

## Review Article

# Experimental systems for the analysis of mutational signatures: no ‘one-size-fits-all’ solution

Dmitri Ivanov<sup>1</sup>, Taejoo Hwang<sup>3</sup>, Lukasz Karol Sitko<sup>1,2</sup>, Semin Lee<sup>3</sup> and  Anton Gartner<sup>1,2</sup>

<sup>1</sup>Center for Genomic Integrity, Institute for Basic Science, Ulsan, Republic of Korea; <sup>2</sup>Department of Biological Sciences, College of Information and Biotechnology, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea; <sup>3</sup>Department of Biomedical Engineering, College of Information and Biotechnology, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

**Correspondence:** Dmitri Ivanov (petersburg01@ibs.re.kr) or Anton Gartner (tgartner@ibs.re.kr)



Cells constantly accumulate mutations, which are caused by replication errors, as well as through the action of endogenous and exogenous DNA-damaging agents. Mutational patterns reflect the status of DNA repair machinery and the history of genotoxin exposure of a given cellular clone. Computationally derived mutational signatures can shed light on the origins of cancer. However, to understand the etiology of cancer signatures, they need to be compared with experimental signatures, which are obtained from the isogenic cell lines or organisms under controlled conditions. Experimental mutational patterns were instrumental in understanding the nature of signatures caused by mismatch repair and BRCA deficiencies. Here, we describe how different cell lines and model organisms were used in recent years to decipher mutational signatures observed in cancer genomes and provide examples of how data from different experimental systems complement and support each other.

## Introduction

Genetic changes affecting cell function and proliferation rate give rise to cancers. Initial analysis of cancer genomes focused on the identification of mutations in oncogenes and tumor suppressors, which drive the process of oncogenesis. With the advent of next-generation sequencing (NGS), it became possible to characterize cancer mutational landscapes in their entirety and to begin to uncover how they were shaped by the accumulation of DNA lesions caused by the action of exogenous and endogenous mutagens and by DNA repair pathways, which either reversed the lesions or converted them into mutations. The concept of mutational signature is based on the classification of single-nucleotide variants (SNVs) into 96 classes (6 possible substitutions multiplied by 16 combinations of 5' and 3' neighboring bases). In addition to SNVs, the patterns of small insertions/deletions (IDs) and structural chromosomal variants (SVs), including inversions, deletions, tandem and inverse duplications, and chromosomal translocations, produce signatures based on the type, size, and sequence context [1,2]. The first mutational signatures were derived from cancer exomes. However, the later switch to the analysis of whole genome sequences greatly increased the number of mutations available for analysis, improving the mathematical power and providing insights into the mutagenesis of the non-transcribed regions, such as repetitive DNA sequences. It is likely that the majority of mutational signatures in human cancer genomes have now been identified and compiled in the Catalogue of Somatic Mutations in Cancer (COSMIC) (<https://cancer.sanger.ac.uk/cosmic>). In principle, each signature results from a defined mutational process. Indeed, some signatures have overt associations with exposure to known mutagens, such as UV light, cigarette smoking, or food contaminants, e.g. aflatoxin and aristolochic acid, or with DNA repair defects, such as DNA mismatch repair (MMR), homologous recombination (HR), or polymerase  $\epsilon$  deficiency. However, when cancer genomes are analyzed by computational best-fit approaches, multiple signatures are typically derived in parallel. This could be due to several mutagenic processes being active in cancer cells. However, signature

Received: 7 April 2023  
 Revised: 22 May 2023  
 Accepted: 24 May 2023

Version of Record published:  
 7 June 2023

extraction is inherently ambiguous. The mutational pattern of any given cancer genome can be explained by multiple combinations of signatures in different proportions and it is possible that a contribution of signatures firmly associated with unrelated mutagenic agents might be computationally inferred. Computational approaches might derive too many signatures and, in many cases, their origins remain unclear. To validate computationally derived signatures and to decipher the complexity of cancer mutational spectra, it is necessary to analyze mutational patterns in simple experimental systems, which make it possible to study the effects of defined variables, such as the inactivation of a given gene, the exposure to a given chemical, or a combination thereof. In addition, the relative contributions of various repair pathways toward mending DNA lesions *in vivo* can be assessed. To emphasize the importance of experimentally obtained signatures, the COSMIC database now highlights whether each signature is experimentally validated or not.

