Proton and alpha radiation-induced mutational profiles in human cells

Tiffany M. Delhomme[1], Manuela Buonanno[2], Veljko Grilj[2], Josep Biayna[1], Fran Supek[1,3]*

[1] Genome Data Science, Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain
[2] Radiological Research Accelerator Facility (RARAF), Columbia University, New York, USA.
[3] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain.

*correspondence to: fran.supek@irbbarcelona.org

Abstract

Ionizing radiation (IR) is known to be DNA damaging and mutagenic, however less is known about which mutational footprints result from exposures of human cells to different types of IR. We were interested in the mutagenic effects of particle radiation exposures on genomes of various human cell types, in order to gauge the genotoxic risks of space travel, and of certain types of tumor radiotherapy. To this end, we exposed cultured cell lines from the blood, breast and lung to intermittent proton and alpha particle (helium nuclei) beams at doses sufficient to affect cell survival. Whole-genome sequencing revealed that mutation rates were not overall markedly increased upon proton and alpha exposures. However, there were changes in mutation spectra and distributions, such as the increases in clustered mutations and of certain types of indels and structural variants. The spectrum of mutagenic effects of particle beams may often be cell-type and/or genetic background specific. Overall, the mutational effects of recurrent exposures to proton and alpha radiation on human cells appear subtle, however further work is warranted to understand effects of chronic, long-term exposures on various human tissues.

Introduction

Humans have a keen interest in travelling in space, from deep space exploration, to extraterrestrial colonization. In addition, the space environment was also considered for biomedical research purposes, for example the “Tumors in space” experiment aims to study the effects of microgravity and exposure to space radiation on tumor organoids, with implications to cancer risk and treatment[1]. Spaceflight, i.e. flight beyond the Earth’s atmosphere[2], is usually defined as crossing the Kármán separation line i.e. 80-100 kilometres altitude above the sea level[2–4]. Above this boundary, the environment dramatically changes, notably in terms of microgravity, temperature and radiation[5]. Studies revealed the pathological consequences on human physiology from travelling in space and reported cardiovascular, musculoskeletal and immune changes[6]. However, a better understanding of space travel impact on the human body at the cellular and sub-cellular levels is needed, for instance how the genome integrity is affected. Important effects of space travel are exerted by the various types of radiation to which tissues are exposed outside the Earth’s atmosphere and magnetosphere. Those include the solar radiation, the radioactive environment of the planet, and galactic cosmic radiation (or rays;

GCR).

GCR originates mainly from the Sun or outside of the Solar system but within the Milky Way galaxy (there exist also the extra-galactic cosmic radiations, which are less relevant as a space travel hazard due to their low flux and extremely high energies). The GCR encompasses photon-based radiation as well as particle radiations, including neutrons and charged particles; among the latter, protons and helium nuclei are the most abundant. Cosmic radiation has emerged as an issue of concern, since exposure to radiation may increase cancer risk in astronauts[7,8]. A better understanding of the mutagenic mechanisms of various cosmic radiation types upon human DNA would be helpful to understand the basis of the increased risk of cancer or of reproductive harm resulting from space travel.

Photons are the better-studied component of the GCR, since they consist of gamma rays and X-rays, whose biological impact was studied in other contexts apart from GCR. For instance, gamma rays were released during the Chernobyl nuclear power plant disaster, prompting studies of how this type of radiation impacts human genome stability. The genomic profiles of thyroid carcinomas in irradiated children post-Chernobyl contained similar driver mutations and gene fusions as in non-radiation-associated thyroid tumors[9]. However there was a radiation dose-dependant increase in the fusion driver events, as well as a general radiation-associated increase in certain types of structural variants (rearrangements)[10]. Consistent with a mutational impact that can generate somatic driver mutations, ionizing radiation-generating incidents (e.g. the Chernobyl disaster, and also the Hiroshima and Nagasaki A-bomb attacks) also increased the risk of developing cancer, proportionally to the dose of exposure[11,12].

