Mutational signatures are jointly shaped by DNA damage and repair

Nadezda V Volkova*[1], Bettina Meier*[2], Víctor González-Huici*[2,3], Simone Bertolini[2], Santiago Gonzalez[1,3], Federico Abascal[4], Iñigo Martincorena[4], Peter J Campbell[4-6], Anton Gartner[#2,7,8] and Moritz Gerstung[#1,9]

* these authors contributed equally
# to whom correspondence should be addressed

1) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK.
2) Centre for Gene Regulation and Expression, University of Dundee, Dundee, Scotland.
3) Current affiliation, Institute for Research in Biomedicine (IRB Barcelona), Parc Científic de Barcelona, Barcelona, Spain.
4) Cancer, Ageing and Somatic Mutation, Wellcome Sanger Institute, Hinxton, UK.
5) Department of Haematology, University of Cambridge, Cambridge, UK.
6) Department of Haematology, Addenbrooke's Hospital, Cambridge, UK.
7) Center for Genomic Integrity, Institute for Basic Science, Ulsan, Republic of Korea
8) Department of Biological Sciences, School of Life Sciences, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
9) European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany.

Correspondence:

Dr Anton Gartner
Centre for Gene Regulation and Expression
University of Dundee
Dow Street Dundee, DD1 5EH
Scotland.
Tel: +44 (0) 1382 385809
E-mail: a.gartner@dundee.ac.uk

Dr Moritz Gerstung
European Molecular Biology Laboratory
European Bioinformatics Institute (EMBL-EBI)
Hinxton, CB10 1SA
UK.
Tel: +44 (0) 1223 494636
E-mail: moritz.gerstung@ebi.ac.uk

Highlights

- Combining exposure to DNA damaging agents and DNA repair deficiency in *C. elegans* leads to altered mutation rates and new mutational signatures
- Mutagenic effects of genotoxic exposures are generally exacerbated by DNA repair deficiency
- Mutagenesis of UVB and alkylating agents is reduced in translesion synthesis polymerase deficiencies
- Human cancer genomes contain examples of DNA damage/repair interactions, but mutations in DNA repair genes usually only associate with moderate mutator phenotypes, in line with evolutionary theory

Summary

Mutations arise when DNA lesions escape DNA repair. To delineate the contributions of DNA damage and DNA repair deficiency to mutagenesis we sequenced 2,721 genomes of 54 *C. elegans* strains, each deficient for a specific DNA repair gene and wild-type, upon exposure to 12 different genotoxins. Combining genotoxins and repair deficiency leads to differential mutation rates or new mutational signatures in more than one third of experiments. Translesion synthesis polymerase deficiencies show dramatic and diverging effects. Knockout of Polκ dramatically exacerbates the mutagenicity of alkylating agents; conversely, Polζ deficiency reduces alkylation- and UV-induced substitution rates. Examples of DNA damage-repair deficiency interactions are also found in cancer genomes, although cases of hypermutation are surprisingly rare despite signs of positive selection in a number of DNA repair genes. Nevertheless, cancer risk may be substantially elevated even by small increases in mutagenicity according to evolutionary multi-hit theory. Overall, our data underscore how mutagenesis is a joint product of DNA damage and DNA repair, implying that mutational signatures may be more variable than currently anticipated.

Introduction

A cell's DNA is constantly altered by a multitude of genotoxic stresses including environmental toxins and radiation, DNA replication errors and endogenous metabolites, all rendering the maintenance of the genome a titanic challenge. Organisms thus evolved an armamentarium of DNA repair processes to detect and mend DNA damage, and to eliminate or permanently halt the progression of genetically compromised cells. Nevertheless, some DNA lesions escape detection and repair or are mended by error prone pathways, leading to mutagenesis — the process that drives evolution but also inheritable disease and cancer.

The multifaceted nature of mutagenesis results in distinct mutational spectra, characterized by the specific distribution of single and multi-nucleotide variants (SNVs and MNVs), short insertions and deletions (indels), large structural variants (SVs), and copy number alterations. Studying the combined mutational patterns can yield insights into the nature of DNA damage and DNA repair processes. Indeed, large scale cancer genome and exome sequencing allowed to computationally deduce more than 50 mutational signatures of base substitutions (Alexandrov et al., 2018, 2013) in cancer and normal cells. Some of these signatures, generally deduced by computational pattern recognition, have evident associations with exposure to known mutagens such as UV light, tobacco smoke, the food contaminants aristolochic acid and aflatoxins (Alexandrov et al., 2016; Helleday et al., 2014; Poon et al., 2013), or with DNA repair deficiency syndromes and compromised DNA replication. The latter include defects in homologous recombination (HR), mismatch repair (MMR), or DNA polymerase epsilon (POLE) proofreading. However, the etiology of many of the computationally extracted signatures still has to be established and it is not clear if these signatures have a one to one relationship to mutagen exposure or DNA repair defects.

The association between mutational spectra and their underlying mutagenic process is further complicated given that mutations arise from the action of primary DNA lesions, that include a multitude of base modifications and DNA adducts, single and double strand breaks, intra- and inter-strand DNA crosslinks, and counteraction of the DNA repair machinery. This implies that there are at least two unknowns that contribute to a mutational spectrum. The effects of such interactions are exemplified by cancers with combined MMR deficiency and *POLE* mutations that lead to DNA replication errors, which display a profoundly different spectrum compared to samples with the same *POLE* deficiency but intact MMR (Haradhvala et al., 2018). The importance of DNA repair capacity is also

evident when considering the apparent tissue specificity of some cancer-associated DNA repair syndromes; for example inherited defects in HR are commonly associated with breast and ovarian cancer, while Lynch syndrome characterized by MMR deficiency is primarily linked to colorectal cancer (Sieber et al., 2005). It is established that thousands of DNA lesions occur during a single cell cycle, even in healthy cells (Tubbs and Nussenzweig, 2017). However, only a tiny proportion of primary DNA lesions ultimately result in mutation, further highlighting the importance of the mechanisms that mend DNA damage in the genesis of mutational spectra.

Here we experimentally investigate the counteracting roles of genotoxic processes and the DNA repair machinery. Using *C. elegans* whole genome sequencing, we determine the mutational spectra resulting from exposure to 12 genotoxic agents in wild-type and 53 DNA repair defective lines, encompassing most known DNA repair and DNA damage response pathways. Experiments combining genotoxin exposure and DNA repair deficiency show signs of altered mutagenesis, signified by either higher or lower rates of mutations as well as altered mutation spectra. These interactions highlight how DNA lesions arising from the same genotoxin are mended by a number of DNA repair pathways, often specific towards a particular type of DNA damage and therefore changing mutation spectra usually in subtle but sometimes also dramatic ways.

In addition, analysing 9,946 cancer exomes from the TCGA collection we reveal that loss of function mutations in DNA repair pathways are common across cancer types. Yet except for rare and extreme cases these mutations seemingly have moderate mutagenic effects, which appear small compared to the observed inter-sample heterogeneity. Still a number of cases displaying genotoxin-repair interactions are found, such as *MGMT* silencing and DNA alkylation by temozolomide. Despite these small effects, DNA repair defects drive carcinogenesis, as evidenced by an excess of missense and truncating mutations in DNA repair and damage response genes. These apparently contradictory findings can be explained by the fact that cancer transformation usually requires a combination of 2-10 driver gene mutations, the chances of which rise steeply even for small increase in mutation rates. It thus emerges that seemingly subtle mutagenic phenomena can be important for cancer development, underscoring the need to understand the contributions to mutagenesis in great detail.

Results

Experimental mutagenesis in *C. elegans*

To determine how the interplay between mutational processes and DNA repair status defines mutational spectra we used 54 *C. elegans* strains, including wild-type and 53 lines defective for DNA repair and DNA damage sensing (**Figure 1A**). These cover all major repair pathways including direct damage reversal (DR), base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), double strand break repair (DSBR), translesion synthesis (TLS), DNA crosslink repair (CLR), DNA damage sensing checkpoints (DS), non-essential components of the DNA replication machinery, helicases that act in various DNA repair pathways, telomere replication genes, apoptosis regulators and components of the spindle checkpoint involved in DNA damage response (Arvanitis et al., 2013; Bertolini et al., 2017; Boulton et al., 2002; Lans and Vermeulen, 2011; Tijsterman et al., 2002) (**Supplementary Table 1**). To inflict DNA damage we used 12 genotoxic agents that included UVB-, X- and γ-radiation; alkylating agents such as the ethylating agent ethyl-methane-sulfonate (EMS), methylating agents dimethyl-sulfate (DMS) and methyl-methane-sulfonate (MMS); aristolochic acid and aflatoxin-B1, which form bulky DNA adducts; hydroxyurea as a replication fork stalling agent; and cisplatin, mechlorethamine (nitrogen mustard) and mitomycin C known to form DNA intra- and inter-strand crosslinks.

To investigate the level of background mutagenesis for each genotype in the absence of exogenous mutagens, we clonally propagated wild-type and DNA repair defective *C. elegans* lines over 5-40 generations in mutation accumulation experiments (**Methods, Figure 1A**). In addition, to measure the effects of genotoxic agents, nematodes were exposed to target both male and female germ cells and mutational spectra were analysed by sequencing clonally derived progeny (Meier et al., 2014). Both approaches take advantage of the *C. elegans* 3 days generation time and its hermaphroditic, self-fertilizing mode of reproduction. In both experimental setups, single zygotes provide a single cell bottleneck, and subsequent self-fertilization yields clonally amplified damaged DNA for subsequent whole genome sequencing of the 100Mbp genome (**Figure 1A**). Experiments were typically done in triplicate with dose escalation for genotoxin treatments.

Overall, we analysed a total of 2,721 whole genome sequenced samples, comprising 478 mutation accumulation samples, 237 samples from genotoxin-treated DNA repair proficient wild-type lines and 2,006 samples from interaction experiments combining genotoxin-treatment with DNA repair deficiency (**Figure 1B**). These harboured a total of 166,624 acquired mutations (average of ~60 mutations per sample) comprising 139,136 SNVs, 1,151 MNVs, 24,337 indels and 2,000 SVs. Attribution of observed mutations to each genotoxin, to distinct DNA repair deficiencies and to their combined action was done using a Negative Binomial regression analysis leveraging the explicit knowledge of genotoxin dose, genotype, and the number of generations across all samples to obtain absolute mutation rates and signatures, measured in units of mutations/generation/dose (**Methods**).

Baseline mutagenic effects of DNA damage and repair deficiency in *C. elegans*

Comparing the genomes of the first and last generations from the 478 mutation accumulation samples, we found a generally low rate of mutagenesis in the range of 0.8 to 2 mutations per generation as reported previously (Meier et al., 2014, 2018) (**Supplementary Figure 1)**. Notable exceptions are lines carrying knockouts of genes affecting MMR, TLS and DSBR (**Figure 2B**). Previously described *mlh-1* MMR defective line showed high base substitution rates combined with a characteristic pattern of 1bp insertions and deletions at homopolymeric repeats (Meier et al., 2018). *polh-1* and *rev-3* TLS polymerases knockouts yielded a characteristic signature dominated by 50-400bp deletions (Roerink et al., 2014), and *smc-6* mutants, defective for the Smc5/6 Cohesin complex involved in HR, showed an increased incidence of SVs (**Figure 2B**). **Supplementary Figure 1** and **Supplementary Table 2** provide a high-level summary of the experimental mutational signatures for all 54 genetic backgrounds.

Laying out the multi-dimensional mutation spectra in two dimensions based on their similarity revealed some general trends: Genotoxins generally appear to have a stronger influence on the mutational spectrum of a sample than its genetic background (**Figure 2A, Supplementary Figure 2**, for genotoxic effects in wild-type). Notable clusters correspond to the methylating agents MMS and DMS, which led to characteristic T>A and T>C mutations, as well as a cluster of samples characterized by C>T mutations caused by the ethylating agent EMS (**Figure 2A, C**). Ionizing radiation (UVB, X- and ɣ-rays) caused single and double nucleotide substitutions, but also deletions and structural variants. UVB

irradiation showed characteristic C>T transitions in a YpTpH context (Y=C/T; H=A/C/T), and also an unexpected prevalence of T>A transversions in a TpTpA context. Bulky adducts created by aristolochic acid and aflatoxin-B1 showed mutational spectra similar to those observed in exposed human cancers and cell lines with typical C>A and T>A mutations, respectively (Huang et al., 2017; Poon et al., 2013). Lastly, cisplatin exposure induced C>A transversions in a CpCpC and CpCpG context as well as deletions and structural variants, in agreement with previous studies (**Figure 2 A,C**) (Boot et al., 2018; Kucab et al., 2019; Meier et al., 2014; Szikriszt et al., 2016). A summary of all mutational spectra associated with the 12 genotoxic exposures in wild-type is shown in **Supplementary Figure 2** and **Supplementary Table 2.** At a mean cosine similarity of 0.63 (range 0.20-0.84) these experimental signatures generally display a good, although not perfect level of similarity with their human experimental counterparts (Kucab et al., 2019) and also with computationally derived cancer signatures with the same suspected origins (**Supplementary Figure 3**). This similarity reflects the fact that the majority of DNA repair pathways are highly conserved among eukaryotes, but also that DNA repair capacity and genotoxin metabolism may differ moderately between nematodes, human cell lines and cancer cells.