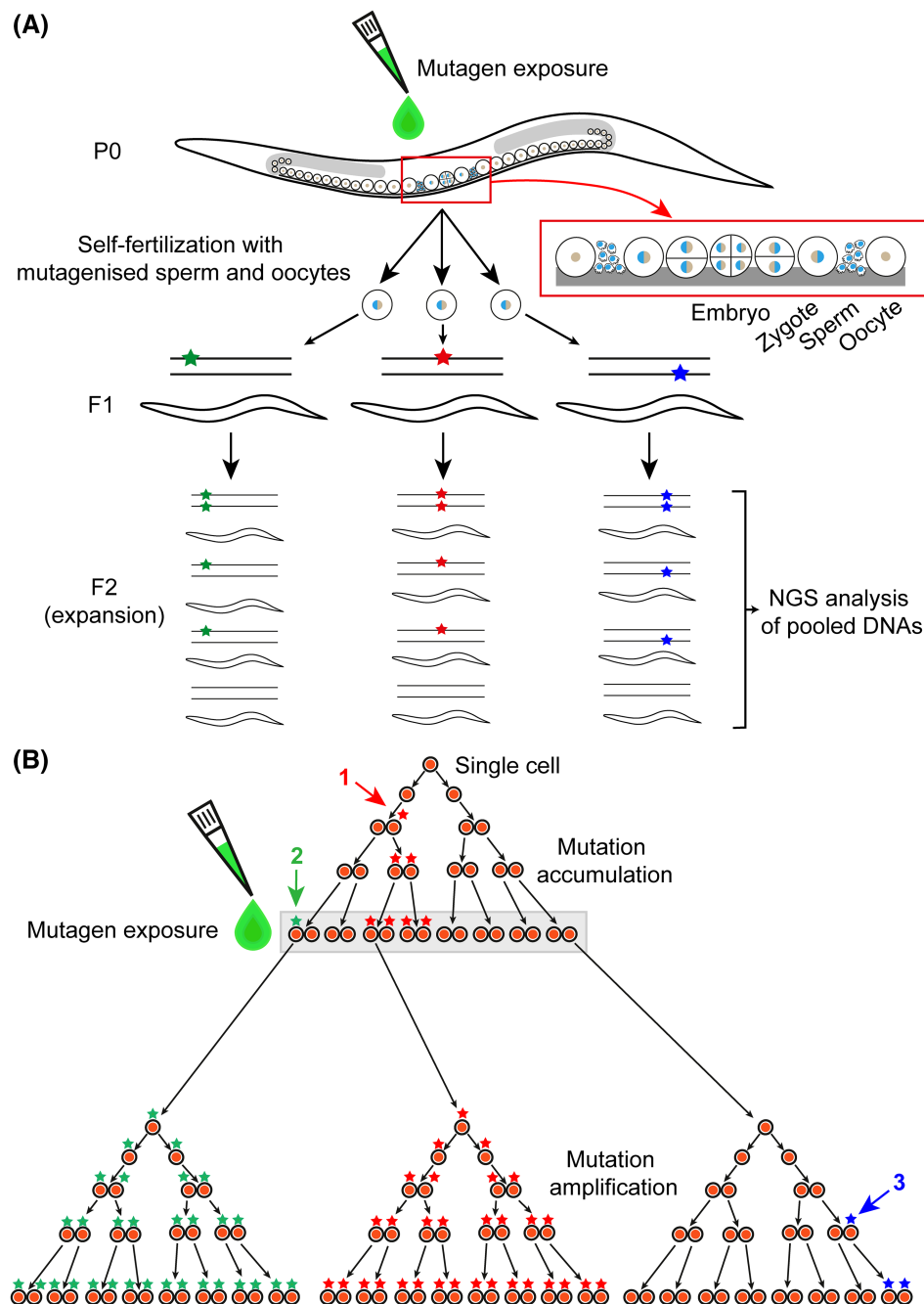
Since mutagenesis happens at a single-cell level, the analysis of mutational spectra necessitates single-cell sequencing, which is subject to the introduction of errors during DNA amplification and other technical artifacts. Alternatively, clonal cell populations derived from single mutagenized cells can be sequenced. If cell mixtures are used for sequencing, mutations will be diluted by genomes of non-mutated cells and fall below the detection limit. At the organismal level, this means either analysis of naturally occurring clonal populations, e.g. intestinal crypts [3], clones derived from single dissected cells [4,5], or taking advantage of the single-cell bottleneck provided by the germ line [6] (Figure 1).

## Mutational signatures in *Caenorhabditis elegans* and budding yeast

Recent studies using budding yeast and *C. elegans* provided a systematic view of mutational signatures accumulating in wild-type (WT) strains and mutants defective in genome stability with and without mutagen exposure. Analysis of genomes of 4732 yeast strains with deletions of non-essential genes involved in DNA repair, cell cycle progression, chromosome segregation, and DNA replication revealed that 10% of them had altered rDNA gene copy number and 2% carried sub-chromosomal deletions or amplifications [7]. In contrast, increased rates of SNVs were detected in only 14 strains, the majority of which were defective in either MMR or HR. Thus, yeast cells appear to tolerate a large number of SVs and copy number variations (CNVs). A separate study examined genomes of 14 single and double DNA repair mutants, which were propagated for 180 single-cell bottleneck passages 25 generations each [8]. WT strains accumulated mainly C > A and C > T substitutions, which were increased ~50 fold in MMR mutants, such as *msh2*. HR mutants demonstrated a more even induction of all types of SNVs as well as SVs. Defects in ordered S-phase progression or HR resulted in aneuploidy.

Analyses of *C. elegans* mutational signatures have been reviewed [9] and here we will only highlight some key findings. Systematic analysis of a panel of 61 WT and DNA repair defective *C. elegans* strains propagated for up to 40 generations revealed a more than two-fold increased mutagenesis in almost half of the strains [6,10]. The passage from one worm generation to the next involves approximately 15 germ cell divisions. Mutations detected in the mutation accumulation experiments might arise in the germ cells or during meiosis, where SV can result from faulty meiotic recombination. The rate of SNV mutagenesis was elevated in HR and nucleotide excision repair (NER) mutants and was the highest in MMR defective strains [6,10]. SVs are predominantly enriched in HR defective lines, but the number of SVs tends to be small, likely because chromosome translocations and aneuploidy compromise ordered meiotic chromosome segregation and animal development.

NGS analyses of *C. elegans* mutation accumulation lines provided important insights into the role and mechanistic action of several DNA repair enzymes. Deletions of G-rich sequences that have the propensity to form G-quadruplexes implicated the DOG-1 (FANCF) helicase in reading through those sequences [11,12]. Not only this, the analysis of the breakpoints of those localized deletions, which typically span 50–300 bp, provided a unique tool to uncover the roles of DNA polymerase  $\theta$  (termed POLQ-1 in the nematode). The authors observed that minute stretches of microhomology, often involving no more than one nucleotide, are common in those lesions and that lesions increase in size and lose microhomology at the breakpoint in *polq-1* defective strains [12,13]. These findings are in line with the role of POLQ-1 in stabilizing base pairing (involving as little as one base pair) to facilitate strand extension, a key step in polymerase  $\theta$ -mediated end joining (TMEJ). However, a small number of G-quadruplex-containing sequences are flanked by longer (20–30 bp) repeats, and these are deleted through the action of HELQ-1 helicase [14]. All in all, POLQ-1 promotes strand annealing of



