



King's Research Portal

DOI:

[10.1016/j.dnarep.2018.08.003](https://doi.org/10.1016/j.dnarep.2018.08.003)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Phillips, D. H. (2018). Mutational spectra and mutational signatures: insights into cancer aetiology and mechanisms of DNA damage and repair. *DNA REPAIR*. Advance online publication. <https://doi.org/10.1016/j.dnarep.2018.08.003>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Mutational spectra and mutational signatures: insights into cancer aetiology and mechanisms of DNA damage and repair

David H. Phillips^{1,2}

¹MRC-PHE Centre for Environment and Health, ²NIHR Health Protection Research Unit in Health Impact of Environmental Hazards at King's College London in Partnership with Public Health England; Department of Analytical, Environmental and Forensic Sciences, School of Public Health and Environmental Sciences, Faculty of Life Sciences and Medicine, King's College London

Address for correspondence

King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

Email: david.phillips@kcl.ac.uk

ORCID iD: 0000-0001-8509-3485

For: Cutting-edge Perspectives in Genomic Maintenance V, a Special Issue of *DNA Repair*, edited by

Philip C. Hanawalt

Key words: mutagen, mutational signature, whole genome sequencing

Abbreviations: 1,2-DCP – 1,2-dichloropropane; DCM – dichloromethane; AAN – aristolochic acid nephropathy; BEN – Balkan endemic nephropathy; TC-NER – transcription coupled nucleotide excision repair; iPS – induced pluripotent stem; ROS – reactive oxygen species

Abstract

Reporter gene assays, in which a single mutation from each experiment can contribute to the assembly of a mutation spectrum for an agent, have provided the basis for understanding the mutational processes induced by mutagenic agents and for providing clues to the origins of mutations in human tumours. More recently exome and whole genome sequencing of human tumours has revealed distinct patterns of mutation that could provide additional clues for the causative origins of cancer. This can be tested by examining the mutational signatures induced in experimental systems by putative cancer-causing agents. Such signatures are now being generated *in vitro* in a number of different mutagen-exposed cellular systems. Results reveal that mutagens induce characteristic mutation signatures that, in some cases, match signatures found in human tumours. Proof of principle has been established with mutational signatures generated by simulated sunlight and aristolochic acid, which match those signatures found in human melanomas and urothelial cancers, respectively. In an analysis of somatic mutations in cancers for which tobacco smoking confers an elevated risk, it was found that smoking is associated with increased mutation burdens of multiple different mutational signatures, which contribute to different extents in different tissues. One of these signatures, mainly found in tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of DNA editing by APOBEC cytidine deaminases and of an endogenous clock-like mutational process. The results are consistent with the proposition that smoking increases cancer risk by increasing the somatic mutation load although direct evidence for this mechanism is lacking in some cancer types. Thus next generation sequencing of exomes or whole genomes is providing new insights into processes underlying the causes of human cancer.

1. Introduction

Cancers arise as a result of somatic mutations. All tumours have been found to contain multiple mutations, including point mutations, insertions, deletions and gene rearrangements. Mutations that occur in critical genes involved in processes that maintain the integrity of cells (the 'driver' mutations) and that when mutated confer a growth advantage, can be considered causative in transforming a normal cell into a malignant one. At the same time, many other mutations are present in tumours (the 'passenger' mutations), accumulated as a result of genomic instability induced by the transformed or malignant state, or as bystander mutations arising from the same influences that gave rise to the 'driver' mutations, but occurring in regions of the genome that do not cause the phenotypic changes causing the emergence of the transformed state. Tumours are clonal, having arisen from a single cell that has acquired abnormal growth characteristics as a result of mutation, but can nevertheless display heterogeneity due to later acquisition of more mutations. A cancer genome can be considered to contain the history of mutagenic processes that have occurred throughout the life of the cancer patient, both before and after the acquisition of a neoplastic transformation phenotype by the progenitor cell of the tumour.

2. Mutational spectra – multiple studies in single genes

Experimental studies have, until recently, focused on the analysis of mutations in a single gene, whereby mutated cells were selected on the basis of a growth advantage under the conditions of the assay, or by producing progeny with a readily identified marker, such as a cellular dye.

Some of these systems arise out of short-term tests to identify mutagens, i.e. those that simply identify whether or not a test compound or agent is mutagenic. However, further investigation to identify the nature of the mutations that have arisen forms the basis of generating a mutation spectrum, by the acquisition of mutations from many experiments, often one at a time. These

include bacterial studies based on the Ames test, identifying mutations in the *his* operon of *Salmonella typhimurium* [1]. Such analyses have been applied to both single chemicals and complex mixtures [2]. A mutagen will produce the same type of substitution mutation in all organisms, which reflects the conservation of DNA repair and replication processes through evolution [3].

An early eukaryotic system for mutagenicity testing focused on the *HPRT* gene, whose mutation in response to mutagens was measured in mouse, hamster and human cells lines; cells in which *HPRT* mutations are rendered resistant to 6-thioguanine and will form colonies when cultured in media containing it. Detection of the mutations requires PCR amplification of the gene and direct sequencing of the transcripts [4].