In addition to nuclear incidents, radiation exposures can result from medical uses. Commonly this involves X-rays, used for both diagnostic purposes, and also in applications to tumor radiotherapy. The impact of therapeutic X-ray exposure on human DNA was examined via genomic analyses of radiation-associated second malignancies and of post-treatment metastatic tumors, reporting an increase of small and large deletions, and also certain mutational signatures[13,14]. For instance, this entails a particular signature of indels (the PCAWG Signature ID8) characterized by larger deletions (≥ 5-bp) without flanking micro-homology; this was previously linked to double-strand break repair[15]. Consistently with their mutagenic effect, medical use of X-rays has been reported to increase cancer risk[16] .

Regardless of the source of the radiation – either the cosmos, nuclear accidents or medical uses – it is not known if ionizing radiation consisting of charged-particle radiations, such as protons or alpha rays, would have similar mutagenic effects to more commonly studied photons (X-rays and gamma-rays). Generally, all types of ionizing radiations, including gamma-rays, X-rays, energetic charged particles (protons, alpha particles [helium nuclei] and heavier ions) and neutrons have been classified as human carcinogens by the World Health Organization (WHO)[17]. This suggests mutagenic impacts of various radiation types, however research on the specific mutational footprints of charged particles on the integrity of the human genome has been limited.

These are of interest because during transit beyond low Earth orbit, every cell nucleus within an astronaut's body is traversed, on average, by an energetic proton every few days, and by a helium nucleus once every few weeks. Mutational effect of proton radiation is of interest also for reasons other than space travel. This type of radiation is increasingly adopted in the clinic to treat cancer[18] , and appears promising for toxicity profiles[19–22]. Proton therapy may reduce the health risks (compared to X-rays) of secondary malignancies[23], and of severe radiation-induced lymphopenia[24,25]. The above provides motivation for studying the mutagenic effects of proton, and of helium nuclei (alpha) radiation exposures, which have the potential to result in long term health effects in exposed individuals.

The activity of mutational processes in DNA, which may arise from both endogenous factors such as DNA repair deficiencies and exogenous factors such as tobacco smoking, can be captured by mutational signatures[26], mathematical constructs that describe the differential frequencies of mutation types across many genomes. Such signatures are commonly based on trinucleotide spectra of single nucleotide variants (SNVs), however they can also be based on indels, structural variation (rearrangements), or mutation clusters[15,27,28]. SNV and indel mutational signatures in human cancer are well characterized and organized in comprehensive catalogues[15,29], including also signatures in tumors pre-treated with ionizing radiation[13,14,30] (here referring to photons).

In addition, some genomic studies have reported mutational signatures on experimental models, including experimentally-induced mutational signatures of (non-ionizing) UV radiation in human cell lines[31], and also in cultured cells from healthy human tissues[32,33]. Additionally, signatures of ionizing radiations were reported in exposed mice[34,35] and in the worm *C. elegans*[36]; both reported SNV signatures enriched in C>T transitions[15].

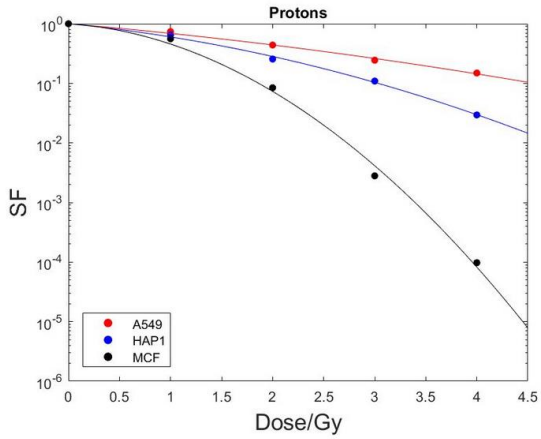
In order to increase our understanding of the potential impact of particle radiation encountered during space travel on human tissues, it is important to systematically, in a genome-wide manner examine the DNA mutations arising from the exposure in different human tissues. To this end, we analyse genomes three human cancer cell lines, A549 (lung adenocarcinoma, epithelial cells), HAP1 (chronic myelogenous leukaemia, blood cells) and MCF7 (breast carcinoma, epithelial cells). Those cell lines were irradiated with two types of particle radiations – proton and helium fluxes – in order to identify the mutational footprints induced and their potential cell type-specificity.

Results

Determining proton and helium ion fraction, and irradiating the cell lines

The dosage of radiation was determined experimentally in order to achieve between 40% and 50% lethality (corresponding to 50%-60% clonogenic survival), independently across the two types of cosmic rays (Figure 1A-B).