Interactions of DNA damage and DNA repair shape the mutational spectrum in *C. elegans*

Mutation rates and signatures can change considerably when combining genotoxin exposure and DNA repair deficiency. These phenomena are exemplified by experiments of MMS exposure and translesion synthesis polymerase *polk-1* knockouts: In wild-type *C. elegans,* DNA methylation by MMS leads to T>M (T>A or T>C) mutations at a rate of approximately 230 mutations/mM/generation (**Figure 3A, top**). Deficiency of polymerase κ increases the total mutation rate 17x to approximately 3,800 mutations/mM/generation (**Figure 3A, right**). This increase also coincides with a distinct change in the mutational spectrum with a prominent peak of T>A transversions in a TpTpT context (**Figure 3A, middle**). Quantifying these changes relative to the wild-type mutation spectrum and accounting for possible genotoxin-independent effects of polymerase κ deficiency (**Methods**) reveals that the rate of T>M mutations is approximately 10-100x higher, especially in TpTpN and CpTpN contexts (**Figure 3A, bottom**). These figures indicate that 90-99% of MMS-induced mutations are avoided by Polκ driven TLS. T>M mutagenesis is likely to be caused by N3-methyladenosine, which stalls replicative polymerases and is mended by several TLS polymerases (Yoon et al., 2017). Our data indicate that TLS by Polκ in wild-type *C. elegans*

occurs at a low error rate; in the *polk-1* mutant, however, bypass has to be achieved by other, error-prone TLS polymerases, resulting in largely increased T>M mutagenesis. A similar change of T>M SNVs in the mutation spectrum of *polk-1* mutants was observed for DNA ethylation by EMS, at a rate of 10x (**Supplementary Figure 4**).

O6-methylguanine, which mispairs with thymine resulting in C>T transitions, is a further DNA modification induced by MMS. Treatment of wild-type *C. elegans* with MMS, however, yields less than 10 C>T mutations/mM (**Figure 3B**). In contrast, combining MMS exposure with alkyl-transferase *agt-1* deficiency increases C>T mutations by a factor of 12, while leaving the rate of T>M mutations entirely unchanged. This demonstrates that AGT-1 specifically reversemost O6-methyl-guanine adducts, thus acting as the functional *C.elegans* ortholog of the human methylguanine DNA methyltransferase *MGMT*. Interestingly, a similar, but weaker interaction of *agt-1* was observed with the ethylating agent EMS, leading to an increase in C>T transitions by a factor of 1.5, indicating that AGT-1 is involved in, but less efficient at, removing ethyl groups, consistent with reports in *E. coli (Taira et al., 2013)* (**Supplementary Figure 4**).

Entirely distinct phenomena occur upon UVB irradiation. Error prone TLS is a key mechanism to overcome UV-induced DNA damage such as cyclobutane pyrimidine dimers, which stall replication polymerases. UVB exposure led to approximately 10 mutations/100J/m^2, comprised mostly of C>T and T>C transitions (**Figure 3C**). Paradoxically, knockout of *rev-3*, which encodes the catalytic subunit of TLS Polζ, reduced UVB mutation rates 1.5-fold, mostly by a reduction in base substitutions, which outweighs a 4-fold increase in deletions > 50bp (**Figure 3C**). To protect the genome from such mutations, which are likely to be deleterious, it is believed that TLS bypasses UV lesions at the cost of a higher base substitution rate (Yoon et al., 2019). A similar phenomenon was observed for *rev-3* mutants under MMS exposure, which led to an approximately 20% decrease of single nucleotide variants but increased deletions (**Supplementary Figure 4**), indicating that DNA repair in some cases adds to overall mutagenicity.

In contrast to the above examples where DNA repair deficiency changes mutational signatures, knockouts of the NER genes *xpf-1* and *xpc-1* uniformly increased the rate of UVB induced mutagenesis by a factor of 20 and 32, respectively, across the entire mutation spectrum (**Figure 3D**). This uniform increase indicates that NER is involved in repairing the large majority of UVB damage including both single- and multi- nucleotide lesions.

Interestingly, *xpf-1* and *xpa-1* knockouts also uniformly increased the mutation burden resulting from EMS and MMS alkylation by a factor of 2 indicating that NER also contributes to error-free repair of alkylation damage (**Supplementary Figure 5**). Similarly, *xpf-1* mutants showed a five-fold increase in mutations after aristolochic acid exposure, demonstrating that NER repairs both small and bulky DNA adducts (**Supplementary Figure 4**).

These examples of genotoxin-repair interactions are not uncommon: In total, 72/196 (37%, at FDR of 10%) of combinations displayed an interaction between DNA repair status and genotoxin treatment (**Figure 4A**). Conversely, more than half of combinations produced mutation spectra which could be fully described as the superposition of the wild-type genotoxin signature and that of the DNA repair deficiency background, indicating that the two processes acted independently. Usually genotoxin-repair interactions increase the numbers of mutations obtained for a given dose of mutagen, leaving the mutational spectrum mostly unchanged, others have a profound impact on mutational spectra (**Figure 4B**, for a comprehensive list of effects see **Supplementary Figure 5** and **Supplementary Table 3)**. The emergence of distinct mutational spectra depending on DNA repair status reflects how multiple repair enzymes and pathways contribute to mending damaged DNA as in the case of DNA alkylation. If DNA repair is predominantly achieved by one pathway such as by NER for UVB damage and bulky DNA adducts, the number of mutations is increased, but the signature tends to remain unchanged.

To illustrate the overall magnitude of these interaction effects, we estimate that of the 141,004 mutations we observed upon treatment with genotoxins in DNA repair deficient strains 23% of mutations were attributed to the endogenous mutagenicity of DNA repair deficiency genotypes independent of genotoxic exposure, and 62% of mutations would be attributed to genotoxic exposures independent of the genetic background. Of these, 2% were not observed due to negative interactions of genotoxins and DNA repair deficiency, such as UVB exposure of TLS knockout strains, and 17% of mutations were added because of the positive interactions, which increased mutagenicity (**Figure 4C**).

Mutagenic consequences of DNA repair and damage sensing deficiency in human cancers

Mutation is the driving force of cancer. Studying mutagenic interactions between DNA damage and repair in primary cancers, however, is more challenging than in controlled mutagenesis experiments because the history of DNA repair pathway deficiency and the cumulative genotoxic exposures of a given sample are generally unknown. To establish a catalogue of potential DNA repair pathway deficiency in human cancers we annotated missense and loss-of-function mutations, as well as epigenetic gene silencing in 81 core genes of the 9 consensus DNA repair pathways (Knijnenburg et al., 2018; Pearl et al., 2015) across 9,946 samples and 30 cancer types available from TCGA (**Supplementary Table 4**). As reported previously, we found that heterozygous mutations with at least one gene of a given repair pathway being mutated are very common across DNA repair pathways, reaching nearly 100% for the DNA damage sensing (DS) pathway in ovarian cancers and uterine carcinosarcomas due to very high rates of *TP53* mutations (Knijnenburg et al., 2018)**; Supplementary Figure 5A, Supplementary Table 4**). In contrast, biallelic DNA repair gene deficiency is generally rare (**Supplementary Figure 5B**), and most prevalent in the DS pathway (11% pan-cancer), but also MMR (7% of endometrial cancers), as well as HR (6% of ovarian cancers). The age-adjusted mutation burden associated with mono- and biallelic DNA repair deficiency broadly displayed the expected hypermutator phenotypes, including biallelic MMR deficiency, monoallelic *POLE* exonuclease mutations, and also biallelic mutations in HR (**Supplementary Figure 5C, D**). These patterns are consistent with the findings of our experimental screen where most DNA repair deficiencies exhibited only a weak mutator phenotype unless further probed by genotoxic exposures (**Figure 2A**).

In order to infer interactions between suspected genotoxin exposures and DNA repair deficiency, we next adapted the Poisson model used in our mutagenesis screen to simultaneously extract signatures and DNA repair effects thereon (see **Methods**). Applying this approach to a range of cancers with suspected genotoxic exposures, we found that the observed DNA repair and damage interaction effects were usually moderate. Yet a number of noteworthy examples exist, illustrating the actions of different DNA repair pathways, including DR, MMR, and TLS, in response to particular types of DNA damage *in vivo* (**Figure 5A**). The strongest interaction, which is even therapeutically exploited, occurs between the human *agt-1* ortholog *MGMT* and temozolomide in temozolomide-treated glioblastomas (**Figure 5B**) (Kim et al., 2015). This is in good agreement with our experimental findings that the nature of mutation spectra detected upon EMS, DMS or MMS alkylation depends on the status of *agt-1* **(Figure 3B)**. We note that the signature associated

with temozolomide exposure in *MGMT* deficient cancers leads to a more characteristic C>T spectrum in a NpCpY context (Y = C or T).

A further noteworthy interaction arises between MMR deficiency and a range of other mutational processes. These include concurrent *POLE* mutations manifesting in an increased C>A mutagenesis in a NpCpT context (**Figure 5C**), as noted previously (Haradhvala et al., 2018; Shlien et al., 2015). Moreover, MMR deficiency changes the rates of mutational indel processes causing mononucleotide insertions and deletions in homopolymeric stretches (**Supplementary Figure 6D**), which arise at variable rate across tissues in MMR proficient samples (Alexandrov et al., 2018). Compared to MMR proficient samples, the rate of single base deletions per year was on average 5-fold higher in MMR deficient samples following the same trend across different tissues; the rate of single base insertions was increased on average 4-fold, and the rate of C>T at CpG sites was 3-fold higher in the samples with MMR defects compared to the tissue baseline. These data also square well with the observation that MMR deficiency is associated with a number of additional mutational signatures of unknown aetiology (Meier et al., 2018), suggesting that these may reflect different mutagenic processes which are exacerbated by MMR deficiency.

A last example of DNA damage-repair interaction is conferred by APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) and error-prone TLS-driven BER. APOBEC deaminates single-stranded cytosines to uracil, which pairs with adenine during replication, leading to C>T mutations. Uracil is thought to be removed by Uracil DNA Glycosylase UNG; subsequent synthesis by error-prone TLS REV1 leads to C>G mutations (Morganella et al., 2016; Roberts and Gordenin, 2014). Indeed, a lack of C>G mutations has been observed in a cancer cell line with *UNG* silencing (Kim et al., 2015; Petljak et al., 2019). In human cancers, evidence of *REV1* and *UNG* mutation or silencing contributing to APOBEC mutagenesis is weaker, but confirms the expected trend: On average, samples with defective *REV1* or *UNG* display an 8% decrease in C>G mutations (**Figure 5D**). However, *REV1* and *UNG* status at diagnosis only explains a small fraction of APOBEC mutations with low rates of C>G transversion, suggesting that either other processes contribute to these variants or that the underlying cause may not be detectable at the time of diagnosis anymore.

Notably no effect was observed for NER variants in lung and skin cancers, although one might expect NER involvement in repairing bulky DNA adducts derived from tobacco smoke and UV light **(Figure 5A)**. Similarly NER-deficient bladder cancers with recurrent

mutations in *ERCC2/XPD* did not show a strong effect on the mutation burden, despite reports of a mild increase of a signature similar to COSMIC signature 5 (Kim et al., 2016). Taken together these data demonstrate that definitive examples of mutagenic interactions between DNA damage and repair can be found in human cancers, but these are restricted to cases with the strongest effects. The unknown lifetime exposure and unknown timing of somatically acquired DNA repair deficiency makes it very challenging to establish associations of weaker effects in cancers. Indeed, weak interactions between primary DNA lesions and repair defects are also most common in our *C. elegans* experimental screen (**Figure 4A**).

Selection analysis confirms a cancer-promoting effect of DNA repair mutations

Given that the association between DNA repair pathway defects and mutation phenotypes is weaker than expected, we took on an evolutionary approach to investigate if changes in DNA repair status nevertheless affect oncogenesis. A high-level of recurrence of non-silent variants in a gene across cancer samples indicates that the given gene is positively selected and promotes cancer progression. To assess whether DNA repair pathways mutations are positively selected in cancers, we calculated the rate of non-silent to silent mutations dN/dS relative to its expected ratio based on silent variants and epigenetic covariates (Greenman et al., 2006; Martincorena et al., 2017) for a set of 248 DNA repair genes and members of DNA repair pathways (**Supplementary Table 4**). A value of dN/dS = 1 indicates no selection, while values greater or less than 1 signify positive and negative selection, respectively. Indeed, a variety of DNA repair genes and pathways analysed showed a high dN/dS ratio, indicating positive selection (**Figure 6A**). Pathway-level analysis confirmed the pan-cancer excess of non-silent mutations in genes encoding for DNA damage sensing (DS) proteins including *TP53*, *ATR* and *ATM* (**Supplementary Table 5**). Equally, nonsense mutations are enriched in MMR genes in uterine and stomach cancers. Overall, 7/248 genes involved in DNA repair displayed significant signs of positive selection in missense mutations, and 17/248 DNA repair genes are enriched in truncating mutations (FDR<10%; **Figure 6A, Supplementary Table 5**). Among these the majority are known cancer genes such as *TP53, IDH1, PTEN, ATM, ATRX, SMARCA4, BRCA1/2*, *MLH1*, but we also detected new genes with mild signs of positive selection, namely the DNA damage sensing protein kinase *PRKDC* (dN/dS = 2 for nonsense variants; q-value = 0.01), the pre-MRNA processing factor *PRPF19*, involved in TCR, (dN/dS = 4.6 for nonsense variants; q-value = 0.005) and the

homologous recombination repair gene *SWI5* (dN/dS = 3.2 for missense and 5.3 for nonsense variants; q-value = 0.05 and 0.07, respectively).

Small changes in mutation rate may cause large cancer risk

The apparent discrepancy between the strong genetic evidence for DNA repair deficiency driving cancer and the commonly moderate effects on mutational spectra can be explained from the perspective of somatic evolution. Given that mutations occur throughout a cancer patient's lifetime (Gerstung et al., 2017; Tomasetti et al., 2013), a DNA repair deficiency acquired at a late stage during tumour development may not have enough time to substantially impact on the mutational burden or signature. In addition, it is important to consider the relationship between mutation rate and cancer risk, mindful that a small number of driver mutations are needed for cancer transformation. Classic studies by (Armitage and Doll, 1954; Nordling, 1953) postulated that there are about 6-7 rate limiting steps needed for cancer formation, based on the observation that cancer risk increases approximately as the fifth power of age, a prediction in good agreement with modern estimates of 2-10 driver gene point mutations in cancer genomes (Martincorena et al., 2017; Vogelstein and Kinzler, 2015). As the chance to independently mutate multiple driver genes in the same cell becomes a power of the mutation rate, a relatively small change in the mutation rate leads to a dramatic increase in the incidence rate of a cancer at any given age: Increasing the mutation rate by a factor of 2 leads to an approximately 2^6=64 fold increased probability of 6 co-occurring driver mutations; clonal expansions reduce the exponent, but preserve the overall exponential scaling between mutation rate and cancer risk (Tomasetti et al., 2015).