**Figure 1. Use of germ line and clonal cell populations to determine experimental mutational signatures.**

**(A)** Single-cell bottleneck in the worm germ line. DNA lesions, which naturally occur or are experimentally induced during meiotic prophase, lead to mutant gametes, which give rise to mutant animals propagated by self-fertilization. While *C. elegans* ‘clones’ are in fact a mix of homozygous and heterozygous individuals, sequencing of the bulk genomic DNA identifies the mutations, which were present in the zygote that has given rise to the ‘clone’. **(B)** Analysis of mutation accumulation in cell lines. Clones are grown from single cells (‘mutation accumulation’ step), followed by an optional mutagen exposure and the second sub-cloning (‘mutation amplification’ step). Different types of mutations are represented as stars. The ‘red mutation’ spontaneously appeared during the first step of clonal expansion and was amplified after the sub-cloning. The ‘green mutation’ was induced by mutagen treatment and amplified during subclonal expansion. The ‘blue mutation’ appeared very late and will not be detected by sequence analysis due to its low prevalence. Only mutations that are present in the final clones but absent from the maternal clone are considered for analysis, ensuring that all mutations appeared *de novo* and were not present in the mother cell before the experiment.

sequences sharing minute microhomology (TMEJ), while HELQ-1 facilitates the annealing of longer repeat sequences through microhomology-mediated end joining (MMEJ) [14].

Nematode genome analyses also yielded insights into the genesis of chromosome-to-chromosome fusions and associated complex SVs resulting from the loss of telomeres in telomerase-defective strains. Fusion chromosomes showed hallmarks of breakage–fusion–bridge cycles followed by final chromosome fusion [15–17]. Other chromosome fusions were associated with copy number gain close to the breakpoint, likely linked to promiscuous template switching [15,16]. Rare lines surviving telomere crisis appear to maintain their chromosome ends by a two-step mechanism [18]. An internal template for alternative lengthening of telomeres (TALT) sequence, which contains telomere repeat units, is first translocated to a subtelomeric region on the same chromosome (cis-duplication) before amplification and ‘trans-duplication’ to the subtelomeric region of other chromosomes to protect chromosome ends [15,18]. Such lines also show excessive numbers of small deletions with microhomology at the endpoints indicative of double-strand breakage and repair by TMEJ at the time of crisis [13,15].

Combining DNA repair deficiency with mutagen exposure reveals how the repair of primary DNA lesions results in mutagenesis. When mutagen causes a distinct mutational pattern, which is uniformly amplified in case of a DNA repair deficiency, a single pathway is involved in mending DNA lesions. This is observed in *C. elegans* NER-defective strains upon exposure to UV, aristolochic acid, or aflatoxin which cause bulky DNA lesions [6]. In all cases, lesions are repaired by a single pathway, e.g. by the excision of the damaged DNA strand. On the contrary, when several pathways are involved in mending lesions caused by a mutagen, changes in signature arise upon repair deficiency [6]. For instance, methyl methanesulfonate (MMS) treatment leads to two main mutagenic base adducts: O<sup>6</sup>-methylguanine and N<sup>3</sup>-methyladenine [19]. O<sup>6</sup>-methylguanine mispairs with thymine and results in increased C > T substitutions. Such MMS-induced C > T transitions rarely occur in WT nematodes but are abundant when direct DNA repair mediated by the AGT-1/MGMT O<sup>6</sup>-methylguanine-DNA methyltransferase is compromised. N<sup>3</sup>-methyladenine leads to replication stalling and when unrepaired causes T > A, and T > C changes. It can be read through by translesion synthesis (TLS) polymerases Pol  $\kappa$  and Rev-3/Pol  $\zeta$  [20]. Pol  $\kappa$  appears to read through N<sup>3</sup>-methyladenine in a largely error-free manner [6]. In *pol-k* mutants, T > A, and T > C substitutions are increased up to 30-fold. Conversely, Rev-3/Pol  $\zeta$  deficiency results in a reduced number of SNVs at the cost of an increased number of small indels and SVs. Thus, upon MMS treatment T > A, and T > C mutations are largely caused by Rev-3/Pol  $\zeta$ -dependent translesion synthesis leading to adenine and cytosine misincorporation, but preventing the formation of small indels and SVs [6]. In summary, mutational signatures are jointly shaped by DNA-damaging events and DNA repair pathways.

Complex SV can be observed when the first and second-generation progeny of worms are analyzed after mutagenizing germ cells in the parental generation. For instance, localized clustered SVs with associated CNV were identified upon exposure to the DNA cross-linking agents cisplatin and mechlorethamine [17]. These rearrangements resemble chromoanagenesis (‘chromo’ for chromosomes and ‘anagenesis’ for reconstitution), a catastrophic genomic event first described as a cause of human inherited disorders. Copy number changes likely result from replication fork collapse at persistent DNA inter-strand cross-links, followed by promiscuous microhomology-driven strand invasions [21].

When *C. elegans* mature sperm cells are treated with ionizing irradiation, error-prone repair by MMEJ, which occurs in the zygote, leads to a large number of chromosome-to-chromosome fusions, which can be readily uncovered by NGS [22]. It appears somewhat unexpectedly that sperm DNA is mended by such an error-prone mechanism, which might also occur in humans [22].

All in all, analyzing mutational signatures in nematodes and yeast provided important insights into the basic mechanisms of mutagenesis and aided our understanding of the roles and mechanisms of repair enzymes, particularly DOG-1/FANCI, POLQ-1/Pol  $\theta$ , and HELQ-1.

## Use of cell lines in mutational pattern research

Human cell lines offer the advantage of easily obtainable cell clones and have been used extensively to obtain experimental mutational signatures. Isogenic collections of DNA repair gene knockouts have been established in human induced pluripotent stem cells (hiPSCs) [23] as well as in the human HAP1 cell line [24]. Human TK6 [25,26] and avian DT40 lymphoblastoid lines [25–30] were also employed in mutational signature studies. Many DNA repair gene knockouts obtained in HAP1 and TK6 lines are made available by Horizon and the TK6 Mutants Consortium, respectively. The important feature of these cell lines is their stable haploid (HAP1)

or near-diploid karyotype. On the other hand, mouse embryonic fibroblasts (MEFs) might exhibit a high variation of chromosomal numbers among clones as a result of the immortalization procedure, which makes them poorly suited for studying mutational signatures. Cell lines differ in their dependence on certain repair pathways for cell viability. It appears that hiPSCs might have a low tolerance for DNA double-strand breaks (DSBs) and some DSB DNA repair gene knockouts are lethal in hiPSCs [23]. On the contrary, using the DT40 cell line, it became possible to establish the only currently available isogenic collection of HR gene knockouts [29].

