

Supplementary Materials and Methods

Mutational signatures are jointly shaped by DNA damage and repair

[1. Experimental design](#)

[2. Data acquisition and filtering](#)

[3. Model specification for signature and interaction effect extraction in *C. elegans*](#)

[3.1. Input formatting](#)

[3.2. First step: using mutation accumulation data to define a prior on signatures of DNA repair deficiencies](#)

[3.3. Second step: simultaneous identification of the signatures of genotoxins and extraction of interaction effects](#)

[3.4. Parameter estimation](#)

[3.5. Batch effect estimation in EMS exposed samples](#)

[3.6 Interaction testing](#)

[4. Comparison between *C. elegans*, iPS cell and cancer derived signatures of genotoxins](#)

[5. Microhomology analysis for deletion breakpoints in *C. elegans*](#)

[6. DNA repair analysis across TCGA](#)

[6.1 Motivation](#)

[6.2. Data description](#)

[6.3. Classification of mutations across pathways and labeling of samples](#)

[6.3.1 Monoallelic and biallelic deficiency](#)

[6.4. Analysis of mutational burden and change in the mutational profile](#)

[6.4.1 Mutation burden](#)

[6.4.2 Mutational profiles](#)

[7. dNdS analysis across DNA repair genes in TCGA](#)

[7.1. Small changes in mutation rate may cause large cancer risk](#)

[7.2. Sample selection](#)

[7.3. Pathway-level analysis](#)

[7.4. Gene-level analysis](#)

[8. Interaction analysis in human data](#)

[8.1 Damage-repair interactions in cancer: motivation](#)

[8.2 Model specification](#)

[8.3. POLE proofreading domain mutations and MMR deficiency](#)

[8.4. Tissue-specific effects in MMR deficiency](#)

[8.5. Temozolomide and MGMT](#)

[8.6. APOBEC mutagenesis and REV1/UNG defects](#)

[8.7. UV exposure and XP patients](#)

[8.8. UV exposure and somatic defects NER](#)

[8.9. Smoking and NER defects](#)

1. Experimental design

The study follows the design of the pilot study¹ using a new set of *C. elegans* wild-type and mutant lines and a larger set of genotoxic agents, both described in **Supplementary Table 1**.

2. Data acquisition and filtering

Filtering of samples was performed using the same criteria as in ². Base substitutions were called using CAVEMAN ³ and subjected to the following filtering criteria:

1. Coverage of the variant site in both the sample of interest and reference should not exceed 150 reads or recede below 15 reads;
2. There should be no reads reporting the variant in the reference sample;
3. There should be at least 20% of reads and at least 5 reads reporting the variants in test sample;
4. There should be at least one read in the test sample reporting the variant in each direction;
5. There should be no indel called at the same position (relevant for homopolymer junctions);
6. All the variants are filtered against the normal panel of 6 wild-type samples; samples of the same genotype and generation higher than 1 are not compared to each other to avoid overfiltering of related samples.

Multiple substitutions which were found at adjacent sites in the same sample are classified as dinucleotide or multi-nucleotide variants, if their VAF is similar (difference less than 5%). Small size insertions and deletions were called using PINDEL ⁴ and subjected to the following filtering criteria:

1. Coverage of the variant site in both the sample of interest and reference should not exceed 150 reads or recede below 10 reads;
2. There should be no more than 1 read reporting the variant in the reference sample;
3. There should be more than 20% of reads and at least 5 reads reporting the variants in test sample;
4. There should be at least one read in test sample reporting the variant in each direction;
5. If the variant falls into a repetitive region, the regions should not be longer than 18 repeats;
6. All the variants are filtered against the normal panel of 6 wild-type samples; samples of the same genotype and generation higher than 1 are not compared to each other to avoid overfiltering of related samples.

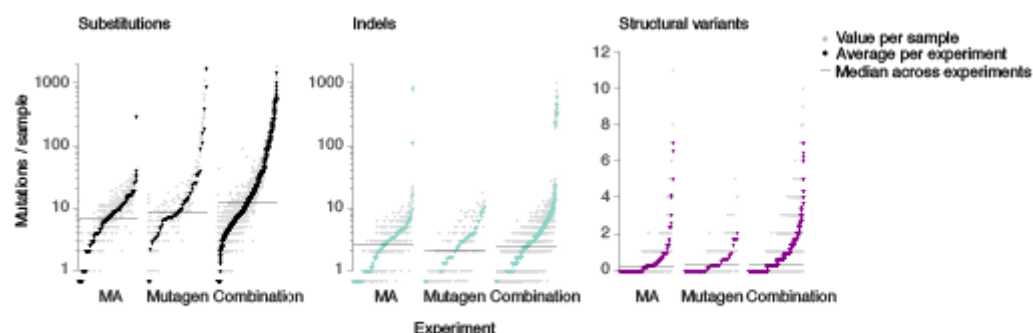
Indels are further classified based on the type (deletions, insertions, and complex indels - or deletions-insertions (DI)) and size (1 bp, 2-5 bp, 5-50 bp, 50-400 bp). Small insertions and deletions are further classified based on the local context: if the indel happened in repetitive sequence or not.

The structural variants were called using DELLY ⁵ with the following filters:

1. Each variant should be supported by more than 10 high-quality reads in test sample and no reads in the reference sample;
2. Each variant should pass the default DELLY quality filter (which relates to mapping quality around the breakpoints);
3. Variants in telomeric regions were removed;
4. Duplicated SVs across unrelated samples were removed; samples of the same genotype and generation higher than 1 are not compared to each other to avoid overfiltering of related samples.

The resulting sets of breakpoints were further classified in line with ⁶ using clustering by proximity and a simplified set of variants:

- Tandem duplications;
- Deletions;
- Inversions;
- Intrachromosomal translocations;
- Interchromosomal events;
- Foldbacks (change of the orientation of sequence without a second breakpoint in close proximity);
- Complex events (everything which was not possible to classify any other way).



Sup. Mat. Figure 1. Observed mutations across different experiment types.

3. Model specification for signature and interaction effect extraction in *C. elegans*

3.1. Input formatting

Variants were classified into 96 SNV classes (6 base substitution types (using pyrimidine reference) per 16 trinucleotide contexts), 2 MNV classes (dinucleotide variants and longer multinucleotide 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). For each of the classes i , $i = 1, \dots, 119$ and samples j , $j = 1, \dots, 2721$ we counted the respective number of mutations per sample Y_j^i .

We were interested in estimating the signatures of 54 genetic backgrounds, 12 mutagen exposures, and 196 interaction factors for each experiment.

3.2. First step: using mutation accumulation data to define a prior on signatures of DNA repair deficiencies

In order to leverage the information from all mutant samples towards the refinement of signatures of genetic backgrounds, we first define a prior based solely on mutation accumulation samples, and calculate the posterior distribution using the rest of the samples. In the first stage, for each of the mutation accumulation samples, the mutation counts $Y \in N^{119}$ were modeled by a negative binomial distribution,

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

with expectation

$$\mu = g^* (G^* \cdot S_G^*).$$

Here g denotes the adjusted generations in the absence of a mutagenic exposure, $g^* = \lambda(N)$, where

$$\lambda(N) = \sum_{i=1}^N \frac{1}{2^{i-1}} + \frac{1}{4} \sum_{i=1}^{N-1} \sum_{j=1}^i \frac{1}{2^{j-1}}$$

is a function reflecting the fraction of all accumulated mutations observed after N generations given the 25%-50%-25% chances of each mutation being fixed as homozygous, remaining heterozygous, or being lost. G^* is a binary indicator matrix of genotypes, and $S_G^* \in M_{119 \times 51}(\mathbb{R}^+)$ is a matrix of signatures of mutation accumulation effects per genotype (star denotes the subset of data where mutation accumulation data is available). We used a log-normal prior distribution for these signatures, $S_G^* \sim \log N(0, \sigma_G^2)$ iid with a scalar variance σ_G^2 . In total, we used 451 samples from mutation accumulation experiments with generation number higher than 1, and obtained signatures for 51 genotypes (*agt-1*, *exo-1*, *rad-51* knockouts did not have mutation accumulation experiments).

3.3. Second step: simultaneous identification of the signatures of genotoxins and extraction of interaction effects

After estimating S_G^* using the mutation accumulation data, we use the posterior draws for it as a prior for the signatures of DNA repair deficiency signatures in the full model (the genotypes without mutation accumulation data are assumed to have the same prior as the wild-type). Now the expected mutation counts are modelled as

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

Here $\mu_G = G \cdot S_G \in \mathbb{R}_+^{2721 \times 119}$ describes the mutations expected due to the genetic background, with G denoting an indicator matrix of genotypes per sample and S_G being a matrix of mutational signatures of the genetic knockouts with a prior $S_G \sim \Gamma(\text{shape}_G, \text{rate}_G)$ fitted to the posterior draws for S_G^* . Similarly, $\mu_M = M \cdot S_M \in \mathbb{R}_+^{2721 \times 119}$ is the expectation for the mutations created by mutagen M at median dose, 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 $S_M^{(j)} \sim \log N(0, \sigma_{Mj}^2)$, where the variances $\sigma_M^2 = (\sigma_{M1}^2, \dots, \sigma_{M12}^2)$ have prior distributions $\sigma_{Mj}^2 \sim \Gamma(1, 1)$ iid. The terms $\alpha_I \in \mathbb{R}_+$ and $\beta_I \in \mathbb{R}_+^{119}$ model how these terms interact. The scalar term for the genotype is linear with respect to the mutagen dose,

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

with $I \in \mathbb{Z}_2^{2721 \times 196}$ and $J \in \mathbb{Z}_2^{2721 \times 208}$ being binary indicator matrices for all interactions and all mutagenesis experiments (i.e. interactions and wild-type mutagenesis experiments), respectively. Rates $b_I \in \mathbb{R}_+^{196 \times 1}$ and offsets $a_I \in \mathbb{R}_+^{208 \times 1}$ were assumed to have an exponential prior distribution, $a_I \sim \text{Exp}(1)$ and $b_I \sim \text{Exp}(1)$. The rationale behind this parameterisation is that the *C. elegans* strain used for the experiment may have slightly diverged, thereby adding $a_I \cdot S_G$ mutations with the same spectrum as in the absence of a mutagen. In addition to that, mutagen exposure can amplify the mutation types specific to the genotype, as in the cases of alkylating agent treatment of TLS deficient samples, and hence the mutagen exposure of dose d may add $d \cdot b_I$ mutations of the mutagen-free spectrum.

The main assumption about interaction effects is that the wild-type spectrum of the mutagen S_M may change by the factors $\beta_I \in \mathbb{R}_+^{119 \times 196}$ measuring the fold-change of each mutation type which is expressed as

$$\beta_I = (I \cdot S_I),$$

where $S_I \in \mathbb{R}_+^{196 \times 119}$ is the fold-change with 119 terms per interaction.

As the number of mutations in individual classes were sometimes too low to capture the effects which were visible when aggregating counts within each mutation category, we used the interaction effects $\bar{S}_I \in \mathbb{R}_+^{196 \times 11}$ derived from an analogous model applied to the counts summed across 11 categories (6 types of SNVs, MNVs, 3 types of indels, SVs), $\bar{Y} \in \mathbb{R}_+^{196 \times 11}$, as the prior for per-class interaction effects. The posterior expected value for \bar{S}_I , $\bar{S}_I^* = E[\bar{S}_I | \bar{Y}]$, was calculated using a Laplace prior $\log \bar{S}_I^{(j)} \sim \text{Laplace}(0, \bar{\sigma}_{Ij}^2)$. Being aware that the fold-change of a sum is not linearly related to the individual mutation class fold-changes, we nevertheless observed a better fit and correspondence between the size and variability of the effects when using this structure. Hence, S_I had a prior distribution $\log S_I^{(j)} \sim \text{Laplace}(\bar{S}_I^* \times C, \sigma_{Ij}^2)$, where C is a matrix spreading the category value across the corresponding individual mutation classes, and all $\bar{\sigma}_I^2 = (\bar{\sigma}_{I1}^2, \dots, \bar{\sigma}_{I196}^2)$ and $\sigma_I^2 = (\sigma_{I1}^2, \dots, \sigma_{I196}^2)$ were assumed to come from $\Gamma(1, 1)$ iid.

3.4. Parameter estimation

The overdispersion parameter $\phi = 100$ was selected based on the estimates of overdispersion in the dataset when fitting a negative binomial distribution with a shared overdispersion parameter to all the experiments in the dataset. It allows to account for additional variation within the replicates as well as for overflow of zeros which could present a problem in a Poisson framework.

The model was specified using R-package “greta”, v. 0.2.3 (<http://CRAN.R-project.org/package=greta>), and the posteriors of S_G , S_M , S_I , a_J , b_I and hyperparameters σ_M^2, σ_I^2 were estimated using Hamiltonian Monte Carlo sampling. We used 2000 steps for warmup and 5000 steps over 4 chains to ensure convergence. The point estimates for the parameters were taken as means of the samples across all chains. The relevant R codes are published on Github under github.com/gerstung-lab/signature-interactions. The estimates are provided in **Supplementary Tables 2-3**, and visualized in **Supplementary Figure 2a**, **Supplementary Note Figures 1-2**.