Indeed, the observed relation of relative risk and mutation rate increases in a number of cancer predisposition syndromes, but also smoking-related lung cancers follow this trend (**Figure 6B**). As noted previously, MMR-deficient colorectal cancers have an ~8-10 fold higher mutation burden than MMR-proficient carcinomas, inherited MMR deficiency increases colorectal cancer risk >115 fold. Similarly, HR-deficient breast cancers display a ~3-fold higher mutation burden, despite 20-40 fold increased risks for *BRCA1/2* mutation carriers (Antoniou et al., 2003). Even more so, XP patients display an 100-fold increase of mutation rate (Zheng et al., 2014), but have an approximately 10,000 fold increased rate of skin cancers (Bradford et al., 2011). Under this model, high-impact SNPs in NER genes,

which lead to ~2-fold increase in skin cancer risk (Wheless et al., 2012) would be expected to cause a 20% increase in mutation rate, in agreement with our observations (**Figure 6B**). While the exact number of driver gene mutations remains unknown in each cancer type an implication of theses data show that small changes in mutation rates can have a large impact on cancer risk, and conversely noticeable risk factors may derive from rather moderate mutagenic effects.

It thus emerges that also small mutagenic effects may be important for cancer development, explaining why genes with no obvious mutator phenotype may be positively selected in cancers. These findings also underscore the need to accurately determine seemingly small and subtle mutagenic effects, which would be challenging to detect in cancer genomes, but are often found in our experimental screen (**Figure 3B,C**).

Discussion

Taken together, our experimental screen and data analysis showed that mutagenesis is fundamentally driven by the counteraction of DNA damage and repair. A consequence of this interplay is that the resulting mutational spectra vary in a number of ways, ranging from a uniform increase of the genotoxic mutation profiles to changes in the mutation spectrum and even a reduction in the rate of mutations for certain combinations of DNA damage and repair. Each of these phenomena sheds light on the mode of action of DNA repair. Uniform increases across the entire mutation spectrum suggest that the repair gene is involved in repairing all the DNA lesions generated by the genotoxin. Conversely, mutation spectra change if divergent repair pathways are involved in repairing specific subsets of DNA lesions introduced by the same genotoxin. This was observed here for alkylating agents and aristolochic acid, implying that the same biochemical adducts at different bases and residues are repaired by distinct pathways, which may be a consequence of their impact on DNA helix shape and stability. In the case of DNA methylation our data corroborate the notion that the mutagenicity of O6-methylguanine stems from mis-pairing with thymine, and is repaired by methylguanine methyltransferases, while N3-methyladenine stalls replication, which is resolved by translesion synthesis polymerase Polκ. Lastly, TLS Polζ deficiency reduced the mutagenicity of UVB – and to a lesser extent also of alkylating agents – seemingly replacing a large number of single nucleotide variants with a smaller number of large deletions. This finding illustrates how functional error-prone TLS prevents the accumulation of potentially more damaging large deletions at the cost of an increased base substitution rate, which may

present the lesser of two evils for a cell having to deal with a large number of genotoxic lesions.

Our analysis of human cancers showed that, while non-silent mutations in DNA repair genes are common, bi-allelic inactivation of both copies, which is generally required for loss of function, appears to be a relatively rare phenomenon. About 10% of DNA repair genes display signals of positive selection, signifying a causal contribution to cancer progression. Surprisingly few DNA repair genes, however, were found to exhibit a measurable mutational phenotype on their own or in relation to other mutational processes. While the frequent lack of strong mutational signature appears puzzling at first sight, it seems less surprising when seen through the lense of somatic evolution. If DNA repair deficiency is acquired at a late stage there may simply not have been enough time to accumulate substantial amounts of mutations. Also relatively large increases of cancer risk – as seen for a number of inherited DNA repair deficiency syndromes – may stem from a relatively moderate increase in mutation burden. These findings agree well with the observation of our screen, where many DNA repair knockouts did not yield dramatic mutation rate increases unless further probed by genotoxic exposures. A notable exception to this trend was mismatch repair deficiency, possibly because MMR mends a large number of endogenous DNA alterations, which arise in the absence of exogenous genotoxin exposures, including T·G mismatches resulting from spontaneous deamination of 5meC and mispairing resulting from replication slippage. Similarly, we observed that APOBEC-induced C>G mutagenesis depends on both UNG-driven uracil excision and translesion synthesis by REV1, leading to a 20% reduction of C>G mutations when either repair component is deficient.

The analysis of mutational signatures has gained much attention in recent years and dramatically deepened our understanding of the range and common types of mutation patterns observed in human cancer and normal tissues. These signatures have revealed the mutagenic consequences of a number of endogenous and exogenous genotoxin exposures and signatures indicative of DNA repair deficiency conditions, which can be therapeutically exploited. Some mutational signatures are very promising biomarker candidates, such as MMR deficiency for immunotherapy or HR deficiency forPARP inhibitor treatment. However, our study suggests that finding signatures of other DNA repair deficiencies may be more challenging as the resulting mutation signal may be faint – because small mutation rate increases can cause a large cancer risk – and variable due to DNA-damage repair interactions. Seemingly, for a number of DNA repair conditions one might have to combine

the analysis of mutational profiles with genetic and molecular testing of repair genes in order to establish whether or not a DNA repair pathway is truly compromised.

Looking ahead it appears important to recognise the variable nature of mutagenesis due to the eternal struggle between DNA damage and repair. Small shifts of these antagonistic forces in favour of mutation may drive cancer progression. Still, our ability to observe the consequences of many genotoxic processes is limited by the overall high efficiency and redundancy of DNA repair processes which only leave a small fraction of damage visible as mutations. Thus, catalogues of mutational signatures reported in cancer cells may only ever represent the most striking exemplars of specific mutagenic constellations requiring further and thorough experimental dissection.

Methods

Data generation

The experimental *C. elegans* data was generated as described previously (Meier et al., 2014), with strains and mutagen doses described in **Supplementary Table 1**. Hermaphrodites were treated with DNA damaging agents at the late L4 and early adult stages to target both male and female germ cells. Resulting zygotes provide a single cell bottleneck where mutations are fixed before being clonally amplified during *C. elegans* development and passed on to the next generation in a mendelian ratio. The baseline of mutagenesis in wild-type and DNA repair defective strains was determined by independently propagating 3-5 self-fertilizing hermaphrodites per genotype typically for 20 or 40 generations as described (Meier et al., 2018). For these mutation accumulation experiments, the P0 (parental) or F1 (1st filial) generation and the final generation were sequenced.

Data acquisition

DNA from *C. elegans* samples were prepared as described previously (Meier et al., 2014). Samples were sequenced using Illumina HiSeq 2000 and X10 short read sequencing platforms at 100 bp paired-end with a mean coverage of 50x. TCGA human cancer data was taken from NCI GDC (https://gdc.cancer.gov, (Grossman et al., 2016)) and respective studies (Cancer Genome Atlas Network, 2015; Cancer Genome Atlas Research Network, 2012, 2014a, 2014b; Cancer Genome Atlas Research Network et al., 2013; Kim et al., 2015; Ng et al., 2017).

Mutation calling

C. elegans samples were run through the Sanger Cancer IT pipeline including CaVEMan for SNV calling (Nik-Zainal et al., 2012), Pindel for indel calling (Ye et al., 2009), and BRASS (Nik-Zainal et al., 2012) and DELLY (Rausch et al., 2012) calling consensus for structural variant identification. Mutation calls were filtered as described (Meier et al., 2014). The clustering and classification of SVs was constructed based on (Li et al., 2017) as described in **Supplementary Methods**. Additionally, duplicated mutations across samples were removed (**Supplementary Methods**). Mutation counts in all samples are listed in **Supplementary Table 1**.

Variants were classified into 96 SNV classes (6 base substitution types per 16 trinucleotide contexts), 2 MNV classes (dinucleotide variants and longer variants), 14 indels classes (6 classes of deletions based on deletion length and local context (repetitive region or not), similar for insertions, and two classes for short and long complex indels), and 7 SV classes (complex variants, deletions, foldbacks, interchromosomal events, inversions, tandem duplications and intrachromosomal translocations).
Data visualisation was performed using *t*-SNE (Maaten and Hinton, 2008) based on the cosine distance of the 119-dimensional mutation spectra, averaged across replicates.

Extraction of interaction effects in experimental data

For each sample $i = 1, \ldots, 2721$ and mutation class $j = 1, \ldots, 119$ we counted the respective number of mutations per sample Y^{ij}. Using matrix notation, the counts $Y \in \mathbb{N}^{2721 \times 119}$ were modeled by a negative binomial distribution,

$$Y \sim NB(\mu, \varphi = 100),$$

with scalar overdispersion parameter $\varphi = 100$ selected based on the estimates of overdispersion in the dataset, and matrix-variate expectation $\mu \in \mathbb{R}_+^{2721 \times 119}$. The basic structure of the expectation is

$$\mu = \mu_G \cdot (g + \alpha_I) + d \cdot \mu_M \cdot \beta_I,$$

where $g \cdot \mu_G \in \mathbb{R}_+^{2721 \times 119}$ describes the expected mutations due to the genotype *G* in the absence of a genotoxin after g generations, while $d \cdot \mu_M \in \mathbb{R}_+^{2721 \times 119}$ denotes the expectation for mutagen *M* at dose *d* in wild-type conditions. The terms $\alpha_I \in \mathbb{R}_+^{2721 \times 1}$ and $\beta_I \in \mathbb{R}_+^{2721 \times 119}$ model how these terms interact (more details follow below). Here the dot product "$\cdot$" is element-wise and we use the convention that a dimension of length 1 is extended by replication to match the length of the corresponding dimension of other factor in the product.
For $\alpha_I = 0$, expected number of mutations for genotype G in wildtype has the structure

$$\mu_G = g \cdot (G \times S_G),$$

where $G \in \mathbb{Z}_2^{2721 \times 54}$ is the binary indicator matrix for denoting which one of the 54 genotypes a given sample has ($G^{ij} = 1$ if sample i has genotype j, otherwise $G^{ij} = 0$). The matrix $S_G \in \mathbb{R}_+^{54 \times 119}$ denotes the mutation spectrum (i.e. signature) across the 119 mutation classes for each of the 54 genotypes in the absence of genotoxic exposures. S_G has a log-normal prior distribution $S_G \sim logN\left(0, \sigma_G^2\right)$ iid with a scalar variance σ_G^2. The vector $g \in \mathbb{R}_+^{2721 \times 1}$ denotes the generation number in every sample, $g = g(N)$, adjusted for the chances of

fixation after N generations given the 25%-50%-25% probability of each mutation becoming homozygous, remaining heterozygous, or being lost in each generation.

Similarly, the expected mutations for a given mutagen at dose $d \in \mathbb{R}_+^{2721\times 1}$ reads

$$\mu_M = d \cdot (M \times S_M) \ ,$$

with $M \in \mathbb{Z}_2^{2721\times 12}$ being the indicator for the 12 mutagens used in this screen and $S_M \in \mathbb{R}_+^{12\times 119}$ being their signatures in wildtype, with per-column prior $S_M^{(j)} \sim logN\left(0, \sigma_{Mj}^2\right)$, where the variances $\sigma_M^2 = \left(\sigma_{M1}^2, \ldots, \sigma_{M12}^2\right)$ have prior distributions $\sigma_{Mj}^2 \sim \Gamma(1,1)$ iid.

Thus, the matrix-variate expected number of mutations can be written as

$$\mu = (g + \alpha_I) \cdot (G \times S_G) + d \cdot \beta_I \cdot (M \times S_M) \ .$$

The vector-variate interaction term α_I measures how the mutations expected for genotype G may uniformly increase under mutagen exposure and is taken to be linear with respect to the mutagen dose,

$$\alpha_I = d \cdot (I \times b_I) + \varepsilon \, ,$$

The symbol $I \in \mathbb{Z}_2^{2721\times 196}$ is the binary indicator matrix for each one of the 196 interactions tested in the screen multiplied by the interaction rate $b_I \in \mathbb{R}_+^{196\times 1}$, which measures how much the mutation signature of the genetic background increases upon genotoxic exposure with dose d. The interaction term is modelled to have a prior log-normal distribution $b_I \sim logN(0, 0.5)$ iid.

Lastly ε is a random offset to allow for a possible divergence of the genotypes between experiments, adding $\varepsilon\mu_G$ mutations with the same spectrum as in the absence of a mutagen in a dosage-independent fashion. The random offset is modelled as $\varepsilon = (J \times a_J)$ where $J \in \mathbb{Z}_2^{2721\times 208}$ is an extended binary indicator matrix for each one the 196 interactions and 12 wild-type exposure experiments, with value $a_I \in \mathbb{R}_+^{208\times 1}$, $a_I \sim logN(0, 0.5)$ iid.

In addition to this scalar interaction, the wild-type spectrum of the mutagen S_M may change by the factors $\beta_I \in \mathbb{R}_+^{2721\times 119}$ measuring the fold-change of each mutation type which is expressed as

$$\beta_I = I \times S_I \, ,$$

where $S_I \in \mathbb{R}_+^{196\times 119}$ is the fold-change for each mutation class in a given interaction, with prior distribution per column $logS_I^{(j)} \sim Laplace\left(0, \sigma_{Ij}^2\right)$, where the variances $\sigma_I^2 = \left(\sigma_{I1}^2, \ldots, \sigma_{I196}^2\right)$ have prior distribution $\sigma_{Ij}^2 \sim \Gamma(1,1)$ iid.