## Global vs cell-type specific signatures

Some of the computationally derived signatures appear to be ‘globally’ present [31] and reflect the pathways of mutagenesis common among all cell types, for example, signature SBS1, which results from deamination of 5-methylcytosine to thymine and is characterized by C > T transitions at CpG [1]. This signature reflects a ‘clock-like’ accumulation of mutations as the organism ages [32]. Other signatures were identified only in certain cancers (so-called ‘local signatures’ [31]) and are probably caused by tissue-specific exposure to exogenous or endogenous DNA-damaging agents and/or the tissue-specific prevalence or absence of certain DNA repair pathways. One of the relevant factors is cell division rate. T/G mismatches resulting from 5-methylcytosine deamination can be corrected by MMR only prior to replication. Consistent with this explanation, cancers originating from tissues with a high cell division rate display a higher contribution of signature SBS1 [32]. Likewise, mutagenesis in adult stem cells (ASCs) from human colon and small intestine is dominated by SBS1, whereas in the liver SBS5 prevails [33]. A higher proportion of C > A substitutions observed in mouse small bowel stem cells compared with other tissues might be due to tissue-specific damage due to reactive oxygen species [34].

Certain cell types might be deficient in specific DNA repair pathways or utilize certain pathways in preference to others. For example, UV-induced 6-4 photoproducts (6-4PP) are generally repaired by NER pathway. However, in NER-deficient cells topoisomerase I (TOP1) cleavage complexes become trapped in the vicinity of 6-4PP, which activates base excision repair (BER) [35]. Primary fibroblasts and other tissues might not express sufficient levels of TOP1 to support this pathway, whereas ectopic expression of TOP1 allows BER-dependent repair of 6-4PP [35]. As another example, single-nucleotide gaps, which are produced by thymidine DNA glycosylase (TDG) in the course of DNA demethylation, are filled in by short-patch BER in differentiating macrophages but by long-patch BER in differentiating neurons [36]. A notion, that the relative contribution of DNA repair pathways and the exposure to endogenous and exogenous mutagens varies between cell types of the same organism, is in line with tumors associated with certain DNA repair deficiencies being often surprisingly tissue-specific. Thus, HR deficiency is associated with breast and ovarian cancers and the lack of MMR is linked to gastrointestinal cancer. Some DNA repair gene deficiencies might produce mutational patterns only in response to certain genotoxic agents or endogenous DNA lesions. For example, epigenetic inactivation of the O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) gene sensitizes gliomas to the methylating chemotherapeutic agent temozolomide leading to C > T substitutions due to mispairing of O<sup>6</sup>-methylguanine with thymine [1]. This mutational pattern cannot be observed in MGMT + cell lines, e.g. hiPSCs, in which O<sup>6</sup>-methylguanine is efficiently demethylated to guanine through the action of the MGMT enzyme [37]. Curiously, endometrial carcinomas with mutationally inactivated MMR genes were reported to carry a higher mutation burden and to be more susceptible to immune checkpoint blockade therapy than carcinomas with epigenetically silenced *MLH1* gene [38]. This suggests that in some cases even the mechanism of how DNA repair deficiency was acquired might influence the mutation accumulation. The mutational spectra induced by DNA cross-linking agent cisplatin differ in DT40, TK6, MCF10A, and HepG2 cells presumably reflecting peculiarities of DNA repair pathways in these cell lines and highlighting the importance of isogenic systems in mutational signature research [26,39,40]. Cell line models of different tissue types and cancers should provide better insight into so-called ‘local’ [31], or cell-type specific mutational signatures. Isogenic CRISPR–Cas9 edited patient-derived cell lines, e.g. gliomas [41], are perhaps the most ‘authentic’ system that approximates cancer mutagenesis in cell culture. Another attractive experimental model are epithelial organoids, which are derived from single stem cells, are not transformed, are genetically stable, and allow for CRISPR–Cas9 genome editing. When subject to only one round of cloning and expansion, organoids enable the analysis of somatic mutations, which occurred in normal cells during mouse development as the small number of subclonal mutations, which happened during *in vitro* expansion, can be easily filtered out. It is worth noting, however, that *in vitro* mutational pattern in organoid cultures differs from somatic mutagenesis in mice [34].



## Background mutagenesis in cell lines

The question of the rate and origin of background mutagenesis in cell lines deserves special consideration. When human ASCs were expanded *in vitro* and subclonal mutations were filtered out, the ‘*in vivo* somatic mutation spectra’ were found to be a combination of signatures SBS1 (deamination of 5-methylcytosines) and SBS5 (a ‘featureless’ or ‘flat’ signature of unclear etiology) [33]. Signatures SBS1 and SBS5 were also proposed to explain mutational spectra in the human germline [42] and were shown to correlate with age at cancer diagnosis [32]. They probably represent the bona fide clock-like spontaneous mutagenesis in humans. On another hand, signature SBS18 (predominantly C > A substitutions, likely resulting from 8-oxo-dG pairing with A), was found to contribute minimally to *in vivo* mutagenesis. However, SBS18, termed ‘culture-related signature’, is the major source of spontaneous mutations in HAP1 [24] and hiPSC cells [23]. It was proposed that SBS18 is an endogenous signature that is amplified under cellular stress conditions associated with increased reactive oxygen species [31]. Noteworthy, SBS18 is not a component of spontaneous mutation spectra in cell lines, which grow in suspension, such as TK6 and DT40 [26,30], suggesting that it could be trypsinization of the attached cells during passaging or absence of cell–cell contacts and lack of autocrine and paracrine factors [43] that leads to the aforementioned cell stress. When hiPSCs were cultured in hypoxic (3% oxygen) rather than normoxic (20% oxygen) conditions, the rate of spontaneous SNVs and the fraction of C > A substitutions (SBS18) decreased while the fraction of C > T, particularly at CpG (SBS1), increased [43]. SBS 18 is greatly augmented in hiPSCs deficient for the 8-oxoguanine glycosylase (*OGG1*) consistent with guanine oxidation being the primary lesion [23]. On the other hand, the origin of the background SBS5-like signature is not clear. Since knockout of the *REVI* gene leads to decreased rate of SNVs in DT40 cells [27], it seems probable that in addition to errors of replicative polymerases, TLS might be the major contributor to spontaneous mutagenesis.