The use of transgenic rodent mutation assays has been a useful and informative way of determining mutational spectra *in vivo* [5]. Mutations are detected by the introduction of the *lacZ* (Muta™mouse) or *lacI* gene (Big Blue® mouse or rat) into animals, which is then recovered from genomic DNA of treated rodents and packaged into a λ phage vector and infected into *E. coli*. Mutant colonies are then identified by means of a colour change from wild-type colonies; for mutagen testing purposes the mutation frequency is sufficient; however sequencing of the *lacZ*, *lacI* or the smaller *cII* gene in these mutant plaques provides the data to enable a mutation spectrum to be compiled. Another transgenic rodent model, *gpt* delta mouse and rat, has been developed that allows detection of point mutations by virtue of resistance to 6-thioguanine toxicity [5].

None of the genes mentioned thus far have anything to do with carcinogenesis. Indeed, in some of the systems, the reporter gene is not even expressed (e.g. *lacZ* and *lacI*), although paradoxically this can be an advantage in that the mutations detected have not undergone any selective pressure.

Attention has thus turned to considering mutations in 'driver' genes. Although mutational spectra based on *Kras* and *Hras* proto-oncogenes will be biased because only some types of mutations at certain codons can result in an activated oncogene, nevertheless different agents induce mouse

tumours with distinctly different patterns of mutations in these genes, which can be informative about mechanism of carcinogenesis and origins of tumours [6]. Differences in mutational spectra in premalignant papillomas and metastatic skin tumours induced in mice by 7,12-dimethylbenz[*a*]anthracene have shed light on the clonal evolution of metastasis in this experimental model [7].

More complete spectra can be obtained from analysis of the tumour suppressor gene *TP53*, which carries mutations in more than 50% of human tumours [8]. Nearly 30,000 *TP53* mutations in human tumours are catalogued in a database (<http://p53.iarc.fr>). From this compendium some distinct patterns have emerged pointing to specific causative agents: UV light in the case of skin cancers [9, 10], aflatoxin B1 in the case of hepatocellular carcinoma [11, 12], benzo[*a*]pyrene and other polycyclic aromatic hydrocarbons in the case of smokers' lung cancer [13] and aristolochic acid (a plant chemical) in the case of urothelial cancer (described in detail later in this article).

The generation of a mouse line with exons 4-9 of the human *TP53* gene (the Hupki mouse) [14] led to the development of an assay to detect mutations in sequences of the human gene in clones of immortalised mouse embryo fibroblasts. Using this approach agents such as UV light induced C to T and CC to TT mutations, benzo[*a*]pyrene (a tobacco carcinogen) induced G to T transversions, and aristolochic acid induced A to T transversions. In each case these mutations and their distribution (spectrum) in the *TP53* gene sequence correlated with those seen in human tumours where these agents were suspected of being causative [15].

Impressive though these associations are, the number of human carcinogens identified by scrutinising one gene only, either experimentally or in human tumours, is limited to just these few cases. Generating mutation spectra by the accumulation of mutations one by one from each experiment or tumour is not an efficient process. Given that it is now known that many tumours contains hundreds to thousands of mutations, the question arises as to what information can be gained, and what patterns can emerge, from an examination of all of them?

3. Mutational signatures – multiple mutations across the exome or whole genome

3.1. Mutational processes in human cancer

From sequence analysis of over 7,000 human cancers (mostly by exome sequencing, but with around 7% by whole genome sequencing) Alexandrov et al. [16] extracted 22 distinct mutational signatures, a signature being a representation of the 96 possible substitution mutations possible within at a base pair in the middle of a trinucleotide (i.e. there are six possible point mutations, four possible bases 5' to it and another four 3' to it, making $6 \times 4 \times 4 = 96$ possible events) (see Figure 1). When the analysis was expanded to ~12,000 human tumours, the number of signatures increased to 30 (<http://cancer.sanger.ac.uk/cosmic/signatures>). Imminently, with a further increase in the database to include more than 23,000 tumours of 71 cancer types, containing more than 83 million mutations, around 49 substitution signatures, together with 17 indel signatures and 11 dinucleotide (tandem) mutation signatures, will be reported (L. Alexandrov, personal communication).

Already, examination of these signatures, in association with the aetiology, where known, of the cancers from which they originate, has revealed a great amount about the mutagenic processes involved in carcinogenic processes. Around a third of the 30 signatures have been attributed to endogenous mutagenic processes, such as the activity of the APOBEC family of deaminases and deamination of 5-methylcytosine, deficiencies in DNA repair processes, including mismatch repair and homologous recombination, or defective DNA polymerases [17]. A further six are attributed to exposure to mutagenic agents, including tobacco carcinogens, UV radiation, aflatoxin B1, alkylating chemotherapy drugs and the plant carcinogen aristolochic acid; these are discussed in more detail below. The causes of the remaining 14 signatures are not yet identified. With the identification of additional signatures, as mentioned above, it still remains the case that approximately half of them are of unknown cause or origin (L. Alexandrov, personal communication).

Identifying the processes underlying the uncharacterised signatures and/or determining the agents that cause them will shed new light on the causes of cancer. In pursuit of this, a number of approaches have been used to generate mutagen-induced signatures *in vitro*. These have been useful in providing experimental evidence to validate or, in some cases question, the causative mechanisms assigned to some of the signatures in human tumours. What follows are examples of these approaches.