Bayesian estimates of the parameters $S_G,\ S_M,\ S_I,\ a_J,\ b_I$ and hyperparameters σ_M^2, σ_I^2 were calculated via Hamiltonian Monte Carlo using the Greta R package (Golding 2018), where first the posterior distribution for S_G^* is estimated on mutation accumulation samples only; all other estimates are subsequently calculated conditional on this distribution (**Supplementary Methods**). Mean estimates along with their credible intervals may be found in **Supplementary Tables 2-3**.

TCGA/PCAWG human cancer data analysis

The susceptibility of DNA pathway to alteration was defined as having altered expression or high impact mutations in relevant genes (**Supplementary Methods, Supplementary Table 4**). The interaction effect was then estimated using Hamiltonian Monte Carlo sampling (Neal and Others, 2011) for the following model for matrix of mutation counts in *N* samples with *R* mutation classes $Y \in Mat_{NxR}(\mathbb{N} \cup \{0\})$:

$$Y \sim NB(\mu, \phi),$$

$$\mu = S^{(-K)} \cdot \alpha_{(-K)} + S^{(K)} \cdot \alpha_{(K)} \cdot f_G,$$

$$f_G = e^{X\beta},$$

$$\alpha \sim Unif(0, M),$$

$$\beta \sim N(0, 0.5),$$

where α is the matrix of exposures (number of mutations assigned to each signature), S is a matrix of *K* signatures (each column is a signature, and is normalized to sum to 1), and f_G is the interaction factor which alters the signature. It consists of *X* - a binary matrix of labels, and β which is a matrix of spectra of the interaction effects. *M* is the maximal number of mutations per sample in the dataset. The overdispersion parameter $\phi = 50$ was chosen based on variability estimates across all cancer samples.

Analysis of selection in DNA repair genes

dN/dS analysis of the ratios of numbers of expected vs observed non-synonymous and synonymous variants were performed using the trinucleotide model according to (Martincorena et al., 2017). Background mutation rates were estimated using all the genes except for consistently undercovered genes (Wang et al., 2018). For single-gene analysis, all samples with less than 1000 coding mutations per exome were selected. For pathway-level analysis, all samples with less than 1000 coding mutations per exome from the selected cancer type were considered. The dN/dS ratio in the gene (for single-gene analysis) or set of genes (for pathway-level analysis) of interest was compared to 1 (the dN/dS ratio under the

neutral evolution model), and the significance of selection was assessed by comparing the $\chi 2$ statistic of the gene's dN/dS ratio, $\frac{log(dN/dSgene)^2}{\sigma(log(dN/dSgene))^2}$, to $\chi 2$ distribution with $df = 1$, with FDR control performed by Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

Acknowledgements

This work was supported by the Wellcome Trust COMSIG consortium grant RG70175, a Wellcome Trust Senior Research award (090944/Z/09/Z), a Worldwide Cancer Research grant (18-0644), and by the Korean Institute for Basic Science (IBS-R022-A1-2019) to AG. We thank the Mitani Lab funded by the National Bio-Resource Project of the MEXT, Japan, and the Caenorhabditis Genetics Center funded by the NIH Office of Research Infrastructure Programs (P40 OD010440) for providing strains. Nadezda Volkova is a member of Lucy Cavendish College, University of Cambridge. We thank Nuria Lopez-Bigas for critical comments on the manuscript.

Data availability

Sequencing data are available under ENA Study Accession Numbers ERP000975 and ERP004086. Code for downstream analyses and figures can be downloaded from github.com/gerstung-lab/signature-interactions.

Author contributions

NV, BM, AG and MG wrote the manuscript, which was approved by all authors. NV analysed all data with input from BM. BM performed mutation accumulation and mutagenesis experiments. VGH and SB performed mutagenesis experiments. SG analysed somatic mutation data. FA and IM contributed to selection analysis. PC and AG conceived the study. AG and MG supervised the analysis.

Figures

Figure 1

Experimental design of the study

A. Left: Wild-type and *C. elegans* mutants of selected genes from 10 different DNA repair pathways, including mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), double-strand break repair (DSBR), DNA crosslink repair (CLR), translesion synthesis (TLS), telomere replication (TR), and direct damage reversal by O-6-Methylguanine-DNA Methyltransferase (MGMT), were propagated over several generations or exposed to increasing doses of different classes of genotoxins. Genomic DNA was extracted from samples before and after accumulation/ without and with treatment and whole genome sequenced to determine mutational spectra. Right: representative experimental mutational profile plots across 119 mutation classes: 96 single base substitutions classified by change and '5 and 3' base context, di- and multi-nucleotide variants (MNV), 6 types of deletions of different length and context, 2 types of complex indels, 6 types of insertions and 7 classes of structural variants.
B. Number of observed mutations per replicate (grey dots) and experiment (average across replicate, black dot) for base substitutions, indels and structural variants across different experimental types: mutation accumulation in wild-type and DNA repair mutants, genotoxin treatment of wild-type *C. elegans*, and interaction experiments investigating the effect of genotoxin treatment in different DNA repair deficient backgrounds. Horizontal lines indicate the medians across experiments.

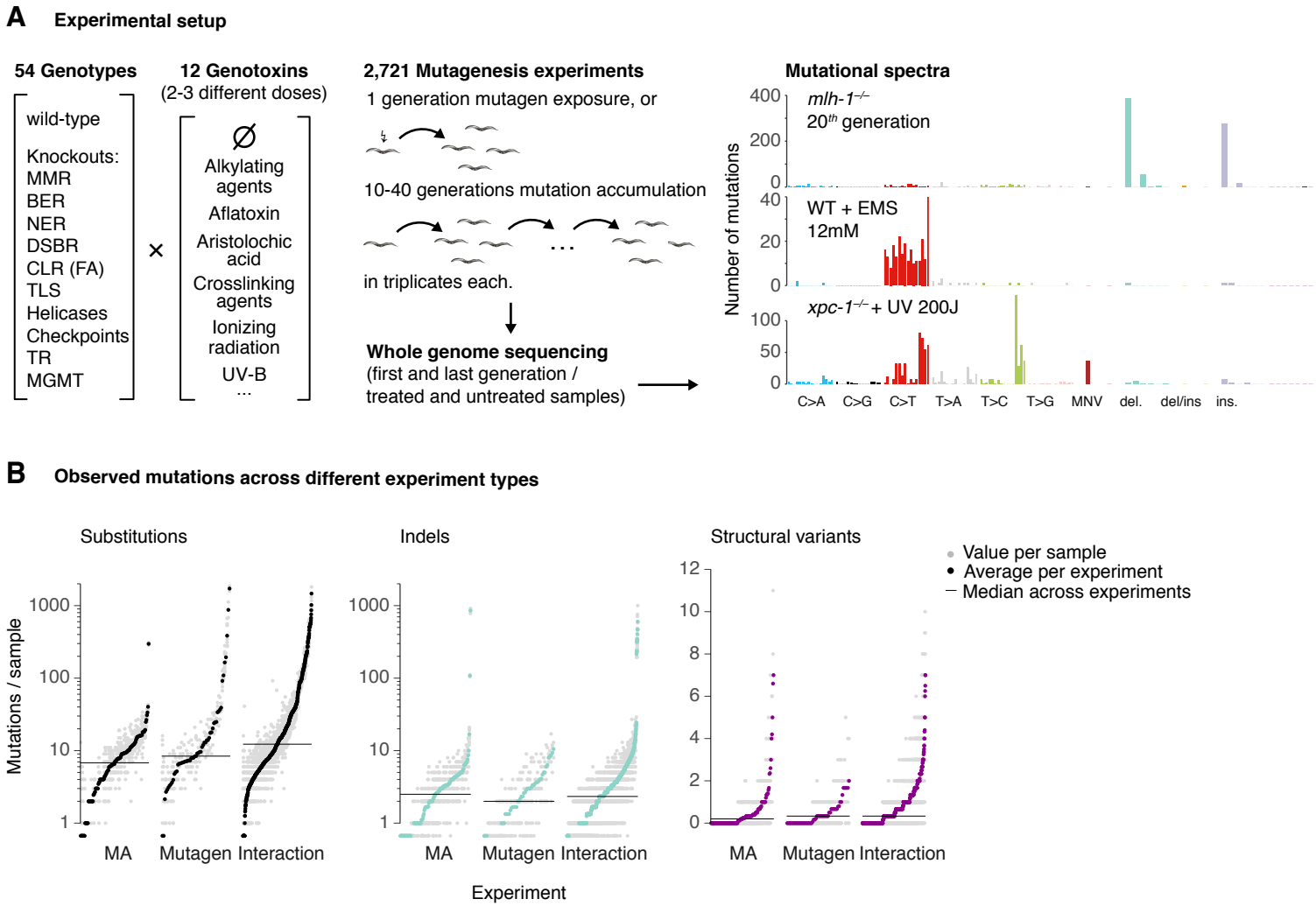


Figure 1. Experimental design of the study

Figure 2

Overview of mutational signatures of DNA repair deficiency and genotoxin exposures.

A. 2D t-SNE representation of all *C. elegans* samples based on the cosine similarity between mutational profiles. Each dot corresponds to one sequenced sample, the size of the dot is proportional to the number of observed mutations. Colors depict the genotoxin exposures, no color represents mutation accumulation samples. Highlighted are selected genotoxins and DNA repair deficiencies. EMS, ethyl methanesulfonate; MMS, methyl methanesulfonate; DMS, dimethyl sulfate.
B. Experimental mutational signatures of selected DNA repair deficiencies in the absence of genotoxic exposures.
C. Experimental mutational signatures of selected genotoxin exposures in wild-type *C. elegans*.

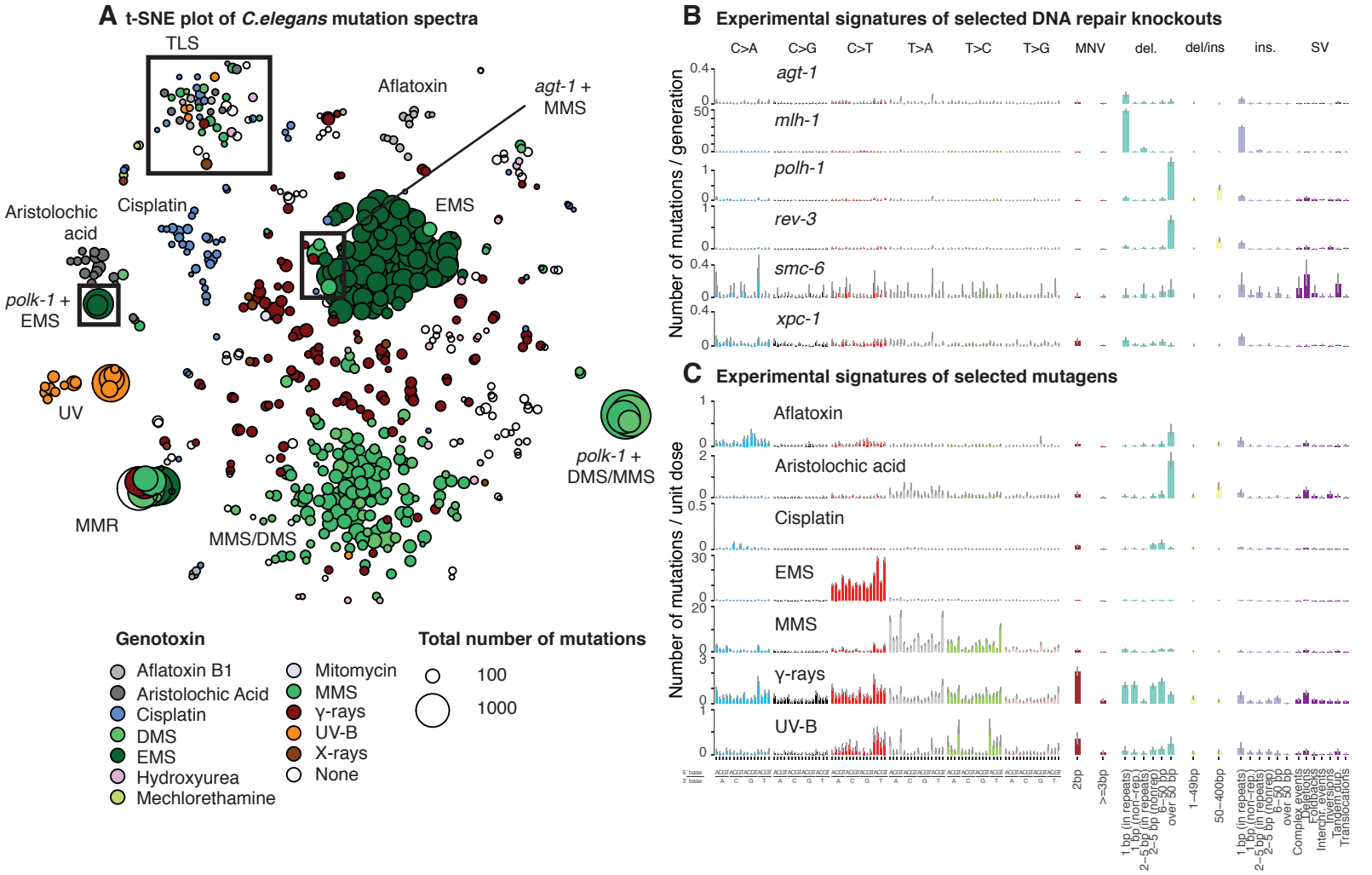


Figure 2. Overview of mutation spectra and signatures of DNA repair deficiency and genotoxic exposures

Figure 3

Examples of experimental genotoxin-repair interaction signatures.