## Mutational signature of mismatch repair deficiency

As of today, the best characterized experimental mutational signatures are those caused by MMR and BRCA deficiencies. Although at least seven different SNV signatures were computationally derived from MMR defective (MMRd) cancer genomes (SBS6, 14, 15, 20, 21, 26, and 44 [1,2]), the experimental mutational spectra, which were observed in HAP1 [24], hiPSCs [23], DT40 [28], and DLD-1 [28] cell lines or in human colon organoids [44] with *MSH2*, *MSH6*, or *MLH1* genes knocked out, as well as signatures derived from MMRd *C. elegans* when adjusted to the base composition of the human genome [45], were remarkably similar, suggesting the existence of a ‘universal’ signature associated with MMR deficiency. While the analysis of this signature was greatly facilitated by the high number of SNVs and indels accumulating in MMRd cells as well as by the well-established link between MMR mutations and colon cancer, this example is the first ‘success story’ of how experimental research corrected and improved the computationally derived signatures (Figure 2). Based on experimental mutational patterns, RefSig MMR1 was proposed as a substitute for the seven computationally derived signatures [23]. Alternatively, a set of two novel MMRd-associated SNV signatures was derived from cancer mutational databases and could describe both cancer and experimental mutational spectra [28]. One of them, dominated by NCG > NTG, was suggested to result from a higher error rate of replicative polymerases on 5-methylcytosine. The second signature, characterized by T > C, C > T, and C > A might be caused by the intrinsic error rate of DNA polymerases and the lack of MMR contribution to oxidative damage repair. Interestingly, the relative contribution of the first signature is much higher in tumors compared with cell lines. Knockouts of *PMS1* and *PMS2* genes resulted in mutational spectra which were distinct from RefSig MMR1 [23]. Replication strand bias, which was previously observed in MMRd cancers, was also confirmed in cell line models [23]. On another hand, microsatellite instability (MSI), which stems from short indels at long repetitive sequences, appears to be more prominent in MMRd cancers compared with cell lines, possibly because cancers were developing over a significant amount of time and had many more replicative cycles to accumulate replication errors. However, the contribution of additional factors or mutations to MSI *in vivo* cannot be excluded [46]. Intriguingly, SBS11, which was computationally derived from TMZ-treated cancers and is characterized by C > T substitutions, was recently experimentally demonstrated to be jointly shaped by TMZ exposure and MMR deficiency, which is selected in tumors over the course of the treatment and provides resistance to TMZ [41].

## Mutational signature of homologous recombination repair deficiency

Genome analysis of *BRCA1* and *BRCA2*-deficient cancers revealed the predominance of short deletions with 1–5 bp microhomologies at breakpoints (signature ID6). These deletions were interpreted as ‘scars’ left by the



**Figure 2. The ‘universal’ mutational spectrum of MMRd cells from different organisms is a combination of computationally derived MMRd cancer signatures.**

Seven computationally derived signatures from cancer genomes [50] are represented followed by the signatures obtained from MMRd mutants of budding yeast [8], *C. elegans* [45], chicken DT40 cell line [28], human HAP1 line [24], and hiPSCs [23].

MMEJ pathway, which plays a more prominent role when HR repair is defective. Microhomology-flanked deletions in *BRC-1* (*BRCA1* orthologue)-deficient worms were demonstrated to be the result of the DNA end joining by the DNA polymerase  $\theta$  [47]. The same conclusion is supported by the analysis of tumor genomes [48] but proved to be difficult to experimentally test in human cells, where double knockout of *BRCA1* and *POLQ* might be lethal. *BRCA*-mutated tumors also displayed a ‘flat’ SNV signature (SBS3), whose origin remained enigmatic [1,2]. Remarkably, both the indel and the SNV patterns were reproduced in HR-deficient (HRd) DT40 cell lines [29,30], as well as in *BRCA*-deficient *C. elegans* [6,47] providing experimental confirmation of computationally derived signatures. The same signatures were observed not only in *BRCA1* and *BRCA2* knockouts but also in DT40 lines deficient in other HR genes, such as *PALB2* and *RAD51* paralogs (*RAD51C*, *XRCC2*, and *XRCC3*) [29]. Similarly, *C. elegans* HRd strains with a broad range of single base substitutions in conjunction with increased indels and/or structural variants include strains deficient for *RAD-51* paralogs, the *MUS-81* structure-specific nuclease and the *SMC-5/6* cohesin-like complex [6]. With regard to the origin of SNVs in HRd cell lines, an important insight emerged from the analysis of close mutation pairs, which were defined as pairs of SNVs separated by less than 100 bps while excluding SNVs immediately adjacent to each other [25]. These mutation pairs were increased in HRd lines and reduced in *BRCA1*<sup>-/-</sup> *PCNA*<sup>K164R</sup> double mutants suggesting that PCNA ubiquitination-dependent recruitment of low-fidelity TLS polymerases contributes to collateral mutagenesis, i.e. insertion of incorrect bases away from the primary lesion. It was proposed that HR is required for bypassing DNA lesions via template switching and, in the absence of this pathway, TLS activity and DNA synthesis errors are increased [25]. This is consistent with data from budding yeast, where an increased rate of SNVs associated with *RAD51* deficiency is likely mediated by Pol  $\zeta$  translesion polymerase and was shown to require its catalytic subunit *REV3* as well as *REV7* and *REV1* [8].