3.2. Aflatoxin mutational signature

In the COSMIC catalogue, signature 24 has been found only in liver cancer and is associated with reported exposure to aflatoxin B1, a known cause of hepatocellular carcinoma. The strong transcriptional bias in the mutations, which are predominantly GC-TA transversions, also supports the hypothesis that these are the result of misreplication of aflatoxin-DNA adducts, which would be subject to transcription-coupled nucleotide excision repair (TC-NER). In order to test the hypothesis, mutations induced by aflatoxin B1 in two different human cells lines, in liver tumours in wild-type mice and also in mice with a hepatitis B surface antigen transgene, were investigated by whole genome sequencing [18]. HepaRG and HepG2 cells that had been chronically exposed to aflatoxin B1 and cloned had predominantly GC-TA substitutions, with a strong bias for the G being on the non-transcribed strand. The bias decreased from the 5' to 3' ends of transcripts, in line with evidence that the effectiveness of TC-NER decreases thus [19]. The most enriched sequences were in both cases TGC, TGG and AGC, the same as found in the signatures, consistent with COSMIC signature 24, of newly sequenced hepatocellular carcinomas from Qidong, China, where exposure to aflatoxin B1 is well documented. There were similarities with the mouse liver tumours, but also greater diversity in the mutational landscapes, likely evidence of additional processes, such as the involvement of the hepatitis B surface antigen in the mouse strain engineered to express it [18]. Taken together, these findings greatly strengthen the causative association between aflatoxin B1 exposure and liver cancer.

3.3. Haloalkane exposure in print workers

Following a report of a high incidence of cholangiocarcinoma among print workers in Japan, whole exome sequencing was carried out on a number of cases [20]. The mutational signatures of each of four printing workers' cholangiocarcinomas were dominated by CG-TA transversions with a substantial strand bias and they all had a prominent trinucleotide signature mutation GCY to GTY (where Y is a pyrimidine, C or T), followed by NCY to NTY and NAY (where N is any base) that were seen only to a minor extent in control cases of cholangiocarcinoma. Two of the solvents to which the workers were exposed, 1,2-dichloropropane (1,2-DCP) and dichloromethane (DCM), were mutagenic to *S. typhimurium* TA100, with CG-TA transversions accounting for 58-71% of the 1,2-DCP-induced mutations and 47% of those induced by DCM. The predominant trinucleotide mutation sequence for 1,2-DCP was NCC to NTC, i.e. overlapping with the 2nd most common pattern in the printing workers, whereas the signature for DCM did not have anything in common with the tumour signatures. It can be concluded from these studies that the mutational signatures in the four print workers display common features that are indicative of exposure to a chemical that forms bulky adducts that are recognised by the nucleotide excision repair pathway (hence the mutational strand bias); among candidate mutagens, 1,2-DCP gives a mutational signature in bacteria that partially recapitulates the signature in the tumours. The discrepancies between the two systems point to more complex underlying processes that have yet to be elucidated, and possibly to additional mutagens that have yet to be identified [20]. The authors attempted to generate a mutational signature for 1,2-DCP in human epithelial cell lines, but were unsuccessful; thus the lack of a suitable mammalian system, thus far, in which to generate mutational signatures from 1,2-DCP also places limitations on the conclusions to be drawn from this study.

3.4. Mutational signatures of tobacco smoking

Tobacco smoke contains at least 60 different carcinogens and smoking is causally linked to 17 different cancers. In an analysis of several thousand cancer genome sequences from both smokers

and non-smokers, it was found that the smokers had significantly higher numbers of base substitutions than the non-smokers for all the 17 cancer types together and for several individual cancer types, namely larynx, liver, kidney and lung adenocarcinoma [21]. In tumours of tissue directly exposed to tobacco smoke (e.g. lung and larynx) the COSMIC signature 4 was prominent. As noted earlier, this signature is similar to that produced by benzo[a]pyrene in cells *in vitro* and this, coupled with the strong strand bias indicative for TC-NER, suggests that the signature arises from the misreplication of DNA damage (adducts) formed by carcinogens present in tobacco smoke. Several other signatures were found to be present at elevated levels of mutation in the smokers' tumours relative to the non-smokers, especially signatures 2, 5, 13 and 16.

Signatures 2 and 13, which feature GC-AT and GC-CG mutations, respectively, at TpC dinucleotides, are considered to be the result of overactivity of APOBEC deaminases in DNA editing. The reason for this overactivity, which is found in many tumour types, is not clear but may be the result of retrotransposon movement, cellular response to foreign DNA and local inflammation [22]. The frequency of mutations of signature 5, which is found in all tumour types and has a predominance of AT-GC (with transcriptional bias) and also GC-AT mutations, correlates with age of diagnosis in non-smokers [23]. The increase in its frequency in smokers' lung squamous and adenocarcinomas, larynx, pharynx, oral cavity, oesophageal squamous, bladder, liver and kidney cancers suggests that the "clocklike" process that generates it is accelerated by smoking. Signature 16 is associated mainly with liver cancer and its origin is unknown. It is characterised by AT-GC mutations at ApT dinucleotides, which exhibit a strong transcriptional strand bias.

The numbers of mutations that accumulate per cell annually per pack year of smoking in various tissues was calculated to be: for lung, 150; for larynx, 97; for pharynx, 39; for oral cavity, 23; for bladder, 18; for liver, 6. Interestingly, this order correlates with the ranking of odds ratios for the risk of smoking-related cancer in these organs. The study demonstrates that tobacco smoking enhances, or accelerates, endogenous mutagenic processes in some susceptible tissues, but not others, and it

also induces mutations directly by carcinogens present in tobacco smoke, although this mechanism appears to be confined mainly to tissues that are directly exposed to tobacco smoke [21]. The findings reveal the complexity of the mechanisms of tobacco carcinogenesis, which can vary between different tumour sites.