3.5. Batch effect estimation in EMS exposed samples

EMS exposure was the only genotoxin for which we observed very different mutation counts across different experiments in wildtype. Respective samples were coming from 5 batches, which are stated in the ‘Comments’ section of the samples description table in Supplementary Table 1. In order to account for this effect, we introduced additional factor $\xi = \{\xi_i\}$, $i = 1, \dots, 4$ which accounted for the log-difference in real dose applied to the samples in batches 2 to 5 compared to batch 1 which was considered as reference. The prior distribution for these adjustments was taken as $\xi_i \sim N(0, 0.5)$ iid. The dose for these samples was then calculated as $d' = d * e^{\text{batch} * \xi}$, where $\text{batch} \in \{0, 1\}^{2721 \times 4}$ is a binary matrix reflecting if a sample belongs to any of the batches 2-5. These adjustments were estimated along with the rest of the coefficients.

3.6 Interaction testing

Combined genotype-mutagen interactions were tested for effect in two settings: altering the total number of base substitutions, and changing the distribution of mutations. The fold-change in the number of single base substitutions was calculated as predicted with interactions versus the one predicted without interactions using 2000 out of 10000 samples draws across 4 chains for all 196 interactions. The change in profile was quantified by calculating the cosine distance between the expected profiles with and without interactions, and those with distance higher than 0.2 were considered different. As all of the interactions which showed a change in signature appearance also came up in burden analysis, we only applied hypothesis testing (chi-square test for $\log(E(FC_s))$ for $s = 1, \dots, 196$) and Benjamini-Hochberg FDR control procedure for the set of substitution burden fold-changes.

4. Comparison between *C. elegans*, iPS cell and cancer derived signatures of genotoxins

We next wished to compare the experimental mutational signatures of genotoxin exposure between *C. elegans* and human iPS cells⁷, and explore their similarity to mutational signatures computationally extracted from tumour mutation spectra⁸. With a mean cosine similarity of 0.63 (range 0.20-0.84), the *C. elegans* experimental signatures (adjusted to the human trinucleotide frequencies (**Methods**)) generally display a good level of similarity with their human experimental counterparts, and at times even better concordance with computationally derived cancer signatures with the same suspected origin (**Supplementary Figure 2b**).

In particular, the *C. elegans* signature of the DNA ethylating agent EMS was strikingly similar (0.90) to the cancer-derived computational signature SBS11 associated with temozolomide treatment, yet different from the temozolomide signature derived from iPS cell lines (**Supplementary Figure 2c**). This may be due to iPSC-specific metabolic activation of temozolomide, or altered use of DNA repair pathways in these cells. The reported mutation spectrum observed in *Salmonella typhimurium* upon EMS exposure is nearly identical to the one observed in *C. elegans*^{9,10}. Similarly, the iPSC-derived experimental signature of the methylating agent DMS demonstrated a preference for C>T changes, a base substitution almost absent in the *C. elegans* DMS signature (**Supplementary Figure 2c**).

Cisplatin treatment yielded a signature which only resembled the C>A part of the spectrum observed iPSCs and cancers^{8,11,12}, similar to that reported in chicken fibroblasts which also consisted primarily of C>A changes¹³. *C. elegans* exposure to UV-B resulted in a mutation spectrum dominated by C>T, similar to that in cell lines and human cancer samples, albeit with an additional fraction of T>C mutations (**Supplementary Figure 2c**). This discrepancy might be due to the difference in UV sources used. Our experiments were performed using a UV-B source, whereas¹⁴ used a mixture of 90% UV-A and 10% UV-B to mimic the actual UV spectrum contained in the sunlight.

Importantly, *C. elegans* spectra resulting from the exposure to ionizing radiation, generally associated with the equal mutagenesis of all bases, share significant similarity with the mutation spectrum observed in secondary tumours from tissues previously exposed to radiotherapy ¹⁵ (**Supplementary Figure 2c**).

Signatures of aflatoxin B1 and aristolochic acid exposure were similar in all systems. 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. Indeed, some of the experimental signatures derived from human cells, such as signatures of aflatoxin and temozolomide (**Supplementary Figure 2c**), were less similar to cancer signatures than *C. elegans* signatures.

5. Microhomology analysis for deletion breakpoints in *C. elegans*

Microhomology search was conducted for all medium-size deletions found in the *C. elegans* dataset. We looked for precise homology at the breakpoints using ce11 *C. elegans* genome build. The expected distribution of microhomologies across deletions was calculated based on the probability of encountering the same k-mer around the breakpoints as described in ¹⁶. We then fitted an additive Poisson GLM to describe the number of deletions with MH of a certain size as a combination of random contribution, 1-bp MH with their associated distribution of higher order MH occurring by chance ($P(MH_{obs} = 2 | MH_{real} = 1) = \frac{P(MH_{obs}=2)}{P(MH_{real}=1)}$) for $j = 2, 3, \dots$), similar terms for 2-bp and 3-bp MH, and for MH longer than 3 as the numbers of deletions with longer MH were too low to permit confident estimation. For the deletions across non-AA exposed samples, the model predicted that 52% (SD = 3.5%) of all deletions were happening in an MH-independent manner, 24% (SD = 3.4%) were assigned to 1-bp MH, 10% (SD = 2.5%) to 2-bp MH, 4% (SD = 1.7%) to 3-bp MH, and 10% (SD = 1.4%) were predicted as dependent on longer MH. For deletions in AA-exposed samples, 40% (SD = 13%) of all deletions were predicted as non-random, with a higher bias towards 1-bp MH: 35% (SD=13%) of deletions were predicted as dependent on 1-bp MH, with no deletions assigned to higher-order MH (**Supplementary Figure 5b**). According to the overall distribution of indels across all samples, aristolochic acid exposed samples, and *polq-1* deficient mutants (**Supplementary Figure 5c**), AA-exposed samples have more indels in the range between 50 and 300 bp than an average sample, whereas *polq-1* mutants had almost no indels longer than 10 bp.

6. DNA repair analysis across TCGA

6.1 Motivation

DNA repair deficiencies have a crucial role in cancer development. However, only a few of them have been assigned an associated mutational signature. We established a catalogue of potential DNA repair pathway deficiencies in human cancers by annotating heterozygous

and homozygous somatic missense and loss-of-function mutations, as well as epigenetic gene silencing, in 81 core genes of the 9 consensus DNA repair pathways^{17,18} across 9,946 patients and 30 cancer types available from TCGA (**Supplementary Table 4**, **Supplementary Figure 1a,b**). Tumours with somatic heterozygous mutations in genes required for DNA repair and DNA damage signaling are very common, reaching nearly 100% for the DNA damage sensing (DS) pathway in ovarian cancers and uterine carcinosarcomas due to very high rates of TP53 mutations¹⁸. Somatic biallelic DNA repair gene deficiency, in contrast, was generally rare (**Supplementary Figure 1b**), suggesting that single DNA repair defects in cancers are rarely associated with a change in mutation rate or spectrum. Exceptions were damage sensing defects (mostly well-studied *TP53* mutations), biallelic MMR defects, HR deficiency, and monoallelic mutations of the polymerase epsilon (POLE) exonuclease domain (**Supplementary Figure 1c,d**).

6.2. Data description

If not stated otherwise, data was obtained from GDC (<https://cancergenome.nih.gov>) and filtered according to¹⁹. Somatic and germline mutations were acquired using cgpCaVEMan (<http://cancerit.github.io/CaVEMan/>) and cgpPindel (<https://github.com/cancerit/cgpPindel>) variant callers. Copy number profiles, purities and ploidies of TCGA samples were estimated using ASCAT²⁰.

6.3. Classification of mutations across pathways and labeling of samples

6.3.1 Monoallelic and biallelic deficiency

DNA repair genes across 9 pathways were selected as core or accessory components based on the combined classifications from^{17,18} excluding genes indicates as consistently undercovered in exome sequencing²¹ (**Supplementary Table 4**). We investigated point mutations, indels and copy number states for each gene. Samples were classified as having monoallelic deficiency of a certain DNA repair pathway if at least one of the core components of this pathway carried a damaging mutation of allele frequency higher than 0.4 (damaging mutation is defined as being a missense, nonsense, frameshift, essential splice site, 'stop gained', 'stop lost' or 'start lost' mutation), or is located in a region with major copy number 1, or is methylated with average methylation score between 0.2 and 0.75 (see below).

Methylation analysis was based on the respective methodology from¹⁸. For each cancer type where methylation and expression data were available, we selected the CpG sites relevant to the gene of interest by proximity (using the annotation from "IlluminaHumanMethylation450kanno.ilmn12.hg19" R package, v. 0.6.0, <https://bioconductor.org/packages/IlluminaHumanMethylation450kanno.ilmn12.hg19/>) and by correlation with the expression of this gene (with a cutoff of -0.5). Then we assessed the median methylation of these CpG sites across the samples within the cancer type. Assuming that the gene of interest with median beta value lower than 0.1 can be considered unmethylated, we calculated z-scores of expression values within the group of samples with

unmethylated gene of interest. Then we iteratively adjusted the group of samples with “methylated” gene of interest by considering first all samples with median beta value of at least 0.2 in the gene of interest, and then removing samples with lowest beta values until the “methylated” group contained less than 3 samples, its mean expression z-score reached 5% quantile of the distribution of expression z-scores in unmethylated group, or until the difference in expression values between the groups fell below 2-fold. Then, if the resulting group of samples with presumably methylated gene of interest is meaningful (larger than 2 samples, with at least one sample with a median beta value >0.5 , with average expression value at least 50% lower than that in unmethylated group, and with mean expression z-score below 5% quantile of normal distribution), we additionally confirm the difference using one-tailed Student t-test on z-scores between “methylated” and unmethylated groups. If it results in a p-value lower than 0.05, the samples from “methylated” group are further labeled as such. Using this approach, we identified significant groups of methylated samples for MLH1, MLH3, MGMT, REV3L and TP53.

Cancer samples were further labelled as carrying biallelic deactivation of a certain pathway if a sample had a high-*VAF* (> 0.8) damaging mutation in any core component of this pathway, or if this core gene was subject to deep deletion (belonged to a region with major copy number 0), or if it was methylated with a median beta value >0.8 .

Monoallelic deficiency in proofreading ability of polymerase epsilon was assessed by screening for damaging mutations in the proofreading domain of POLE (residues 268 to 471) with *VAF* > 0.2 .

6.4. Analysis of mutational burden and change in the mutational profile

6.4.1 Mutation burden

The change in mutation burden between samples with functional and defective DNA repair pathways was performed per cancer type. In total, we analyzed 8797 samples which had more than 50 but less than 20000 mutation per exome. We further looked at the mutation rates per year per Megabase in the samples without any defects in a certain pathway, those with only monoallelic defects, and those with biallelic. Each group was only considered if it contained more than 4 samples (samples with mutations in POLE proofreading domain mutations were removed from all other analyses but POLE and POLE+MMR). The increase in average mutation rate between the wild-type and heterozygous (and, when available, wild-type and homozygous) groups was tested using one-tailed Wilcoxon test. P-values were further adjusted for multiple testing correction using Benjamini-Hochberg procedure (**Supplementary Figure 1c**).

In addition to the Wilcoxon test, for the cases with significant change in the mutation rates we also tested how likely was it for the mutation in the respective pathway to be a consequence of elevated mutational burden. For that, we compared the odds ratios between mutation rates within the respective pathway to the ploidy-adjusted global mutation rates in the non-mutated and mutated groups. It allowed for discriminating the cases such as

damage sensing pathway mutations in melanomas (SKCM), which is according to this test very likely to be a consequence of elevated mutation burden, whereas MMR deficient uterine cancers (UCEC) have an odds ratio of having mutations in the pathway which is much higher than that for MMR proficient samples.

6.4.2 Mutational profiles

In order to identify a change in the mutation spectrum between DNA repair proficient and deficient samples, we regressed out several endogenous processes (signatures SBS1 - deamination of methylated cytosines at CpG sites, SBS2 and SBS13 - APOBEC mutations, SBS5 and SBS30 - flat clock-like signatures associated with patient age) using the NMF procedure minimizing KL divergence with the signature matrix being fixed at values from ²². We then selected only the samples where at least 100 mutations were left in order to get a reliable profile of mutations. Then we calculated cosine distances between the groups of non-mutated samples within certain cancer type, samples with monoallelic defects and samples with biallelic defects in a given pathway (**Supplementary Figure 1d**).

7. dNdS analysis across DNA repair genes in TCGA

7.1. Small changes in mutation rate may cause large cancer risk

In order to infer interactions between suspected genotoxin exposure and DNA repair deficiency in cancer genomes, we adapted the Negative Binomial model used for the *C. elegans* experimental system to cancer genomes (see main **Methods**). Applying this approach to a range of cancers with suspected genotoxic exposures and stratifying based on DNA repair status, we were able to detect cases of DNA repair and damage interaction leading to increased mutagenesis and/or altered signature. We found that such interactions were rare and typically only revealed moderate effects. Nevertheless, a number of noteworthy examples exist, illustrating the action of different DNA repair pathways, including DR, MMR, and TLS, upon exposure to distinct types of DNA damage conferred by APOBEC enzymes, polymerase proofreading deficiency, UV exposure and temozolomide treatment (**Figure 2d**).