A. Signatures of MMS exposure in wild-type and *polk-1* deficient *C. elegans*. Top: Mutation spectra of MMS exposure in wild-type and *polk-1*$^{-/-}$ background. Error bars denote 95% confidence intervals. Bottom: Signature fold-change per mutation type between the wild-type mutagen and interaction spectra. Black lines denote the maximum a posteriori point estimates; faded color boxes depict 95% CIs. Point estimates and CIs are highlighted if different from 1. Right: Dose response for MMS in wild-type (upper part) and *polk-1* deficient (lower part) samples, measuring the total number of mutations (*x*-axis) versus dose (*y*-axis) for each replicate. *polk-1* deficiency leads to a 10-100x increase of T>A transversions under MMS treatment. Mutation types: MNV refer to multinucleotide variants, del/ins. to deletions with inserted sequences, SV to structural variants.
B. Signatures of MMS exposure in wild-type and *agt-1* deficient *C. elegans*. Same layout as in A. MMS treatment induced substantially more C>T transitions in an *agt-1* deficient background.
C. Signatures of UVB exposure in wild-type and TLS polymerase ζ deficient (*rev-3*) deficient *C. elegans*. Same layout as in A. *rev-3* knockout leads to a 3-fold reduction in the number of substitutions and a 10-fold increase in the number of indels compared to the values expected without interaction.
D. Signatures of UVB exposure in wild-type and *xpc-1* NER deficient *C. elegans*. The knockout of *xpc-1* leads to 40-50 fold increase in C>T and T<C mutations compared to the values expected without interaction.

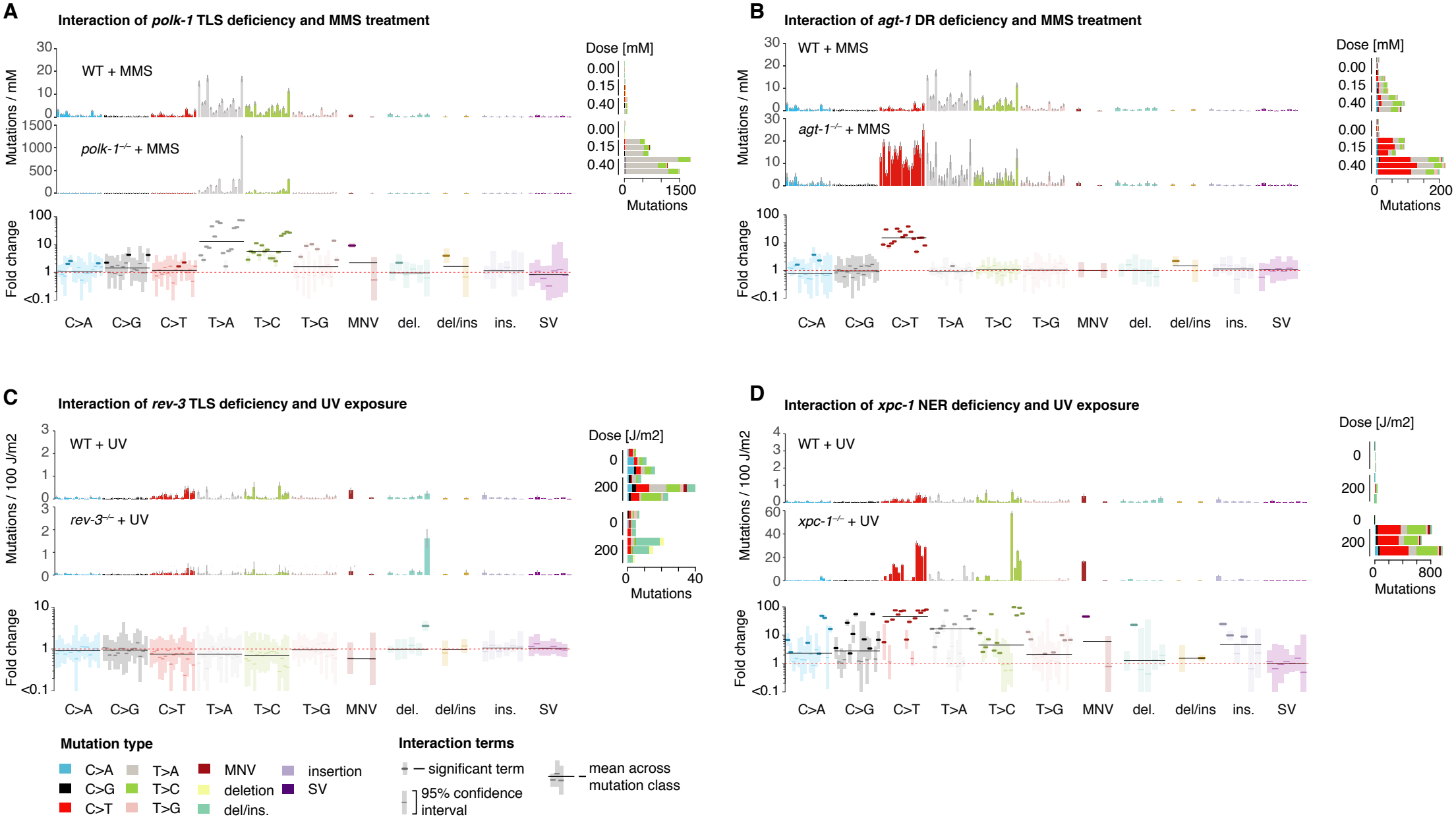


Figure 3. Examples of repair-genotoxin interactions

Figure 4

Widespread and diverse genotoxin-repair interactions in *C. elegans*

A. Estimated fold changes between observed numbers of base substitutions relative to expected numbers based on additive genotoxin and DNA repair deficiency effects. A value of 1 indicates no change. Interactions with fold change significantly over or below 1 are shown in dark grey, the rest as light grey. Overall, about 37% of interactions produced a significant fold change. AA, aristolochic acid.
B. Mutation signatures changes in interaction experiment. Black lines denote point estimates. Interactions with cosine distances larger than 0.2 are shown with dark blue confidence intervals, others are shown in light blue. Around 10% of interactions produced a significant change in mutational spectra.
C. Numbers of mutations attributed to different causes. Overall, about 17% of mutations in the dataset are attributed to genotoxin-repair interactions, including a reduction of −2% of mutations which would be expected under wild-type repair conditions.

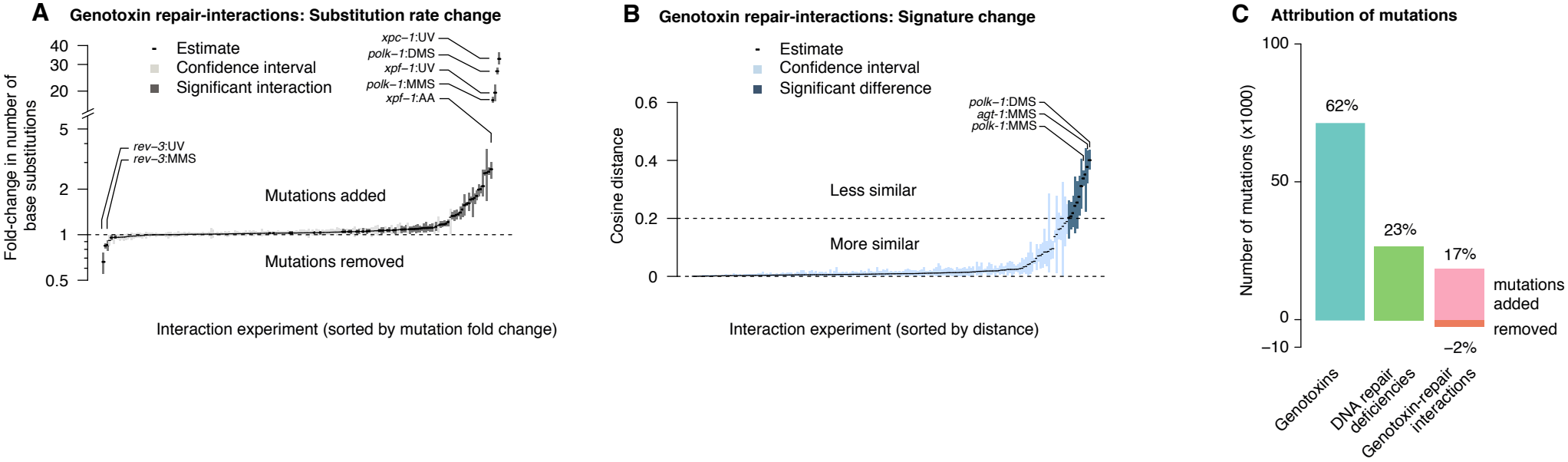


Figure 4. Widespread and diverse genotoxin-repair interactions in *C. elegans*

Figure 5

DNA damage-repair interactions in human cancers

A. Summary of interaction effects of selected DNA repair pathway deficiencies with DNA damage-associated mutational signatures. Top: Changes in age-adjusted mutation burden. Bottom: Change in mutational signatures.
B. Interaction of temozolomide treatment in glioblastoma samples with MGMT. Mutational signatures of expressed (top) and silenced *MGMT* (middle). Bottom: Mutation rate fold change per mutation type. Right: Total mutation burden in temozolomide-treated samples with silenced and expressed *MGMT*. DR, direct damage reversal.
C. Interactions of DNA damage caused by polymerase epsilon (POLE) exonuclease deficiency and mismatch repair. Mutational signatures of MMR deficiency, POLE deficiency, and combined POLE and MMR deficiency. Microsatellite instability is used as a read-out for MMR deficiency. Right: Mutation burden of POLE exonuclease deficiency in MMR proficient (n=40) and deficient (n = 15) uterine cancers. MSI, microsatellite instable; MSS, microsatellite stable)
D. Interactions of APOBEC-induced DNA damage with REV1 and UNG BER. APOBEC mutation signatures in *REV1* and *UNG* wildtype (top) or deficient (middle) samples. Bottom: Fold-change of variant types. Right: Observed difference between C>G and C>T mutations in a TCN context.

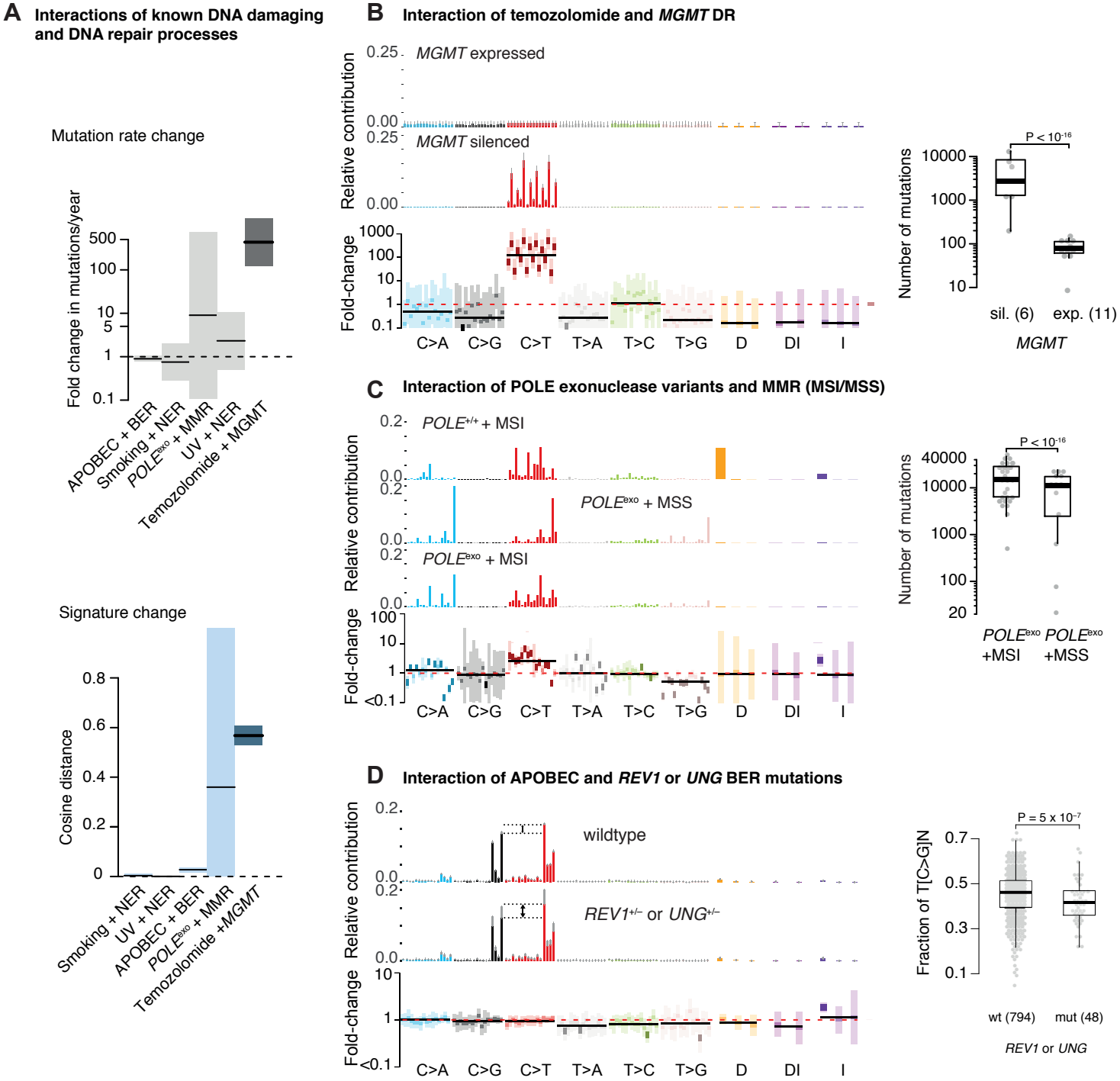


Figure 5. DNA damage-repair interactions in human cancers

Figure 6

DNA repair deficiency drives cancer evolution

A. dN/dS values for 248 DNA repair-associated genes across pan-cancer types. dN/dS measures the excess of observed non-silent mutations relative to the expected number based on observed silent mutations. A value of 1 denotes neutrality, while a value greater than 1 signifies a surplus of non-silent changes, indicating a cancer-promoting effect. dN/dS values per gene were calculated separately for missense and nonsense mutations. dN/dS ratios were also calculated for 9 DNA repair pathways (aggregating all genes) across 30 cancer types (right). Black/colored dots reflect results which reached significance compared to the background (FDR 10% within the respective group) with respective genes and DNA repair pathways indicated on the plot.
B. Relationship between relative cancer risk and point mutation rate change for different DNA repair pathway gene mutations. Black lines depict the expected power law depency in an evolutionary model with different number of driver gene mutations (Tomasetti et al., 2015).