3.5. *The aristolochic acid mutational signature*

Aristolochic acid is a natural compound found in many plants that are used as traditional herbal remedies and health supplements. Although it was known to be carcinogenic to animals, it was not until it was accidentally included in a slimming regimen at high doses in Belgium in the 1990s, with the consequence that many of those treated succumbed to renal failure and urothelial cancer, that the cancer risk to humans was appreciated. Within a short time it was noted that the pathology of aristolochic acid nephropathy (AAN), as it became known, was similar to that of Balkan endemic nephropathy (BEN), a hitherto idiopathic condition occurring in many rural regions of Southeast Europe [24], but where human exposure to aristolochic acid now appears to result from contamination of cereal crops by *Aristolochia clematitis* growing wild in the regions [25].

Evidence that aristolochic acid is the causative agent of both AAN and BEN came first from investigation of mutation spectra and subsequently from mutational signatures. In tumours associated with both the human diseases, *TP53* mutations were frequently AT-TA transversions, a mutation that is rare in urothelial cancers in general. In experiment systems aristolochic acid forms DNA adducts predominantly with adenine residues in DNA, leading to AT-TA transversions [24, 26]. Comparison of the sites in *TP53* at which AT-TA transversions occurred in mouse embryo fibroblasts and in tumours from BEN patients showed a significant similarity [27].

Progressing from this single gene analysis, it has now been demonstrated conclusively from whole exome and whole genome sequencing of urothelial tumours that aristolochic acid gives rise to a distinctive and diagnostic mutational signature (COSMIC signature 22) that is dominated by AT-TA

transversions [28, 29]. The evidence that the tumours have been caused by aristolochic acid-induced mutations is further supported by the detection of DNA adducts on aristolochic acid in the tumour and tumour-adjacent tissue, and the experimental evidence that aristolochic acid produces the same signature in mouse embryo fibroblasts [30] and human induced pluripotent stems cells [31].

There is now evidence that aristolochic acid exposure in the Balkans may not be confined to the predominantly rural areas where BEN occurs. Whole genome sequencing of 14 Romanian renal cell carcinoma cases from non-BEN areas revealed that they all exhibited mutational signatures consistent with aristolochic acid exposure, a result also supported by the presence of aristolochic acid-DNA adducts in non-tumour renal tissue [32].

AAN is now recognised as a worldwide health issue. More recently, evidence has appeared for the involvement of aristolochic acid in cancer at additional sites. Analysis of bladder tumours from Asia found that the aristolochic acid signature was present in a number of them, including 2 out of 2 from Taiwan, 1 out of 11 from Singapore and 3 out of 99 from China [33]. Exome analysis of hepatocellular carcinomas revealed the presence of signature 22 in 78% of those from Taiwan, 49% of those from China and in 29% of those from elsewhere in Southeast Asia [34]. In addition, the signature was also detected in 13% of cases from Korea and 2.7% of those from Japan.

It should be noted that the AT-TA component of the mutational signature that aristolochic acid induces is not unique. A very similar signature is produced *in vitro* by a polycyclic aromatic hydrocarbon, dibenzo[*a,l*]pyrene [31]. However, in the case of aristolochic acid, the AT-TA mutations account for >85% of the mutations detected, whereas for dibenzo[*a,l*]pyrene it accounts for only 50%. Nevertheless, this sounds a note of caution that overreliance on evidence from mutational signatures, in the absence of evidence of exposure, could lead to misclassification of the cause(s) of cancer cases.

3.6. A compendium of experimentally induced mutational signatures

In order to examine systematically the mutational patterns associated with treatment of a comprehensive selection of environmental mutagens generated under highly controlled conditions, 79 chemicals/agents were selected on the basis that they are environmental or therapeutic agents that have been classified by the International Agency for Research on Cancer (IARC) as human carcinogens (Group 1), probable human carcinogens (Group 2A) or possible human carcinogens (Group 2B) [31]. Human induced pluripotent stem (iPS) cells were then exposed to each of them, clonally expanded and subjected to whole genome sequencing. In all cases, including untreated controls, the mutational signatures contained a component that was very similar to COSMIC signature 18 (Figure 1), which has been attributed to mutagenesis by reactive oxygen species (ROS) in combination with base excision repair defects from *MUTYH* germline mutations [35, 36]. After subtraction of the background signature from the pooled mutations for each treatment, the remaining mutations of approximately half of the agents tested also gave rise to distinct mutational signatures. Three of these, produced by simulated sunlight, aristolochic acid and benzo[a]pyrene, correlated closely with signatures 7, 22 and 4, respectively, reproducing what had previously been observed with mouse cells (Figure 1) [30].