Given that the association between DNA repair pathway defects and mutation phenotypes in cancer genomes appears weaker than expected, we took an evolutionary approach to investigate if changes in DNA repair status nevertheless affect oncogenesis. Using the rate ratio of non-silent (N) to silent (S) mutations dN/dS in a given gene as an established method for quantifying positive selection in cancer ¹⁹, we found that overall 7/248 genes involved in DNA repair displayed significant signs of positive selection for missense mutations, and 17/248 DNA repair genes were enriched for truncating mutations (FDR<10%; **Supplementary Figure 3h, Supplementary Table 5**). Among these the majority were 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).

7.2. Sample selection

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 ¹⁹. Background mutation rates were estimated using all the genes except for consistently undercovered genes ²¹. The dN/dS ratios of interest were compared to 1 (the dN/dS ratio under the neutral evolution model), and the significance of selection was assessed by comparing the relevant χ^2 statistic of the dN/dS ratio to χ^2 distribution with $df = 1$.

7.3. Pathway-level analysis

For pathway-level analyses, we ran dNdSCV analysis separately within each cancer type using the samples with less than 1000 coding mutations. For each combination of pathway and cancer type, we estimated the global dNdS ratio over the list of core genes within that pathway, and compared them to 1 using χ^2 statistic of the squared z-score of

$\left(\log(dN/dS_{pathway/cancer}) \right)$, with FDR control performed using Benjamini-Hochberg procedure. The resulting values are depicted in **Supplementary Figure 3h** and provided in **Supplementary Table 5**.

7.4. Gene-level analysis

For gene-level analyses, we estimated the global and per-gene dNdS values across all cancer types using all samples with less than 1000 coding mutations. dNdSCV analysis was performed for the extended set of 248 DNA repair associated genes (**Supplementary Table 4**) across all cancer types. For each gene, we estimated the dNdS ratio for missense and nonsense mutations, and compared them to 1 using χ^2 statistic of the squared z-score of

$\left(\log(dN/dS_{gene/cancer}) \right)$, with FDR control performed using Benjamini-Hochberg procedure. The resulting values are depicted in **Supplementary Figure 3h** and provided in **Supplementary Table 5**.

8. Interaction analysis in human data

8.1 Damage-repair interactions in cancer: motivation

Following up on previous work on uterine cancers that revealed distinct mutational signatures associated with MMR or POLE exonuclease activity deficiency, and the combination of both ²³, we developed a model to quantify non-additive signature changes between different groups of samples assuming that it is the spectrum associated with POLE proofreading deficiency that is altered in MMR+POLE deficient samples (**Methods**). Using this approach, we confirmed the increased C>A mutagenesis in a NpCpT context in samples with concurrent POLE mutations and MMR defects (**Supplementary Figure 3a,b**).

8.2 Model specification

The sampling model for human data was set up as described in Methods using MCMC sampling via R package “greta”, v. 0.2.3 (<http://CRAN.R-project.org/package=greta>). All models were run in 4 chains up to 1000 or 2000 warm-up and 1000 post warm-up samples to ensure convergence. The codes for all the analyses are available on GitHub: <http://github.com/gerstung-lab/signature-interactions> .

We run 4 chains of sampling, and claimed an effect being real if it was consistently assigned to the same signature. The final number of signatures was selected based on the convergence, similarity between signatures and feasibility of effect assignment, as the model tends to fluctuate or duplicate most variable signals when the dimensions chosen is too high.

8.3. POLE proofreading domain mutations and MMR deficiency

The effects of interaction between POLE and MMR defects were investigated using the uterine cancer cohort from TCGA (UCEC TCGA project) as indicated in ²³. Samples classified as MSI-H by Bethesda protocol (available in TCGA Clinical Explorer ²⁴) were considered to be MMR deficient, and samples with missense mutations in POLE proofreading domain (amino acids 267-472) were considered to have compromised proofreading activity. Overall, we analyzed 546 samples, 167 of which were labeled as MMR deficient, 55 - as having a POLE mutations, and 15 - as having both deficiencies. We identified 5 signatures, with a signature similar to COSMIC SBS10 signature (the one associated with POLE proofreading domain defects) being subject to interactions with different POLE mutations and POLE+MMR factor. In line with associations suggested previously, the transformation of POLE signature upon MMR deficiency closely resembled COSMIC signature SBS14, and signatures C1-2 from ²³ (**Supplementary Figure 3a**).

In line with the initial study, we further looked into differences in POLE signature across samples with different mutations in POLE proofreading domain, and found the two hotspot mutations P286R and V411L producing different profiles compared to each other and the rest of POLE mutations. With our approach, were able to better refine the modulation related to each mutation (**Supplementary Figure 3b**). Our findings confirm the concept of different mutations across POLE proofreading domain introducing different changes to the activity of the polymerase through variability in structural changes of the protein.

8.4. Tissue-specific effects in MMR deficiency

Previous studies also suggested an interaction of MMR with other processes such as the repair of deaminated 5-methylcytosines ². Using the same approach, we assessed the variability of MMR deficiency signature in cancer types originating from different tissues.

We screened all TCGA samples for the defects in MMR genes MLH1, PMS2, MSH2, MSH3 and MSH6 (we excluded MLH3 for the lack of samples with just the MLH3 mutation and visible MMR deficiency phenotype). Samples were labeled as having MMR defects if they had an impactful somatic mutation in MMR pathway (**Supplementary Table 4**) or were clinically determined to have MSI-H status as per TCGA Clinical Explorer ²⁴.

The samples were further filtered to exclude ones with less than 100 mutations or more than 20000 mutations, and the ones whose mutational profile was dominated by a strong mutagenic process: APOBEC, POLE, UV or tobacco (i.e. we excluded those with cosine similarity higher than 0.7 to the respective signatures). From the remaining samples, we selected data from 8 cancer types: breast (BRCA), cervical (CESC), head and neck (HNSC), stomach (STAD), colorectal (COAD and READ), liver (LIHC), lung (LUAD and LUSC), prostate (PRAD), and uterine cancers (UCEC).

Using this dataset of 782 samples, we first looked at the effects of each of the MMR genes being mutated (using 6 covariates, one for each gene), and did not find any significant effect. Next, we looked at the tissue effects (using colorectal as default and fitting 7 interaction effects for tissues - breast, cervix, head and neck, lung, stomach, uterus, liver), where we observed changes in the rates of mononucleotide variants, insertions and deletions in homopolymeric stretches, and variation in C>T mutations in a CpG context across different tissues (**Supplementary Figure 3c,d**). Compared to MMR proficient tumours, MMR deficient tumours showed on average a 5-fold higher rate of single base deletions per year following the same trend across different tissues. Moreover, the single base insertion rate 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 square well with the observation that MMR deficiency is associated with a number of additional mutational signatures of unknown aetiology ², suggesting that these may reflect distinct mutagenic processes which are exacerbated by MMR deficiency.

8.5. Temozolomide and MGMT

The MGMT methylation data as well as the somatic mutations from exomes of GBM samples treated with TMZ were taken from previous studies ²⁵. The effect estimation model was run on 96 substitution types considering 17 samples treated with TMZ: 11 with wild-type MGMT and 6 with MGMT hypermethylation. The model showed best results for 2 signatures, one similar to COSMIC signature 1 and another one flat (**Figure 3d**), and suggested that the estimated change of the second signature is higher than 100-fold.

8.6. APOBEC mutagenesis and REV1/UNG defects

In order to estimate the modulation of APOBEC related signature in response to defects in REV1/UNG machinery, we analyzed 842 TCGA tumour samples with significant APOBEC contribution (the ones which showed cosine similarity of at least 0.8 to a combination of COSMIC signatures SBS2 and SBS13) which had between 50 and 15000 mutations per exome. We further stratified these samples into REV1/UNG wild-type (794 samples) and mutated group (48 samples with at least monoallelic defects in *REV1* or *UNG*). Analysis of APOBEC-specific T[C>A/C/T]N mutations showed that defects in REV1 and/or UNG lead to 8% decrease in the relative fraction of C>G mutations related to C>T transversions. Signature extraction with one signature showed a slight relative increase of C>T part of the signature (**Figure 4c**).

8.7. UV exposure and XP patients

UV exposure signature in NER-proficient and deficient skin tumours was assessed using the whole genome data from cutaneous squamous cells carcinomas of 5 XP patients with in the age between 2 and 10, and 8 adult individuals in the age between 61 and 87 ²⁶ (**Figure 5d**). The variant calls were obtained from ²⁷. Signature extraction with one signature demonstrated a shift in the signature and more then 100-fold change in the age-adjusted mutation rate (**Figure 5d**).

8.8. UV exposure and somatic defects NER

NER effects in UV-associated melanomas were estimated using samples from TCGA SKCM project where the similarity to a combination of COSMIC signatures SBS7a and SBS7b (previously associated with UV) was higher than 0.8. 397 patients were tested for having either expression of any of the NER genes in the tumour sample below 20% of the median level in the dataset, or a somatic biallelic deactivation of a core NER gene in the tumour sample as described above in section 4.2.1. As the numbers were low, we also included samples with a high impact germline variant (homozygous or heterozygous) as predicted using Ensembl VEP ²⁸.

The final set contained 9 samples labeled as NER defective.

We estimated the effects of having defects in NER pathway using a single signature. NER did not show a dramatic change in the profile of UV signature, only produced a 5% increase in C>T mutations, but led to an overall 2-fold increase in the mutation burden per year (**Supplementary Figure 3e**).

In order to check if this is an effect of NER defects, or simply a consequence of increased mutational burden, we simulated UV-induced mutations in NER genes according to each sample's total burden of synonymous mutations, and tested if the observed number of mutations in these genes across the dataset is in line with expected. The observed number of nonsynonymous mutations across NER genes followed the distribution we generated via simulations, which indicated that most of the signal was coming from the difference in

mutational burden (**Supplementary Figure 3f**). Together with the high variability of this fold change effect, it led to the conclusion that there is no detectable effect of mutations in NER genes on the appearance of UV exposure signature.

We have also tested 59 samples with defects in TLS machinery but did not see any effect on either exposure coefficient or appearance of UV signature.

8.9. Smoking and NER defects

Effects of DNA repair machinery defects on the signature of tobacco smoke were estimated using samples from LUAD and LUSC projects ^{28,29}. Out of 905 samples across the two datasets, we selected 219 samples with high presence of COSMIC signature SBS4 (cosine similarity over 0.8), associated with tobacco smoking. Of these, 82 samples were labeled as being NER deficient, and 92 were labeled as having TLS defects based on somatic and germline damaging mutations, both homozygous and heterozygous. Having extracted two signatures most similar to smoking and APOBEC associated signatures, we did not find a significant effect of either NER or TLS on the exposure coefficient or mutational distribution of the signature associated with smoking (**Supplementary Figure 3g**).

References

1. Meier, B. & Gartner, A. Having a direct look: analysis of DNA damage and repair mechanisms by next generation sequencing. *Exp. Cell Res.* **329**, 35–41 (2014).
2. Meier, B. *et al.* Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and human cancers. *Genome Res.* **28**, 666–675 (2018).
3. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
4. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. 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 (2009).
5. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
6. Li, Y., Roberts, N., Weischenfeldt, J., Wala, J. A. & Shapira, O. Patterns of structural variation in human cancer. *bioRxiv* (2017).
7. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* (2019). doi:10.1016/j.cell.2019.03.001
8. Alexandrov, L. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* (2018). doi:10.1101/322859
9. Matsumura, S., Fujita, Y., Yamane, M., Morita, O. & Honda, H. A genome-wide mutation analysis method enabling high-throughput identification of chemical mutagen signatures. *Scientific Reports* **8**, (2018).
10. Flibotte, S. *et al.* Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics* **185**, 431–441 (2010).
11. Boot, A. *et al.* In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res.* **28**, 654–665 (2018).
12. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents.

Cell (2019). doi:10.1016/j.cell.2019.03.001

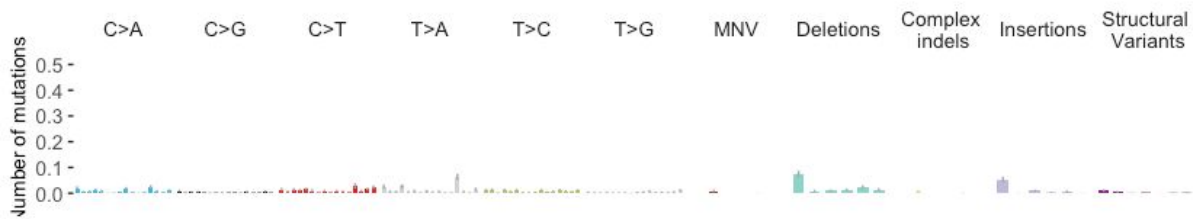
13. Szikriszt, B. *et al.* A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol.* **17**, 99 (2016).
14. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* (2019). doi:10.1016/j.cell.2019.03.001
15. Behjati, S. *et al.* Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
16. Roth, D. B., Porter, T. N. & Wilson, J. H. Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.* **5**, 2599–2607 (1985).
17. Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. & Pearl, F. M. G. Therapeutic opportunities within the DNA damage response. *Nature Reviews Cancer* **15**, 166–180 (2015).
18. Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep.* **23**, 239–254.e6 (2018).
19. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–1041.e21 (2017).
20. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16910–16915 (2010).
21. Wang, V. G., Kim, H. & Chuang, J. H. Whole-exome sequencing capture kit biases yield false negative mutation calls in TCGA cohorts. *PLoS One* **13**, e0204912 (2018).
22. Alexandrov, L. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* 322859 (2018). doi:10.1101/322859
23. Haradhvala, N. J. *et al.* Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.* **9**, 1746 (2018).
24. Lee, H., Palm, J., Grimes, S. M. & Ji, H. P. The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical–genomic driver associations. *Genome Med.* **7**, 112 (2015).

25. Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25**, 316–327 (2015).
26. Bradford, P. T. *et al.* Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J. Med. Genet.* **48**, 168–176 (2011).
27. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
28. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
29. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

Supplementary Note

Supplementary Note Figure 1. Experimental mutational signatures extracted for 54 genetic backgrounds (including the N2 wild-type strain) in *C. elegans*. 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**. Mutations are roughly ordered by pathways (**Supplementary Table 1**). The numbers for mutations per generation in each mutation class can be found in **Supplementary Table 2**.

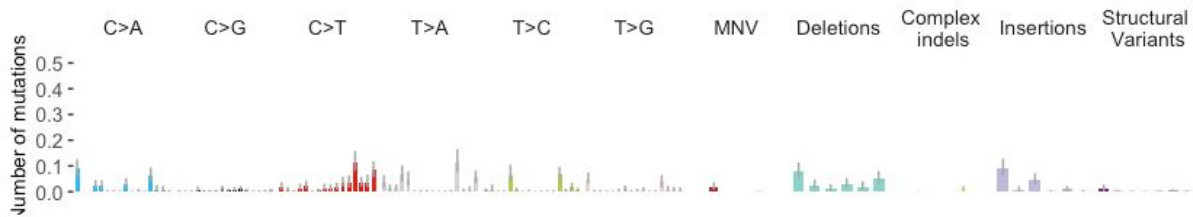
Effects for N2; 1 (0.6-1.5) het. mut-s per gen., max 40 gen. with 15 (11-24) mut-s



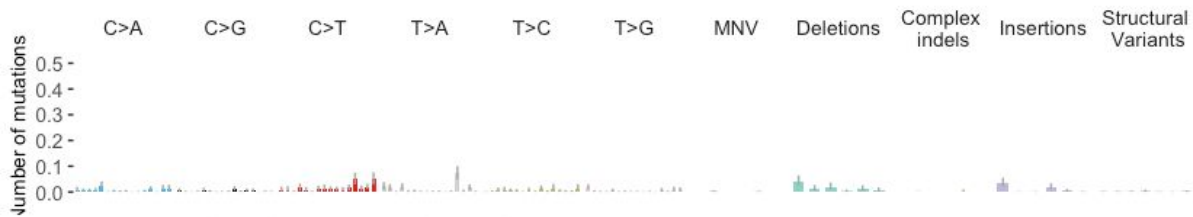
Estimated effects for agt-1; 0.9 (0.4-1.9) het. mut-s per gen., max 1 gen.



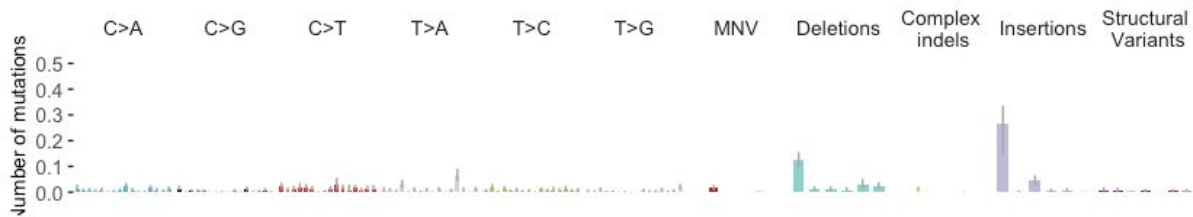
Effects for agt-2; 1.9 (0.9-3.4) het. mut-s per gen., max 20 gen. with 19 (8-35) mut-s



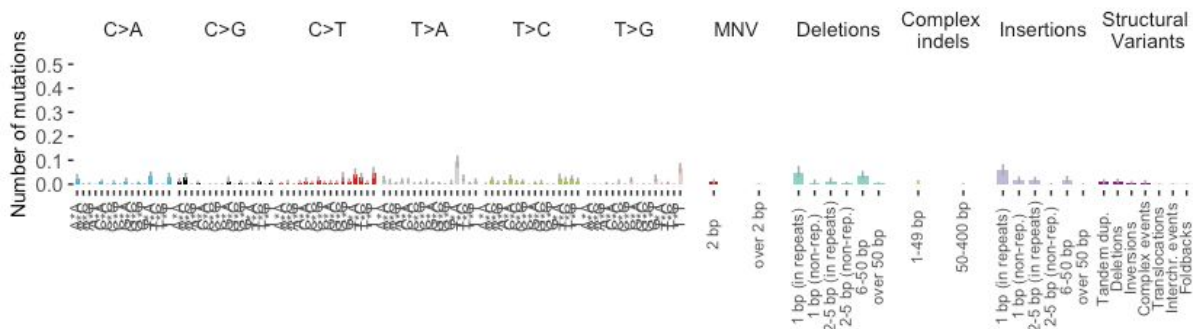
Effects for apn-1; 0.8 (0.3-1.8) het. mut-s per gen., max 40 gen. with 15 (12-20) mut-s



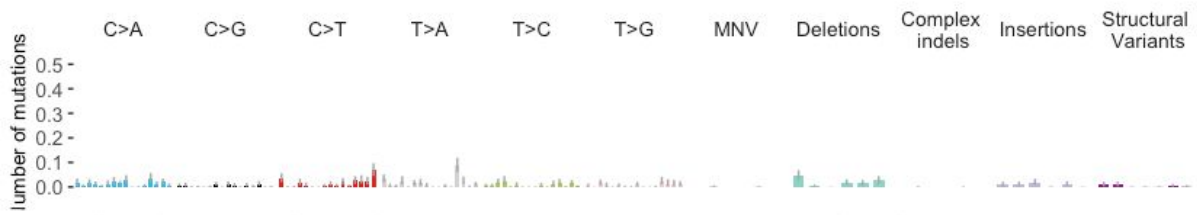
Estimated effects for exo-1; 1.3 (0.6-2.4) het. mut-s per gen., max 1 gen.



Effects for exo-3; 1.2 (0.5-2.4) het. mut-s per gen., max 40 gen. with 23 (21-27) mut-s



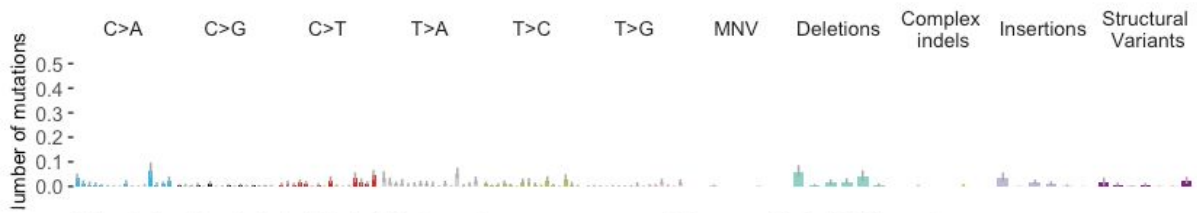
Effects for *ndx-4*; 1.2 (0.5-2.3) het. mut-s per gen., max 40 gen. with 24 (10-36) mut-s



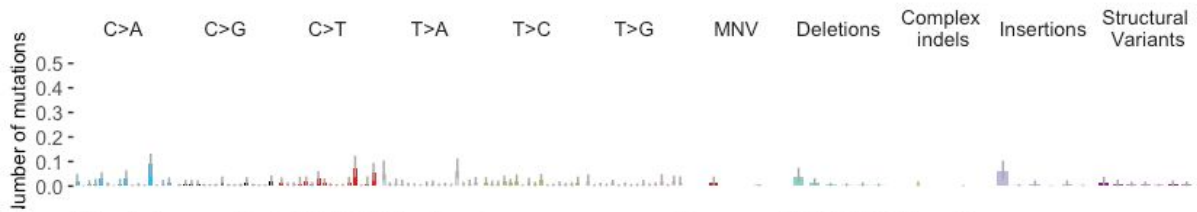
Effects for *parp-1*; 1 (0.4-1.9) het. mut-s per gen., max 40 gen. with 15 (13-18) mut-s



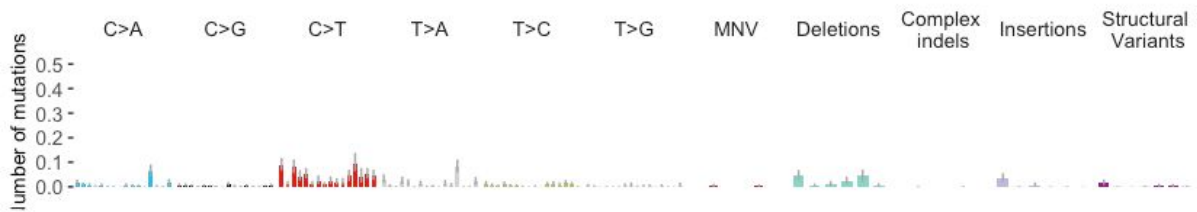
Effects for *parp-2*; 1.1 (0.4-2.2) het. mut-s per gen., max 40 gen. with 19 (12-25) mut-s



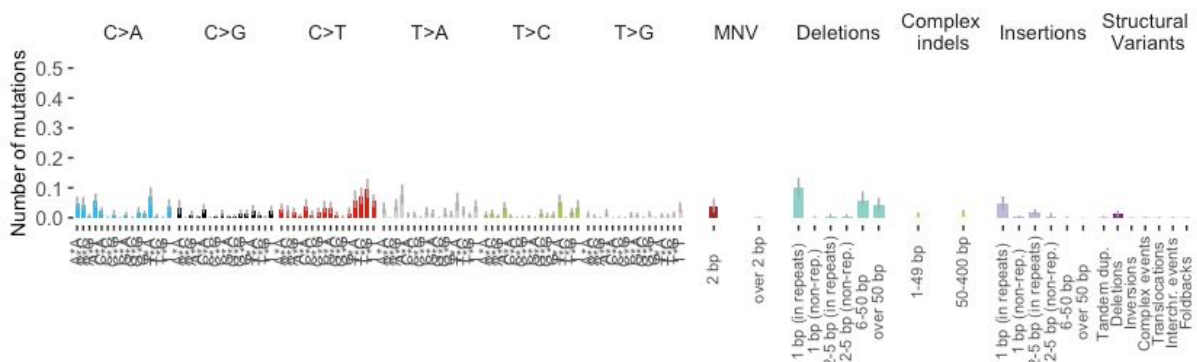
Effects for *tdp-1*; 1.2 (0.3-3.1) het. mut-s per gen., max 20 gen. with 9 (6-11) mut-s



Effects for *ung-1*; 1.4 (0.7-2.6) het. mut-s per gen., max 40 gen. with 28 (22-35) mut-s



Effects for *xpa-1*; 2 (1-3.6) het. mut-s per gen., max 40 gen. with 40 (34-44) mut-s



Effects for xpc-1; 1.6 (0.7-3) het. mut-s per gen., max 20 gen. with 16 (4-26) mut-s



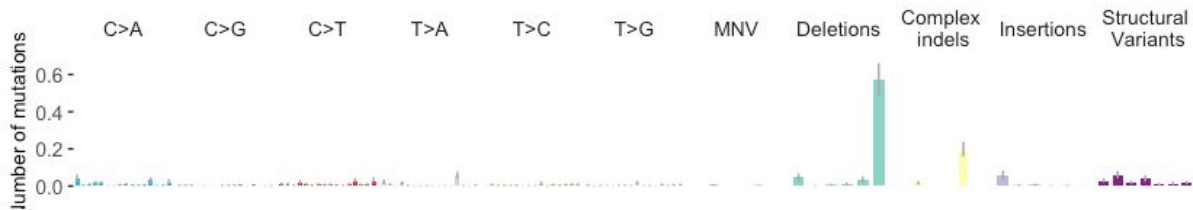
Effects for xpf-1; 1.4 (0.7-2.5) het. mut-s per gen., max 40 gen. with 23 (5-33) mut-s



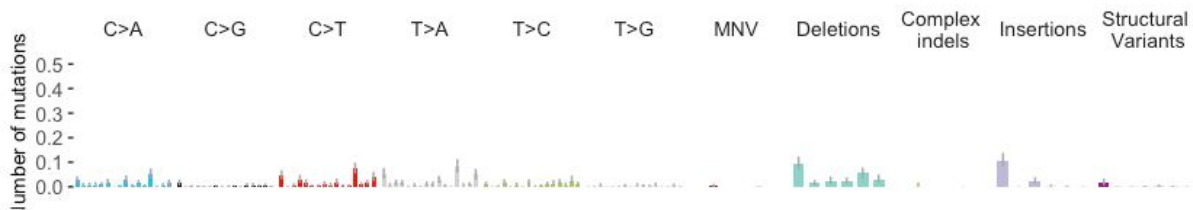
Effects for csb-1; 1 (0.4-2.1) het. mut-s per gen., max 20 gen. with 10 (4-18) mut-s