A

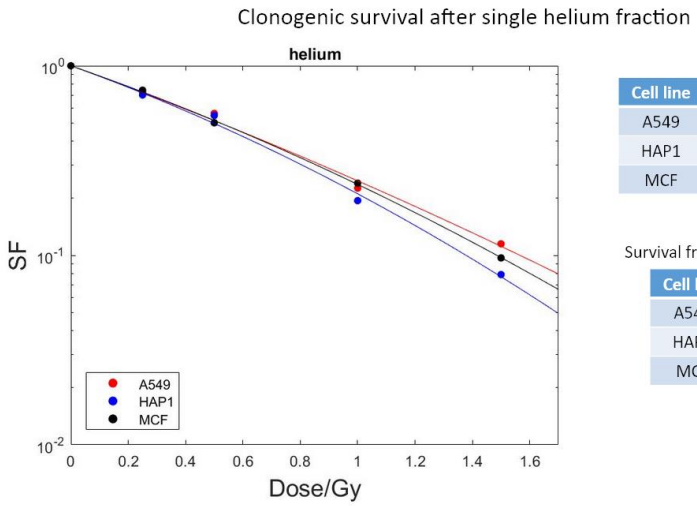


Cell line	alpha	beta
A549	0.34	0.04
HAP1	0.39	0.12
MCF	0.26	0.52

Survival fraction after 1 Gy:

Cell line	SF
A549	0.69
HAP1	0.60
MCF	0.46

B



Cell line	alpha	beta
A549	1.27	0.13
HAP1	1.25	0.31
MCF	1.22	0.22

Survival fraction after 0.5 Gy:

Cell line	SF
A549	0.51
HAP1	0.50
MCF	0.51

C

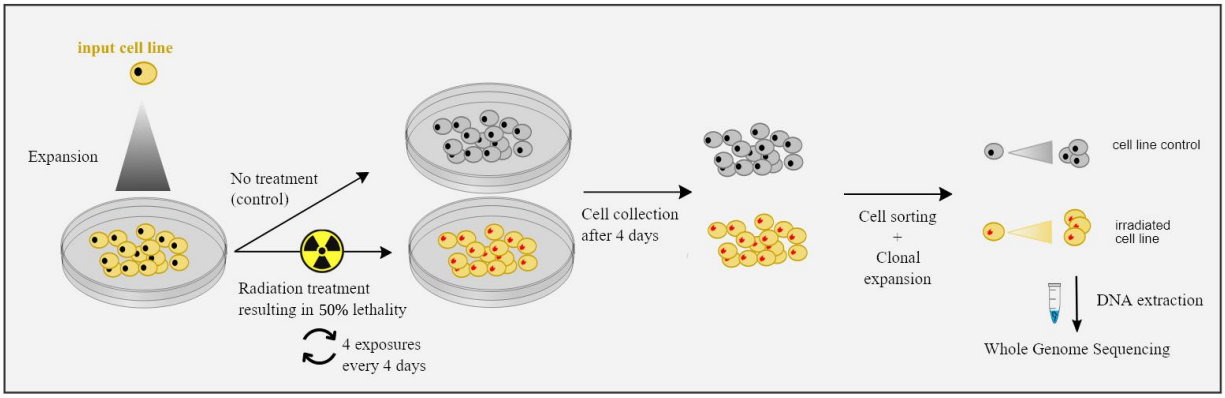


Figure 1. Overview of the experimental design. A-B. Experimental identification of radiation dose in order to achieve ~50% of cell lethality (i.e. ~50% of cell survival) independently across the three cell lines and across the two types of radiations. SF, survival fraction. C: Schematic overview of the experimental design of the study. A549, MCF7 and HAP-1 cell lines were exposed to GCR (protons or helium ions). GCR: Galactic Cosmic Radiation. Single cells from treated and untreated conditions were separated by FACS and clonally expanded. DNA from single-cell derived populations was extracted and subjected to whole genome sequencing.

This led to a dose of 0.5 Gy for helium fluxes and a dose of 1 Gy for proton fluxes. After an expansion of the three cell lines collected (A549, HAP1 and MCF7), the cells were irradiated using the 5.5 MV Singletron accelerator at the Radiological Research Accelerator Facility (RARAF; see Methods for details about the procedure, with the corresponding dosage). We exposed each cell line four times every 4 days to allow cell recovery between each exposure to facilitate mutation accumulation.

