occurred after treatment with three dissimilar

compounds, with or without S9, these signatures appear to be organoid-specific rather than agent-specific. These signatures were also very similar to the D88 control signature (cossim = 0.86–0.91). Additionally, an indel signature was extracted from acetaldehyde-treated D88 samples but not from D95 samples, despite the very high concentrations used in both (D88: 25–50 mM and D95: 50–75 mM). The acetaldehyde signature in D88 organoids consisted mainly of 1-bp deletions of T or C. Final signatures that were found in only a single organoid model and/or were not consistent between at least three samples from the same treatment group are not considered stable and are presented in Figure S9.

Concentration-dependent effect on mutagenesis and variation between organoid lines

Across this study most compounds were tested at two concentrations, which were selected according to the model-specific cytotoxicity of each compound (Figure 1). In many cases, the individual organoid models showed marked differences in IC_{50} for the same compound; thus, a range of concentrations were used to induce mutagenesis across the organoid lines, making it challenging to compare compound-associated mutation burdens in the different models. Relationships between SBS, DBS and indel burdens and compound concentration are shown in Figures S10–S12 (refer also to Tables S8 and S10).

SBS burden exhibited some degree of concentration-dependence within particular models for a few agents (i.e., AAI, MAM, and MNNG), where organoids treated with a higher concentration accumulated more mutations. This was observed for all models treated with AAI except liver D4 and pancreas D39. For example, the kidney model JD021 had 1,226 SBS treated with 50 μM AAI and 6,568 SBS at 100 μM AAI. Linear concentration-response curves were possible for MNNG (the only agent tested at three concentrations per organoid line), with greater potency observed in gastric D88 compared to gastric D95, likely due to genetic variability between the donors. For many agents, however, there was little evidence of concentration-dependence; similar mutation burdens were induced by the two concentrations tested in each model (e.g., AFB_1, BaP, glycidamide, $KBrO_3$, and PO). For example, AFB_1 induced 4,358 SBS at 37.5 μM and 4,241 SBS at 75 μM in gastric D88. PO showed significant increases in mutation burden across all samples, but a 1.5-fold increase in concentration did not lead to an increase in mutation burden (D88: 1,845 SBS at 40,000 μM and 1,842 SBS at 60,000 μM; D95: 1,419 SBS at 40,000 μM and 1,443 SBS at 60,000 μM).

Interestingly, differences in mutation burden were often specific to the organoid model tissue of origin. For example, the number of mutations generated by a given concentration of AAI or PhIP+S9 differed dramatically between organoid lines (Figure S10; Table S8). For PhIP+S9, in colon organoid lines SC311 and SC351, 12.5–50 μM PhIP generated between 252 and 599 SBS (10–32 SBS/μM). However, in gastric organoid lines D88 and D95, a similar range of concentrations (17.5–50 μM PhIP+S9) generated between 1,356 and 2,328 SBS (27.1–125 SBS/μM). For AAI, both colon organoid lines were highly sensitive to treatment, and AAI concentrations of 0.25–1 μM induced 14,112–34,231 SBS (23,522–84,236 SBS/μM). The two kidney lines JD021 and JD050 were the least sensitive to AAI, and concentrations of 50–100 μM generated only 1,226–7,934 SBS (mean: 30.7–72.5 SBS/μM). The mutagenicity of AAI varied over 1,000-fold between the most and the least sensitive models. A similar relationship between concentration, tissue type, and mutation burden for AAI was also observed for DBS and indels; the colon organoids treated with AAI showed the greatest increase in DBS burden and the kidney organoids showed the least, with mutagenicity ranging nearly 7,000-fold from 0.05 to 348 DBS/μM (Figure S11).

Gene expression, transcriptional strand bias, and mutation distribution

Gene expression can influence mutation rates, where lower mutation burdens may be observed following DNA damage in highly expressed genes.[43] This, in part, is associated with the activity of transcription-coupled nucleotide excision repair (TC-NER), which preferentially repairs DNA damage on the template (transcribed) strand and results in mutational strand asymmetry.[44] We investigated the effect of gene expression on mutation burdens and assessed transcription strand bias of carcinogen-induced mutations in the organoids using methods described by Abascal et al.[11]

Gene expression

SBS and indel burdens in each sample were calculated for four subsets of the genome, according to publicly available gene expression data from the GTEx project[45] with regions Q1 (low) to Q4 (high) corresponding to the genes in each expression level quartile.

SBS burdens in samples treated with many of the compounds assessed in this study showed clear differences according to gene expression (Figures S13 and S14). For the majority of mutagens, SBS burden was the highest in the lowest-expressed genes in most samples, decreasing with increasing gene expression (e.g., AAI, BaP, *N*-OH-PhIP, ENU, glycidamide, $KBrO_3$, PO, and SO), including treatments for which we did not extract signatures (e.g., *o*-TOL). Conversely, for most of the samples treated with MAM or MNNG, SBS burden was *higher* with increasing gene expression. As both mutagens cause O^6-meG lesions that can be repaired by MGMT, these data suggest that MGMT may be less efficient in highly expressed genes.

The low numbers of indel calls resulted in lower confidence of the estimates of burden; thus, clear relationships between indel burden and gene expression were not discernible for most compounds (Figures S15 and S16). AAI, BaP, $KBrO_3$, and *N*-OH-PhIP generated some of the highest indel burdens in this study. As observed for SBS burden, the indel burden in most samples treated with AAI, BaP, or *N*-OH-PhIP was lower in subsets with more highly expressed genes. For most samples treated with $KBrO_3$, however, the indel burden did not decrease in subsets of more highly expressed genes.

Transcriptional asymmetry

During the extraction of mutational signatures, mutations are conventionally defined with respect to the pyrimidine base of a base pair; for example, an SBS is described as C>A as opposed to G>T, which occurs on the complementary strand. It is not possible from sequencing data alone to ascertain for individual mutation calls whether the true underlying mutation occurred

due to a DNA lesion and subsequent misrepair originating on the cytosine (C>A) or guanine (G>T) of the DNA duplex. However, this may be inferred through the analysis of asymmetry of mutations between the coding (non-transcribed) and template (transcribed) strands, providing information about whether certain mutagenic processes may involve transcription-associated damage or TC-NER.

To investigate transcriptional asymmetry of SBS, DBS, and indels in organoids, we examined the mutation spectra from each control or treatment group within genes from the same transcribed regions described above, considering whether the reference base was located on the transcribed or the untranscribed strand (Figures 5A–5C; Figures S17–S20;and see data and code availability). For SBS, the SBS96 spectra expand to "SBS192". Numbers of SBS for each of the six substitution types (C>A, C>G, C>T, T>A, T>C, and T>G) were summed for transcribed and untranscribed strands to test for significant strand bias for these mutation classes. Figure 5A shows the SBS192 spectra for the treatment groups that had refined SBS signatures in the organoids, restricted to mutations occurring in genes in the *highest* quartile (Q4) of gene expression. Spectra for AAI, AFB_1, BaP, glycidamide, and *N*-OH-PhIP all showed marked transcription strand bias for at least one mutation type. For example, T>A mutations in AAI-treated samples occurred more frequently on the transcribed strand, as did C>A mutations in BaP, AFB_1, and *N*-OH-PhIP-treated samples. The SBS192 spectra for ENU, on the other hand, exhibited an untranscribed strand bias (e.g., T>A mutations). Inspection of DBS and indel spectra did not show significant strand bias (Figures 5B and 5C).

Interestingly, SBS192 spectra from mutations occurring in the *lowest* quartile of expressed genes (Q1) showed no strand bias (Figure S20). Thus, the strand asymmetries of mutations observed within Q4 genes are assumed to be the result of TC-NER, which more efficiently repairs damaged bases on the transcribed strand in actively transcribed regions. Our results indicate that the predominantly damaged bases for these treatments were adenine for AAI; guanine for AFB_1, BaP, and activated PhIP; and thymine for ENU.

Comparison to COSMIC and reference signatures

Observed carcinogen-specific signatures from the organoid models were compared to reference signatures in the COSMIC[1,6] and Signal[7,46] databases by calculating cosine similarities (Figure 6; Figures S21 and S22). The reference signatures included all of those found in human tumors as well as relevant signatures from experimental models.

The AAI SBS signature in organoids was highly similar to COSMIC SBS22 (cossim = 0.97) identified in urothelial, liver, and kidney tumors associated with exposure to AAI[47,48] (Figure 6); both signatures display the dominant context of CAG. Additionally, the DBS and indel signatures in AAI-treated organoids match those of COSMIC DBS20 (cossim = 0.91) and ID23 (cossim = 0.89) recently found in kidney cancers reported from Serbia, Romania, and Thailand.[48] The SBS signatures for both BaP and activated PhIP in organoids were strongly similar to that of COSMIC SBS4 (cossim = 0.92), which is associated with tobacco smoking and likely caused by tobacco carcinogens such as BaP.[49] Further, we find similarity between COSMIC DBS2 and our DBS signature for BaP (cossim = 0.88), as well as COSMIC ID3 and our indel signatures for BaP and activated PhIP (cossim >0.93); both COSMIC signatures are linked to tobacco smoking. The AFB_1 SBS signature in organoids was similar to those of COSMIC SBS24 (cossim = 0.85), which is linked to aflatoxin exposure, and SBS29 (cossim = 0.89), which is associated with tobacco chewing. Both SBS24 and SBS29 predominantly feature C>A mutations; however, we found that the dominant sequence context for C>A mutations in SBS24 differs slightly from that found in our AFB_1-treated samples (GCA in the organoids and GCC in SBS24).

The MAM and MNNG SBS signatures in organoids were nearly identical to that of COSMIC SBS11 (cossim >0.96), a signature associated with temozolomide chemotherapy. Furthermore, we found that the SBS signature induced by $KBrO_3$ was like the ROS-associated COSMIC SBS18 (cossim = 0.85), and we observed similarity between COSMIC SBS92 signature and the signatures induced by glycidamide and SO in organoids (cossim = 0.84). SBS92 was identified in normal urothelium samples and bladder cancers that may be associated with tobacco smoking, and features mainly T>C, C>A, and C>T mutations.[50] No similarity to COSMIC signatures was found for the ENU and PO SBS signatures in organoids.

The mutational signatures in our organoids were also compared to signatures previously identified in other experimental systems (Figure S22). We observed strong similarities between SBS signatures from organoids and hiPSCs treated with AAI (cossim = 0.98), BaP (cossim = 0.96), PhIP+S9 (cossim = 0.93), $KBrO_3$ (cossim = 0.91), ENU (cossim = 0.85), and PO (cossim = 0.92). The AFB_1 SBS organoid signature did not match that observed in hiPSCs but was similar to a signature previously extracted from AFB_1-treated mice (cossim >0.91).[51] The SBS signature of glycidamide in our study showed not only some similarity to signatures previously found in glycidamide-treated MEFs but also some differences (cossim = 0.61–0.85)[52,53] (Figure S23). The predominant mutation type in glycidamide-treated organoids was T>C, contrasting with T>A in MEFs; there were also more C>A and C>T mutations in organoids. However, within some mutation types, the prevalence of particular sequence contexts was similar between the glycidamide signatures observed in organoids and MEFs (see highlighted contexts for C>A, T>A, and T>C in Figure S23). The reason(s) for differences between these glycidamide-induced SBS signatures is not clear, but it could be related to the use of different experimental models or the number of mutations detected (Zhivagui et al.[53] detected ~300 excess SBS/sample in MEFs, whereas we observed >1,000 excess SBS/sample in organoids before correcting for sequencing coverage).