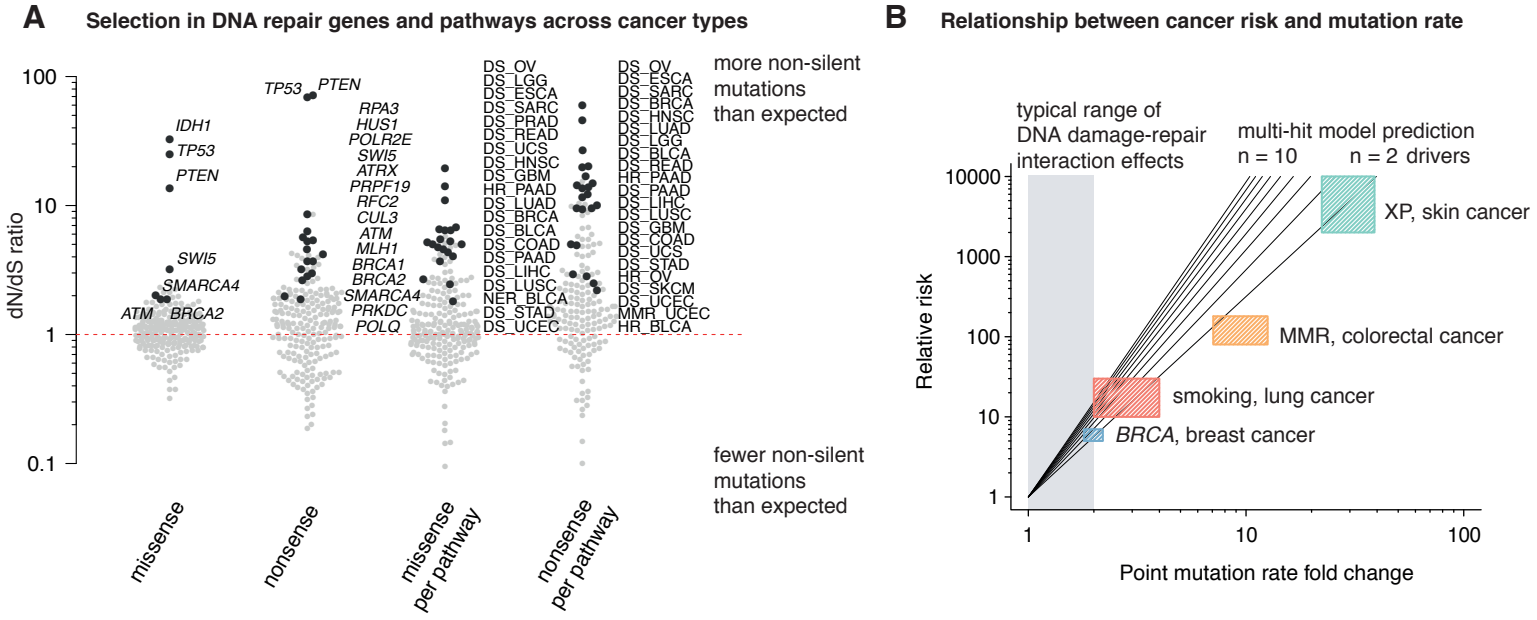


Figure 6. Mutations in DNA repair genes drive cancer evolution

Supplementary Figures

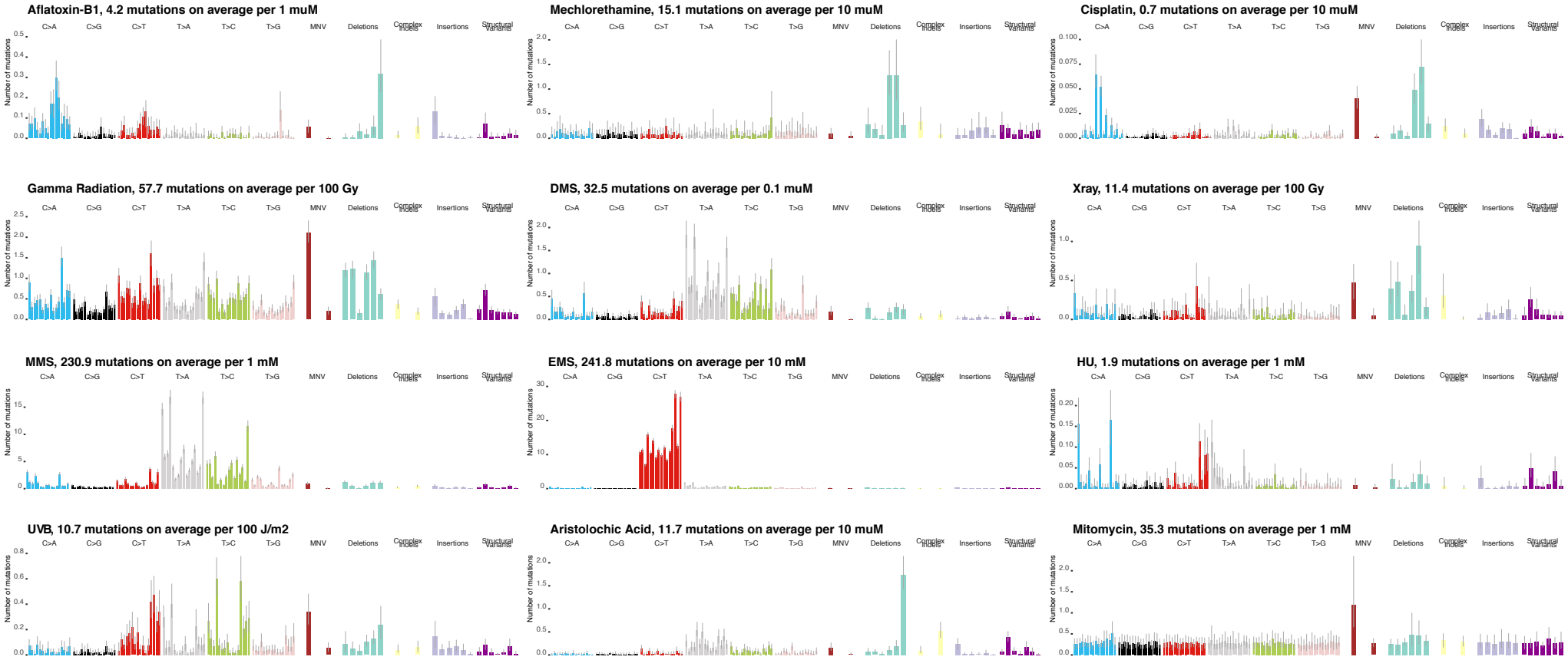
Supplementary Figure 1

Experimental mutational signatures extracted for 54 genetic backgrounds. Each barplot reflects the average number of mutations per generation, the headers contain the estimate for the total number of mutations per generation, and the maximal generation number with respective mutation burden across the samples used in the model. The information on gene knockouts is given in **Supplementary Table 1**.

Supplementary Figure 1. Experimental signatures of DNA repair deficiencies in *C.elegans*

Supplementary Figure 2

Mutational signatures extracted for 12 genotoxins in wild-type *C. elegans*. Each barplot reflects the number of mutations per average dose of indicated mutagen, headers contain the estimate for the total number of mutations per average dose. For details on doses and mutagen specification see **Supplementary Table 1**.



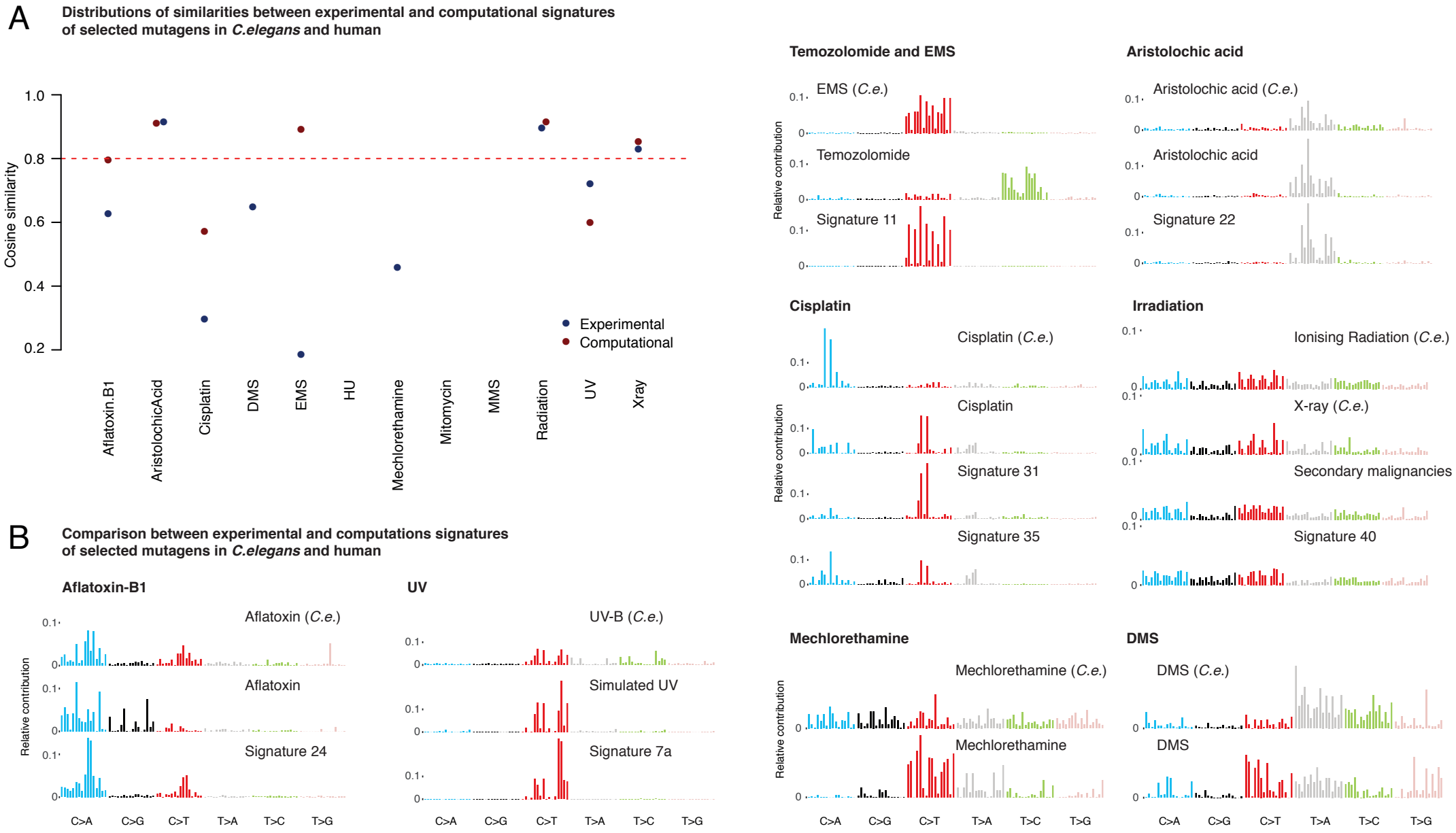
Supplementary Figure 2. Experimental signatures of mutagens in *C.elegans*

Supplementary Figure 3

Comparison between experimental signatures of 12 mutagens extracted from *C.elegans* experiments to the catalogue of mutational signatures of environmental carcinogens from (Kucab et al., 2019), and the catalogue of mutational signatures in cancer (Alexandrov et al., 2018). *C. elegans* signatures were adjusted to the human genome trinucleotide frequency.

A. Cosine similarities between the humanized *C. elegans* signatures and experimental (blue) and computational (red) signatures of same or related mutagens in human cells or cancers. "Experimental" signature for ionizing radiation was estimated as an averaged spectrum of 12 radiotherapy-associated secondary malignancies from (Behjati et al., 2016). No counterparts from human data were found for HU, Mitomycin C and MMS. Red line denotes the cut-off of 0.8 beyond which the spectra are considered similar.

B. Visual comparison of experimental signatures for selected mutagens. Signature of aflatoxin was similar to the computationally extracted signature 24 (similarity 0.8) but different from experimental one (0.62). *C. elegans* UVB exposure showed a C>T mutation spectrum somewhat similar to that in cell line experiments and in cancer, albeit with an additional fraction of T>C mutations. *C. elegans* EMS signature is very different from temozolomide signature in cell lines, but very similar (0.90) to cancer-derived computational signature SBS11 associated with temozolomide. Interestingly, similar phenomena were observed upon exposure with alkylating agents in *Salmonella typhimurium* (Matsumura et al., 2018). Aristolochic acid signatures were consistent between both experimental systems and cancer. Signature of cisplatin was different from the one identified in cancer or human cell lines. Ionizing radiation and X-rays yielded profiles similar to those in human cancers. Exposures to mechlorethamine and DMS exhibited different mutational spectra in the two model systems.