A noteworthy example of how the use of yeast as a model organism was instrumental in discovering the origin of cancer mutational signature is the ID4 indel signature, which is characterized by 2–5 bp deletions in

short tandem repeat sequences and small deletions at sequences with microhomology, especially at TNT sequences. This ID4 signature is enriched in RNase H2-null HeLa cells and cancers and is caused by topoisomerase 1 (TOP1) cleavage at ribonucleotides incorporated into DNA. Since TOP1 is essential in mammalian cells, the direct demonstration that ID4 is TOP1-dependent was possible only in ribonucleotide excision repair (RER)-deficient yeast, which display a very similar signature and where deletion of *TOP1* is viable [49].

## Signatures caused by mutagens

DNA-damaging physical factors (UV light and ionizing radiation) and chemical agents have long been known to cause cancers. Some cancer mutational signatures are associated with exposure to exogenous factors, e.g. SBS4 and tobacco smoking, SBS7 and UV irradiation, and SBS22 and aristolochic acid [1,2]. A comprehensive study of hiPSCs treated with a large array of genotoxins confirmed these associations and detected SNV signatures in the case of about half of the agents that were used [37]. However, the results of this study further highlighted the need to examine mutational signatures in multiple genetic backgrounds. For example, methylnitrosoguanidine (MNNG) and TMZ, which are both known to induce O<sup>6</sup>-methylguanine, resulted, respectively, in no mutational pattern and a spectrum dominated by T > C substitutions. This result is in stark contrast with a C > T dominated signature SBS11 from the genomes of TMZ-treated tumors [41,50] and is probably due to hiPSC cells expressing the MGMT enzyme, which directly reverses O<sup>6</sup>-guanine methylation.

## Future challenges and directions

In summary, *in vitro* studies using cell lines are instrumental in obtaining experimental mutational signatures caused by DNA repair defects and exogenous factors. However, limitations and challenges posed by this experimental model came to the fore. Perhaps, the most significant problem is the rarity of SVs in cultured cells. Current cell line-based systems do not accumulate sufficient numbers of, e.g. large deletions to calculate HRDetect [51] scores in HRd lines [29]. Similarly, in the study of hiPSC cells treated with various mutagens, the numbers of rearrangements and copy number alterations were too small to detect any signatures associated with SVs [37].

Interestingly, the majority of mutagen treatments in hiPSC cells resulted in robust activation of the DNA damage checkpoint, even though treatment with many genotoxic agents led to little or no mutagenesis [37]. The number of SNVs, observed upon treatment of hiPSC cells with mutagens, peaks at ~2000 per genome, although typically mutation counts are much lower. It is likely that the extent of mutagenesis that can be observed by NGS is limited by cell death triggered by DNA damage checkpoint activation and/or the accumulation of large-scale genome alterations that compromise cell proliferation. Since cancers develop over months and years, even rare events can be detected when sequencing tumors and deriving mutagenic signatures by computational approaches is possible.

Another challenge is to study mutagenesis in non-dividing cells, e.g. neurons and cardiomyocytes. Recent advances in the methodology of DNA amplification from single cells, such as primary template-directed amplification (PTA) [52], reduce technical artifacts and make it possible to follow the accumulation of mutations in non-dividing neurons [53,54].

It can be concluded from *in vitro* studies that different cell lines might employ different DNA repair pathways in response to the same type of DNA damage, which might result in different mutational signatures. This heterogeneity of responses observed *in vitro* can serve as a model for various cell types and cancers *in vivo*. Thus, in order to obtain reliable results, it is imperative to conduct the same test in multiple cell lines.

All in all, mutational signatures research needs to take advantage of human cell-based systems, model organisms, and the analysis of cancer genomes. While some pathways, such as MMR and HR are overall conserved, future studies would greatly benefit from using diverse experimental systems to account for differences in DNA repair pathways, metabolism, and the level of exposure to mutagenic factors.

## Perspectives

Mutational signatures computationally derived from cancer genomes shed light on the DNA repair deficiencies and history of mutagen exposure of the tumor. Experimental mutational signatures obtained under controlled conditions are instrumental in revealing the etiology of cancer signatures.



Initial studies were focused on WGS analysis of isogenic collections of yeast and worms. More recent work was dedicated to the investigation of mutagenesis in isogenic human cell lines building upon many insights obtained from the genomes of model organisms.

Mutational signatures are jointly shaped by the nature of the primary DNA damage and the status of DNA repair pathways. Future research should employ cells from different tissues, including non-dividing cells, to account for cell-type specificity of mutagen exposure and DNA repair competence.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This study was supported by the Korean Institute for Basic Science (IBS-R022-A2-2023) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1A6A1A03025810).

## Acknowledgements

We would like to thank Shannon Payne Bourke for helping with the preparation of the figures. We apologize to the many authors whose contributions we were not able to cite due to space limitations.

## Abbreviations

ASCs, adult stem cells; BER, base excision repair; CNVs, copy number variations; DSBs, DNA double-strand breaks; HR, homologous recombination; HRd, HR-deficient; IDs, insertions/deletions; MMEJ, microhomology-mediated end joining; MMR, mismatch repair; MMS, methyl methanesulfonate; MSI, microsatellite instability; NER, nucleotide excision repair; NGS, next-generation sequencing; SNVs, single-nucleotide variants; SVs, structural chromosomal variants; TLS, translesion synthesis; TOP1, topoisomerase I; WT, wild-type.