DISCUSSION

It is increasingly evident that carcinogens may or may not leave a characteristic mutational signature in the tumors they cause. For instance, COSMIC SBS4, associated with tobacco smoking, is found in respiratory tract tumors of smokers but not in tumors from other organs, despite elevated cancer risk in those tissues.[49] Also, no difference in mutational signatures is observed between esophageal cancers from high- and low-incidence

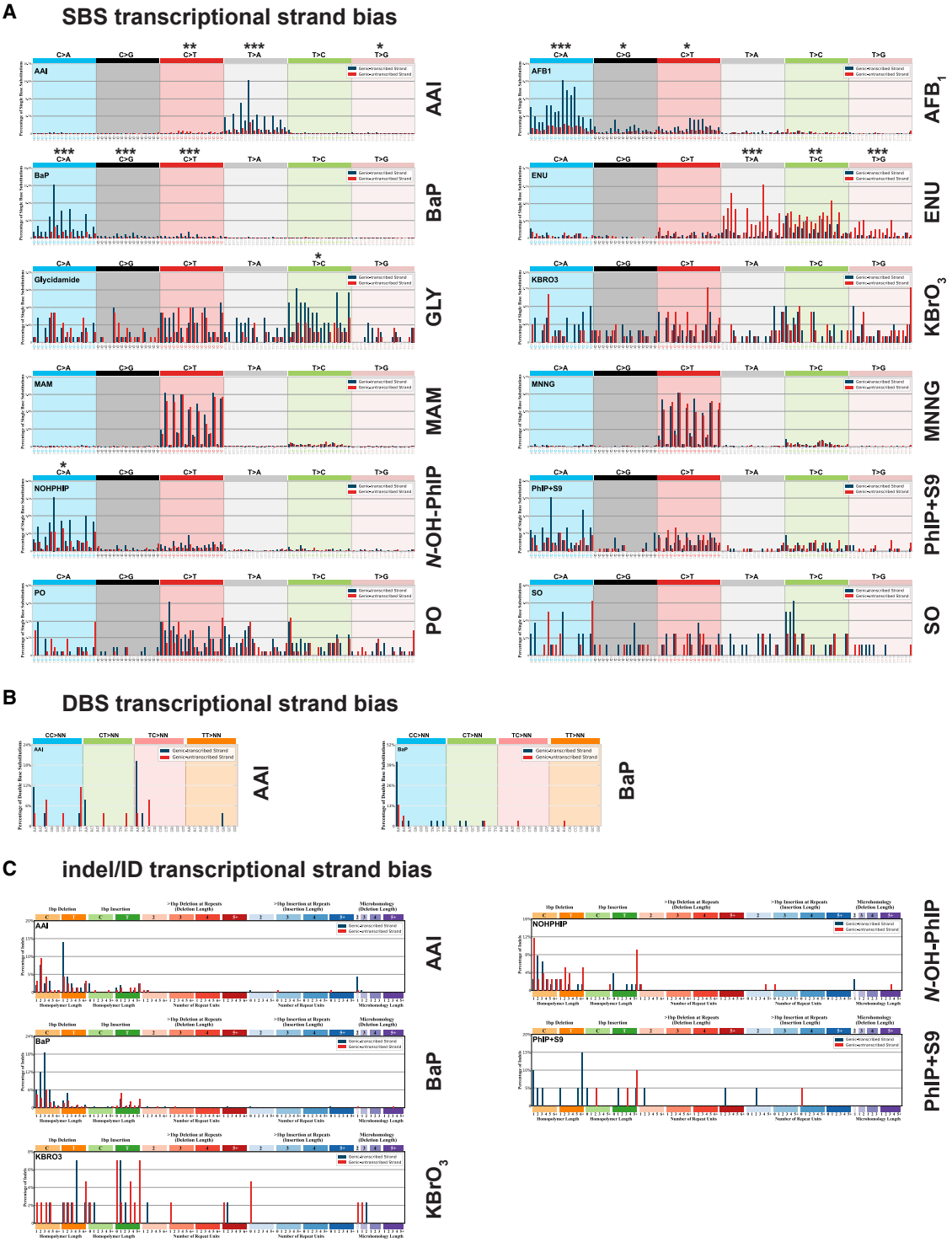


Figure 5. Investigation of transcriptional asymmetry of SBS, DBS, and indel signatures in mutagen-treated organoids

(A) SBS, (B) DBS, or (C) indel spectra of carcinogen-treated organoid samples, separated by reference base(s) location on the transcribed or untranscribed strand. For each treatment, the spectrum shown was compiled from all samples included in the respective final, refined signature, restricted to the mutations occurring in genes in the highest quartile of gene expression (quartile 4). The *y* axis refers to the percentage of mutations occurring at each channel, calculated from the number of raw SBS/DBS/indel calls (uncorrected for sample duplex coverage). Only DBS occurring at two pyrimidine bases were included. Significant strand bias is indicated by $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, and $^{***}p \leq 0.001$, calculated using a Poisson test.

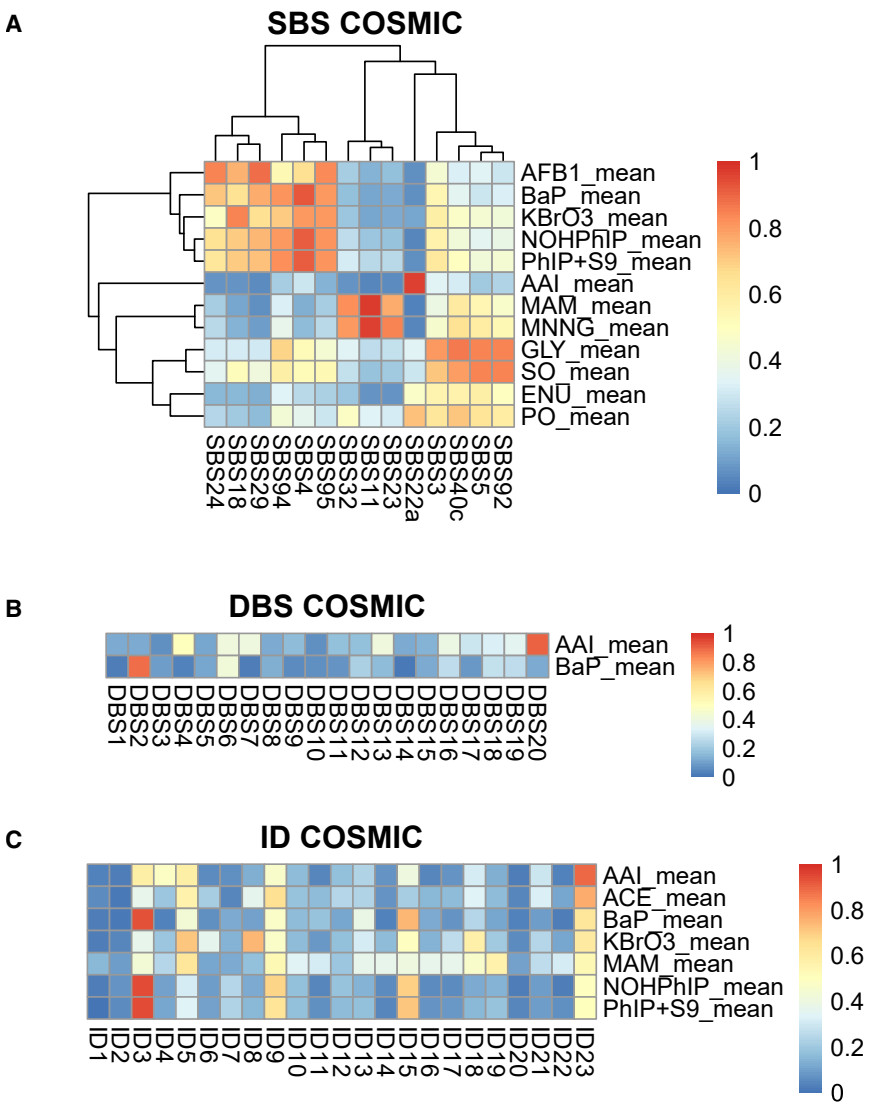


Figure 6. Comparison of mutational signatures in cancers and carcinogen-treated organoids

These heatmaps (plotted using pheatmap v.1.0.12) depict cosine similarities calculated between the mutational signatures of the COSMIC reference set v3.4 (https://cancer.sanger.ac.uk/cosmic/signatures) and final signatures extracted from carcinogen-treated organoids, including (A) SBS, (B) DBS, (C) indels. For SBS, COSMIC reference signatures for which none of the signatures in the experimental dataset show cossim >0.8 are excluded from the plot.

regions.[54] In animal models, some genotoxic carcinogens have failed to induce recognizable signatures.[55] Conversely, renal cancers reported from Romania, Serbia, and Thailand show SBS22 (linked to AAI exposure), while cases from Japan display signature SBS12, suggesting exposure to an as-yet-unidentified agent.[48] Colorectal cancers also show regional variation in signatures, reflecting diverse mutagenic exposures.[56]

Despite such complexities, experimentally defining mutational signatures of carcinogens remains crucial, both for tracing unknown signatures and for validating associations inferred from epidemiology. In this study, we used organoid models derived from normal human tissues and applied duplex sequencing to characterize signatures from a panel of environmental carcinogens.

Previous *in vitro* studies have relied on clonal expansion after mutagen exposure to overcome the limitations of standard sequencing sensitivity. However, isolating single-cell clones from organoids is laborious and inefficient, particularly with mutagen-treated cells, which show poor cloning efficiency. To overcome these limitations, we used duplex sequencing, which sequences both DNA strands and filters out single-strand errors arising from PCR amplification. This approach, specifically NanoSeq,[11] enables detection of mutations without subcloning, covering ~30% of the genome and providing a representative mutational profile. As such, duplex sequencing offers significant promise for advancing genetic toxicology and nonclinical safety studies.[57–59]

Using this platform, we identified distinct mutational signatures for 11 out of 17 tested carcinogens. Some were

confirmatory of previously described signatures (e.g., SBS signatures for AAI and BaP, and DBS/indel signatures for BaP),[2,5,9] while others were new. For example, we present an SBS signature for MAM, and we show that AAI produced DBS and indel signatures *in vitro* that match those recently reported in human tumors.[48] Our AFB_1-induced SBS signature closely resembled COSMIC SBS24, although the dominant trinucleotide context for C>A mutations in organoids was G-C-A, while in tumors it is G-C-C. This is consistent with prior observations from cell-line studies[51] and possibly reflects differences in aflatoxin mixtures in real-world exposures.

Activated PhIP (with S9 or as *N*-OH-PhIP) produced an SBS signature similar to that of BaP, echoing the findings in hiPSCs.[9] While both compounds are found in tobacco smoke and both induce signatures that resemble that of SBS4, only BaP induced a distinct DBS signature, and their indel profiles differed. This suggests that full mutational spectra can distinguish exposures more effectively than SBS patterns alone.

The alkylating agents MNNG and MAM, which act via methyldiazonium ions, yielded SBS signatures matching that of COSMIC SBS11, which is associated with temozolomide exposure. Temozolomide also exerts its genotoxicity following conversion to methyldiazonium ions.[60] Notably, MAM is a genotoxic product of cycasin, found in cycad seeds, which have been implicated in the etiology of Western Pacific amyotrophic lateral sclerosis and parkinsonism-dementia complex (ALS-PDC).[61] Our findings suggest that populations with cycad exposure could be screened for this MAM-associated signature. MAM also produces an animal model of schizophrenia[62,63] that could be similarly investigated.

The mutational signatures of AAI, BaP, activated PhIP, and AFB_1 were the same in the five different tissue organoids we tested, but there were quantitative differences in burden. Interestingly, mutation burden did not align with known target tissues *in vivo* (liver for AFB_1, kidney for AAI, colon for PhIP, and, putatively, pancreas for BaP). For instance, AFB_1-induced mutations were not the highest in the liver, and AAI-induced burden peaked in the colon, not the kidney. This may reflect differences in dosing (acute vs. chronic) and the absence of physiological variables such as adsorption and excretion, cautioning against overinterpreting mutation burden *in vitro* as a predictor of tissue-specific carcinogenicity.

Compared to our previous hiPSC studies,[9] organoids demonstrated greater metabolic competence. For instance, organoids from all tested tissues activated BaP and AFB_1 without S9. Furthermore, the organoids generally showed higher mutation burdens than hiPSCs after equitoxic (IC_{50}) exposures, facilitating clearer signature extraction. For example, the gastric organoid model D88 showed ~19,000 SBS/genome after AAI treatment versus ~900 in hiPSCs at comparable toxicity. Similarly, after treatment with AFB_1 ~4,000 SBS/genome were detected in the gastric organoids compared to ~100 in hiPSCs.

Sixteen of the 17 carcinogens were previously tested in hiPSCs (not MAM). SBS signatures for AAI, BaP, PhIP, $KBrO_3$, and PO were consistent between models. No SBS signature was observed in either system for acetaldehyde, KChr, MeIQx, or *o*-TOL, possibly due to incomplete activation or insufficient dosing. Signatures were detected in organoids but not in hiPSC for glycidamide, MNNG, and SO, likely due to higher permissible concentrations. AFB_1 and ENU showed divergent signatures across systems, highlighting model-dependent variability and the importance of validation across platforms.