After the final latency period of 4 days, the cells were collected, sorted in a 96-well plate, and a clonal expansion was performed to obtain colonies with a near-identical genome to facilitate identifying mutations (Figure 1C). Finally, the DNA was extracted from those colonies and whole-genome sequenced (WGS) on an Illumina NovaSeq 6000 machine. Details of the following bioinformatics analysis are described in Methods.

Overall burden of various mutation types generated by cosmic radiations

After the GCR treatments, the cell line genomes contained newly-acquired point mutations (single nucleotide variants, SNVs), short insertions and deletions (indels), and structural variants (SVs) in all three cell lines assayed, HAP1, A549 and MCF7, and regardless of the radiation type, protons or helium ions. As a control, we considered sham-irradiated (i.e. untreated) cells for HAP1 and A549 cell lines, which also accumulated certain numbers of SNVs, small indels and SVs, consistent with them undergoing cell divisions and also potentially being exposed to various stresses during handling.

The global number of observed mutation events (SNVs, indels and SVs) does not markedly differ between the conditions, suggesting that proton and alpha radiation exposure is not highly mutagenic to cultured human cells using the fractionated dosing regimen applied herein. Nevertheless, we detected roughly double the amount of SNVs in the MCF7 cell line than in the two other cell lines after exposure to either particle (with the exception of one helium clone, Figure 2A).