## References

- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V. et al. (2013) Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 <https://doi.org/10.1038/nature12477>
- Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y. et al. (2020) The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 <https://doi.org/10.1038/s41586-020-1943-3>
- Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L. et al. (2019) The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 <https://doi.org/10.1038/s41586-019-1672-7>
- Park, S., Mali, N.M., Kim, R., Choi, J.W., Lee, J., Lim, J. et al. (2021) Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* **597**, 393–397 <https://doi.org/10.1038/s41586-021-03786-8>
- Franco, I., Helgadottir, H.T., Moggio, A., Larsson, M., Vrtacnik, P., Johansson, A. et al. (2019) Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol.* **20**, 285 <https://doi.org/10.1186/s13059-019-1892-z>
- Volkova, N.V., Meier, B., Gonzalez-Huici, V., Bertolini, S., Gonzalez, S., Vohringer, H. et al. (2020) Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 <https://doi.org/10.1038/s41467-020-15912-7>
- Puddu, F., Herzog, M., Selivanova, A., Wang, S., Zhu, J., Klein-Lavi, S. et al. (2019) Genome architecture and stability in the *Saccharomyces cerevisiae* knockout collection. *Nature* **573**, 416–420 <https://doi.org/10.1038/s41586-019-1549-9>
- Loeillet, S., Herzog, M., Puddu, F., Legoix, P., Baulande, S., Jackson, S.P. et al. (2020) Trajectory and uniqueness of mutational signatures in yeast mutators. *Proc. Natl Acad. Sci. U.S.A.* **117**, 24947–24956 <https://doi.org/10.1073/pnas.2011332117>
- Meier, B., Volkova, N.V., Gerstung, M. and Gartner, A. (2020) Analysis of mutational signatures in *C. elegans*: implications for cancer genome analysis. *DNA Repair (Amst.)* **95**, 102957 <https://doi.org/10.1016/j.dnarep.2020.102957>
- Meier, B., Volkova, N.V., Hong, Y., Bertolini, S., Gonzalez-Huici, V., Petrova, T. et al. (2021) Protection of the *C. elegans* germ cell genome depends on diverse DNA repair pathways during normal proliferation. *PLoS One* **16**, e0250291 <https://doi.org/10.1371/journal.pone.0250291>
- Cheung, I., Schertzer, M., Rose, A. and Lansdorff, P.M. (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.* **31**, 405–409 <https://doi.org/10.1038/ng928>
- Koole, W., van Schendel, R., Karambelas, A.E., van Heteren, J.T., Okihara, K.L. and Tijsterman, M. (2014) A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* **5**, 3216 <https://doi.org/10.1038/ncomms4216>

- 13 Roerink, S.F., van Schendel, R. and Tijsterman, M. (2014) Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res.* **24**, 954–962 <https://doi.org/10.1101/gr.170431.113>
- 14 Kamp, J.A., Lemmens, B., Romeijn, R.J., Changoer, S.C., van Schendel, R. and Tijsterman, M. (2021) Helicase Q promotes homology-driven DNA double-strand break repair and prevents tandem duplications. *Nat. Commun.* **12**, 7126 <https://doi.org/10.1038/s41467-021-27408-z>
- 15 Kim, E., Kim, J., Kim, C. and Lee, J. (2021) Long-read sequencing and de novo genome assemblies reveal complex chromosome end structures caused by telomere dysfunction at the single nucleotide level. *Nucleic Acids Res.* **49**, 3338–3353 <https://doi.org/10.1093/nar/gkab141>
- 16 Wolden, M.R., Flibotte, S., Moerman, D.G. and Ahmed, S. (2011) DNA synthesis generates terminal duplications that seal end-to-end chromosome fusions. *Science* **332**, 468–471 <https://doi.org/10.1126/science.1199022>
- 17 Meier, B., Cooke, S.L., Weiss, J., Bailly, A.P., Alexandrov, L.B., Marshall, J. et al. (2014) *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* **24**, 1624–1636 <https://doi.org/10.1101/gr.175547.114>
- 18 Seo, B., Kim, C., Hills, M., Sung, S., Kim, H., Kim, E. et al. (2015) Telomere maintenance through recruitment of internal genomic regions. *Nat. Commun.* **6**, 8189 <https://doi.org/10.1038/ncomms9189>
- 19 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](https://doi.org/10.1016/0027-5107(90)90173-2)
- 20 Yoon, J.H., Roy Choudhury, J., Park, J., Prakash, S. and Prakash, L. (2017) Translesion synthesis DNA polymerases promote error-free replication through the minor-groove DNA adduct 3-deaza-3-methyladenine. *J. Biol. Chem.* **292**, 18682–18688 <https://doi.org/10.1074/jbc.M117.808659>
- 21 Liu, P., Erez, A., Nagamani, S.C., Dhar, S.U., Kolodziejaska, K.E., Dharmadhikari, A.V. et al. (2011) Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **146**, 889–903 <https://doi.org/10.1016/j.cell.2011.07.042>
- 22 Wang, S., Meyer, D.H. and Schumacher, B. (2023) Inheritance of paternal DNA damage by histone-mediated repair restriction. *Nature* **613**, 365–374 <https://doi.org/10.1038/s41586-022-05544-w>
- 23 Zou, X., Koh, G.C.C., Nanda, A.S., Degasperi, A., Urgo, K., Roumeliotis, T.I. et al. (2021) A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 <https://doi.org/10.1038/s43018-021-00200-0>
- 24 Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I. and Nik-Zainal, S. (2018) Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun.* **9**, 1744 <https://doi.org/10.1038/s41467-018-04052-8>
- 25 Poti, A., Szikriszt, B., Gervai, J.Z., Chen, D. and Szuts, D. (2022) Characterisation of the spectrum and genetic dependence of collateral mutations induced by translesion DNA synthesis. *PLoS Genet.* **18**, e1010051 <https://doi.org/10.1371/journal.pgen.1010051>
- 26 Szikriszt, B., Poti, A., Nemeth, E., Kanu, N., Swanton, C. and Szuts, D. (2021) A comparative analysis of the mutagenicity of platinum-containing chemotherapeutic agents reveals direct and indirect mutagenic mechanisms. *Mutagenesis* **36**, 75–86 <https://doi.org/10.1093/mutage/geab005>
- 27 Chen, D., Gervai, J.Z., Poti, A., Nemeth, E., Szeltner, Z., Szikriszt, B. et al. (2022) BRCA1 deficiency specific base substitution mutagenesis is dependent on translesion synthesis and regulated by 53BP1. *Nat. Commun.* **13**, 226 <https://doi.org/10.1038/s41467-021-27872-7>
- 28 Nemeth, E., Lovrics, A., Gervai, J.Z., Seki, M., Rospo, G., Bardelli, A. et al. (2020) Two main mutational processes operate in the absence of DNA mismatch repair. *DNA Repair (Amst.)* **89**, 102827 <https://doi.org/10.1016/j.dnarep.2020.102827>
- 29 Poti, A., Gyergyak, H., Nemeth, E., Ruzs, O., Toth, S., Kovacs, C. et al. (2019) Correlation of homologous recombination deficiency induced mutational signatures with sensitivity to PARP inhibitors and cytotoxic agents. *Genome Biol.* **20**, 240 <https://doi.org/10.1186/s13059-019-1867-0>
- 30 Zamborszky, J., Szikriszt, B., Gervai, J.Z., Pipek, O., Poti, A., Krzystanek, M. et al. (2017) Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene* **36**, 746–755 <https://doi.org/10.1038/nc.2016.243>
- 31 Koh, G., Degasperi, A., Zou, X., Momen, S. and Nik-Zainal, S. (2021) Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **21**, 619–637 <https://doi.org/10.1038/s41568-021-00377-7>
- 32 Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S. et al. (2015) Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 <https://doi.org/10.1038/ng.3441>
- 33 Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N. et al. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 <https://doi.org/10.1038/nature19768>
- 34 Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D.C., Tamuri, A.U. et al. (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 <https://doi.org/10.1038/nature13448>
- 35 Saha, L.K., Wakasugi, M., Akter, S., Prasad, R., Wilson, S.H., Shimizu, N. et al. (2020) Topoisomerase I-driven repair of UV-induced damage in NER-deficient cells. *Proc. Natl Acad. Sci. U.S.A.* **117**, 14412–14420 <https://doi.org/10.1073/pnas.1920165117>
- 36 Wang, D., Wu, W., Callen, E., Pavani, R., Zolnerowich, N., Kodali, S. et al. (2022) Active DNA demethylation promotes cell fate specification and the DNA damage response. *Science* **378**, 983–989 <https://doi.org/10.1126/science.add9838>
- 37 Kucab, J.E., Zou, X., Morganello, S., Joel, M., Nanda, A.S., Nagy, E. 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>
- 38 Chow, R.D., Michaels, T., Bellone, S., Hartwich, T.M.P., Bonazzoli, E., Iwasaki, A. et al. (2023) Distinct mechanisms of mismatch-repair deficiency delineate two modes of response to anti-PD-1 immunotherapy in endometrial carcinoma. *Cancer Discov.* **13**, 312–331 <https://doi.org/10.1158/2159-8290.CD-22-0686>
- 39 Boot, A., Huang, M.N., Ng, A.W.T., Ho, S.C., Lim, J.Q., Kawakami, Y. et al. (2018) In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res.* **28**, 654–665 <https://doi.org/10.1101/gr.230219.117>
- 40 Szikriszt, B., Poti, A., Pipek, O., Krzystanek, M., Kanu, N., Molnar, J. et al. (2016) A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol.* **17**, 99 <https://doi.org/10.1186/s13059-016-0963-7>
- 41 Touat, M., Li, Y.Y., Boynton, A.N., Spurr, L.F., Iorgulescu, J.B., Bohrsen, C.L. et al. (2020) Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 <https://doi.org/10.1038/s41586-020-2209-9>
- 42 Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Turki, S.A. et al. (2016) Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 <https://doi.org/10.1038/ng.3469>
- 43 Kuijk, E., Jager, M., van der Roest, B., Locati, M.D., Van Hoeck, A., Korzelius, J. et al. (2020) The mutational impact of culturing human pluripotent and adult stem cells. *Nat. Commun.* **11**, 2493 <https://doi.org/10.1038/s41467-020-16323-4>