Variability in mutation burden seen between donor organoids for indirect-acting mutagens (those requiring metabolic activation) may be due to differences in xenobiotic metabolism capability,[64,65] whereas variability seen with direct-acting mutagens (those not requiring metabolic activation) could be due to differences in DNA repair capacity of the donor cells.[66]

It is noteworthy that the primary base mutation of many agents is similar in bacteria, mammalian cells, transgenic mice, and humans.[67] As noted by DeMarini,[68] this is likely due to the highly conserved nature of the DNA repair/replication system across species.

Overall, this study expands the catalog of experimentally defined mutational signatures and demonstrates that organoids derived from normal human tissues, when combined with ecNGS, can efficiently capture carcinogen-specific signatures. Given their superior metabolic competence relative to other models, organoids offer a powerful platform for future studies of additional compounds and complex mixtures. More broadly, this work represents an important advancement in experimental mutagenesis, using modern sequencing technologies and physiologically relevant human models to sensitively assess mutagenic potential.

Limitations of the study

For the final mutation signatures presented in the main figures, we have only included stable signatures for which at least three replicates showed consistent spectra (cossim $\geq$0.8), and only the consistent samples were included in the analysis. Additional putative signatures that did not share similarity with others from the same treatment group are not included in the main figures, but can be found elsewhere (see data and code availability). The protocol used here is somewhat costly and labor intensive, which may limit its immediate application for regulatory purposes. If costs decrease over time and medium/high-throughput methods, already used for anti-cancer drug screening in organoids,[69,70] can be applied, then the use of organoids for regulatory purposes may become more feasible.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to the lead contact, David H. Phillips (david.phillips@kcl.ac.uk).

Materials availability

Availability of all organoids is subject to materials transfer agreement. In the first instance, contact the lead contact.

Data and code availability

- Raw NanoSeq sequencing data generated in this study have been deposited in the European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) under accession number EGA: EGAD00001015616. Raw standard sequencing data generated in this study for bulk, matched normals have been deposited in the EGA under accession number EGA: EGAD00001015630. Substitution and indel calls, as well as other data required to reproduce most of the main plots from this study are publicly available at Zenodo (Zenodo: https://doi.org/10.5281/zenodo.15911686).

- The code for the bioinformatics pipeline to process NanoSeq output data, including processing VCF files, generating mutation catalogs and analysis of mutation burden, is available at Zenodo: https://doi.org/10.5281/zenodo.4604537. Those steps and additional code used in this study for mutational signature analysis, refinement of *de novo* experimental mutational signatures, fitting of reference signatures to background signatures in organoids, comparison of organoid signatures with COSMIC and Signal signatures and reproduction of most of the main plots from this study are also available at Zenodo: https://doi.org/10.5281/zenodo.15911686.
- All other computational tools utilized in this publication have been referenced in the key resources table and can be accessed through their respective publications.

ACKNOWLEDGMENTS

We gratefully acknowledge Federico Abascal and Sarah Moody (Wellcome Sanger Institute), Xueqing Zou (AstraZeneca, Cambridge), and Adrian Baez-Ortega (Department of Genetics and Zoology, University of Cambridge) for their help and advice with NanoSeq and mutation signature analysis. We also acknowledge the contributions of the CASM Support team at the Wellcome Sanger Institute, in particular Laura O'Neill, Kirsty Roberts, and Nilan Yilmaz and of the Sanger Sequencing Operations teams, in particular Siobhan Austin-Guest. We thank Juliet Chen (University of Warwick) for lab assistance at King's College. R.S.S.B. and A.L.C.G. acknowledge support from the NIHR Health Protection Research Unit in Chemical and Radiation Threats and Hazards. D.H.P. is a member of the NIHR Health Protection Research Unit in Chemical and Radiation Threats and Hazards. BioRender was used to prepare the graphical abstract. This work was delivered as part of the Mutographs of Cancer team supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (C98/A24032). It was also part funded at King's College London by the National Institute for Health and Care Research Health Protection Research Unit in Chemical and Radiation Threats and Hazards (NIHR 200922), a partnership between UK Health Security Agency and Imperial College London. Additional funding to the Wellcome Sanger Institute was provided by Wellcome Trust grants 206194 and 220540/Z/20/A. The views expressed are those of the authors and not necessarily those of the NIHR, UK Health Security Agency or the Department of Health and Social Care. This work was also supported in part by the US National Institute of Health grants R01ES032547, R01ES036931, R01CA269919, R01CA296974, P01CA281819, and U01CA290479 to L.B.A. and by L.B.A.'s Packard Fellowship for Science and Engineering. The research was also supported by UC San Diego Sanford Stem Cell Institute. The computational development reported in this manuscript utilized the Triton Shared Computing Cluster at the San Diego Supercomputer Center of UC San Diego. Funders had no roles in study design, data collection and analysis, decision to publish, or preparation of the manuscript. For open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

AUTHOR CONTRIBUTIONS

J.E.K., H.A.-S., R.S.S.B., A.L.C.G., and E.C.W. performed the experimental work for the study. S.P.N., E.D., and S.S. performed the bioinformatics analyses. H.F., M.J.G., K.S.-P, M.H., J.D., and M.Z. developed organoids for the study. B.F. and F.Y. carried out karyotyping of organoids. G.K. provided MAM. J.E.K., V.M.A., and D.H.P. directed research at King's College, M.R.S. at Sanger, and L.B.A. at UCSD. L.H. was project manager for Mutographs of Cancer. J.E.K. and D.H.P. drafted the manuscript, with contributions from E.D., S.P.N., and L.B.A.

DECLARATION OF INTERESTS

M.R.S is founder, consultant and stockholder for Quotient Therapeutics. L.B.A. is a co-founder, CSO, scientific advisory member, and consultant for io9 (now Acurion); has equity; and receives income. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. L.B.A. is a compensated member of the scientific advisory board of Inocras. L.B.A.'s spouse is an employee of Hologic, Inc. L.B.A. declares US provisional applications with serial numbers: 63/289,601; 63/269,033; 63/366,392; and 63/412,835, as well as international patent application PCT/US2023/010679. L.B.A. is also an inventor of a US Patent 10,776,718 for source identification by non-negative matrix factorization. L.B.A. further declares a European patent application with application number EP25305077.7. M.H. is an inventor in several patents related to organoids.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Organoids derived from normal human tissues
  - Organoid culture
- METHOD DETAILS
  - Harvest of metaphase chromosomes and multiplex-FISH karyotyping
  - Mutagen treatment
  - Cell viability assessment
  - Organoid treatment for mutation induction
  - Western blotting
  - DNA extraction from organoids
  - DNA library preparation for NanoSeq
  - DNA library preparation for whole-genome sequencing of matched-normal samples
  - Next-generation sequencing
  - Data processing: NanoSeq
  - Data processing: Whole-genome sequencing of matched-normal samples
  - Variant calling and filtering
  - Sample quality control
  - Mutation catalog generation
  - Mutation burden analysis
  - Mutational signature analysis
  - Refinement of final *de novo* experimental mutational signatures
  - Fitting of reference signatures to background signatures in organoids
  - Comparison of organoid signatures with COSMIC and signal signatures
  - Mutation distribution according to genomic region
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2026.117406.

Received: October 30, 2025
Revised: March 3, 2026
Accepted: April 27, 2026

REFERENCES

1. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.
2. Nik-Zainal, S., Kucab, J.E., Morganella, S., Glodzik, D., Alexandrov, L.B., Arlt, V.M., Weninger, A., Hollstein, M., Stratton, M.R., and Phillips, D.H. (2015). The genome as a record of environmental exposure. Mutagenesis *30*, 763–770. https://doi.org/10.1093/mutage/gev073.

3. Zavadil, J., and Rozen, S.G. (2019). Experimental delineation of mutational signatures is an essential tool in cancer epidemiology and prevention. Chem. Res. Toxicol. *32*, 2153–2155. https://doi.org/10.1021/acs.chemrestox.9b00339.

4. Zhivagui, M., Korenjak, M., and Zavadil, J. (2017). Modelling mutation spectra of human carcinogens using experimental systems. Basic Clin. Pharmacol. Toxicol. *121*, 16–22. https://doi.org/10.1111/bcpt.12690.

5. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421. https://doi.org/10.1038/nature12477.

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

7. SIGNAL. https://signal.mutationalsignatures.com/.

8. Phillips, D.H. (2018). Mutational spectra and mutational signatures: Insights into cancer aetiology and mechanisms of DNA damage and repair. DNA Repair *71*, 6–11. https://doi.org/10.1016/j.dnarep.2018.08.003.

9. Kucab, J.E., Zou, X., Morganella, S., Joel, M., Nanda, A.S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S.P., et al. (2019). A compendium of mutational signatures of environmental agents. Cell *177*, 821–836.e16. https://doi.org/10.1016/j.cell.2019.03.001.

10. Caipa Garcia, A.L., Arlt, V.M., and Phillips, D.H. (2022). Organoids for toxicology and genetic toxicology: applications with drugs and prospects for environmental carcinogenesis. Mutagenesis *37*, 143–154. https://doi.org/10.1093/mutage/geab023.

11. Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., Russell, A.J.C., Alcantara, R.E., Baez-Ortega, A., Wang, Y., et al. (2021). Somatic mutation landscapes at single-molecule resolution. Nature *593*, 405–410. https://doi.org/10.1038/s41586-021-03477-4.

12. Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993. https://doi.org/10.1016/j.cell.2012.04.024.

13. Pfau, W., Schmeiser, H.H., and Wiessler, M. (1990). Aristolochic acid binds covalently to the exocyclic amino group of purine nucleotides in DNA. Carcinogenesis *11*, 313–319. https://doi.org/10.1093/carcin/11.2.313.

14. Arlt, V.M., Stiborová, M., Henderson, C.J., Thiemann, M., Frei, E., Aimová, D., Singh, R., Gamboa da Costa, G., Schmitz, O.J., Farmer, P.B., et al. (2008). Metabolic activation of benzo[a]pyrene in vitro by hepatic cytochrome P450 contrasts with detoxification in vivo: experiments with hepatic cytochrome P450 reductase null mice. Carcinogenesis *29*, 656–665. https://doi.org/10.1093/carcin/bgn002.

15. Baird, W.M., Hooven, L.A., and Mahadevan, B. (2005). Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. Environ. Mol. Mutagen. *45*, 106–114. https://doi.org/10.1002/em.20095.

16. Turesky, R.J. (2002). Heterocyclic aromatic amine metabolism, DNA adduct formation, mutagenesis, and carcinogenesis. Drug Metab. Rev. *34*, 625–650. https://doi.org/10.1081/dmr-120005665.

17. Rushing, B.R., and Selim, M.I. (2019). Aflatoxin B1: A review on metabolism, toxicity, occurrence in food, occupational exposure, and detoxification methods. Food Chem. Toxicol. *124*, 81–100. https://doi.org/10.1016/j.fct.2018.11.047.

18. Sugimura, T., and Terada, M. (1998). Experimental chemical carcinogenesis in the stomach and colon. Jpn. J. Clin. Oncol. *28*, 163–167. https://doi.org/10.1093/jjco/28.3.163.

19. Fiala, E.S. (1977). Investigations into the metabolism and mode of action of the colon carcinogens 1,2-dimethylhydrazine and azoxymethane. Cancer *40*, 2436–2445. https://doi.org/10.1002/1097-0142(197711)40:5+<2436::aid-cncr2820400908>3.0.co;2-u.

20. Fahrer, J., and Christmann, M. (2023). DNA alkylation damage by nitrosamines and relevant DNA repair pathways. Int. J. Mol. Sci. *24*, 4684. https://doi.org/10.3390/ijms24054684.

21. Galtress, C.L., Morrow, P.R., Nag, S., Smalley, T.L., Tschantz, M.F., Vaughn, J.S., Wichems, D.N., Ziglar, S.K., and Fishbein, J.C. (1992). Mechanism for the solvolytic decomposition of the carcinogen N-methyl-N'-nitro-N-nitrosoguanidine in aqueous solutions. J. Am. Chem. Soc. *114*, 1406–1411. https://doi.org/10.1021/ja00030a042.

22. Beranek, D.T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat. Res. *231*, 11–30. https://doi.org/10.1016/0027-5107(90)90173-2.

23. Grevatt, P.C., Solomon, J.J., and Bhanot, O.S. (1992). In vitro mispairing specificity of O2-ethylthymidine. Biochemistry *31*, 4181–4188. https://doi.org/10.1021/bi00132a005.