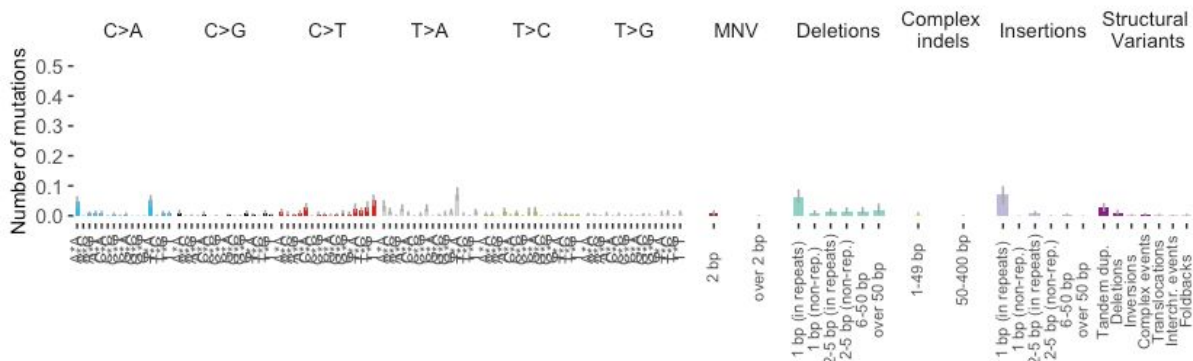
Effects for polh-1; 1.7 (1.1-2.7) het. mut-s per gen., max 20 gen. with 9 (8-12) mut-s



Effects for polk-1; 1.4 (0.7-2.6) het. mut-s per gen., max 40 gen. with 23 (10-37) mut-s



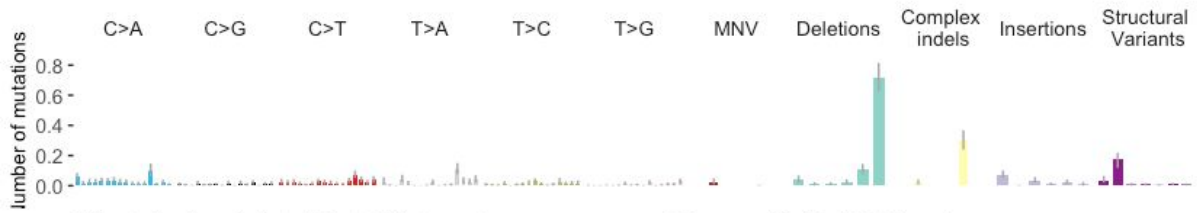
Effects for rev-1; 1 (0.4-1.9) het. mut-s per gen., max 40 gen. with 17 (13-21) mut-s



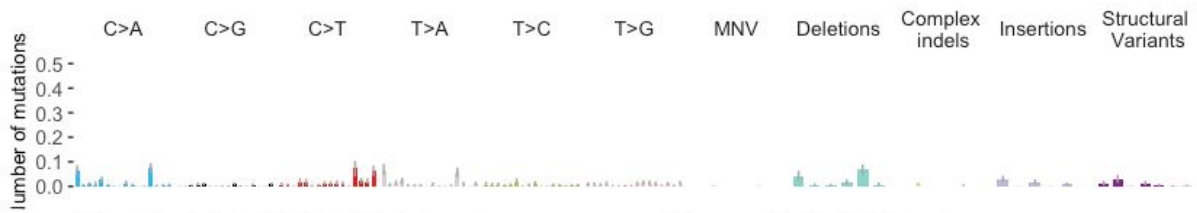
Effects for rev-3; 1.3 (0.6-2.4) het. mut-s per gen., max 20 gen. with 11 (4-14) mut-s



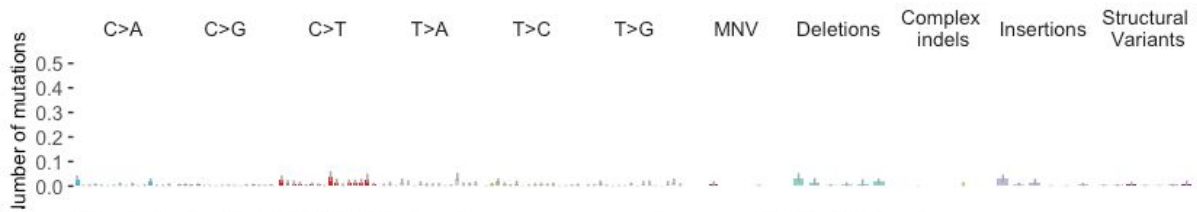
Effects for dog-1; 3.2 (1.9-5.2) het. mut-s per gen., max 20 gen. with 35 (25-41) mut-s



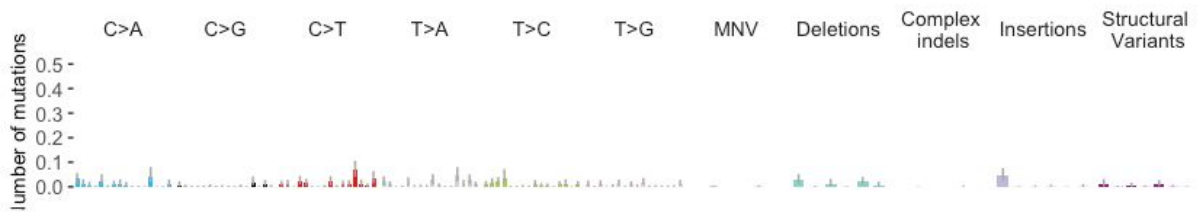
Effects for fan-1; 1.2 (0.6-2.2) het. mut-s per gen., max 20 gen. with 11 (3-22) mut-s



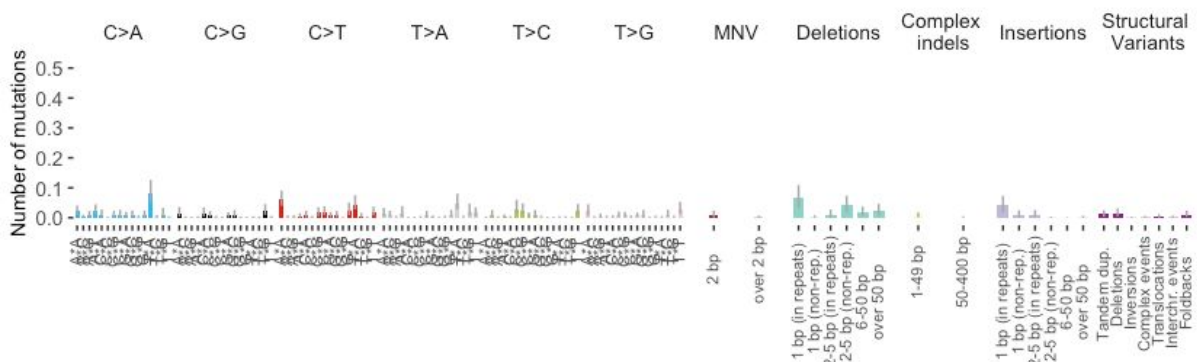
Effects for fcd-2; 0.7 (0.2-1.7) het. mut-s per gen., max 40 gen. with 11 (3-15) mut-s



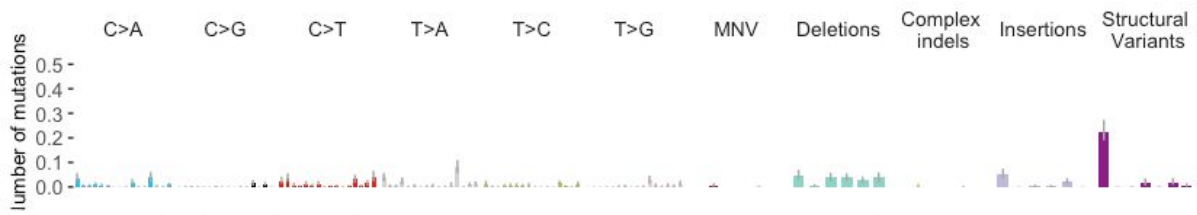
Effects for fnci-1; 1 (0.3-2.4) het. mut-s per gen., max 20 gen. with 10 (5-14) mut-s



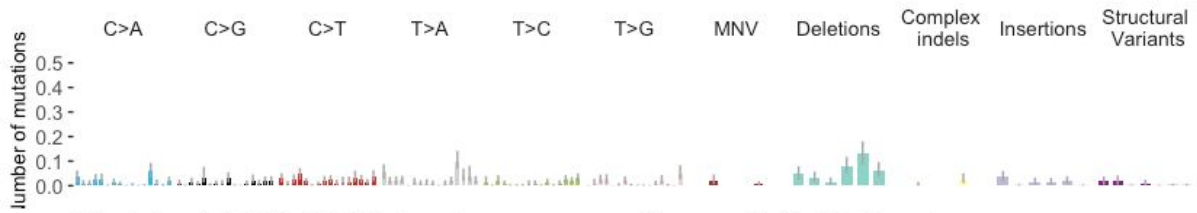
Effects for fncm-1; 1.2 (0.4-2.7) het. mut-s per gen., max 20 gen. with 12 (9-18) mut-s



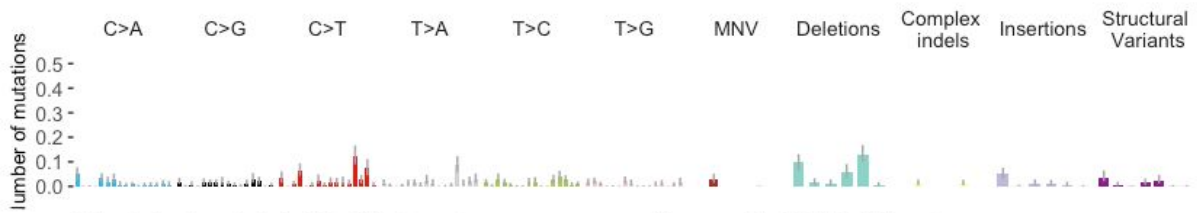
Effects for *helq-1*; 1.3 (0.7-2.2) het. mut-s per gen., max 40 gen. with 20 (16-24) mut-s



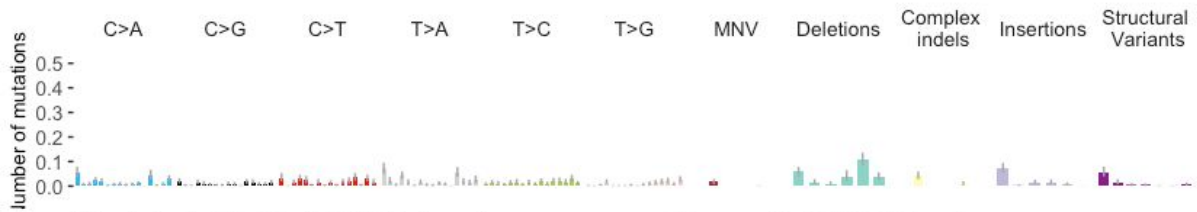
Effects for *rfs-1*; 2 (0.8-3.9) het. mut-s per gen., max 20 gen. with 22 (14-27) mut-s



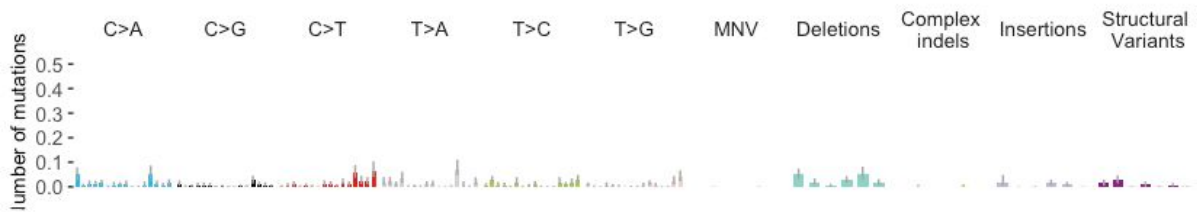
Effects for *rip-1*; 2 (0.9-3.6) het. mut-s per gen., max 40 gen. with 40 (29-46) mut-s



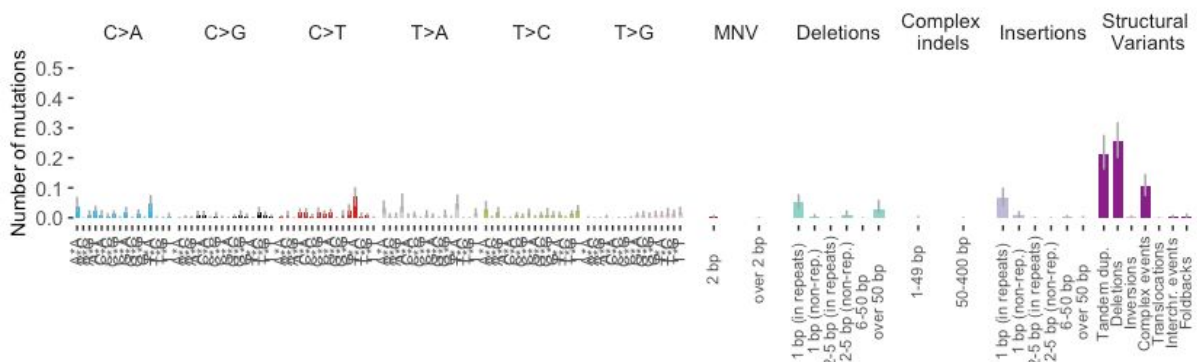
Effects for *brc-1*; 1.9 (1-3.1) het. mut-s per gen., max 40 gen. with 36 (26-45) mut-s



Effects for *brd-1*; 1.3 (0.5-2.7) het. mut-s per gen., max 40 gen. with 21 (21-21) mut-s



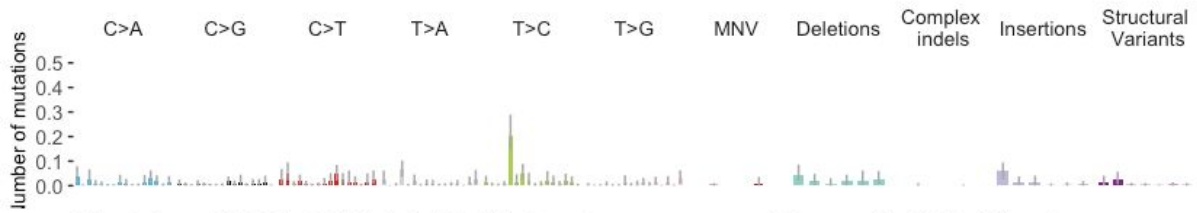
Effects for *mus-81*; 1.7 (0.8-3.2) het. mut-s per gen., max 40 gen. with 26 (26-26) mut-s



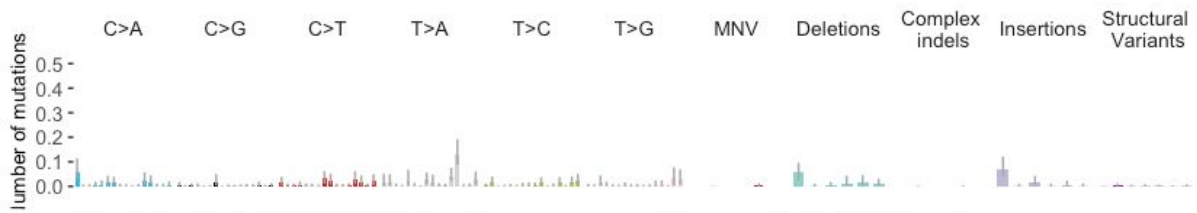
Estimated effects for rad-51; 0.9 (0.3-1.9) het. mut-s per gen., max 1 gen.