In all, some 40 new signatures have been generated from mutagenic agents [31]. Some of the agents gave similar signatures, as might be expected from their known patterns of DNA adduct formation. For example, the *N*-ethyl-*N*-nitrosourea, *N*-methyl-*N*-nitrosourea and temozolomide signatures were markedly similar to each other; however they were markedly different from COSMIC signature 11 (Figure 1), which has been attributed to temozolomide therapy. The reason for this difference is not yet clear, but it may be that there are DNA repair processes that are more prominent in stem cells than in somatic cells, or vice versa. Another case of similar compounds giving similar signatures was the platinum complexes cisplatin and carboplatin, whose DNA adducts are identical to each other. However there are also examples for members of different classes giving rise to quite similar signatures, for example benzo[a]pyrene (a polycyclic aromatic hydrocarbon) and 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP, a heterocyclic amine). Compounds of the same chemical class

could also give rise to significantly different signatures, such as the polycyclic aromatic hydrocarbons, ostensibly due to differences in the proportions of adducts formed at guanine and adenine residues in DNA. Nevertheless there was great similarity between the signatures of each of three polycyclic aromatic hydrocarbons and their respective bay-region diol-epoxides, providing evidence that the latter are indeed the major intermediates in the metabolic activation of PAHs in human cells [31].

The agents that give rise to bulky DNA adducts, subject to TC-NER, nearly all showed significant strand bias in their mutational signatures. An exception is simulated sunlight, which did not display any strand bias, even though COSMIC signature 7, which it matches, does have the strand bias expected of TC-NER of UV-induced DNA lesions. A similar lack of strand bias was found in the UV-induced signature, also highly similar to signature 7, in mouse cells [30]. The reason for this anomaly is not clear at present.

4. Concluding remarks

The rapidly expanding database of mutational landscapes in human tumours is providing significant new information on the molecular events underlying the hallmarks of cancer. Until recently reporter gene assays, in which a single mutation (usually) from each experiment can contribute to the assembly of a mutation spectrum for an agent, have provided a basis of our understanding of mutational processes in human tumours. However we have now entered an era in which technologies for whole genome sequencing have advanced to the point where it is economically feasible to observe all the mutations produced in such assays, not just those in the reporter gene, and also to go further by generating mutational signatures in clones generated in systems without a selection pressure [30, 37, 38]. A number of different approaches are already evolving using, for example, cloned chicken cells [39], human induced pluripotent stem cells [31], and *C. elegans* [40] and showing the potential to reveal the role of exogenous and endogenous electrophiles, and or DNA repair and DNA damage response processes, in carcinogenesis.

With this massive increase in data acquisition and in statistical power coupled with the wealth of data that can be obtained from whole genome sequencing of human tumours, it is anticipated that mutational signatures produced experimentally will provide new insights into the agents and processes underlying the causes of human cancer.

Conflict of interest statement

The author has no conflict of interest

Acknowledgements

Work in the author's laboratory has been supported by Cancer Research UK (Grants C313/A14329 and C313/A25041), Wellcome Trust (Grants 101126/Z/13/Z and 101126/B/13/Z), and in part by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Health Impact of Environmental Hazards at King's College London in partnership with Public Health England (PHE). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

References

- [1] D.M. DeMarini, D.A. Bell, J.G. Levine, M.L. Shelton, A. Abu-Shakra, Molecular analysis of mutations induced at the hisD3052 allele of Salmonella by single chemicals and complex mixtures, *Environ Health Perspect*, 101 Suppl 3 (1993) 207-212.
- [2] D.M. DeMarini, Mutation spectra of complex mixtures, *Mutat Res*, 411 (1998) 11-18.
- [3] D.M. DeMarini, Influence of DNA repair on mutation spectra in Salmonella, *Mutat Res*, 450 (2000) 5-17.
- [4] P. Keohavong, L. Xi, S.G. Grant, Molecular analysis of mutations in the human HPRT gene, *Methods Mol Biol*, 1105 (2014) 291-301.
- [5] I.B. Lambert, T.M. Singer, S.E. Boucher, G.R. Douglas, Detailed review of transgenic rodent mutation assays, *Mutat Res*, 590 (2005) 1-280.
- [6] P.M. Westcott, K.D. Halliwill, M.D. To, M. Rashid, A.G. Rust, T.M. Keane, R. Delrosario, K.Y. Jen, K.E. Gurley, C.J. Kemp, E. Fredlund, D.A. Quigley, D.J. Adams, A. Balmain, The mutational landscapes of genetic and chemical models of Kras-driven lung cancer, *Nature*, 517 (2015) 489-492.
- [7] M.Q. McCreery, K.D. Halliwill, D. Chin, R. Delrosario, G. Hirst, P. Vuong, K.Y. Jen, J. Hewinson, D.J. Adams, A. Balmain, Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers, *Nat Med*, 21 (2015) 1514-1520.
- [8] M. Hollstein, D. Sidransky, B. Vogelstein, C.C. Harris, p53 mutations in human cancers, *Science*, 253 (1991) 49-53.
- [9] D.E. Brash, J.A. Rudolph, J.A. Simon, A. Lin, G.J. McKenna, H.P. Baden, A.J. Halperin, J. Ponten, A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma, *Proc Natl Acad Sci U S A*, 88 (1991) 10124-10128.
- [10] A. Ziegler, D.J. Leffell, S. Kunala, H.W. Sharma, M. Gailani, J.A. Simon, A.J. Halperin, H.P. Baden, P.E. Shapiro, A.E. Bale, et al., Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers, *Proc Natl Acad Sci U S A*, 90 (1993) 4216-4220.
- [11] F. Staib, S.P. Hussain, L.J. Hofseth, X.W. Wang, C.C. Harris, TP53 and liver carcinogenesis, *Hum Mutat*, 21 (2003) 201-216.
- [12] R. Montesano, P. Hainaut, C.P. Wild, Hepatocellular carcinoma: from gene to public health, *J Natl Cancer Inst*, 89 (1997) 1844-1851.
- [13] G.P. Pfeifer, M.F. Denissenko, M. Olivier, N. Tretyakova, S.S. Hecht, P. Hainaut, Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers, *Oncogene*, 21 (2002) 7435-7451.
- [14] J.L. Luo, Q. Yang, W.M. Tong, M. Hergenbahn, Z.Q. Wang, M. Hollstein, Knock-in mice with a chimeric human/murine p53 gene develop normally and show wild-type p53 responses to DNA damaging agents: a new biomedical research tool, *Oncogene*, 20 (2001) 320-328.
- [15] J.E. Kucab, D.H. Phillips, V.M. Arlt, Linking environmental carcinogen exposure to TP53 mutations in human tumours using the human TP53 knock-in (Hupki) mouse model, *FEBS J*, 277 (2010) 2567-2583.
- [16] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.L. Borresen-Dale, S. Boyault, B. Burkhardt, A.P. Butler, C. Caldas, H.R. Davies, C. Desmedt, R. Eils, J.E. Eyfjord, J.A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilcic, S. Imbeaud, M. Imielinski, N. Jager, D.T. Jones, D. Jones, S. Knappskog, M. Kool, S.R. Lakhani, C. Lopez-Otin, S. Martin, N.C. Munshi, H. Nakamura, P.A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J.V. Pearson, X.S. Puente, K. Raine, M. Ramakrishna, A.L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T.N. Schumacher, P.N. Span, J.W. Teague, Y. Totoki, A.N. Tutt, R. Valdes-Mas, M.M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L.R. Yates, I. Australian Pancreatic Cancer Genome, I.B.C. Consortium, I.M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P.A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S.M. Grimmond, R. Siebert, E. Campo, T. Shibata, S.M. Pfister, P.J. Campbell, M.R. Stratton, Signatures of mutational processes in human cancer, *Nature*, 500 (2013) 415-421.