24. Klein, J.C., Bleeker, M.J., Lutgerink, J.T., van Dijk, W.J., Brugghe, H.F., van den Elst, H., van der Marel, G.A., van Boom, J.H., Westra, J.G., Berns, A.J., et al. (1990). Use of shuttle vectors to study the molecular processing of defined carcinogen-induced DNA damage: mutagenicity of single O4-ethylthymine adducts in HeLa cells. Nucleic Acids Res. *18*, 4131–4137. https://doi.org/10.1093/nar/18.14.4131.

25. Guttenplan, J.B. (1990). Mutagenesis by N-nitroso compounds: relationships to DNA adducts, DNA repair, and mutational efficiencies. Mutat. Res. *233*, 177–187. https://doi.org/10.1016/0027-5107(90)90161-v.

26. Loechler, E.L., Green, C.L., and Essigmann, J.M. (1984). In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. Proc. Natl. Acad. Sci. USA *81*, 6271–6275. https://doi.org/10.1073/pnas.81.20.6271.

27. Rosenberg, D.W., Giardina, C., and Tanaka, T. (2009). Mouse models for the study of colon carcinogenesis. Carcinogenesis *30*, 183–196. https://doi.org/10.1093/carcin/bgn267.

28. Venkatachalam, K., Vinayagam, R., Arokia Vijaya Anand, M., Isa, N.M., and Ponnaiyan, R. (2020). Biochemical and molecular aspects of 1,2-dimethylhydrazine (DMH)-induced colon carcinogenesis: a review. Toxicol. Res. *9*, 2–18. https://doi.org/10.1093/toxres/tfaa004.

29. Solomon, J.J., Mukai, F., Fedyk, J., and Segal, A. (1988). Reactions of propylene oxide with 2'-deoxynucleosides and in vitro with calf thymus DNA. Chem. Biol. Interact. *67*, 275–294. https://doi.org/10.1016/0009-2797(88)90064-6.

30. Rueff, J., Teixeira, J.P., Santos, L.S., and Gaspar, J.F. (2009). Genetic effects and biotoxicity monitoring of occupational styrene exposure. Clin. Chim. Acta *399*, 8–23. https://doi.org/10.1016/j.cca.2008.09.012.

31. Wallace, L., Pellizzari, E., Hartwell, T.D., Perritt, R., and Ziegenfus, R. (1987). Exposures to benzene and other volatile compounds from active and passive smoking. Arch. Environ. Health *42*, 272–279. https://doi.org/10.1080/00039896.1987.9935820.

32. Vodicka, P., Koskinen, M., Arand, M., Oesch, F., and Hemminki, K. (2002). Spectrum of styrene-induced DNA adducts: the relationship to other biomarkers and prospects in human biomonitoring. Mutat. Res. *511*, 239–254. https://doi.org/10.1016/s1383-5742(02)00012-1.

33. EFSA Panel on Contaminants in the Food Chain CONTAM (2015). Scientific Opinion on acrylamide in food. EFSA J. *13*, 41104. https://doi.org/10.2903/j.efsa.2015.4104.

34. Besaratinia, A., and Pfeifer, G.P. (2004). Genotoxicity of acrylamide and glycidamide. J. Natl. Cancer Inst. *96*, 1023–1029. https://doi.org/10.1093/jnci/djh186.

35. Gamboa da Costa, G., Churchwell, M.I., Hamilton, L.P., Von Tungeln, L.S., Beland, F.A., Marques, M.M., and Doerge, D.R. (2003). DNA adduct formation from acrylamide via conversion to glycidamide in adult and neonatal mice. Chem. Res. Toxicol. *16*, 1328–1337. https://doi.org/10.1021/tx034108e.

36. Besaratinia, A., and Pfeifer, G.P. (2007). A review of mechanisms of acrylamide carcinogenicity. Carcinogenesis *28*, 519–528. https://doi.org/10.1093/carcin/bgm006.

37. Joye, I.J., Lagrain, B., and Delcour, J.A. (2009). Use of chemical redox agents and exogenous enzymes to modify the protein network during breadmaking – A review. J. Cereal. Sci. *50*, 11–21. https://doi.org/10.1016/j.jcs.2009.04.001.

38. IARC (1986). Some Naturally Occurring and Synthetic Food Components, Furocoumarins and Ultraviolet Radiation.IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans 40.

39. Umemura, T., Kanki, K., Kuroiwa, Y., Ishii, Y., Okano, K., Nohmi, T., Nishikawa, A., and Hirose, M. (2006). In vivo mutagenicity and initiation following oxidative DNA lesion in the kidneys of rats given potassium bromate. Cancer Sci. *97*, 829–835. https://doi.org/10.1111/j.1349-7006.2006.00248.x.

40. Speit, G., Haupter, S., Schütz, P., and Kreis, P. (1999). Comparative evaluation of the genotoxic properties of potassium bromate and potassium superoxide in V79 Chinese hamster cells. Mutat. Res. *439*, 213–221. https://doi.org/10.1016/s1383-5718(98)00200-9.

41. Degtyareva, N.P., Placentra, V.C., Gabel, S.A., Klimczak, L.J., Gordenin, D.A., Wagner, B.A., Buettner, G.R., Mueller, G.A., Smirnova, T.I., and Doetsch, P.W. (2023). Changes in metabolic landscapes shape divergent but distinct mutational signatures and cytotoxic consequences of redox stress. Nucleic Acids Res. *51*, 5056–5072. https://doi.org/10.1093/nar/gkad305.

42. Luan, Y., Suzuki, T., Palanisamy, R., Takashima, Y., Sakamoto, H., Sakuraba, M., Koizumi, T., Saito, M., Matsufuji, H., Yamagata, K., et al. (2007). Potassium bromate treatment predominantly causes large deletions, but not GC>TA transversion in human cells. Mutat. Res. *619*, 113–123. https://doi.org/10.1016/j.mrfmmm.2007.02.029.

43. Pleasance, E.D., Stephens, P.J., O'Meara, S., McBride, D.J., Meynert, A., Jones, D., Lin, M.L., Beare, D., Lau, K.W., Greenman, C., et al. (2010). A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature *463*, 184–190. https://doi.org/10.1038/nature08629.

44. Haradhvala, N.J., Polak, P., Stojanov, P., Covington, K.R., Shinbrot, E., Hess, J.M., Rheinbay, E., Kim, J., Maruvka, Y.E., Braunstein, L.Z., et al. (2016). Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell *164*, 538–549. https://doi.org/10.1016/j.cell.2015.12.050.

45. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

46. Degasperi, A., Zou, X., Amarante, T.D., Martinez-Martinez, A., Koh, G.C.C., Dias, J.M.L., Heskin, L., Chmelova, L., Rinaldi, G., Wang, V.Y.W., et al. (2022). Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. Science *376*, abl9283. https://doi.org/10.1126/science.abl9283.

47. Poon, S.L., Pang, S.T., McPherson, J.R., Yu, W., Huang, K.K., Guan, P., Weng, W.H., Siew, E.Y., Liu, Y., Heng, H.L., et al. (2013). Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. Sci. Transl. Med. *5*, 197ra101. https://doi.org/10.1126/scitranslmed.3006086.

48. Senkin, S., Moody, S., Díaz-Gay, M., Abedi-Ardekani, B., Cattiaux, T., Ferreiro-Iglesias, A., Wang, J., Fitzgerald, S., Kazachkova, M., Vangara, R., et al. (2024). Geographic variation of mutagenic exposures in kidney cancer genomes. Nature *629*, 910–918. https://doi.org/10.1038/s41586-024-07368-2.

49. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. Science *354*, 618–622. https://doi.org/10.1126/science.aag0299.

50. Lawson, A.R.J., Abascal, F., Coorens, T.H.H., Hooks, Y., O'Neill, L., Latimer, C., Raine, K., Sanders, M.A., Warren, A.Y., Mahbubani, K.T.A., et al. (2020). Extensive heterogeneity in somatic mutation and selection in the human bladder. Science *370*, 75–82. https://doi.org/10.1126/science.aba8347.

51. Huang, M.N., Yu, W., Teoh, W.W., Ardin, M., Jusakul, A., Ng, A.W.T., Boot, A., Abedi-Ardekani, B., Villar, S., Myint, S.S., et al. (2017). Genome-scale mutational signatures of aflatoxin in cells, mice, and human tumors. Genome Res. *27*, 1475–1486. https://doi.org/10.1101/gr.220038.116.

52. Hölzl-Armstrong, L., Kucab, J.E., Moody, S., Zwart, E.P., Loutkotová, L., Duffy, V., Luijten, M., Gamboa da Costa, G., Stratton, M.R., Phillips, D.H., and Arlt, V.M. (2020). Mutagenicity of acrylamide and glycidamide in human TP53 knock-in (Hupki) mouse embryo fibroblasts. Arch. Toxicol. *94*, 4173–4196. https://doi.org/10.1007/s00204-020-02878-0.

53. Zhivagui, M., Ng, A.W.T., Ardin, M., Churchwell, M.I., Pandey, M., Renard, C., Villar, S., Cahais, V., Robitaille, A., Bouaoun, L., et al. (2019). Experimental and pan-cancer genome analyses reveal widespread contribution of acrylamide exposure to carcinogenesis in humans. Genome Res. *29*, 521–531. https://doi.org/10.1101/gr.242453.118.

54. Moody, S., Senkin, S., Islam, S.M.A., Wang, J., Nasrollahzadeh, D., Cortez Cardoso Penha, R., Fitzgerald, S., Bergstrom, E.N., Atkins, J., He, Y., et al. (2021). Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. Nat. Genet. *53*, 1553–1563. https://doi.org/10.1038/s41588-021-00928-6.

55. Riva, L., Pandiri, A.R., Li, Y.R., Droop, A., Hewinson, J., Quail, M.A., Iyer, V., Shepherd, R., Herbert, R.A., Campbell, P.J., et al. (2020). The mutational signature profile of known and suspected human carcinogens in mice. Nat. Genet. *52*, 1189–1197. https://doi.org/10.1038/s41588-020-0692-4.

56. Díaz-Gay, M., Dos Santos, W., Moody, S., Kazachkova, M., Abbasi, A., Steele, C.D., Vangara, R., Senkin, S., Wang, J., Fitzgerald, S., et al. (2025). Geographic and age variations in mutational processes in colorectal cancer. Nature *643*, 230–240. https://doi.org/10.1038/s41586-025-09025-8.

57. LeBlanc, D.P.M., Meier, M., Lo, F.Y., Schmidt, E., Valentine, C., 3rd, Williams, A., Salk, J.J., Yauk, C.L., and Marchetti, F. (2022). Duplex sequencing identifies genomic features that determine susceptibility to benzo(a)pyrene-induced in vivo mutations. BMC Genom. *23*, 542. https://doi.org/10.1186/s12864-022-08752-w.

58. Lynch, A.M., Zanoni, T.B., Salk, J.J., Martincorena, I., Young, R.R., Kucab, J., Valentine, C.C., Yauk, C., Escobar, P.A., Witt, K.L., et al. (2023). Next Generation Sequencing Workshop at the Royal Society of Medicine (London, May 2022): how genomics is on the path to modernizing genetic toxicology. Mutagenesis *38*, 192–200. https://doi.org/10.1093/mutage/gead012.

59. Marchetti, F., Cardoso, R., Chen, C.L., Douglas, G.R., Elloway, J., Escobar, P.A., Harper, T., Jr., Heflich, R.H., Kidd, D., Lynch, A.M., et al. (2023). Error-corrected next-generation sequencing to advance nonclinical genotoxicity and carcinogenicity testing. Nat. Rev. Drug Discov. *22*, 165–166. https://doi.org/10.1038/d41573-023-00014-y.

60. Stupp, R., Gander, M., Leyvraz, S., and Newlands, E. (2001). Current and future developments in the use of temozolomide for the treatment of brain tumours. Lancet Oncol. *2*, 552–560. https://doi.org/10.1016/S1470-2045(01)00489-2.

61. Spencer, P.S., Garner, C.E., Palmer, V.S., and Kisby, G.E. (2015). Chapter 11 - Environmental Neurotoxins Linked to a Prototypical Neurodegenerative Disease. In Environmental Factors in Neurodevelopmental and Neurodegenerative Disorders, M. Aschner and L.G. Costa, eds. (Academic Press), pp. 211–252. https://doi.org/10.1016/B978-0-12-800228-5.00011-X.

62. Qi, M., Zhu, P., Wang, H., He, Q., and Huo, C. (2024). Abnormalities in behavior relevant to schizophrenia in embryonic day 17 MAM-exposed rodent models: A systematic review and meta-analysis. Pharmacol. Biochem. Behav. *245*, 173888. https://doi.org/10.1016/j.pbb.2024.173888.