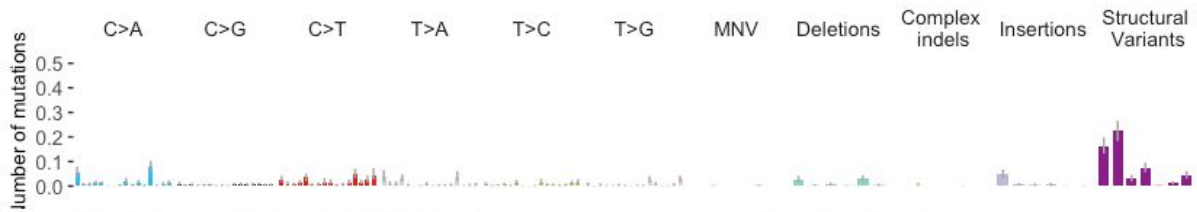
Effects for rad-54B (gt3308); 1.5 (0.4-3.9) het. mut-s per gen., max 20 gen. with 13 (11-16) mut-s



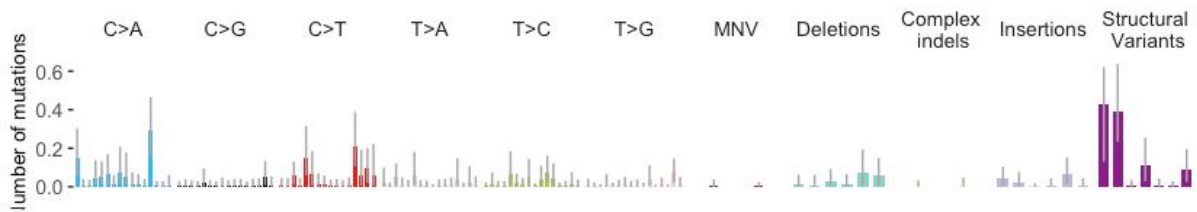
Effects for rad-54B (gt3312); 1.1 (0.3-3) het. mut-s per gen., max 20 gen. with 10 (8-11) mut-s



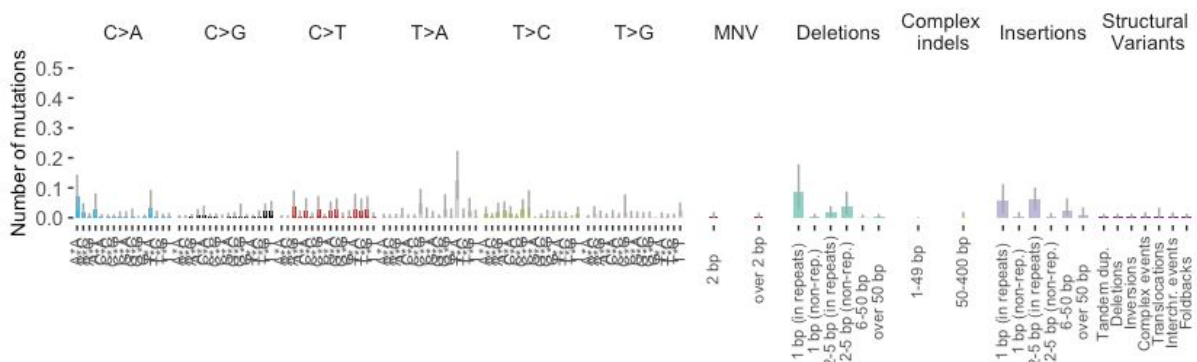
Effects for slx-1; 1.6 (0.9-2.7) het. mut-s per gen., max 40 gen. with 28 (28-29) mut-s



Effects for smc-6; 3.9 (1-10.8) het. mut-s per gen., max 5 gen. with 8 (5-14) mut-s



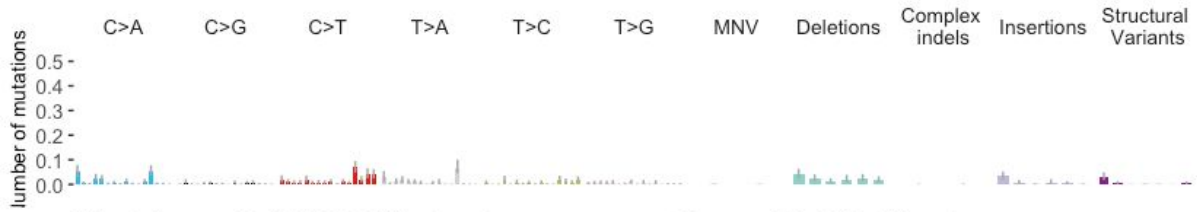
Effects for dna-2; 1.4 (0.3-4.2) het. mut-s per gen., max 20 gen. with 11 (6-15) mut-s



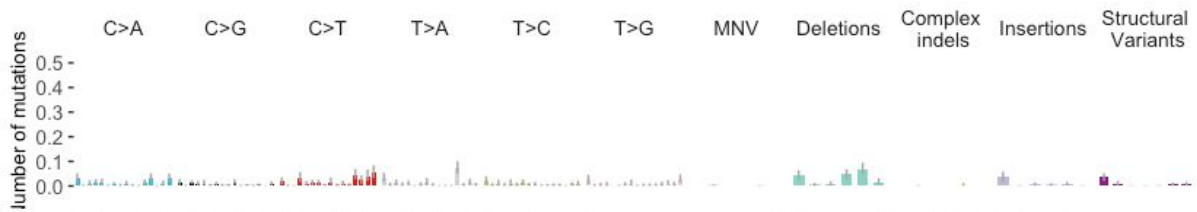
Effects for *him-6*; 1.6 (0.8-2.8) het. mut-s per gen., max 40 gen. with 20 (20-21) mut-s



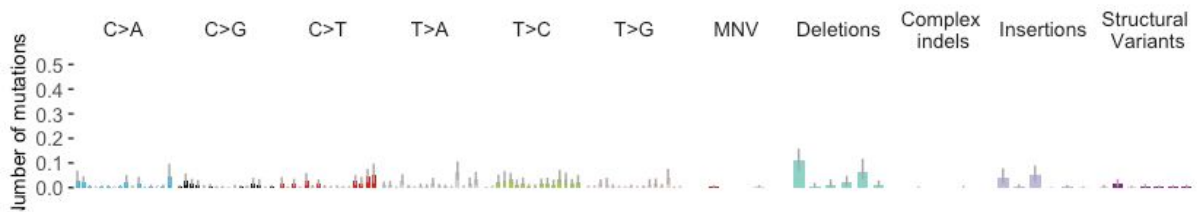
Effects for *rcq-5*; 1.1 (0.5-2.2) het. mut-s per gen., max 20 gen. with 11 (7-16) mut-s



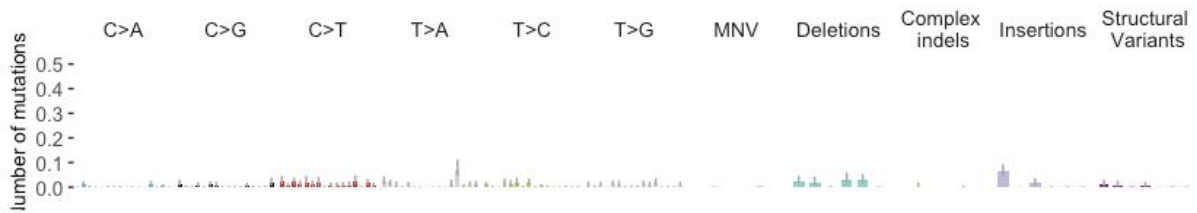
Effects for *wrn-1*; 1.3 (0.5-2.5) het. mut-s per gen., max 20 gen. with 13 (8-21) mut-s



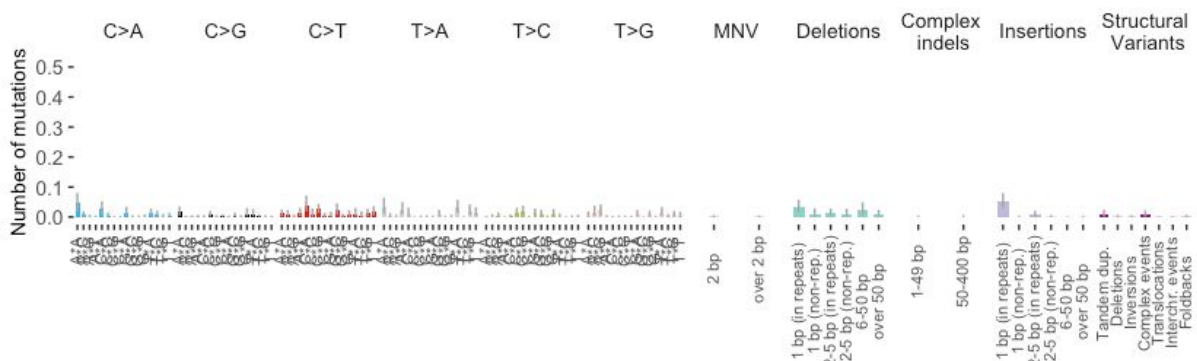
Effects for *bub-3* (gt2000); 1.4 (0.4-3.3) het. mut-s per gen., max 20 gen. with 14 (11-17) mut-s



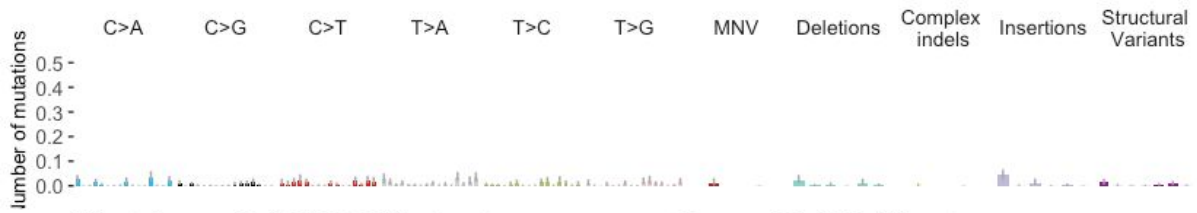
Effects for *bub-3* (ok3437); 0.9 (0.3-2.1) het. mut-s per gen., max 20 gen. with 9 (3-16) mut-s



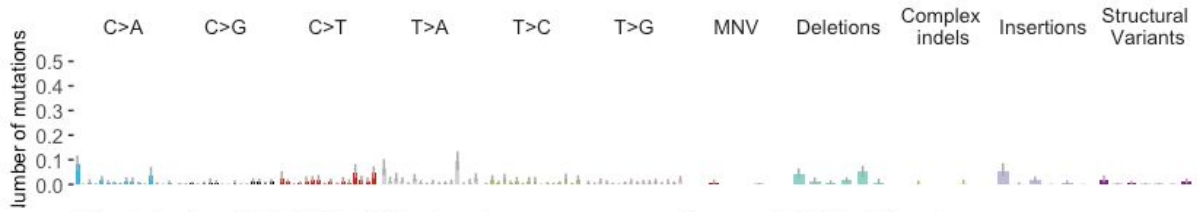
Effects for *ced-3*; 0.9 (0.2-2.3) het. mut-s per gen., max 20 gen. with 8 (6-12) mut-s