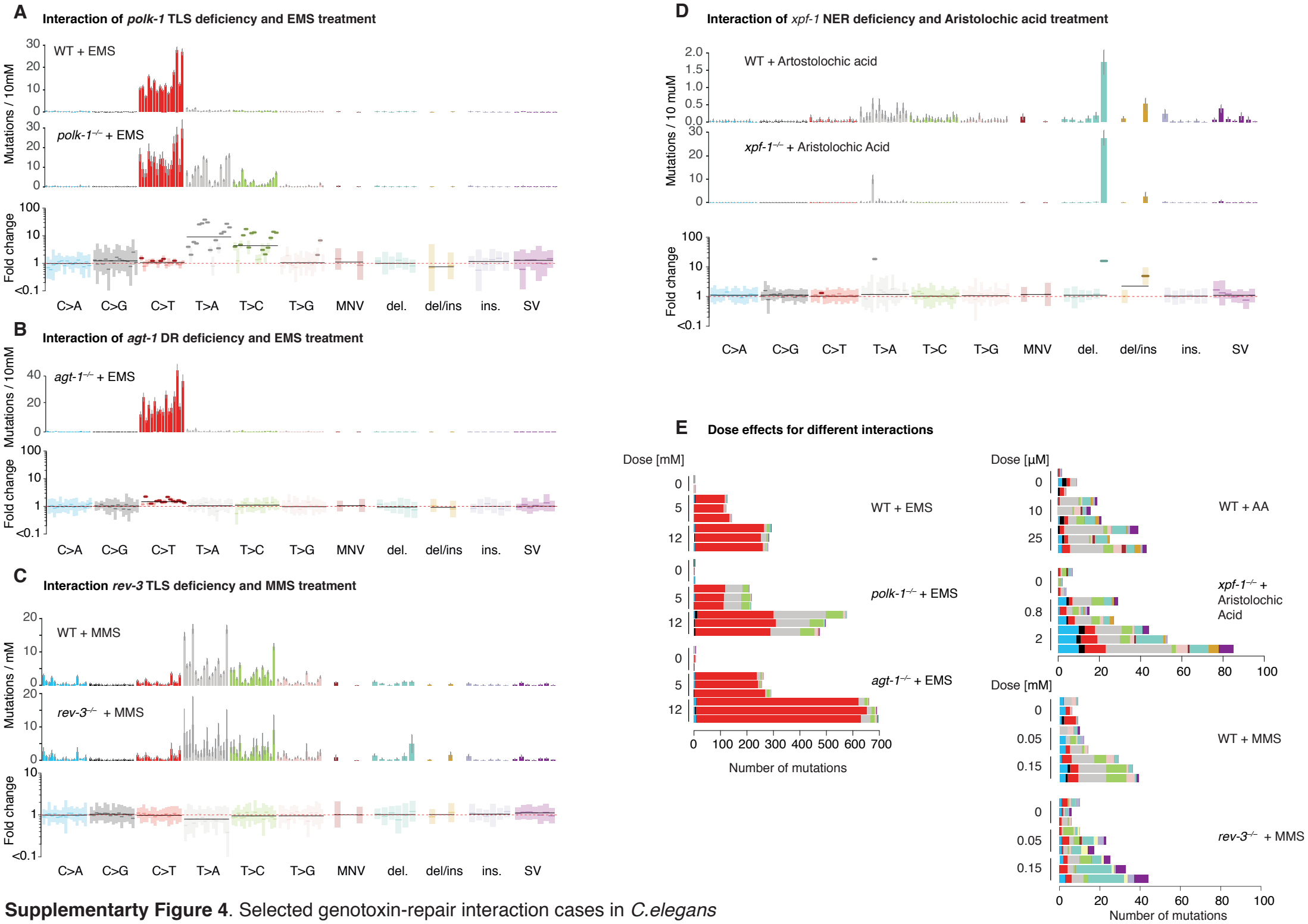


Supplementary Figure 3. Comparison between genotoxin signatures in *C. elegans* and in human cells

Supplementary Figure 4

Selected genotoxin-repair interaction cases.

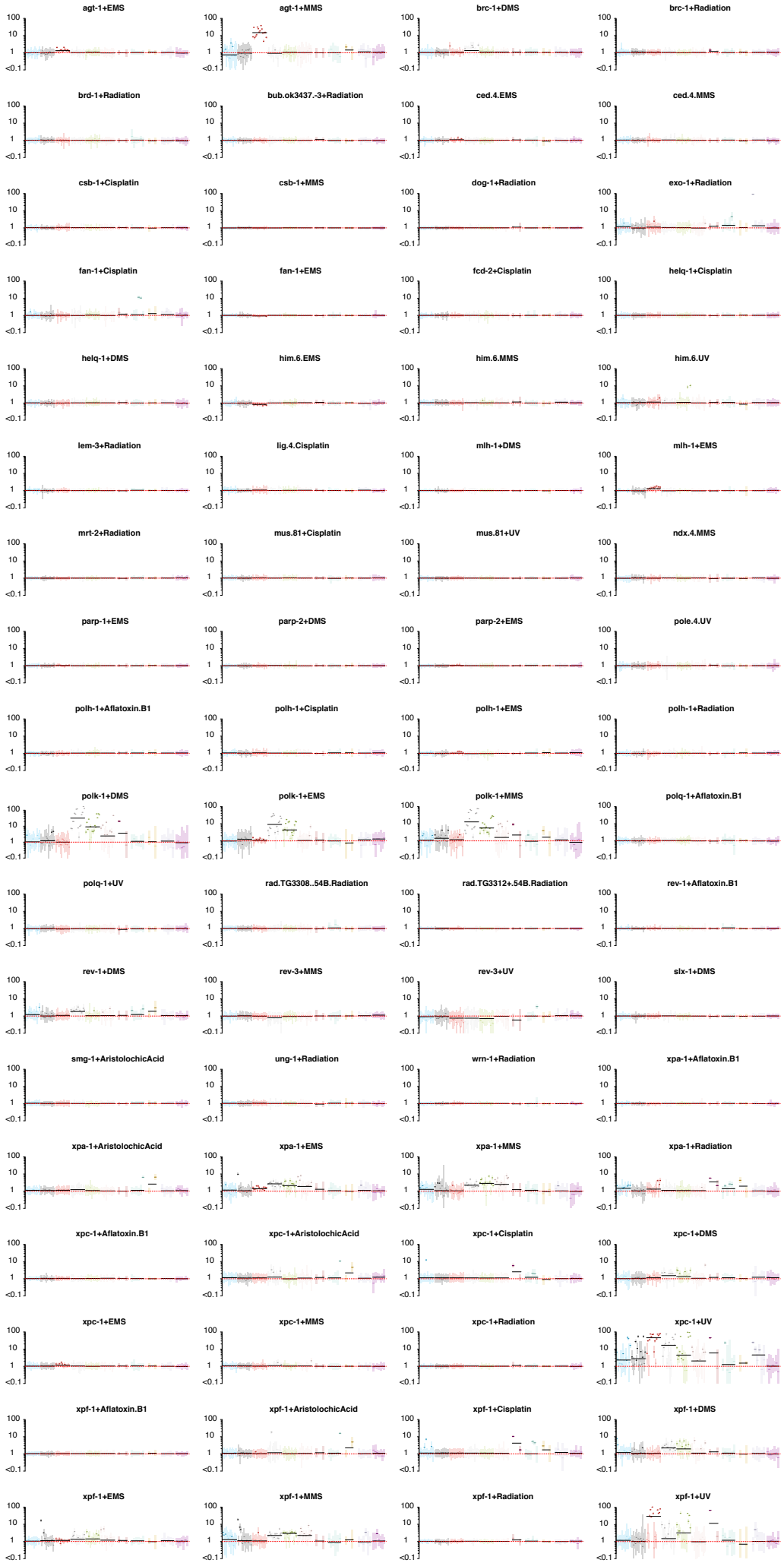
A. Mutational signature and fold-change per mutation type for EMS exposure under *polk-1* deficiency.
B. Mutational signature and fold-change per mutation type for EMS exposure under *agt-1* deficiency.
C. Mutational signature and fold-change per mutation type for MMS exposure under *rev-3* deficiency.
D. Mutational signature and fold-change per mutation type for aristolochic acid exposure under *xpf-1* deficiency.
E. Dose responses for experiments described in **A-D**. The y-axis reflects the dose while the x-axis reflects the number of mutations per sample.



Supplementarty Figure 4. Selected genotoxin-repair interaction cases in *C.elegans*

Supplementary Figure 5

Estimated interaction effects for 72 combinations of genetic backgrounds and genotoxins which showed a significant effect either on the total burden or the mutational profile. Each barplot reflects the fold-change in the number of mutations of different types per average dose or mutagen. The information on the genetic background, doses and mutagen specification is given in **Supplementary Table 1**.

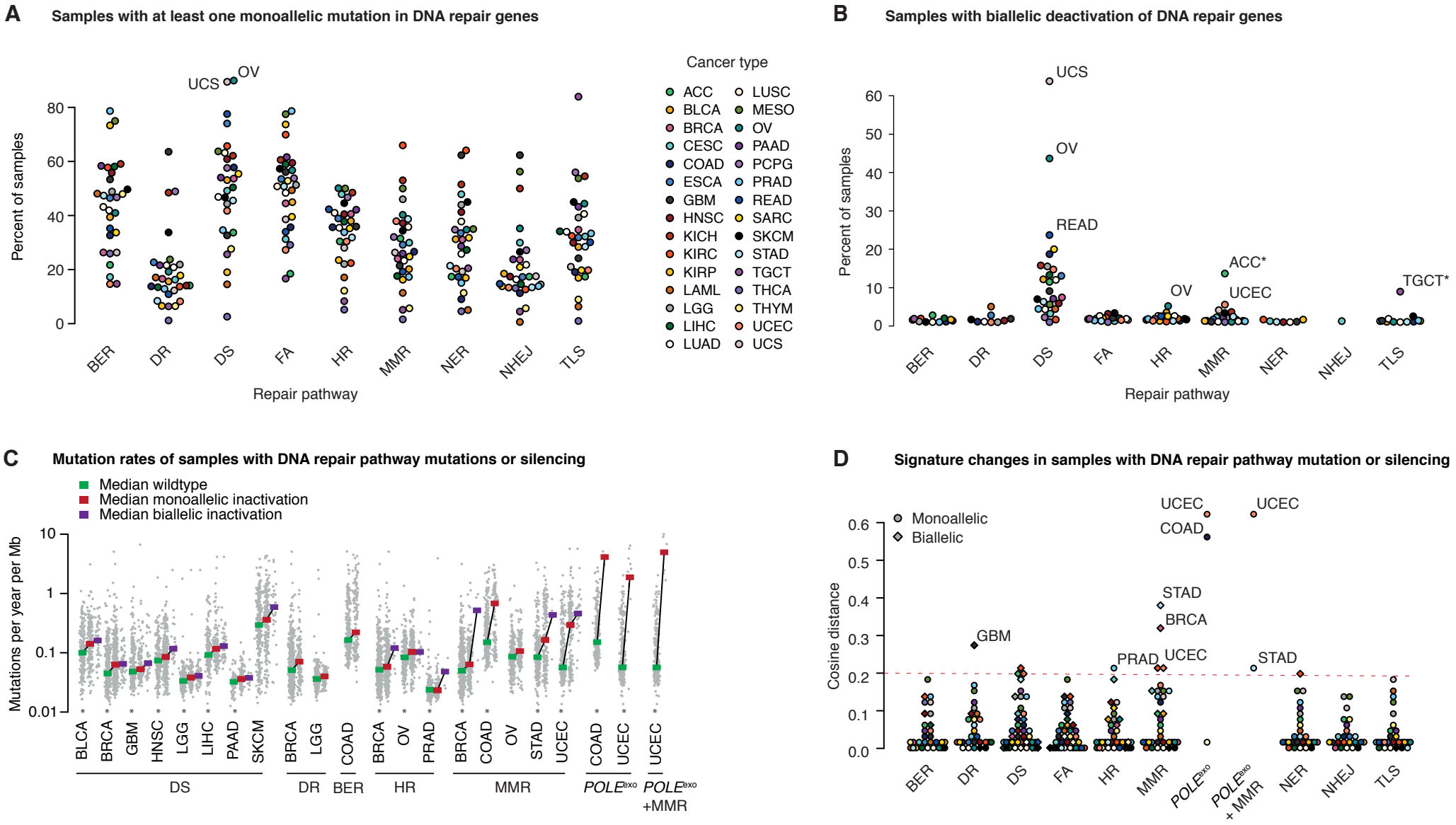


Supplementary Figure 5. Selected genotoxin-repair interactions in *C.elegans*

Supplementary Figure 6

Somatic mutations and silencing of DNA repair pathway genes in human cancers.

A. Percentages of samples with mono- or biallelic defects in core components of the DNA repair pathway genes by cancer type for 9,948 cancers from TCGA.
B. Percentages of samples with biallelic inactivation of core components of the respective DNA repair pathway by cancer type.
C. Age-adjusted mutation rates in samples without or with mono- or biallelic DNA repair pathway deficiency. Each dot represents the number of mutations per year per Megabase, colored bars denote the medians in each group. Only combinations with Wilcoxon rank sum test FDR under 5% are shown. Stars denote the cancer types where the presence of damaging mutations in the pathway is unlikely to be explained just by the increase in mutational burden (see **Supplementary Methods**).
D. Median mutational signature change between samples with DNA repair pathway with respect to median wild-type signature.