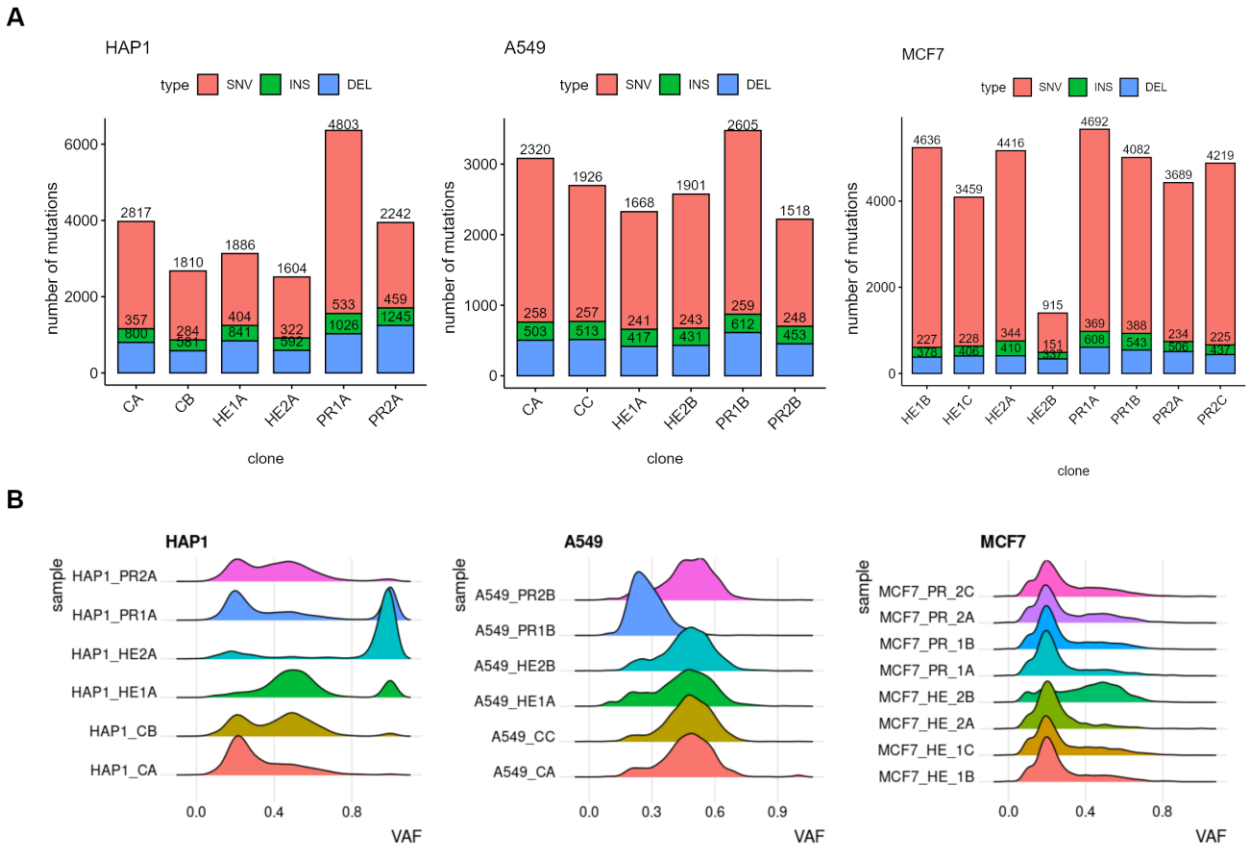


Figure 2. General description of point mutations detected in the clones. A. Distribution of the number of SNVs, small insertions and small deletions per clone in each of the three analysed cell lines. Proton radiation generated the highest number of point mutations and small deletions. In HAP1 cell line, this treatment generated also more insertions compared to the other conditions. B. Distribution of the variant allelic fraction (VAF) of detected point mutations for each clone of each cell lines. A VAF peak around 50% or 100% would correspond to a selection of precisely one cell from the cell culture, whereas a peak close de 25% would correspond to the isolation of two cells.

This may indicate a mutator phenotype in our MCF7 cell line, possibly due to DNA mismatch repair (MMR) failure (see below). However this is difficult to ascertain since the absolute number of mutations between cell lines and also between conditions cannot be compared precisely, due to a potentially variable number of cell cycles the cells have undergone.

In all irradiated cell lines, the proton treatment generated both the highest number of SNVs (around 4.8 thousands) and the highest number of small deletions (around 500), compared to the other conditions (Figure 2A). An enrichment of deletions, as we observed in the proton-treated clones, has recently been highlighted in ionizing radiation-associated tumors[13,14] which are typically treated with X-rays. The globally most-mutated clone is a single proton-irradiated sample of the HAP1 cell line, which contains 1,986 more SNVs (Figure 2A) than the most-mutated control (non-irradiated) HAP1 sample. Proton flux generated also the highest number of insertions in the HAP1 cell line (~500 in both replicates), whereas, for this type of alterations, there was not a clear impact of the proton treatment in the A549 and MCF7 cell lines

(Figure 2A). This suggests that mutagenic impact of proton irradiation on SNVs, insertions and deletions may be specific to cell-type and/or the genetic background.

Variant allelic fractions (VAFs), a metric corresponding to the proportion of sequencing reads that contain a particular variant, is a proxy for the proportion of cells in the sequenced population that harbour the mutation. In our experiments, we aimed to select a single treated cell and expand it into a clone, thus it is expected that the VAFs observed should be centered at 0.5 for heterozygous diploid genome segments or at 1 to homozygous mutations e.g. in LOH segments (or also genome-wide in the haploid cell line HAP1; note that this cell line can grow as haploid or diploid and may switch spontaneously). In practice, three different types of VAF peaks were observed, one at 0.25 which suggests that two cells were collected at the time of bottlenecking (or alternatively, that the genome segment may be tetraploid), one peak at 0.5 which suggests that one cell was collected, and one peak at 1 for the HAP1 cell line, as expected due to the possible haploid state of that cell line (Figure 2B).

Identification of SNV trinucleotide mutational signatures

Upon classifying SNVs into the 6 different categories, tallying DNA strand-symmetrically, i.e. C>A, C>G, C>T, T>A, T>C, T>G, the most prevalent class of SNVs in all cell lines/clones was C>A. One important cause of C>A transversion mutations in cancer genomes is oxidative damage to guanine (Signatures SBS18 and SBS36 as reported previously[26]). Moreover, an abundance of C>A mutations was observed in recent mutation-accumulation experiments on human cell lines similar to ours[31,37], plausibly due to exposure to atmospheric oxygen during cell culture.

Additionally, we also observed an enrichment of the C>T and T>C transition mutation types (Figure 3A), which is consistent with commonly observed mutational signatures in dividing cells, namely the ‘clock-like’ mutational signatures SBS1 and SBS5/SBS40[15].