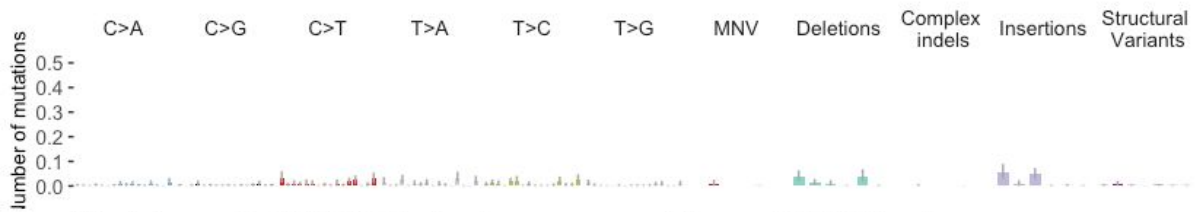
Effects for *ced-4*; 0.8 (0.3-2) het. mut-s per gen., max 20 gen. with 6 (2-10) mut-s



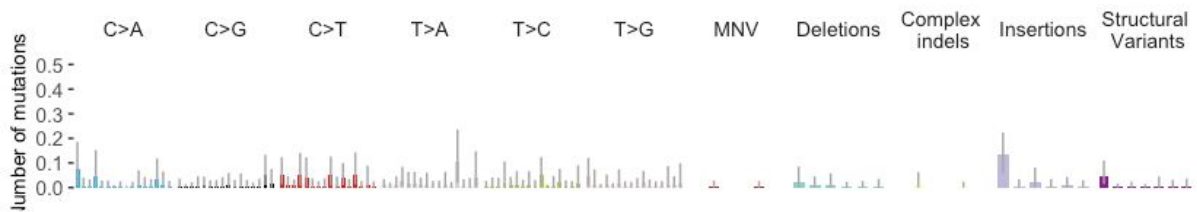
Effects for *cep-1*; 1.3 (0.5-2.8) het. mut-s per gen., max 20 gen. with 12 (8-15) mut-s



Effects for *lem-3*; 0.8 (0.3-1.9) het. mut-s per gen., max 20 gen. with 8 (3-12) mut-s



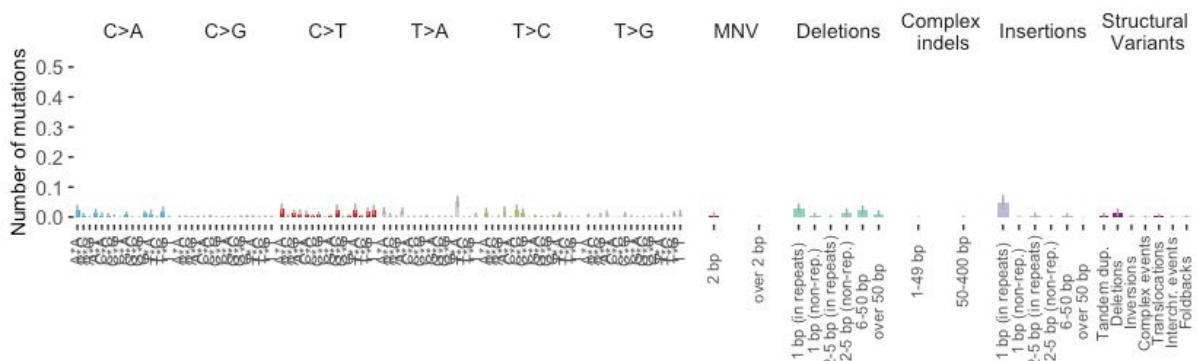
Effects for *san-1*; 1.7 (0.2-6.6) het. mut-s per gen., max 20 gen. with 9 (9-9) mut-s



Effects for *cku-80*; 1.3 (0.3-3.9) het. mut-s per gen., max 20 gen. with 11 (7-13) mut-s



Effects for *lig-4*; 0.7 (0.2-1.6) het. mut-s per gen., max 40 gen. with 12 (8-17) mut-s



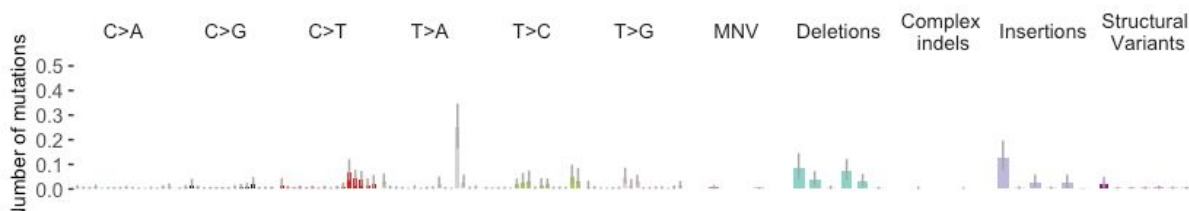
Effects for polq-1; 0.9 (0.4-1.9) het. mut-s per gen., max 40 gen. with 16 (12-21) mut-s



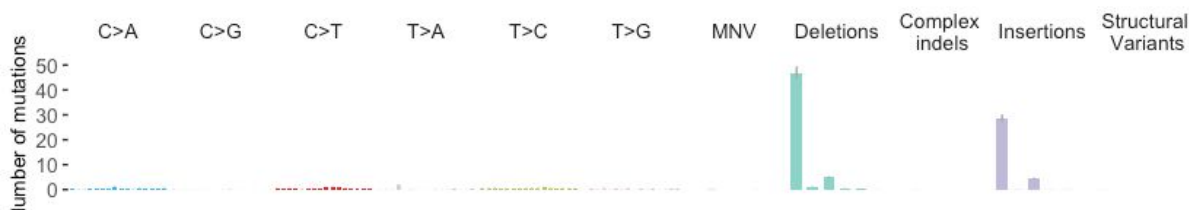
Effects for mrt-2; 2 (0.9-3.9) het. mut-s per gen., max 20 gen. with 16 (16-16) mut-s



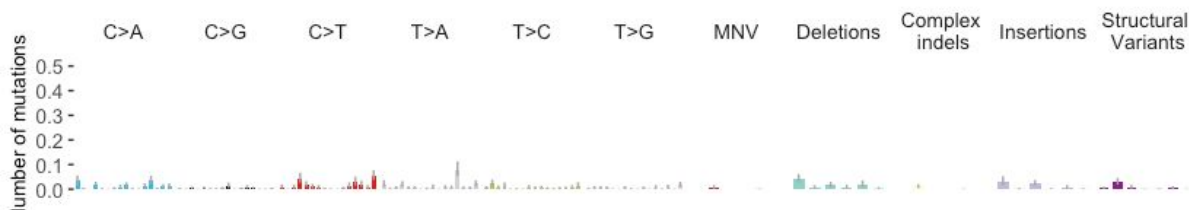
Effects for rif-1; 1.4 (0.5-3.4) het. mut-s per gen., max 20 gen. with 15 (11-19) mut-s



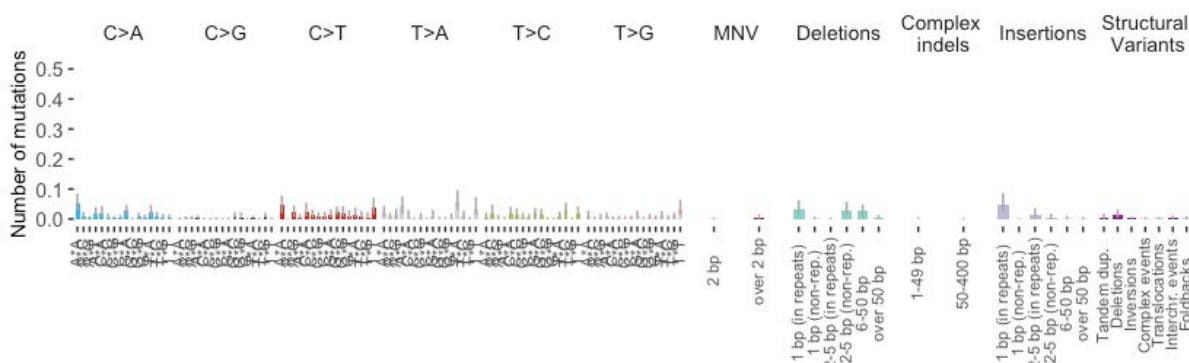
Effects for mlh-1; 114.6 (104.2-126.8) het. mut-s per gen., max 20 gen. with 1174 (1081-1236) mut-s



Effects for pole-4; 1 (0.4-2) het. mut-s per gen., max 40 gen. with 18 (17-21) mut-s



Effects for smg-1; 1.2 (0.3-2.9) het. mut-s per gen., max 20 gen. with 8 (6-11) mut-s



Supplementary Note Figure 2. Estimated interaction effects for 88 combinations of genetic backgrounds and genotoxins which showed a significant effect either on the total burden or the mutational profile in *C. elegans* experiments. 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**, while the numbers for interaction effects are provided in **Supplementary Table 3**.