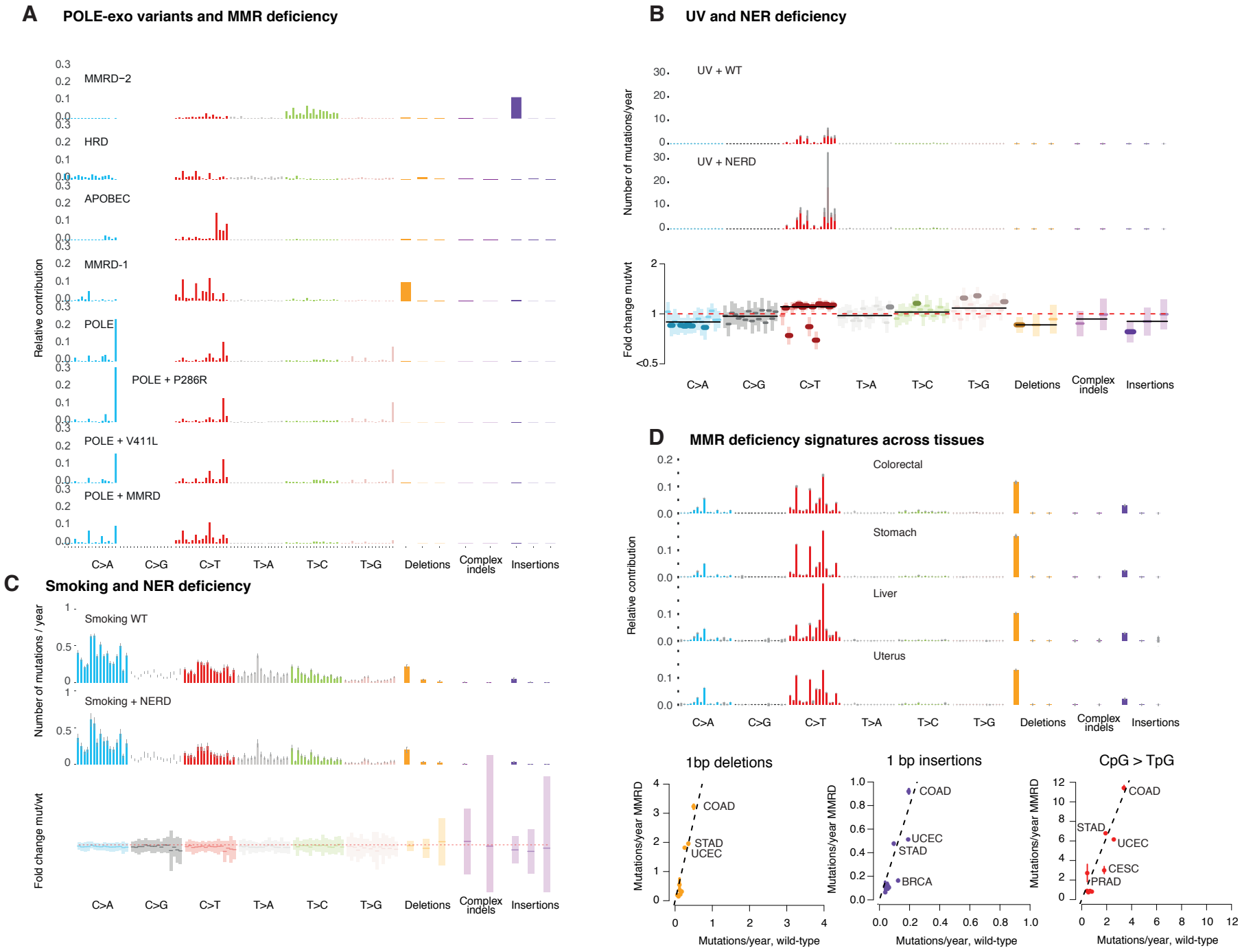


Supplementary Figure 6. Somatic mutations and silencing of DNA repair pathway genes in human cancers

Supplementary Figure 7

Selected DNA damage-repair deficiency interactions in human cancer.

A. 5 mutational signatures extracted from Uterine Cancers, and the POLE exonuclease variant signature with effects of different POLE hotspot mutations (P286R and V411L) and interaction with MMR deficiency (POLE + MMRD).
B. Mutational signature of UV light in melanoma samples with functioning (top panel) and mutated NER (lower panel). The barplot underneath shows the fold-change per mutation type.
C. Mutational signature of tobacco smoking in lung cancer samples with functioning (top panel) and mutated NER (lower panel). The barplot underneath shows the fold-change per mutation type.
D. Changes of MMR deficiency signatures across different tissues. Upper panel represents the appearance of the most widespread MMR signature in different tissues. Lower panel represents the scatterplot of per-year rates of CpG>TpG mutations, 1-bp insertions, and 1-bp deletions in MMR deficient versus MMR proficient samples across tissues. Dotted lines represent the linear slope fitted to the rates of MMR deficient versus proficient samples.



Supplementary Figure 7. Selected damage-repair deficiency interactions in human cancer

Supplementary Tables

Supplementary Table 1

Description of genetic backgrounds used in the study, sample annotations, mutation counts and design matrix for all *C. elegans* samples.

Supplementary Table 2

Experimental mutational signatures for genetic (including wild-type and DNA repair knockouts) and mutagenic (12 genotoxins used in the study) factors in *C. elegans* along with their 95% credible intervals.

Supplementary Table 3

Genotoxin-repair interaction effects: log-fold changes of mutagen signature per mutation type, and dose-dependent genotype amplification factors along with their 95% credible intervals.

Supplementary Table 4

Map of DNA repair genes with pathways, TCGA cancer names abbreviations, and TCGA samples with mutations in DNA pathways.

Supplementary Table 5

dN/dS values for DNA repair related genes across all cancers types and for DNA repair pathways in different cancer types.

References

Alexandrov, L., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E., Lopez-Bigas, N., et al. (2018). The Repertoire of Mutational Signatures in Human Cancer. bioRxiv.

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

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

Antoniou, A., Pharoah, P.D.P., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am. J. Hum. Genet. *72*, 1117–1130.

Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. Br. J. Cancer *8*, 1–12.

Arvanitis, M., Li, D.-D., Lee, K., and Mylonakis, E. (2013). Apoptosis in C. elegans: lessons for cancer and immunity. Frontiers in Cellular and Infection Microbiology *3*.

Behjati, S., Gundem, G., Wedge, D.C., Roberts, N.D., Tarpey, P.S., Cooke, S.L., Van Loo, P., Alexandrov, L.B., Ramakrishna, M., Davies, H., et al. (2016). Mutational signatures of ionizing radiation in second malignancies. Nat. Commun. *7*, 12605.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) *57*, 289–300.

Bertolini, S., Wang, B., Meier, B., Hong, Y., and Gartner, A. (2017). Caenorhabditis elegans BUB-3 and SAN-1/MAD3 Spindle Assembly Checkpoint Components Are Required for Genome Stability in Response to Treatment with Ionizing Radiation. G3 *7*, 3875–3885.

Boot, A., Huang, M.N., Ng, A.W.T., Ho, S.-C., Lim, J.Q., Kawakami, Y., Chayama, K., Teh, B.T., Nakagawa, H., and Rozen, S.G. (2018). In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome Res. *28*, 654–665.

Boulton, S.J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D.E., and Vidal, M. (2002). Combined functional genomic maps of the C. elegans DNA damage response. Science *295*, 127–131.

Bradford, P.T., Goldstein, A.M., Tamura, D., Khan, S.G., Ueda, T., Boyle, J., Oh, K.-S., Imoto, K., Inui, H., Moriwaki, S.-I., et al. (2011). Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. J. Med. Genet. *48*, 168–176.

Cancer Genome Atlas Network (2015). Genomic Classification of Cutaneous Melanoma. Cell

161, 1681–1696.

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Cancer Genome Atlas Research Network (2014a). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

Cancer Genome Atlas Research Network (2014b). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507*, 315–322.

Cancer Genome Atlas Research Network, Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., et al. (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73.

Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S.C., Gonzalez, S., Mitchell, T.J., Rubanova, Y., Anur, P., Rosebrock, D., Yu, K., et al. (2017). The evolutionary history of 2,658 cancers. bioRxiv.

Golding, N. (2018). greta: Simple and Scalable Statistical Modelling in R. R-package v. 0.2.3. http://CRAN.R-project.org/package=greta.

Greenman, C., Wooster, R., Futreal, P.A., Stratton, M.R., and Easton, D.F. (2006). Statistical analysis of pathogenicity of somatic mutations in cancer. Genetics *173*, 2187–2198.

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., and Staudt, L.M. (2016). Toward a Shared Vision for Cancer Genomic Data. New England Journal of Medicine *375*, 1109–1112.

Haradhvala, N.J., Kim, J., Maruvka, Y.E., Polak, P., Rosebrock, D., Livitz, D., Hess, J.M., Leshchiner, I., Kamburov, A., Mouw, K.W., et al. (2018). Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. Nat. Commun. *9*, 1746.

Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. Nat. Rev. Genet. *15*, 585–598.

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

Kim, H., Zheng, S., Amini, S.S., Virk, S.M., Mikkelsen, T., Brat, D.J., Grimsby, J., Sougnez, C., Muller, F., Hu, J., et al. (2015). Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. *25*, 316–327.

Kim, J., Mouw, K.W., Polak, P., Braunstein, L.Z., Kamburov, A., Kwiatkowski, D.J., Rosenberg, J.E., Van Allen, E.M., D'Andrea, A., and Getz, G. (2016). Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat. Genet. *48*, 600–606.

Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S., et al. (2018). Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. *23*,

239–254.e6.

Kucab, J.E., Zou, X., Morganella, S., Joel, M., Nanda, A.S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S.P., et al. (2019). A Compendium of Mutational Signatures of Environmental Agents. Cell.

Lans, H., and Vermeulen, W. (2011). Nucleotide Excision Repair inCaenorhabditis elegans. Molecular Biology International *2011*, 1–12.

Li, Y., Roberts, N., Weischenfeldt, J., Wala, J.A., and Shapira, O. (2017). Patterns of structural variation in human cancer. bioRxiv.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell *171*, 1029–1041.e21.

Matsumura, S., Fujita, Y., Yamane, M., Morita, O., and Honda, H. (2018). A genome-wide mutation analysis method enabling high-throughput identification of chemical mutagen signatures. Scientific Reports *8*.

Meier, B., Cooke, S.L., Weiss, J., Bailly, A.P., Alexandrov, L.B., Marshall, J., Raine, K., Maddison, M., Anderson, E., Stratton, M.R., et al. (2014). C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. Genome Res. *24*, 1624–1636.

Meier, B., Volkova, N.V., Hong, Y., Schofield, P., Campbell, P.J., Gerstung, M., and Gartner, A. (2018). Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers. Genome Res. *28*, 666–675.

Morganella, S., Alexandrov, L.B., Glodzik, D., Zou, X., Davies, H., Staaf, J., Sieuwerts, A.M., Brinkman, A.B., Martin, S., Ramakrishna, M., et al. (2016). The topography of mutational processes in breast cancer genomes. Nat. Commun. *7*, 11383.

Neal, R.M., and Others (2011). MCMC using Hamiltonian dynamics. Handbook of Markov Chain Monte Carlo *2*, 2.

Ng, A.W.T., Poon, S.L., Huang, M.N., Lim, J.Q., Boot, A., Yu, W., Suzuki, Y., Thangaraju, S., Ng, C.C.Y., Tan, P., et al. (2017). Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. Sci. Transl. Med. *9*.

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

Nordling, C.O. (1953). A new theory on cancer-inducing mechanism. Br. J. Cancer *7*, 68–72.

Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B., and Pearl, F.M.G. (2015). Therapeutic opportunities within the DNA damage response. Nat. Rev. Cancer *15*, 166–180.

Petljak, M., Alexandrov, L.B., Brammeld, J.S., Price, S., Wedge, D.C., Grossmann, S., Dawson, K.J., Ju, Y.S., Iorio, F., Tubio, J.M.C., et al. (2019). Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell *176*,

1282–1294.e20.

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

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

Roberts, S.A., and Gordenin, D.A. (2014). Hypermutation in human cancer genomes: footprints and mechanisms. Nat. Rev. Cancer *14*, 786.

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.

Shlien, A., Campbell, B.B., de Borja, R., Alexandrov, L.B., Merico, D., Wedge, D., Van Loo, P., Tarpey, P.S., Coupland, P., Behjati, S., et al. (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat. Genet. *47*, 257–262.

Sieber, O.M., Tomlinson, S.R., and Tomlinson, I.P.M. (2005). Tissue, cell and stage specificity of (epi)mutations in cancers. Nat. Rev. Cancer *5*, 649–655.

Szikriszt, B., Póti, Á., Pipek, O., Krzystanek, M., Kanu, N., Molnár, J., Ribli, D., Szeltner, Z., Tusnády, G.E., Csabai, I., et al. (2016). A comprehensive survey of the mutagenic impact of common cancer cytotoxics. Genome Biol. *17*, 99.

Taira, K., Kaneto, S., Nakano, K., Watanabe, S., Takahashi, E., Arimoto, S., Okamoto, K., Schaaper, R.M., Negishi, K., and Negishi, T. (2013). Distinct pathways for repairing mutagenic lesions induced by methylating and ethylating agents. Mutagenesis *28*, 341–350.

Tijsterman, M., Pothof, J., and Plasterk, R.H.A. (2002). Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient Caenorhabditis elegans. Genetics *161*, 651–660.

Tomasetti, C., Vogelstein, B., and Parmigiani, G. (2013). Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. Proc. Natl. Acad. Sci. U. S. A. *110*, 1999–2004.

Tomasetti, C., Marchionni, L., Nowak, M.A., Parmigiani, G., and Vogelstein, B. (2015). Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc. Natl. Acad. Sci. U. S. A. *112*, 118–123.

Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell *168*, 644–656.

Vogelstein, B., and Kinzler, K.W. (2015). The Path to Cancer --Three Strikes and You're Out. N. Engl. J. Med. *373*, 1895–1898.

Wang, V.G., Kim, H., and Chuang, J.H. (2018). Whole-exome sequencing capture kit biases yield false negative mutation calls in TCGA cohorts. PLoS One *13*, e0204912.

Wheless, L., Kistner-Griffin, E., Jorgensen, T.J., Ruczinski, I., Berthier-Schaad, Y., Kessing, B., Hoffman-Bolton, J., Francis, L., Shugart, Y.Y., Strickland, P.T., et al. (2012). A

community-based study of nucleotide excision repair polymorphisms in relation to the risk of non-melanoma skin cancer. J. Invest. Dermatol. *132*, 1354–1362.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871.

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.

Yoon, J.-H., McArthur, M.J., Park, J., Basu, D., Wakamiya, M., Prakash, L., and Prakash, S. (2019). Error-Prone Replication through UV Lesions by DNA Polymerase θ Protects against Skin Cancers. Cell *176*, 1295–1309.e15.

Zheng, C.L., Wang, N.J., Chung, J., Moslehi, H., Sanborn, J.Z., Hur, J.S., Collisson, E.A., Vemula, S.S., Naujokas, A., Chiotti, K.E., et al. (2014). Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. Cell Rep. *9*, 1228–1234.