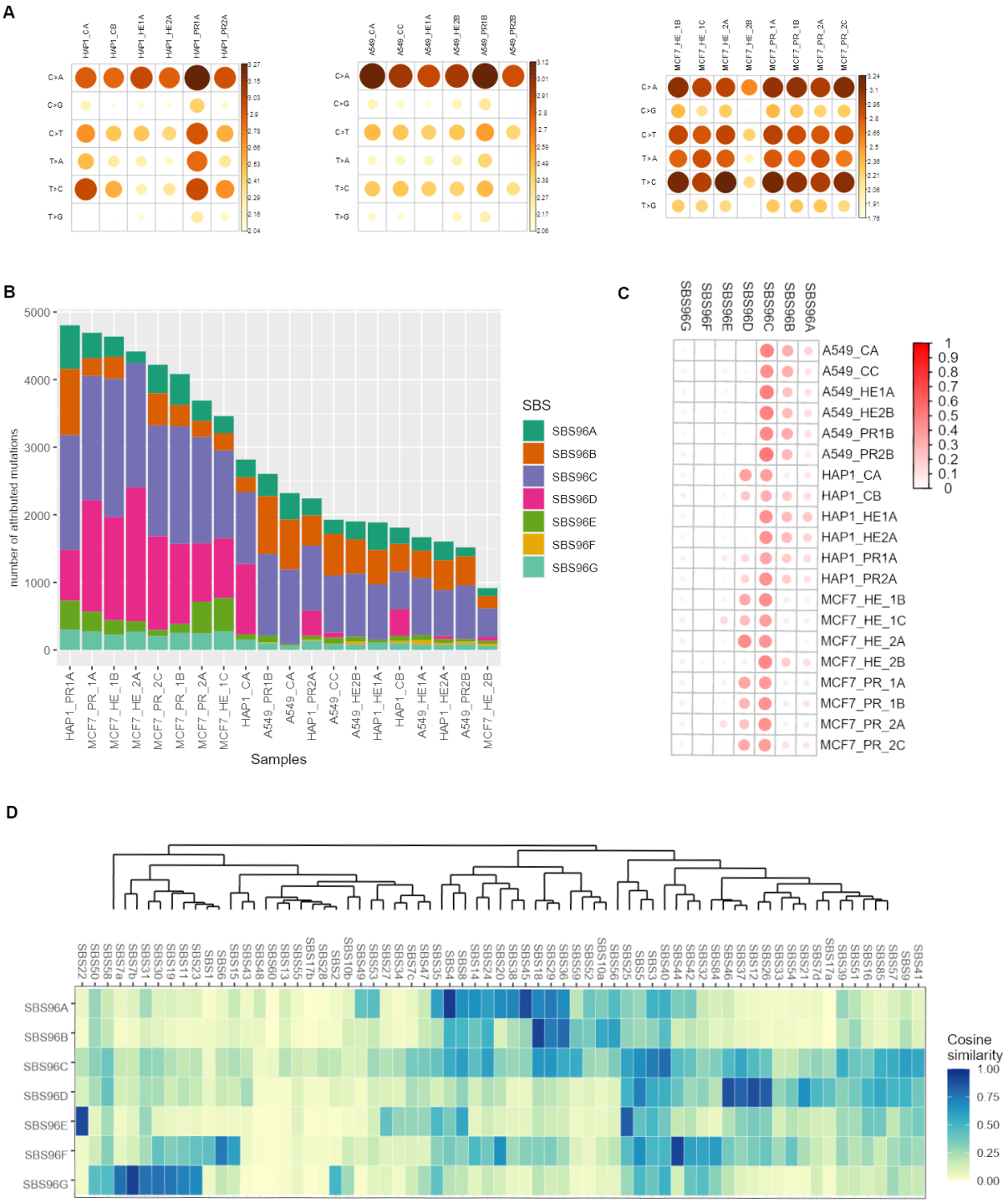


Figure 3. Mutational signatures of SNVs in irradiated human cell lines. A. Break-down of point mutations into the 6 main type of mutations (*i.e.* C>A, C>G, C>T, T>A, T>C, T>G), per clone and per cell line. Values are shown $\log_{10}$ transformed. All genomes were enriched for C>A (possibly due to oxidative damage to DNA), and C>T and T>C transitions (likely due to 'clock-like' mutational processes). B. Number of point mutations attributed to each of the 7 different extracted NMF mutational signatures, for each sample. A signature was considered to be active in a sample if the proportion of total number of mutations attributed to this particular signature is at least 10%. The SBS96A, SBS96B and SBS96C (all