Effects for agt.1.EMS



Effects for agt.1.MMS



Effects for agt.1.Radiation



Effects for agt.2.MMS



Effects for apn.1.MMS



Effects for apn.1.Radiation



Effects for *brc.1.DMS*



Effects for *brc.1.Radiation*



Effects for *brc.1.Xray*



Effects for *brd.1.DMS*



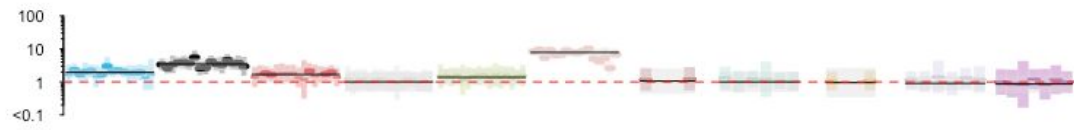
Effects for *ced.4.MMS*



Effects for *cep.1.MMS*



Effects for csb.1.Aflatoxin.B1



Effects for csb.1.DMS



Effects for exo.3.Radiation



Effects for fan.1.Cisplatin



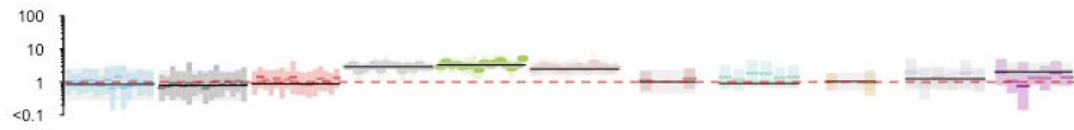
Effects for fan.1.DMS



Effects for fan.1.EMS



Effects for helq.1.DMS



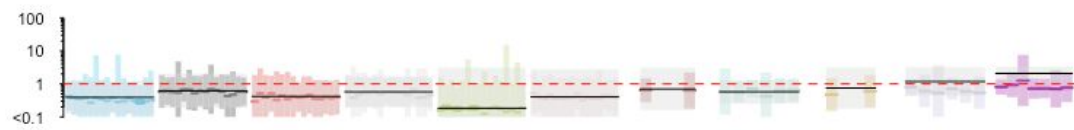
Effects for him.6.EMS



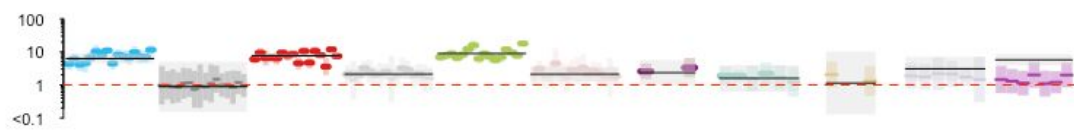
Effects for him.6.MMS



Effects for him.6.Radiation



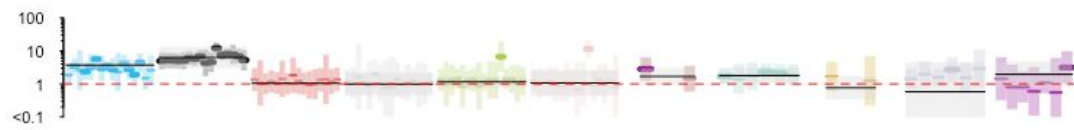
Effects for him.6.UV



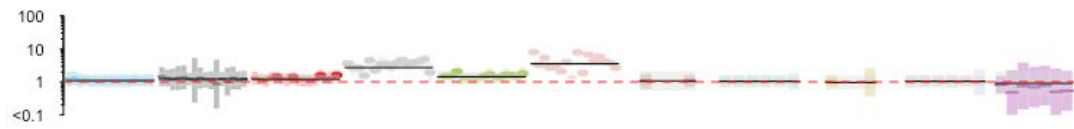
Effects for lem.3.Cisplatin



Effects for lig.4.Cisplatin



Effects for mlh.1.DMS



Effects for mlh.1.EMS



Effects for mlh.1.MMS



Effects for mrt.2.MMS



Effects for mrt.2.Radiation



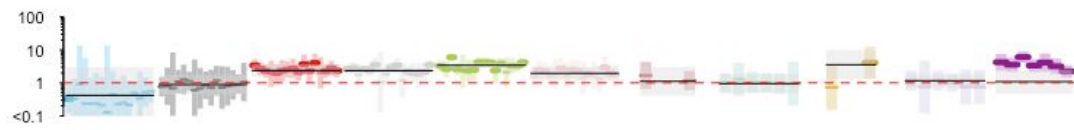
Effects for ndx.4.Radiation



Effects for parp.1.Radiation



Effects for parp.2.DMS



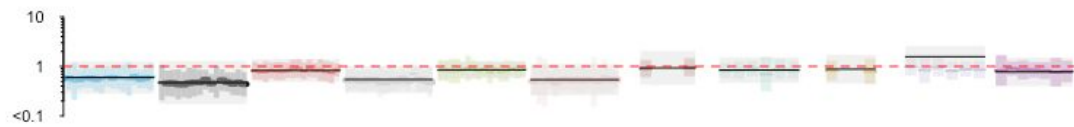
Effects for pole.4.DMS



Effects for pole.4.MMS



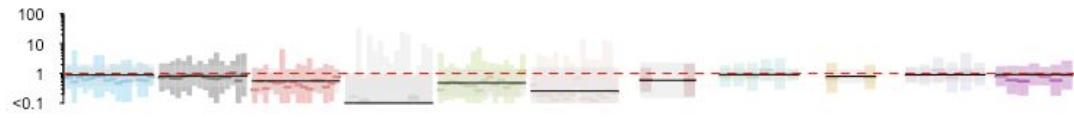
Effects for pole.4.Radiation



Effects for polh.1.Aflatoxin.B1



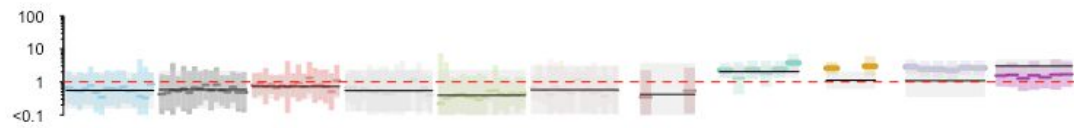
Effects for polh.1.AristolochicAcid



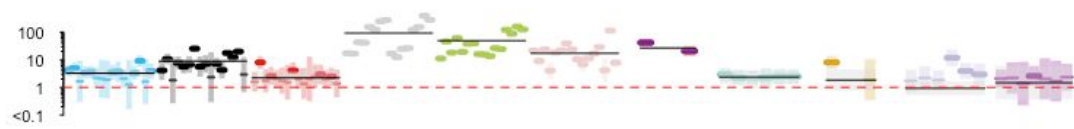
Effects for polh.1.EMS



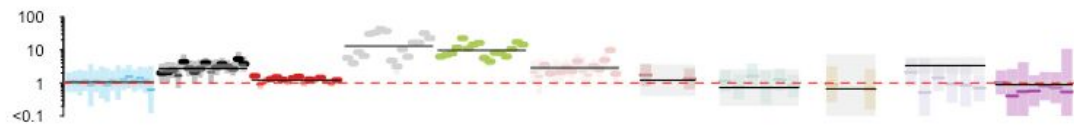
Effects for polh.1.Radiation



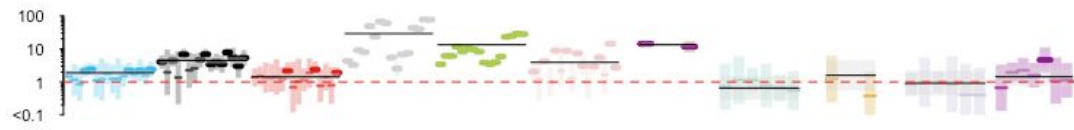
Effects for polk.1.DMS



Effects for polk.1.EMS



Effects for polk.1.MMS



Effects for polq.1.AristolochicAcid



Effects for polq.1.Cisplatin



Effects for polq.1.DMS



Effects for polq.1.EMS



Effects for polq.1.UV



Effects for rad.54B..gt3308..Radiation



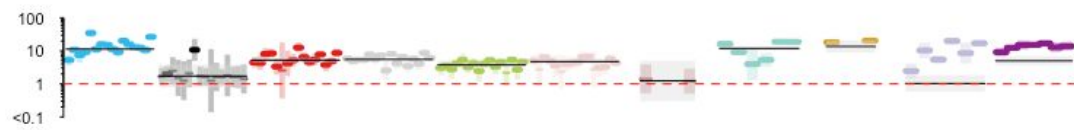
Effects for rcq.5.Radiation



Effects for rev.1.AristolochicAcid



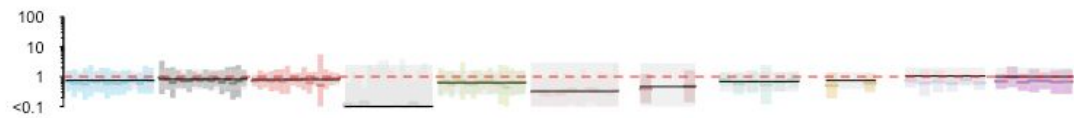
Effects for rev.1.DMS



Effects for rev.3.Aflatoxin.B1



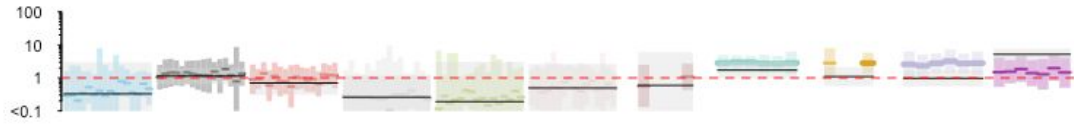
Effects for rev.3.AristolochicAcid



Effects for rev.3.Cisplatin



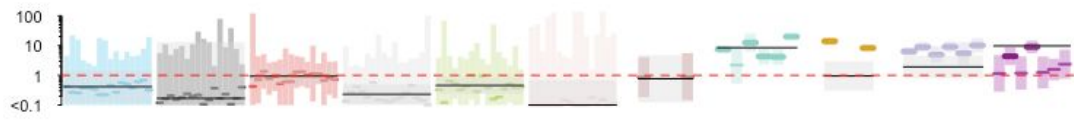
Effects for rev.3.DMS



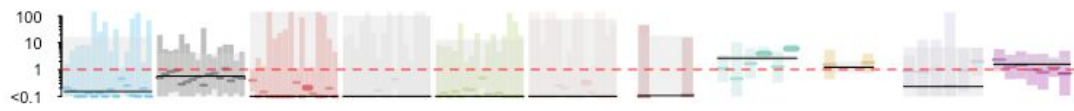
Effects for rev.3.EMS



Effects for rev.3.MMS



Effects for rev.3.UV



Effects for rip.1.Radiation



Effects for slx.1.DMS



Effects for slx.1.MMS



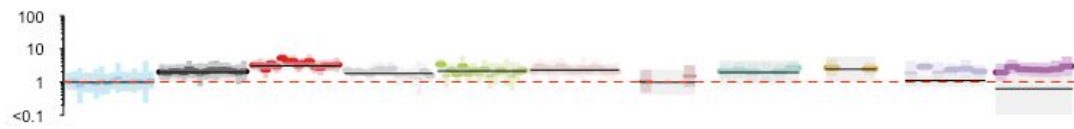
Effects for slx.1.Radiation



Effects for tdp.1.Radiation



Effects for ung.1.DMS



Effects for ung.1.EMS



Effects for ung.1.Radiation



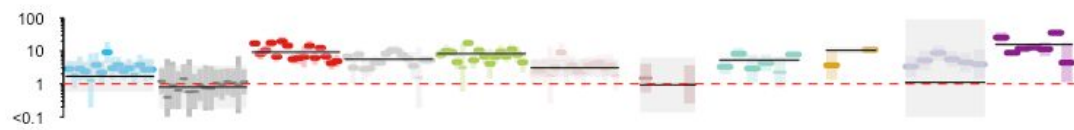
Effects for wrn.1.DMS



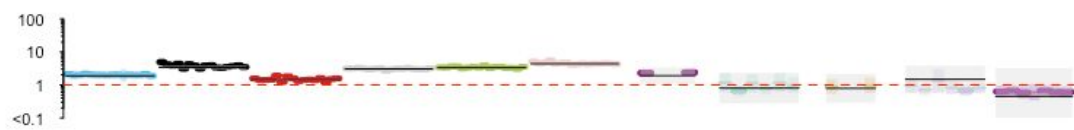
Effects for wrn.1.EMS



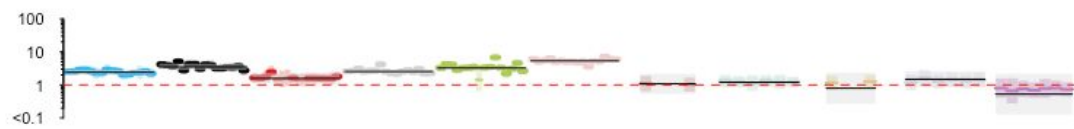
Effects for xpa.1.AristolochicAcid



Effects for xpa.1.EMS



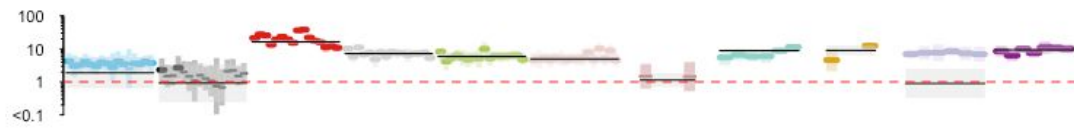
Effects for xpa.1.MMS



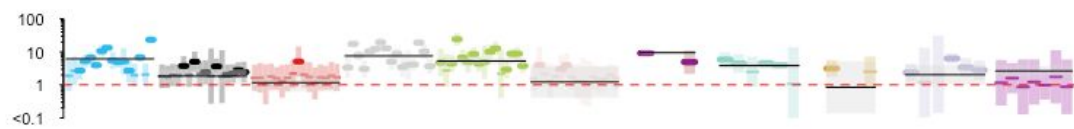
Effects for xpa.1.Radiation



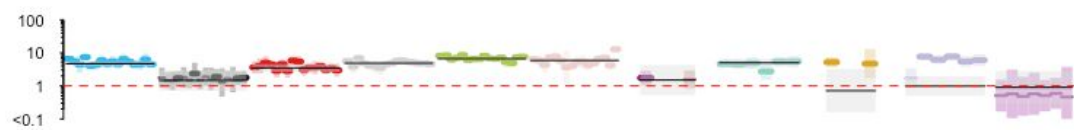
Effects for xpc.1.AristolochicAcid



Effects for xpc.1.Cisplatin



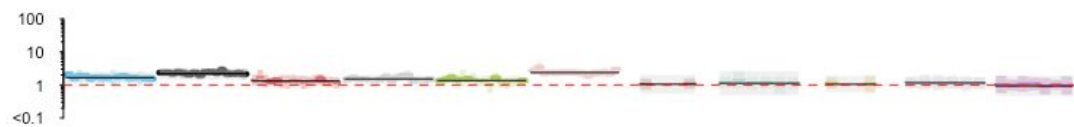
Effects for xpc.1.DMS



Effects for xpc.1.EMS



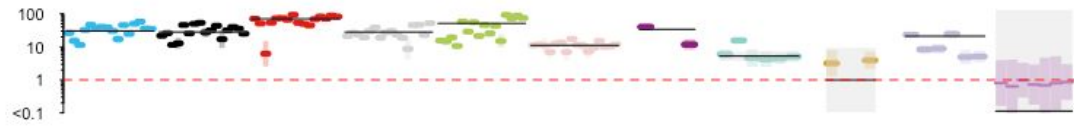
Effects for xpc.1.MMS



Effects for xpc.1.Radiation



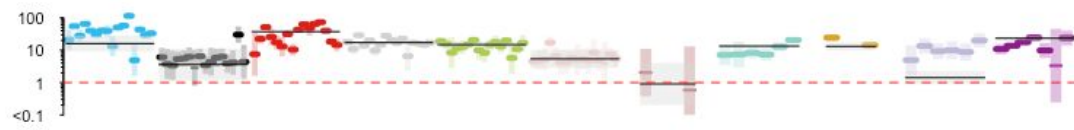
Effects for xpc.1.UV



Effects for xpf.1.Aflatoxin.B1



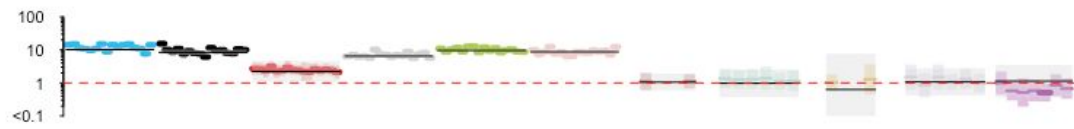
Effects for xpf.1.AristolochicAcid



Effects for xpf.1.Cisplatin



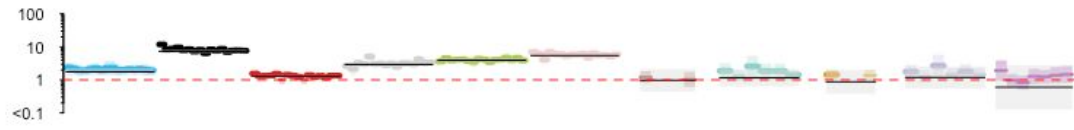
Effects for xpf.1.DMS



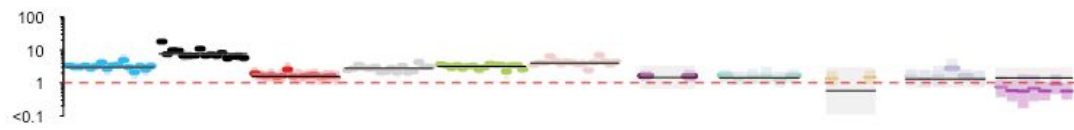
Effects for xpf.1.DMS



Effects for xpf.1.EMS



Effects for xpf.1.MMS



Effects for xpf.1.Radiation



Effects for xpf.1.UV