- 44 Drost, J., van Boxtel, R., Blokzijl, F., Mizutani, T., Sasaki, N., Sasselli, V. et al. (2017) Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **358**, 234–238 <https://doi.org/10.1126/science.aao3130>
- 45 Meier, B., Volkova, N.V., Hong, Y., Schofield, P., Campbell, P.J., Gerstung, M. et al. (2018) Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and human cancers. *Genome Res.* **28**, 666–675 <https://doi.org/10.1101/gr.226845.117>
- 46 Hayashida, G., Shioi, S., Hidaka, K., Fujikane, R., Hidaka, M., Tsurimoto, T. et al. (2019) Differential genomic destabilisation in human cells with pathogenic MSH2 mutations introduced by genome editing. *Exp. Cell Res.* **377**, 24–35 <https://doi.org/10.1016/j.yexcr.2019.02.020>
- 47 Kamp, J.A., van Schendel, R., Dilweg, I.W. and Tijsterman, M. (2020) BRCA1-associated structural variations are a consequence of polymerase theta-mediated end-joining. *Nat. Commun.* **11**, 3615 <https://doi.org/10.1038/s41467-020-17455-3>
- 48 Hwang, T., Reh, S., Dunbayev, Y., Zhong, Y., Takata, Y., Shen, J. et al. (2020) Defining the mutation signatures of DNA polymerase theta in cancer genomes. *NAR Cancer* **2**, zcaa017 <https://doi.org/10.1093/narcan/zcaa017>
- 49 Reijns, M.A.M., Parry, D.A., Williams, T.C., Nadeu, F., Hindshaw, R.L., Rios Szwed, D.O. et al. (2022) Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* **602**, 623–631 <https://doi.org/10.1038/s41586-022-04403-y>
- 50 Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N. et al. (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 <https://doi.org/10.1093/nar/gky1015>
- 51 Davies, H., Glodzik, D., Morganello, S., Yates, L.R., Staaf, J., Zou, X. et al. (2017) HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 <https://doi.org/10.1038/nm.4292>
- 52 Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y. et al. (2021) Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl Acad. Sci. U.S.A.* **118**, e2024176118 <https://doi.org/10.1073/pnas.2024176118>
- 53 Luquette, L.J., Miller, M.B., Zhou, Z., Bohrsen, C.L., Zhao, Y., Jin, H. et al. (2022) Single-cell genome sequencing of human neurons identifies somatic point mutation and indel enrichment in regulatory elements. *Nat. Genet.* **54**, 1564–1571 <https://doi.org/10.1038/s41588-022-01180-2>
- 54 Miller, M.B., Huang, A.Y., Kim, J., Zhou, Z., Kirkham, S.L., Maury, E.A. et al. (2022) Somatic genomic changes in single Alzheimer's disease neurons. *Nature* **604**, 714–722 <https://doi.org/10.1038/s41586-022-04640-1>