likely related to experimental procedures) were found in the majority of the samples. SBS96D (consistent with an MMR deficiency) was found in many MCF7 clones, but also in the control samples of the HAP1 cells. One particular signature, the SBS96E, a signature enriched in T>A transversions, was found in MCF7 but was not active in non-radiated samples (of other cell line; MCF7 did not have a corresponding unirradiated sample available). C. Activity matrix of extracted mutational signatures in various clones herein. D. Cosine similarities between PCAWG mutational signatures and our 7 extracted SBS signatures. The SigsPack R package was used to obtain the mutational profiles of PCAWG signatures.

To further classify SNV mutation types, as is usually performed, the mutations were classified into 96 categories, taking into account both the observed class of mutation and the trinucleotide context (A_A, A_C, ..., T_T), in order to infer the 96-component spectrum of SNVs for each cell line and each clone.

We extracted *de* novo mutational signatures from our irradiated and control clones via the SigProfiler tool[15]. To better identify similarities with known mutational processes from other controlled experiments on human cell line models, our mutation data from proton and alpha particle irradiated clones was pooled with previous genomic data: (i) 54 cell line genomes treated with various mutagens, largely chemicals, in which a specific SNV signature was able to be identified[31], and (ii) additional 173 cell line genomes generated from CRISPR-Cas9 gene knockouts of 43 DNA repair genes[37]. By a joint analysis of our and previous cell line mutation accumulation data, we identified 7 independent mutational signatures (see Methods) based on SNVs (here called SBS [single base substitution]) (Supp. Figure 2). In order to attribute a potential aetiology to each signature, we computed the similarity with the catalogue of known somatic signatures, referred as PCAWG signatures[15] (Figure 3C; of note, these signatures are often referred to in the literature as “COSMIC”, by the name of the database, however this is unrelated with “cosmic radiation” and this previous catalogue does not contain known signatures of cosmic radiation). We further computed the similarity to the signatures of deficient DNA repair or of exposure to mutagenic chemicals from two previous studies[31,37].

The SBS96A was assigned to (Figure 3D) both PCAWG SBS4 (a tobacco smoking-associated signature), SBS45 (C>A-rich signature, likely an experimental artefact due to 8-oxoG generated during preparation of DNA for sequencing[38]), and also to experimental exposure to PAHs (polycyclic aromatic hydrocarbons, a component of tobacco smoke) in cell lines[31]. The SBS96B was assigned to the PCAWG signature SBS18, which was reported to result from a damage to guanines by reactive oxygen species[15,39,40]. This signature was also found to be significantly active in a *OGG1* knockout cell line genome (Supp. Figure 1A), consistent with the known function of base excision repair protein OGG1 in mending nucleobase oxidative damage[41]. The SBS96C was not assigned to a known reported signature in PCAWG catalog (Figure 3D), but 45% of mutations in the previous *EXO1*-/- cell line WGS[37] was attributed to this mutational signature. EXO1 is a nuclease with various roles in the MMR, translesion synthesis, and homologous recombination repair pathway (reviewed in ref. [42]), suggesting that a defect in one or more of the DNA repair pathways generates the signature SBS96C. The SBS96D was assigned to a set of four PCAWG signatures: SBS46, SBS37, SBS12, SBS26. Of those, SBS26 is known to be associated with defective DNA mismatch repair (MMR), and is dominated by C>T transitions as the three other signatures, suggesting that SBS96D may be signature of

MMR deficiency. Consistently, SBS96D was the most active signature in the four *PMS2*-/- cell lines that were available[37], suggesting that a particular type of MMR deficiency may be generating this mutational footprint. The SBS96E was assigned to the PCAWG signature SBS22, resulting from an exposure to aristolochic acid, an agent generating bulky nucleotide adducts and consisting of T>A transversions, and SBS25, associated with a chemotherapy treatment[43,44], again rich in T>A changes. The SBS96F was assigned to the PCAWG signature SBS44 and to the *MLH1* gene knockout[37], indicating another signature of a defective DNA mismatch repair, thus mechanistically related to but distinct from SBS96D mentioned above. Finally, the SBS96G was assigned both to the known mutational signature associated with ultraviolet light exposure (SBS7b), and also to a simulated solar (i.e. non-ionizing, UV-containing) radiation treatment of a cell line[31].