63. Santos-Silva, T., Dos Santos Fabris, D., de Oliveira, C.L., Guimarães, F.S., and Gomes, F.V. (2024). Prefrontal and hippocampal parvalbumin interneurons in animal models for schizophrenia: a systematic review and meta-analysis. Schizophr. Bull. *50*, 210–223. https://doi.org/10.1093/schbul/sbad123.

64. Caipa Garcia, A.L., Kucab, J.E., Al-Serori, H., Beck, R.S.S., Bellamri, M., Turesky, R.J., Groopman, J.D., Francies, H.E., Garnett, M.J., Huch, M., et al. (2024). Tissue Organoid Cultures Metabolize Dietary Carcinogens Proficiently and Are Effective Models for DNA Adduct Formation. Chem. Res. Toxicol. *37*, 234–247. https://doi.org/10.1021/acs.chemrestox.3c00255.

65. Caipa Garcia, A.L., Kucab, J.E., Al-Serori, H., Beck, R.S.S., Fischer, F., Hufnagel, M., Hartwig, A., Floeder, A., Balbo, S., Francies, H., et al. (2022). Metabolic Activation of Benzo[a]pyrene by Human Tissue Organoid Cultures. Int. J. Mol. Sci. *24*, 606. https://doi.org/10.3390/ijms24010606.

66. Au, W.W., Giri, A.K., and Ruchirawat, M. (2010). Challenge assay: A functional biomarker for exposure-induced DNA repair deficiency and for risk of cancer. Int. J. Hyg Environ. Health *213*, 32–39. https://doi.org/10.1016/j.ijheh.2009.09.002.

67. Cross, K.P., and DeMarini, D.M. (2023). Analysis of chemical structures and mutations detected by Salmonella TA98 and TA100. Mutat. Res. *827*, 111838. https://doi.org/10.1016/j.mrfmmm.2023.111838.

68. DeMarini, D.M. (2000). Influence of DNA repair on mutation spectra in Salmonella. Mutat. Res. *450*, 5–17. https://doi.org/10.1016/s0027-5107(00)00013-0.

69. Francies, H.E., Barthorpe, A., McLaren-Douglas, A., Barendt, W.J., and Garnett, M.J. (2019). Erratum to: Drug Sensitivity Assays of Human Cancer Organoid Cultures. Methods Mol. Biol. *1576*, 353. https://doi.org/10.1007/7651_2018_138.

70. Francies, H.E., Barthorpe, A., McLaren-Douglas, A., Barendt, W.J., and Garnett, M.J. (2019). Drug Sensitivity Assays of Human Cancer Organoid Cultures. Methods Mol. Biol. *1576*, 339–351. https://doi.org/10.1007/7651_2016_10.

71. Cancer-IT. NanoSeq. GitHub. https://github.com/cancerit/NanoSeq.

72. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

73. Tischler, G., and Leonard, S. (2014). biobambam: tools for read pair collation based algorithms on BAM files. Source Code Biol. Med. *9*, 13. https://doi.org/10.1186/1751-0473-9-13.

74. Trinh, M.K., Pacyna, C.N., Kildisiute, G., Thevanesan, C., Piapi, A., Ambridge, K., Anderson, N.D., Khabirova, E., Prigmore, E., Straathof, K., et al. (2022). Precise identification of cancer cells from allelic imbalances in single cell transcriptomes. Commun. Biol. *5*, 884. https://doi.org/10.1038/s42003-022-03808-9.

75. Zhang, F., Flickinger, M., Taliun, S.A.G., InPSYght Psychiatric Genetics Consortium; Abecasis, G.R., Scott, L.J., McCaroll, S.A., Pato, C.N., Boehnke, M., and Kang, H.M. (2020). Ancestry-agnostic estimation of DNA sample contamination from sequence reads. Genome Res. *30*, 185–194. https://doi.org/10.1101/gr.246934.118.

76. Bergstrom, E.N., Huang, M.N., Mahto, U., Barnes, M., Stratton, M.R., Rozen, S.G., and Alexandrov, L.B. (2019). SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. BMC Genom. *20*, 685. https://doi.org/10.1186/s12864-019-6041-2.

77. Gori, K., and Baez-Ortega, A. (2020). sigfit: flexibleBayesian inferenceof mutational signatures. Preprint at bioRxiv. https://doi.org/10.1101/372896.

78. Li, X., Francies, H.E., Secrier, M., Perner, J., Miremadi, A., Galeano-Dalmau, N., Barendt, W.J., Letchford, L., Leyden, G.M., Goffin, E.K., et al. (2018). Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. Nat. Commun. *9*, 2983. https://doi.org/10.1038/s41467-018-05190-9.

79. Howell, K.J., Kraiczy, J., Nayak, K.M., Gasparetto, M., Ross, A., Lee, C., Mak, T.N., Koo, B.K., Kumar, N., Lawley, T., et al. (2018). DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. Gastroenterology *154*, 585–598. https://doi.org/10.1053/j.gastro.2017.10.007.

80. Broutier, L., Andersson-Rolf, A., Hindley, C.J., Boj, S.F., Clevers, H., Koo, B.K., and Huch, M. (2016). Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. Nat. Protoc. *11*, 1724–1743. https://doi.org/10.1038/nprot.2016.097.

81. Huch, M., Gehart, H., van Boxtel, R., Hamer, K., Blokzijl, F., Verstegen, M.M.A., Ellis, E., van Wenum, M., Fuchs, S.A., de Ligt, J., et al. (2015). Long-term culture of genome-stable bipotent stem cells from adult human liver. Cell *160*, 299–312. https://doi.org/10.1016/j.cell.2014.11.050.

82. Calandrini, C., Schutgens, F., Oka, R., Margaritis, T., Candelli, T., Mathijsen, L., Ammerlaan, C., van Ineveld, R.L., Derakhshan, S., de Haan, S., et al. (2020). An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. Nat. Commun. *11*, 1310. https://doi.org/10.1038/s41467-020-15155-6.

83. Georgakopoulos, N., Prior, N., Angres, B., Mastrogiovanni, G., Cagan, A., Harrison, D., Hindley, C.J., Arnes-Benito, R., Liau, S.S., Curd, A., et al. (2020). Long-term expansion, genomic stability and in vivo safety of adult human pancreas organoids. BMC Dev. Biol. *20*, 4. https://doi.org/10.1186/s12861-020-0209-5.

84. Bartfeld, S., Bayram, T., van de Wetering, M., Huch, M., Begthel, H., Kujala, P., Vries, R., Peters, P.J., and Clevers, H. (2015). In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. Gastroenterology *148*, 126–136.e6. https://doi.org/10.1053/j.gastro.2014.09.042.

85. Schutgens, F., Rookmaaker, M.B., Margaritis, T., Rios, A., Ammerlaan, C., Jansen, J., Gijzen, L., Vormann, M., Vonk, A., Viveen, M., et al. (2019). Tubuloids derived from human adult kidney and urine for personalized disease modeling. Nat. Biotechnol. *37*, 303–313. https://doi.org/10.1038/s41587-019-0048-8.

86. Banerjee, R., Sotero-Caio, C.G., Fu, B., and Yang, F. (2022). Chromosomal instability (CIN) in HAP1 cell lines revealed by multiplex fluorescence in situ hybridisation (M-FISH). Mol. Cytogenet. *15*, 46. https://doi.org/10.1186/s13039-022-00625-x.

87. Oman, M., Alam, A., and Ness, R.W. (2022). How sequence context-dependent mutability drives mutation rate variation in the genome. Genome Biol. Evol. *14*, evac032. https://doi.org/10.1093/gbe/evac032.

88. Zou, X. ExpSigfinder. GitHub. https://github.com/xqzou/ExpSigfinder.

89. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-phospho-CHK2 (Thr68)	Cell Signaling Technology	Cat.# 2197; RRID: AB_2080501
Mouse monoclonal anti-Glyceraldehyde-3-PDH (GAPDH)	Millipore	Cat.# MAB374; RRID: AB_2107445
Rabbit monoclonal anti-phospho-H2A.X (Ser139)	Cell Signaling Technology	Cat.# 9718; RRID: AB_2118009
Rabbit polyclonal anti-phospho-p53 (Ser15)	Cell Signaling Technology	Cat.# 9284; RRID: AB_331464
Mouse monoclonal anti-p21	BD Bioscience	Cat. # 556431; RRID: AB_396415
Immun-Star Goat Anti-Rabbit-HRP Conjugate	Bio-Rad	Cat# 170–5046, RRID: AB_11125757
Immun-Star Goat Anti-Mouse-HRP Conjugate	Bio-Rad	Cat# 170–5047, RRID: AB_11125753
Biological samples		
Human gastric organoids, D88 (WTSI-OESO_088; HCM-SANG-0297-C15)	Wellcome Sanger Institute; see Table S1	N/A
Human gastric organoids, D95 (WTSI-OESO_095; HCM-SANG-0299-C15)	Wellcome Sanger Institute; see Table S1	N/A
Human sigmoid colon organoids, SC311	University of Cambridge; see Table S1	N/A
Human sigmoid colon organoids, SC351	University of Cambridge; see Table S1	N/A
Human kidney organoids, JD021	Princess Maxima Center for Pediatric Oncology, Utrecht, NL; see Table S1	N/A
Human kidney organoids, JD050	Princess Maxima Center for Pediatric Oncology, Utrecht, NL; see Table S1	N/A
Human pancreas organoids, D39	University of Cambridge; see Table S1	N/A
Human liver organoids, D4	University of Cambridge; see Table S1	N/A
Chemicals, peptides, and recombinant proteins		
Basement Membrane Extract, Type 2 (BME2), reduced growth factor	Cultrex	Cat.# 3533-010-02
Matrigel	Corning	Cat.# 356231
Recombinant human Noggin	Peprotech	Cat.# 120-10C
Recombinant mouse Noggin	Peprotech	Cat.# 250-38-250ug
Recombinant human EGF	Gibco	Cat.# PHG0313
Recombinant mouse EGF	Gibco	Cat.# PMG8041
Recombinant human FGF-10	Peprotech	Cat.# 100-26
Recombinant human HGF	Peprotech	Cat.# 100-39
B27 Supplement (+vitamin A)	Invitrogen	Cat.# 17504001
B27 Supplement (-vitamin A)	Invitrogen	Cat.# 12587001
N2 Supplement	Gibco	Cat.# 17502001
Gastrin-I, human	Sigma	Cat.# G9020
Prostaglandin E2 (PGE2)	Tocris	Cat.# 2296
Forskolin	Tocris	Cat.# 1099
N-acetyl cysteine	Sigma	Cat.# A9165
A83-01	Tocris	Cat.# 2939
Nicotinamide	Sigma	Cat.# N0636
SB202190	Tocris	Cat.# 1264/10