- [17] L.B. Alexandrov, M.R. Stratton, Mutational signatures: the patterns of somatic mutations hidden in cancer genomes, *Curr Opin Genet Dev*, 24 (2014) 52-60.
- [18] M.N. Huang, W. Yu, W.W. Teoh, M. Ardin, A. Jusakul, A.W.T. Ng, A. Boot, B. Abedi-Ardekani, S. Villar, S.S. Myint, R. Othman, S.L. Poon, A. Heguy, M. Olivier, M. Hollstein, P. Tan, B.T. Teh, K. Sabapathy, J. Zavadil, S.G. Rozen, Genome-scale mutational signatures of aflatoxin in cells, mice, and human tumors, *Genome Res*, 27 (2017) 1475-1486.
- [19] J. Hu, S. Adar, C.P. Selby, J.D. Lieb, A. Sancar, Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution, *Genes & development*, 29 (2015) 948-960.
- [20] S. Mimaki, Y. Totsuka, Y. Suzuki, C. Nakai, M. Goto, M. Kojima, H. Arakawa, S. Takemura, S. Tanaka, S. Marubashi, M. Kinoshita, T. Matsuda, T. Shibata, H. Nakagama, A. Ochiai, S. Kubo, S. Nakamori, H. Esumi, K. Tsuchihara, Hypermutation and unique mutational signatures of occupational cholangiocarcinoma in printing workers exposed to haloalkanes, *Carcinogenesis*, 37 (2016) 817-826.
- [21] L.B. Alexandrov, Y.S. Ju, K. Haase, P. Van Loo, I. Martincorena, S. Nik-Zainal, Y. Totoki, A. Fujimoto, H. Nakagawa, T. Shibata, P.J. Campbell, P. Vineis, D.H. Phillips, M.R. Stratton, Mutational signatures associated with tobacco smoking in human cancer, *Science*, 354 (2016) 618-622.
- [22] C. Swanton, N. McGranahan, G.J. Starrett, R.S. Harris, APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity, *Cancer Discov*, 5 (2015) 704-712.
- [23] L.B. Alexandrov, P.H. Jones, D.C. Wedge, J.E. Sale, P.J. Campbell, S. Nik-Zainal, M.R. Stratton, Clock-like mutational processes in human somatic cells, *Nat Genet*, 47 (2015) 1402-1407.
- [24] M.R. Gokmen, J.P. Cosyns, V.M. Arlt, M. Stiborova, D.H. Phillips, H.H. Schmeiser, M.S. Simmonds, H.T. Cook, J.L. Vanherweghem, J.L. Nortier, G.M. Lord, The epidemiology, diagnosis, and management of aristolochic acid nephropathy: a narrative review, *Ann Intern Med*, 158 (2013) 469-477.
- [25] A.T. Gruia, C. Oprean, A. Ivan, A. Cean, M. Cristea, L. Draghia, R. Damiescu, N.M. Pavlovic, V. Paunescu, C.A. Tatu, Balkan endemic nephropathy and aristolochic acid I: an investigation into the role of soil and soil organic matter contamination, as a potential natural exposure pathway, *Environ Geochem Health*, (2017).
- [26] V.M. Arlt, M. Stiborova, J.V. Brocke, M.L. Simoes, G.M. Lord, J.L. Nortier, M. Hollstein, D.H. Phillips, H.H. Schmeiser, Aristolochic acid mutagenesis: molecular clues to the aetiology of Balkan endemic nephropathy-associated urothelial cancer, *Carcinogenesis*, 28 (2007) 2253-2261.
- [27] T. Nedelko, V.M. Arlt, D.H. Phillips, M. Hollstein, TP53 mutation signature supports involvement of aristolochic acid in the aetiology of endemic nephropathy-associated tumours, *Int J Cancer*, 124 (2009) 987-990.
- [28] M.L. Hoang, C.H. Chen, V.S. Sidorenko, J. He, K.G. Dickman, B.H. Yun, M. Moriya, N. Niknafs, C. Douville, R. Karchin, R.J. Turesky, Y.S. Pu, B. Vogelstein, N. Papadopoulos, A.P. Grollman, K.W. Kinzler, T.A. Rosenquist, Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing, *Science translational medicine*, 5 (2013) 197ra102.
- [29] S.L. Poon, S.T. Pang, J.R. McPherson, W. Yu, K.K. Huang, P. Guan, W.H. Weng, E.Y. Siew, Y. Liu, H.L. Heng, S.C. Chong, A. Gan, S.T. Tay, W.K. Lim, I. Cutcutache, D. Huang, L.D. Ler, M.L. Nairismagi, M.H. Lee, Y.H. Chang, K.J. Yu, W. Chan-On, B.K. Li, Y.F. Yuan, C.N. Qian, K.F. Ng, C.F. Wu, C.L. Hsu, R.M. Bunte, M.R. Stratton, P.A. Futreal, W.K. Sung, C.K. Chuang, C.K. Ong, S.G. Rozen, P. Tan, B.T. Teh, Genome-wide mutational signatures of aristolochic acid and its application as a screening tool, *Science translational medicine*, 5 (2013) 197ra101.
- [30] S. Nik-Zainal, J.E. Kucab, S. Morganella, D. Glodzik, L.B. Alexandrov, V.M. Arlt, A. Wenginger, M. Hollstein, M.R. Stratton, D.H. Phillips, The genome as a record of environmental exposure, *Mutagenesis*, 30 (2015) 763-770.
- [31] J.E. Kucab, X. Zou, S. Morganella, M. Joel, V.M. Arlt, D.H. Phillips, S. Nik-Zainal, Mutational signatures induced in human stem cells by environmental mutagens, (2018) submitted for publication.