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Y27632 (rho-associated protein kinase (ROCK) inhibitor)	Tocris	Cat.# 1254
hES Cell Cloning and Recovery	Stemgent	Cat.# 01-0014-500
S9 SD rat liver, aroclor-induced	Moltox (dist. Trinova Biochem)	Cat.# 11-01L.2
1,2-Dimethylhydrazine	Sigma	Cat.# D161802; CAS# 540-73-8
1-Methyl-3-nitro-1-nitrosoguanidine; MNNG	TCI Chemicals	Cat.# M0527; CAS# 70-25-7
2-Amino-3,8-dimethylimidazo[4,5-f]quinoxaline; MeIQx	Toronto Research Chemicals Inc.	Cat.# A606600; CAS# 77500-04-0
2-Amino-1-methyl-6-phenylimidazo(4,5-*b*)pyridine; PhIP	Synthesized at the Biochemical Institute for Environmental Carcinogens (Grosshansdorf, Germany); Krais et al., 2016. https://doi.org/10.1002/ijc.29836	CAS# 105650-23-5
2-Hydroxyamino-1-methyl-6-phenylimidazo(4,5-*b*)pyridine; N-OH-PhIP	Synthesized at the Biochemical Institute for Environmental Carcinogens (Grosshansdorf, Germany); Krais et al., 2016. https://doi.org/10.1002/ijc.29836	CAS# 124489-20-9
Aristolochic acid I	Sigma	Cat.# A9451; CAS# 313-67-7
Acetaldehyde	Sigma	Cat.# 402788; CAS# 75-07-0
Aflatoxin B_1	Enzo	Cat.# ALX-630-093; CAS# 1162-65-8
Benzo[*a*]pyrene	Sigma	Cat.# B1760; CAS# 50-32-8
Ethylnitrosourea	Sigma	Cat.# N3385; CAS# 759-73-9
Glycidamide	Sigma	Cat.# 4704; CAS# 5694-00-8
Methylazoxymethanol acetate	MRIGlobal	Cat.# 213, Batch 03; CAS# 592-62-1
Mutagen X; 3-chloro-4-(dichloromethyl)-5-hydroxy-5H-furan-2-one	Toronto Research Chemicals Inc.	Cat.# C365665; CAS# 117823-31-1
o-Toluidine HCl	Toronto Research Chemicals Inc.	Cat.# T536215; CAS# 95-53-4
Potassium bromate	Sigma	Cat.# A18258; CAS# 7758-01-2
Potassium chromate	Santa Cruz Biotechnology Inc.	Cat.# sc-203351; CAS# 7789-00-6
Propylene oxide	Merck	Cat.# 82320; CAS# 75-56-9
Styrene oxide	Sigma	Cat.# S5006; CAS# 96-09-3
Critical commercial assays		
CellTiter-Glo 3D Cell Viability Assay	Promega	Cat.# G9683
DNeasy Blood and Tissue Kit	QIAGEN	Cat.# 69504
Deposited data		
raw whole genome sequencing data for matched normal organoid cultures	this study	EGA: EGAD00001015630
raw NanoSeq sequencing data	this study	EGA: EGAD00001015616
substitution and indel calls, as well as other data required to reproduce most of the main plots from this study	this study	Zenodo: https://doi.org/10.5281/zenodo.15911686
Experimental models: Cell lines		
Mouse: L Wnt-3A (for conditioned medium)	ATCC	CRL-2647
Mouse: HA-R-Spondin1-Fc 293T (for conditioned medium)	Cultrex	3710-001-01
Software and algorithms		
Prism versions 8-10	GraphPad	https://www.graphpad.com
R software, v4.3.1	R Core Team	https://www.r-project.org
NanoSeq data analysis pipeline v2.3.0–2.3.2	Abascal et al.[11,71]	https://github.com/cancerit/NanoSeq.
BWA-MEM v0.7.17	Li et al.[72]	https://bio-bwa.sourceforge.net/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
biobambam2 v2.0.86	Tischler et al.[73]	https://github.com/gt1/biobambam2
alleleIntegrator	Trinh et al.[74]	https://github.com/constantAmateur/alleleIntegrator
VerifyBamID2 v1.0.6	Zhang et al.[75]	https://github.com/Griffan/VerifyBamID
SigProfilerMatrixGenerator v1.3.1	Bergstrom et al.[76]	https://github.com/SigProfilerSuite/SigProfilerMatrixGenerator
ExpSigfinder	N/A	https://github.com/xqzou/ExpSigfinder
sigfit	Gori et al.[77]	https://github.com/kgori/sigfit
additional code used for mutational signature analysis and to generate the main plots in this study	this study	https://doi.org/10.5281/zenodo.15911686

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Organoids derived from normal human tissues

Gastric organoids were obtained from Mathew Garnett at the Wellcome Sanger Institute, Hinxton, UK, where they were derived from normal tissue of the upper stomach[78] from two male donors (D88 and D95) undergoing biopsy for an esophageal tumor, in accordance with the London Camden and King's Cross Research Ethics Committee (REC#16/L0/1110). Normal, sigmoid colon organoids were received from Matthias Zilbauer at the Department of Pediatrics, University of Cambridge, UK, and generated from biopsies taken from two donors (SC311, male and SC351, female)[79] in accordance with the East of England Cambridge South REC (REC#17/EE/0265). Normal liver and pancreas organoids were obtained from Meritxell Huch (Max Planck Institute of Molecular Cell Biology and Genetics) and derived from tissues of one donor each of liver (D4, female) and pancreas (D39, male) as described.[80,81] Liver and pancreas tissues were removed from deceased organ transplant donors at Addenbrooke's Hospital, Cambridge, UK, in accordance with the NRES Committee East of England, Cambridge Central (REC#15/EE/0152). Kidney organoids were from Jarno Drost (Princess Maxima Center for Pediatric Oncology, Utrecht, NL) and derived from normal tissue of two donors (JD021 and JD050, both female) as described[82]; tissues were taken during biopsy or nephrectomy, in accordance with the Medical Ethical Committee of the Erasmus Medical Center (Rotterdam, NL; REC#MEC-2016-739). All donor tissue was taken after obtaining written informed consent from the donor or their family for research approved and described above. Organoid model details are summarised in Table S1.

Organoid culture

For routine culture, organoids were embedded in extracellular matrix (ECM) consisting of Cultrex RGF Basement Membrane Extract, Type 2 (BME2; Cultrex, #3533-010-02) or Matrigel (colon organoids only; Corning, #356231), and seeded in 50 μL gels on 24-well plates, overlaid with 500 μL of organoid type-specific growth medium (summarised in Table S2) as described.[79,80,83–85] Organoid growth medium consisted of Advanced DMEM/F12 (Gibco, #12634-010), 10 mM HEPES (Gibco, #15630-056), 1X GlutaMAX (Gibco, #35050-038) and the following supplements: for gastric – 50% Wnt-3A conditioned medium, 10% R-Spondin-1 conditioned medium, 1.25 mM N-acetyl cysteine (Sigma, #A9165), 150 ng/mL recombinant human Noggin (Peprotech, #120-10C), 50 ng/mL recombinant human EGF (Gibco, #PHG0313), 100 ng/mL recombinant human FGF-10 (Peprotech, #100-26), 1 nM Gastrin I (Sigma, #G9020), 1% B27 Supplement + vitamin A (Invitrogen, #17504001), 2 μM A83-01 (Tocris, #2939), 10 mM nicotinamide (Sigma, #N0636); for colon – 50% Wnt-3A conditioned medium, 20% R-Spondin-1 conditioned medium, 1.25 mM N-acetyl cysteine, 100 ng/mL recombinant mouse Noggin (Peprotech, #250-38-250ug), 50 ng/mL recombinant mouse EGF (Gibco, #PMG8041), 1% B27 Supplement + vitamin A, 0.5 μM A83-01, 10 mM nicotinamide, 10 μM SB202190 (Tocris, #1264/10); for kidney – 10% R-Spondin-1 conditioned medium, 1 mM N-acetyl cysteine, 50 ng/mL recombinant human EGF, 100 ng/mL recombinant human FGF-10, 1.5% B27 Supplement + vitamin A, 5 μM A83-01; for pancreas – 10% R-Spondin-1 conditioned medium, 1.25 mM N-acetyl cysteine, 25 ng/mL recombinant human Noggin, 50 ng/mL recombinant human EGF, 100 ng/mL recombinant human FGF-10, 10 nM Gastrin I, 1% B27 Supplement - vitamin A (Invitrogen, #12587001), 5 μM A83-01, 3 μM PGE2 (Tocris, #2296), 1% N2 Supplement (Gibco, #17502001), 10 μM forskolin (Tocris, #1099), 10 mM nicotinamide; for liver – 10% R-Spondin-1 conditioned medium, 1.25 mM N-acetyl cysteine, 50 ng/mL recombinant human EGF, 100 ng/mL recombinant human FGF-10, 25 ng/mL recombinant human HGF (Peprotech, #100-39), 10 nM Gastrin I, 1% B27 Supplement - vitamin A, 5 μM A83-01, 1% N2 Supplement, 10 μM forskolin, 10 mM nicotinamide. During the first few days after thawing early passage frozen stocks of liver organoids, the liver growth media was further supplemented with 30% Wnt-3A conditioned medium, 25 ng/mL recombinant human Noggin and 1X Stemolecule hES Cell Cloning & Recovery Supplement (Stemgent, #01-0014-500). Conditioned medium was prepared as described previously,[79,80,83–85] using the cell lines L Wnt-3A (ATCC, #CRL-2647) and HA-R-Spondin1-Fc 293T (Cultrex, #3710-001-01).

Cultures were maintained at 37°C, 5% CO_2 and growth medium was replaced every 2–3 days. Organoids were passaged every 7–10 days, depending on density and size, by mechanical shearing or digestion with TrypLE and reseeded in ECM at a 3- to 8-fold dilution. After passaging media was supplemented with 10 μM of the rho-associated protein kinase (ROCK) inhibitor Y27632 (Tocris, #1254).

We found that the stomach, colon and kidney organoids (two donor lines each) could be consistently expanded for at least 20 passages (split every 7 days at 1:3–1:5). Organoids from pancreas and liver tissue could be split at 1:3 every 7–10 days up to passage 10, but then gradually slowed in growth (leading to a maximum split ratio of 1:2 every 7–10 days from passage 11–15).

METHOD DETAILS

Harvest of metaphase chromosomes and multiplex-FISH karyotyping

Metaphase chromosomes from organoids were harvested as previously described[78] with slight modifications. Recently passaged organoid cultures were incubated for 2–3 h with 0.1 μg/mL colcemid (KaryoMax; Gibco, #15212012) diluted in fresh growth media. Organoids were then dissociated into single cells by incubating with TrypLE (Gibco, #12605028) or Accumax (Merck, #SCR006) at 37°C for 10 min, mixing every three min by pipetting, washed in DMEM and pelleted at 350 *g*. The pellet was transferred to a 1.8-mL microtube in media and repelleted. The pellet was resuspended in ~25 μL of supernatant and then incubated in 1 mL of hypotonic solution (0.4% KCl in 10 mM HEPES [pH 7.4]) for 15–30 min at room temperature. The cells were then fixed by slowly adding 100 μL of a 4:1 (v/v) methanol:glacial acetic acid fixative, followed by two washes in the fixative, resuspended in 100 μL of fixative and then stored at −20°C until further processing. Throughout the fixation process, cells were pelleted at 850–3400 *g* in a microcentrifuge and were fully resuspended before adding solutions.

M-FISH karyotyping was performed as previously described.[78,86] The cell suspension was applied to slides, which were fixed in acetone for 10 min before baking at 62°C for 30 min. Metaphase spreads were then denatured in an alkaline solution (0.5 M NaOH, 1.0 M NaCl) for 7 ½—8 min, followed by two washes in 1 M Tris–HCl (pH 7.4) and 1× PBS, for 4 min each. Slides were then dehydrated in a series of 70%, 90% and 100% ethanol and air dried.

The probe mix of 24 color human M-FISH paint was denatured at 65°C for 10 min before applying onto the denatured slides. Hybridisation was carried out over two nights at 37°C. After hybridisation slides were washed for 30 min in 2× SSC at 37°C for the removal of coverslips, followed by a 5 min stringent wash in 0.5× SSC at 75°C, and two 5 min washes in 2× SSC +0.05% Tween20 (VWR) and 1× PBS, at room temperature. Slides were mounted in Vectashield Vibrance Antifade mounting medium containing DAPI (4′,6-diamidino-2-phenylindole; Vector Laboratories). Imaging was carried out using a 63× objective on a Zeiss AxioImager D1 fluorescent microscope equipped with a Hamamatsu CCD camera and narrow bandpass filters for DAPI, DEAC, FITC, Cy3, Texas Red (Cy3.5) and Cy5 fluorescence. Metaphases were imaged using SmartCapture software and karyotyped using SmartType Karyotyper (both Digital Scientific, UK). Twenty metaphase spreads from each organoid culture were fully karyotyped based on M-FISH classification.

Mutagen treatment

The mutagenic agents examined in this study were: 1,2-dimethylhydrazine (Sigma, #D161802); 2-amino-1-methyl-6-phenylimidazo(4,5-*b*)pyridine (PhIP) and 2-hydroxyamino-1-methyl-6-phenylimidazo(4,5-*b*)pyridine (N-OH-PhIP), which were both synthesised at the Biochemical Institute for Environmental Carcinogens, Grosshansdorf, Germany; 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline (MeIQx; Toronto Research Chemicals Inc., #A606600); 1-methyl-3-nitro-1-nitrosoguanidine (MNNG; TCI Chemicals, #M0527); 3-chloro-4-(dichloromethyl)-5-hydroxy-5H-furan-2-one (Mutagen X; Toronto Research Chemicals Inc., #C365665); aristolochic acid I (Sigma, #A9451); acetaldehyde (Sigma, #402788); aflatoxin B_1 (Enzo, #ALX-630- 093); benzo[*a*]pyrene (Sigma, #B1760); ethylnitrosourea (Sigma, #E3385); glycidamide (Sigma, #4704); methylazoxymethanol acetate (MRIGlobal, #213); o-toluidine HCl (Toronto Research Chemicals Inc., #T536215); potassium bromate (Sigma, #A18258); potassium chromate (Santa Cruz Biotechnology Inc., #sc-203351); propylene oxide (Merck, #82320); styrene oxide (Sigma, #S5006). Further information can be found in Table S3 (e.g., preparation of working stocks and solvents used).