- [32] R.J. Turesky, B.H. Yun, P. Brennan, D. Mates, V. Jinga, P. Harnden, R.E. Banks, H. Blanche, M.T. Bihoreau, P. Chopard, L. Letourneau, G.M. Lathrop, G. Scelo, Aristolochic acid exposure in Romania and implications for renal cell carcinoma, *Br J Cancer*, 114 (2016) 76-80.
- [33] S.L. Poon, M.N. Huang, Y. Choo, J.R. McPherson, W. Yu, H.L. Heng, A. Gan, S.S. Myint, E.Y. Siew, L.D. Ler, L.G. Ng, W.H. Weng, C.K. Chuang, J.S. Yuen, S.T. Pang, P. Tan, B.T. Teh, S.G. Rozen, Mutation signatures implicate aristolochic acid in bladder cancer development, *Genome Med*, 7 (2015) 38.
- [34] A.W.T. Ng, S.L. Poon, M.N. Huang, J.Q. Lim, A. Boot, W. Yu, Y. Suzuki, S. Thangaraju, C.C.Y. Ng, P. Tan, S.T. Pang, H.Y. Huang, M.C. Yu, P.H. Lee, S.Y. Hsieh, A.Y. Chang, B.T. Teh, S.G. Rozen, Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia, *Science translational medicine*, 9 (2017).
- [35] A. Viel, A. Bruxelles, E. Meccia, M. Fornasarig, M. Quaia, V. Canzonieri, E. Policicchio, E.D. Urso, M. Agostini, M. Genuardi, E. Lucci-Cordisco, T. Venesio, A. Martayan, M.G. Diodoro, L. Sanchez-Mete, V. Stigliano, F. Mazzei, F. Grasso, A. Giuliani, M. Baiocchi, R. Maestro, G. Giannini, M. Tartaglia, L.B. Alexandrov, M. Bignami, A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer, *EBioMedicine*, 20 (2017) 39-49.
- [36] C. Pilati, J. Shinde, L.B. Alexandrov, G. Assie, T. Andre, Z. Helias-Rodzewicz, R. Ducoudray, D. Le Corre, J. Zucman-Rossi, J.F. Emile, J. Bertherat, E. Letouze, P. Laurent-Puig, Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas, *J Pathol*, 242 (2017) 10-15.
- [37] M. Olivier, A. Weninger, M. Ardin, H. Huskova, X. Castells, M.P. Vallee, J. McKay, T. Nedelko, K.R. Muehlbauer, H. Marusawa, J. Alexander, L. Hazelwood, G. Byrnes, M. Hollstein, J. Zavadil, Modelling mutational landscapes of human cancers in vitro, *Scientific reports*, 4 (2014) 4482.
- [38] M. Zhivagui, M. Korenjak, J. Zavadil, Modelling Mutation Spectra of Human Carcinogens Using Experimental Systems, *Basic & clinical pharmacology & toxicology*, 121 (2017) 16-22.
- [39] B. Szikriszt, A. Poti, O. Pipek, M. Krzystanek, N. Kanu, J. Molnar, D. Ribli, Z. Szeltner, G.E. Tusnady, I. Csabai, Z. Szallasi, C. Swanton, D. Szuts, A comprehensive survey of the mutagenic impact of common cancer cytotoxics, *Genome Biol*, 17 (2016) 99.
- [40] B. Meier, S.L. Cooke, J. Weiss, A.P. Bailly, L.B. Alexandrov, J. Marshall, K. Raine, M. Maddison, E. Anderson, M.R. Stratton, A. Gartner, P.J. Campbell, C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency, *Genome Res*, 24 (2014) 1624-1636.

Figure legend

Figure 1. Mutational signatures generated experimentally and extracted from human tumours by whole genome sequencing. Each signature has 96 components, comprised of six possible base pair substitutions, each of which has four possible 5' neighbouring bases and four possible 3' neighbouring bases. The experimental signatures, shown on the left, were produced from induced human pluripotent stem cells exposure to mutagens [31]. The human tumour signatures, on the right, are from the COSMIC database (<http://cancer.sanger.ac.uk/cosmic/signatures>). The figure shows that propagation of cells in culture without treatment generates a signature similar to the COSMIC signature 18; simulated sunlight produces a signature similar to COSMIC signature 7; aristolochic acid produces a signature similar to COSMIC signature 22; benzo[*a*]pyrene produces a signature similar to COSMIC signature 4 (the “smoking” signature); temozolomide produces a signature that is distinctly different from COSMIC signature 11.

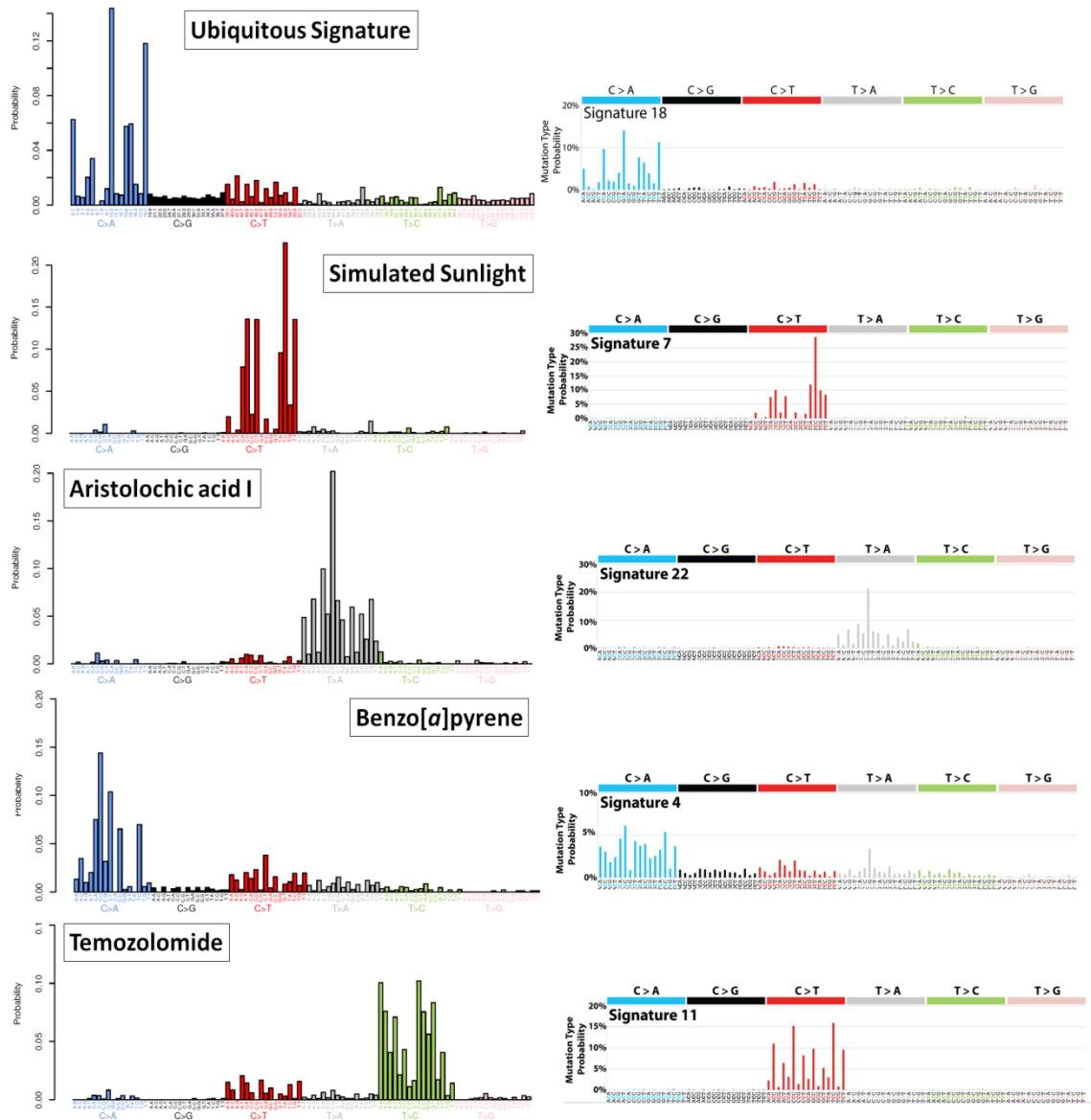


Figure 1