Mutagens were diluted in growth media immediately prior to treating cultures. Control cultures were treated with solvent at the same final percentage used in the mutagen treatments. Organoids were incubated with mutagen treatment media or solvent control media for 48 h, unless S9 mix was required. Compounds requiring cytochrome P450-mediated metabolic conversion to DNA-reactive intermediates were tested with the inclusion of S9 mix, which consisted of 0.5–2.5% S9 fraction from Aroclor-1254-induced male Sprague-Dawley rat liver (Moltox/Trinova, #11-01L.2), 3 mM NADP (Roche) and 15 mM DL-isocitric acid trisodium salt hydrate (Sigma) in growth media. Cells were exposed to compounds in the presence of S9 mix for 4–24 h; treatment media was then replaced with fresh growth media. The percentage of S9 and treatment duration was optimised to ensure maximal activation of the compound (judged by toxicity), with minimal cell toxicity of S9 mix alone. Gastric organoids were treated with PhIP, Mutagen X and MeIQx using 2.5% S9 fraction for 4 h, and were treated with 1,2-DMH and o-toluidine using 1% S9 fraction for 24 h. Colon organoids were treated with PhIP and 1,2-DMH using 0.5% S9 fraction for 24 h. Treatments using S9 mix are summarised in Table S4.

Cell viability assessment

Organoid stock cultures were briefly disaggregated with TrypLE into small clumps of cells, diluted 3- to 8-fold in 70% BME2 (or Matrigel for treatments of colon organoids that included S9 mix) and seeded as 10 μL domes on 96-well plates 48–72 h before treatment. Each chemical was tested at a range of concentrations (≥5) to establish dose-response curves spanning 0–100% cytotoxicity, where possible. Chemicals were tested at least three times in independent experiments, wherein each treatment condition was assayed in ≥3 technical replicates (i.e., wells of a 96-well plate). Organoids were treated for 48 h (or 4–24 h with S9 mix), then treatment media was replaced with fresh growth media and the organoids were incubated for a further 48 h (a recovery period to allow cell division and processing of DNA damage).

Cell viability after treatment was measured using the CellTiter-Glo 3D Cell Viability Assay, which assesses relative cellular ATP levels (Promega, #G9683). The reagent, containing lysis buffer, luciferin and Ultra-Glo rLuciferase, was added either directly after the 48-h treatment or after the 48-h recovery period in a 1:2 ratio with the media. Samples were mixed by scraping the bottom of the wells to dislodge the gels and pipetting up and down several times and then incubated for 30 min at room temperature. Fifty μL from each well was transferred to a white assay plate and luminescence was measured using a GloMax Explorer microplate reader (Promega). Viability post-treatment was calculated as the % ATP in treated wells versus solvent control wells. IC_{50} values were calculated using Prism 7 software.

Organoid treatment for mutation induction

Organoids were disaggregated and diluted as for cell viability assays, seeded on 24-well plates in 50 μL gels with 500 μL of growth media and treated 2–3 days later. Each chemical and organoid line was tested on a separate plate. Cultures were treated (2 wells per condition) with each compound diluted in growth medium at three concentrations, including one that induced 50% cytotoxicity, as well as one higher and one lower concentration (all within the range of 25–75% cytotoxicity). A solvent control-treated culture was included on each plate. Organoids were treated for 48 h without S9, or 4–24 h with S9, then left to recover in normal growth medium for a further 48 h. After recovery cultures were passaged without TrypLE and reseeded at 1:1–1:3 (into 50 μL gels on 2–3 wells of a 24-well plate, per treatment condition), depending on the level of toxicity, then allowed to expand for up to one week. After expansion, a pellet was frozen for DNA extraction and two frozen stocks were prepared (using Recovery Cell Culture Freezing Medium, Gibco) from each control or treatment. Generally, the samples treated with a 50% toxic concentration and a higher concentration were used for mutation analysis. However, in some cases further cell death occurred during the one-week expansion period, and samples treated with lower concentrations of mutagen were used for mutation analysis.

Western blotting

Organoids were seeded in 50 μL of 70% ECM on 24-well plates with 500 μL of growth media. When cultures reached a density of 70–80% they were treated for up to 48 h with or without PhIP ± S9 at the same conditions used for mutation induction. Organoids were then harvested, incubated in TrypLE for 10 min at 37°C to dissolve the ECM, and pelleted. The pellet was washed with cold PBS and then lysed in a buffer of 62.5 mM Tris (pH 6.8), 1 mM EDTA (pH 8.0), 2% sodium dodecyl sulfate, 10% glycerol, and 1X Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, #78442). Western blotting was carried out as previously described.[52] Protein (15 μg) was separated on NuPAGE 4–12% gradient polyacrylamide gels (Thermo Fisher Scientific, #NP0336BOX) and subsequently transferred to a nitrocellulose membrane. The membrane was probed with the following primary antibodies: anti-phospho-CHK2 (T68, 1:1000; Cell Signaling, #2197), anti-phospho-p53 (S15, 1:1000; Cell Signaling, #9284), anti-p21 (1:2000; BD Bioscience, #BD556431), anti-phospho-H2AX (S139, 1:2000; Cell Signaling, #9718) and anti-GAPDH (1:25000; Millipore, #MAB374). Following washes membranes were incubated with secondary HRP conjugated antibodies (#170–5046 goat anti-rabbit HRP 1:20000 or #170–5047 goat anti-mouse HRP 1:10000, both BioRad). ECL (Amersham # RPN2105) was detected with film.

DNA extraction from organoids

DNA was extracted from pelleted cells using the QIAGEN DNeasy Blood and Tissue Kit (#69504).

DNA library preparation for NanoSeq

DNA was submitted to the NanoSeq pipeline (Wellcome Sanger Institute (WSI) Cancer, Aging and Somatic Mutation lab) for library preparation and duplex sequencing using the NanoSeq protocol described by Abascal et al.[11] DNA samples (1–5 ng) were purified using 100 μL of a 50:50 water:AMPure XP bead (Beckman Coulter, #A63881) mixture and eluted in 20 μL of nuclease-free water (NFW). A 20-μL volume of the bead suspension was used for on-bead fragmentation. This reaction was performed in a final volume of 25 μL, using 2.5 μL 10× CutSmart buffer (500 mM potassium acetate, 200 mM Tris-acetate, 100 mM magnesium acetate, 1 mg/mL BSA, pH 7.9 at 25°C), 0.5 μL 5 U/μL HpyCH4V and 2 μL NFW. Reactions were incubated at 37°C for 15 min, purified using 2.5× AMPure XP beads and resuspended in 15 μL of NFW. A 10-μL volume of fragmented DNA was A-tailed in 15-μL reactions containing 1.5 μL 10× NEBuffer 4 (500 mM potassium acetate, 200 mM Tris-acetate, 100 mM magnesium acetate, 10 mM DTT, pH 7.9 at 25°C), 0.15 μL 5 U/μL Klenow fragment (3′–5′ exo-, New England Biolabs [NEB], #M0212L), 1.5 μL 1 mM equimolar dATP/ddBTPs (ddTTP, ddCTP and ddGTP) and 1.85 μL NFW. Reactions were incubated at 37°C for 30 min. The 15-μL A-tailing reaction product was combined with 22.4 μL ligation mix, containing 2.24 μL 10× NEBuffer 4, 3.74 μL 10 mM ATP, 0.33 μL 15 μM xGen Duplex Seq Adapters (IDT, 1080799), 0.56 μL 400 U/μL T4 DNA ligase (NEB, #M0201S) and 15.53 μL NFW. Reactions were incubated at 20°C for 20 min,

purified using 1× AMPure XP beads and resuspended in 50 μL of NFW. DNA was quantified using a quantitative polymerase chain reaction (qPCR)-based method (Library Quantification Kit, KAPA #KK4835). Samples were diluted in NFW to give 0.3 fmol of DNA in 25 μL. This optimises the sequencing duplicate rate to maximize the number of read bundles (families of PCR duplicates) with at least two duplicate reads derived from each original strand when sequencing 150 million 150-bp paired-end reads, equivalent to ∼15× coverage for standard human WGS.[11] Libraries were PCR-amplified in 50-μL reactions containing 25 μL sample, 25 μL NEBNext Ultra II Q5 Master Mix (NEB, #M0544X) and lyophilised unique dual index (UDI)-containing primers, cycled as follows: Step 1, 98°C 30 s; Step 2, 98°C 10 s; Step 3, 65°C 75 s; Step 4, return to step 2 n times; Step 5, 65°C for 5 min; Step 6, hold at 4°C. The number of PCR cycles (n) was set according to the amount of input library DNA (16 cycles for 0.1–0.3 fmol, 14 cycles for 0.3 fmol, 13 cycles for 0.6 fmol). The PCR product underwent two consecutive 0.7× AMPure XP clean-ups. Samples were quantified using the AccuClear Ultra High Sensitivity dsDNA Quantification kit (Biotium, #31028) according to the manufacturer's instructions and pooled for sequencing.

DNA library preparation for whole-genome sequencing of matched-normal samples

DNA samples extracted from early passage, untreated cultures of each organoid line were submitted for standard whole-genome sequencing (WGS) at Wellcome Sanger Institute to provide matched-normal samples for the filtering of germline variants. Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs #E7103L) according to the manufacturer's instructions.

Next-generation sequencing

Both NanoSeq and matched-normal WGS libraries were sequenced using the Illumina NovaSeq 6000 platform with 150-bp, paired-end reads, according to the manufacturer's instructions.

Data processing: NanoSeq

Sequencing data generated from NanoSeq libraries were processed as described by Abascal et al.[11] using the NanoSeq pipeline at the Wellcome Sanger Institute (v2.3.0–2.3.2).[71] This analysis is summarised as follows. De-multiplexed FASTQ files were processed by extracting the random three-nucleotide barcode, clipping the remaining four adaptor bases and appending barcode sequences to the FASTQ header. Reads were aligned to hs37d5 (Full 1000genomes Phase2 Reference Genome Sequence, based on National Center for Biotechnology Information GRCh37) using BWA-MEM v0.7.17,[72] with barcode sequences appended to alignments. Alignments were sorted by coordinate, duplicates were marked, and reads were annotated with the read coordinate, mate coordinate and optical duplicate auxiliary tags, using biobambam2 v2.0.86.[73] Reads were filtered out if they were not marked as proper pairs, or if they were marked as optical duplicates, quality control failures, un-mapped or secondary alignments. Each read was marked with an auxiliary tag consisting of the reference name, and the mapping coordinates and barcodes of the read pair.

Data processing: Whole-genome sequencing of matched-normal samples

Reads were aligned to hs37d5 using BWA-MEM v0.7.17[72] and duplicates were marked using biobambam2 v2.0.86.[73]

Variant calling and filtering

SBS and indels

The calling of SBS and indels from NanoSeq data is described by Abascal et al.[11] Variant filtering steps are applied consistently to the calling of sites as both reference and mutant, allowing accurate calculation of sample mutation burden by dividing the number of mutant calls by the total number of reference or mutant calls (the total duplex coverage). The matched-normal samples were used to filter out clonal mutations in the sample, such as donor germline single-nucleotide polymorphisms (SNPs).

DBS

DBS were called by annotating runs of adjacent SBS calls from variant call format (VCF) files generated during SBS calling.

Sample quality control

Following variant calling, samples were checked for 1) evidence of incorrect pairing to their corresponding matched-normal sample and 2) presence of more than one human genome, indicating sample mixing or contamination during laboratory processing. Correct pairing of samples to their corresponding matched-normal samples was confirmed using the matchBAM function of the R package alleleIntegrator,[74] which uses SNP concordance to assess whether two samples originate from the same individual. Samples containing reads derived from more than one human genome were identified using VerifyBamID2 v1.0.6,[75] an approach for estimating inter-sample DNA contamination between samples from different individuals. Samples with 'freemix' values over 0.01 (>1% of reads originating from a foreign human genome) were excluded.

Mutation catalog generation

Mutation catalogs were generated from the counts of different SBS, DBS and indel subtypes for each sample. These mutation subtypes have been described previously by Alexandrov et al.[1] SBS were split into 96 channels, according to the reference and mutated nucleotide and trinucleotide context (SBS96), and DBS were split into 78 channels, according to the reference and mutated dinucleotide (DBS78). SBS catalogs were generated using the WSI NanoSeq pipeline[11,71] and DBS catalogs were generated using R. Indels

were split into 83 channels according to their size, nucleotides affected, and presence of repetitive or microhomology regions (ID83) and catalogs were generated from VCF files using SigProfilerMatrixGenerator v1.3.1.[76]

Mutation burden analysis

Correction for sequence composition

Mutation rates are influenced by sequence composition[87]; therefore, biases in genome coverage need to be accounted for when comparing sequencing results generated using different library preparation methods. NanoSeq data are restricted in coverage to approximately 28% of the human genome.[11] To compare these results to SBS/DBS burdens and/or spectra generated from standard WGS methods with more even genome coverage, differences in di/trinucleotide composition between the portion of the genome covered by NanoSeq libraries and the human genome as a whole must be considered.

For each of the 96 SBS subtypes considered, the mutation count was corrected using the ratio of the frequency of the corresponding trinucleotide in the human genome to the trinucleotide frequency observed in the sequencing data for the sample in question, as demonstrated in the example below for the subtype ACG>ATG:

$$M_{corrected}\ (ACG > ATG) = M_{observed}\ (ACG > ATG) \times \frac{F_{genome}(\mathrm{ACG})}{F_{observed}\ (\mathrm{ACG})}$$

where M is the mutation subtype count and F is the frequency of the corresponding trinucleotide. These corrected SBS counts project the observed trinucleotide spectra from the fraction of the genome covered by NanoSeq onto the human reference genome. DBS catalogs were also corrected according to the observed and genomic frequencies of the 78 dinucleotides considered, following the approach described above.

Calculation of mutation burden per base pair

Estimates of SBS burden/bp were calculated as described by Abascal et al.[11] Briefly, for each sample, the number of SBS calls was divided by the total duplex coverage, corrected for trinucleotide composition as described above. Estimates of DBS burden/bp were calculated by dividing the number of DBS calls by the total duplex coverage, corrected for dinucleotide composition.

Estimates of indel burden/bp were calculated by dividing the number of indel calls per sample by the total duplex coverage. Correction for the relative coverage of the sites affected by the 83 different indel subtypes considered is prohibitively complex due to the greater diversity of variant types within this class. However, opportunities for indel acquisition are less susceptible to distortion due to sequence composition, as they are less affected by the depletion of certain di/trinucleotides due to the use of a sequence-specific restriction enzyme in library preparation.

Poisson confidence intervals for estimates of SBS burdens/bp were calculated using the NanoSeq pipeline.[11] For DBS and indels, Poisson confidence intervals were calculated following the same method using R.

Calculation of estimated mutation burden per diploid human genome

For all variant classes, the estimated mutation burden per genome was calculated by multiplying the burden/bp by the approximate number of base pairs in a diploid human genome (6 billion).

Mutational signature analysis

De novo mutational signature extraction

Mutational signature extraction was performed using ExpSigfinder, an algorithm which identifies background mutation processes present in *in vitro* models using the relevant control samples, before removing this background profile from the treated samples to identify putative mutational signatures.[88] This algorithm was implemented as part of an approach to identifying significant and robust treatment-associated mutational signatures in this dataset.

SBS signatures

Putative SBS signatures were identified using a stepwise approach.

1. First, the background SBS burden/bp present in each model was calculated using the control samples, calculating the mean and 95% confidence interval.
2. The SBS burden of each treated sample was compared to the corresponding controls. Where the 95% confidence interval of the SBS burden of a treated sample did not overlap the 95% confidence interval for the SBS burden of the controls, the treated sample was considered to show a significant increase in SBS burden compared to background levels.
3. Samples that did not show a significant increase in SBS burden were not taken forward for SBS signature analysis.
4. Samples with fewer than 100 mutation calls were discarded due to insufficient data to determine any underlying mutational signature.
5. Mutation catalogs for the remaining samples were adjusted to give the estimated burden/genome by dividing observed counts by sample total duplex coverage and multiplying by 6 billion.
6. The background mutational spectrum for each model was calculated by taking the mean count of each SBS subtype across all control samples from the model.

7. This background spectrum was subtracted from the mutation catalog for each treated sample to give the excess mutational spectrum using the 'Wrap_KOSig' function from ExpSigfinder[88] (parameters: sampling_number = 100, start_num = 0, boundary = 2).
8. These excess mutational spectra were examined for evidence of a robust mutational signature, distinct from the background mutagenesis using a method adapted from Abascal et al.[11] ('Methods: Bootstrapped cosine similarity'):
   a) According to the probability distribution of the excess mutational spectrum, 50,000 random SBS catalogs of size n were generated, where n is the number of raw calls in the original sample (before correction for total duplex coverage), minus those attributed to the background signature.
   b) Cosine similarities between each of these samples and the background signature were calculated, obtaining a bootstrapped estimate of the 99.98% confidence interval (Bonferroni-adjusted 95% confidence interval; n = 128).
   c) The observed cosine similarity between the excess mutational spectrum and background signature was calculated; where this fell within the estimated confidence interval, the excess spectrum was considered not to show significant difference to the background and the sample was not taken forward for further analysis.
9. For the remaining samples, the background-subtracted spectrum was considered to represent the excess mutagenesis in response to the applied treatment – a putative mutational signature present in this sample.

DBS signatures

Samples with fewer than 20 DBS calls were discarded. Due to the low DBS call numbers in the control samples, it was not possible to generate robust estimates of DBS background spectra for these models. For DBS signatures, treated samples that showed a significant increase in total DBS burden were identified, and the DBS spectra of these samples were considered to represent the putative signature present. The influence of background variants on these signatures is assumed to be minimal given the low DBS burdens of control samples.

Indel signatures

Samples with fewer than 20 indel calls were discarded. Putative indel signatures were identified in the same way as SBS signatures, using the indel burdens to test for a significant increase in treated samples compared to the relevant controls, followed by subtraction of and comparison to the background indel spectrum.

Refinement of final *de novo* experimental mutational signatures

Putative mutational signatures were inspected to identify samples with shared signatures. Samples treated with the same agent were examined to assess any variation in compound-associated signatures according to model, presence of S9 or compound concentration.

1. Pairwise cosine similarities were calculated between all putative mutational signatures identified.
2. Hierarchical clustering by cosine similarity was implemented using the pheatmap v1.0.12 R package ('complete' method). These results were inspected for evidence of clustering within the group according to the variables above.
3. For each treatment group, putative signatures sharing cossim $\geq$0.8 with at least one other sample in the same treatment group were included in the refined, final mutational signature for that treatment. A weighted average of the putative signatures was calculated and taken as the refined signature of that treatment. Averages were weighted according to the estimated numbers of raw, compound-associated mutation calls each of the putative signatures were derived from. Mutational signatures and spectra were plotted with Python 3.9.23 using SigProfilerMatrixGenerator v1.3.3 (ref.[76]).

Fitting of reference signatures to background signatures in organoids

The sources of background mutagenesis in the organoids were explored by fitting COSMIC reference signatures[1] to the mutation catalogs. Subsets of the SBS96, DBS78 and ID83 COSMIC reference signatures were taken *a priori* to reduce them to signatures that may feasibly be present in these models. SBS reference signatures: SBS1,2,4,5,12,13,16,17a,17b,18,40,88,92; DBS reference signatures: DBS2,4,11; indel reference signatures: ID1,2,3,5,18. Signatures were excluded where they met one or more of the following criteria: 1. possible sequencing artifact, 2. associated with DDR deficiency (unlikely to be present in models derived from normal cells), 3. associated with exposure to an exogenous mutagen unlikely to affect these samples either due to low frequency in the population or tissue restriction, 4. unknown etiology, with the exception of SBS5 and SBS40, which have been observed extensively in normal tissues; SBS12, which has been observed widely in liver cancers and in normal liver; SBS16, which has been associated with aldehyde exposure; SBS17a and SBS17b, which have been observed in normal gastric samples; ID5, which is suspected to be ubiquitous (for further details, see Figure S4). For SBS and indel background signatures, aggregated mutation catalogs were generated for control samples from each model by summing the counts of each mutation subtype. Reference signatures were fitted to the aggregated catalogs using the 'fit_signatures' function from sigfit[77] (parameters: iter = 5000, warmup = 2000, chains = 1). Signatures were considered to be present in the sample set at a significant level if there was any sample for which the 95% highest posterior density (HPD) interval did not overlap 0.05, indicating that the signature showed significant attribution at a level of $\geq$5% in one or more samples. The selected signatures were then refitted to the catalogs (fit_signatures: iter = 10000, warmup = 2000, chains = 1). Exposure data were obtained using 'retrieve_pars', followed by correction for total duplex coverage

per sample. Due to the low DBS call numbers in control samples, mutations were aggregated across all eight organoid lines for signature analysis, with reference signatures fitted to this single catalog as described above.

Comparison of organoid signatures with COSMIC and signal signatures

Signatures from the COSMIC database were compared to the final, refined SBS, DBS, and ID signatures from the organoids using cosine similarity. For SBS, only COSMIC signatures that had cossim $\geq$0.8 for at least one refined signature are shown in Figure 6; all comparisons are shown in Figure S21. Additionally, cosine similarity was computed between COSMIC and Signal signatures for SBS and DBS. Signal reference signatures sharing cossim $\leq$0.95 with COSMIC signatures were compared to the refined signatures from organoids. Again, for SBS, only Signal signatures having cossim $\geq$0.8 for at least one refined signature from the treated organoids are shown in Figure 6; all comparisons are shown in Figure S21. Refined signatures from the organoids were also compared to SBS COSMIC experimental signatures compiled from the literature on mutational signatures generated by exogenous mutagens (v1.0), with the same cossim $\geq$0.8 threshold.

Mutation distribution according to genomic region

Mutations in the organoid models were also analyzed according to genomic regions based on gene expression and transcriptional strand asymmetry. For this analysis, the appropriate genome regions were defined using browser extensible data (BED) files, followed by calculation of mutation burdens and catalogs for each of the defined regions. For these analyses, correction for sequence composition was performed as described before, using the trinucleotide frequencies calculated according to the genomic regions analyzed. Mutation burden was also corrected for duplex sequencing coverage.

Gene expression

Gene expression data were sourced from the Genotype-Tissue Expression (GTEx Analysis V7) project.[45] For each model, the most appropriate tissue type available was used (D88 and D95: Stomach; SC311 and SC351: Colon – Sigmoid; JD021 and JD050: Kidney; D39: Pancreas; D4: Liver). Mutation rates were calculated according to four regions containing the genes in each expression level quartile (1 low, 4 high).

Transcriptional strand asymmetry

Mutations within genes for each quartile of expression were annotated according to the location of the pyrimidine base on the transcribed or untranscribed strand using data from the RefSeq database[89] excluding mutations at sites transcribed from both strands. Mutation spectra were derived from raw mutation calls, not corrected for duplex sequencing coverage.

Strand bias

To evaluate strand bias, matrices were computed for each of the gene expression quantiles, averaged for each treatment. For SBS, the SBS384 matrix was separated into transcribed and untranscribed mutations, and both were collapsed into SBS6 contexts ([C>A], [C>G], [C>T], [T>A], [T>C], [T>G]) and rounded. A Poisson test was conducted between the number of transcribed and untranscribed mutations for each of the 6 contexts for each treatment. *p*-value thresholds of * = 0.05, ** = 0.01, and *** = 0.001 were used to determine significant difference between the transcribed and untranscribed strands.

The same method was followed to evaluate strand bias for indels. In this case, the matrices were collapsed from ID415 to ID28 contexts and rounded before performing the Poisson tests.

It was not possible to determine strand bias significance for DBS mutations since the values were too small. Assessment of DBS strand bias requires orienting the mutations based on pyrimidine base pairs and can only be applied when both bases are pyrimidines, such as comparing CC>TT on the transcribed strand vs. GG>AA on the untranscribed strand. For purine-pyrimidine doublets, such as AC>TT, strand bias cannot be performed because it is not possible to determine whether there is more damage on the purine or pyrimidine. Therefore, the strand location of only 4 of the 10 DBS mutation types was examined.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests (assessing confidence intervals, cosine similarity and Poisson test for strand bias) were performed using R v4.3.1 (RStudio v2023.06.1 + 524) or Python 3.9.23. Details of statistical analyses are provided, along with sample number, *p* value and replicates, in the Figure Legends or respective section of the STAR Methods. Unless otherwise specified, plots were generated using the R package ggplot2 v3.4.1 or pheatmap v1.0.12 for heatmaps.