SNV mutational signatures predominantly active in certain samples

To identify the mutational processes operative in our cell line samples, we considered a signature to be active in a particular sample if the proportion of mutations attributed to that signature is at least 10% of the total amount of mutations. The signatures SBS96A, SBS96B and SBS96C were found in all the three cell lines and in the majority of the samples of each line and were also abundant in the non-irradiated control samples (Figure 3C). This means that signatures of adducts at C:G pairs (SBS96A), of oxygen reactive damage (SBS96B), and possibly deficient various repair pathways (see above for SBS96C) are not a result of the effects of radiation, because there were detected also in the non-irradiated clones at similar proportions. The signature of DNA oxidation reaction, SBS96B, has a higher activity in the A549 lung cancer cell line (Figure 3C), suggesting either a more rapid formation of reactive oxygen species in that cell line, and/or a higher rate of resulting DNA damage, and/or a lower rate (or accuracy) of repair of oxidized DNA in that cell line.

The signature associated with MMR deficiency via the *PMS2* gene and/or SBS26 (SBS96D) was found in almost all samples of the MCF7 cell line (7/8 samples), and also in the two non-irradiated clones from the HAP1 cell line. This suggests that a form of MMR deficiency was probably acquired during the cell growth of HAP1, and is present in all our MCF7 clones, and so this signature did not result from radiation (we note there is no available MCF7 non-irradiated WGS data to compare). The second, “canonical” MMR-associated signature that associated to the *MLH1* gene loss, the SBS96F, was not active in our samples; instead, it was seen only in the external data sets we integrated. Thus GCR exposure does not commonly select for cells with a deficient MMR system, as is the case for some chemical exposures that generate mismatch-like DNA lesions, such as drugs TMZ or 6-TG; reviewed in ref. [45]. Overall, signatures of deficient MMR in our data appear unrelated with alpha or proton radiation exposure.

The mutational signature of the non-ionizing UV radiation SBS96G, was, expectedly, not detected at notable levels in our clones. The minor fraction (around ~5%) of observed mutations attributed to this process was similarly active in the non-irradiated conditions, and so is not linked with ionizing radiation, but is more likely to result from imprecisions in the mutational

signature inference methodology.

One particular signature was associated with ionizing radiation exposure, specifically, the T>A transversion-rich signature SBS96E (Figure 3B-C). This was detected in MCF7 irradiated samples, both in helium-treated and proton-treated clones, without being active in the non-irradiated samples of various cell lines (note that non-irradiated MCF7 WGS was not available). This observation suggests two novel insights. Firstly, the SNV mutational consequences of particle radiations are likely to be tissue-specific and/or genetic background specific, since they were observed consistently across clones in one of the three cell lines (breast), but not in the other two (lung or blood). Secondly, the proton/helium radiation effect can mimic the mutational footprint of mutagens generating bulky adducts at A:T nucleotide pairs, similar to aristolochic acid or certain chemotherapies. Speculatively, such a bulky DNA lesion might be generated by cross-linking nucleotides, or by cross-linking DNA with proteins. Since this mutational signature of point mutations was found in both types of radiation we tested, it does appear broadly relevant to various particle radiation types present in GCR. Moreover it is expected to result from other sources of alpha or proton radiation (e.g. in cancer radiotherapy) in the susceptible cell types and/or genetic backgrounds.

Mutational signatures of indels across cell types and conditions

Small insertions and deletions (indels) can be classified into 83 types (the ID83 categorization, based on PCAWG indel signatures[15]). In brief, the indels are separated into 1-bp or longer indels, insertions or deletions, at repeats or not, and finally deletions with micro-homology are counted separately; these classes can further be divided into subclasses based on the gained/lost nucleotide, or based on the length of the alteration. Each observed indel is assigned to one category, thus providing the mutational spectrum of indels for each cell line and each condition (Figure 4A).

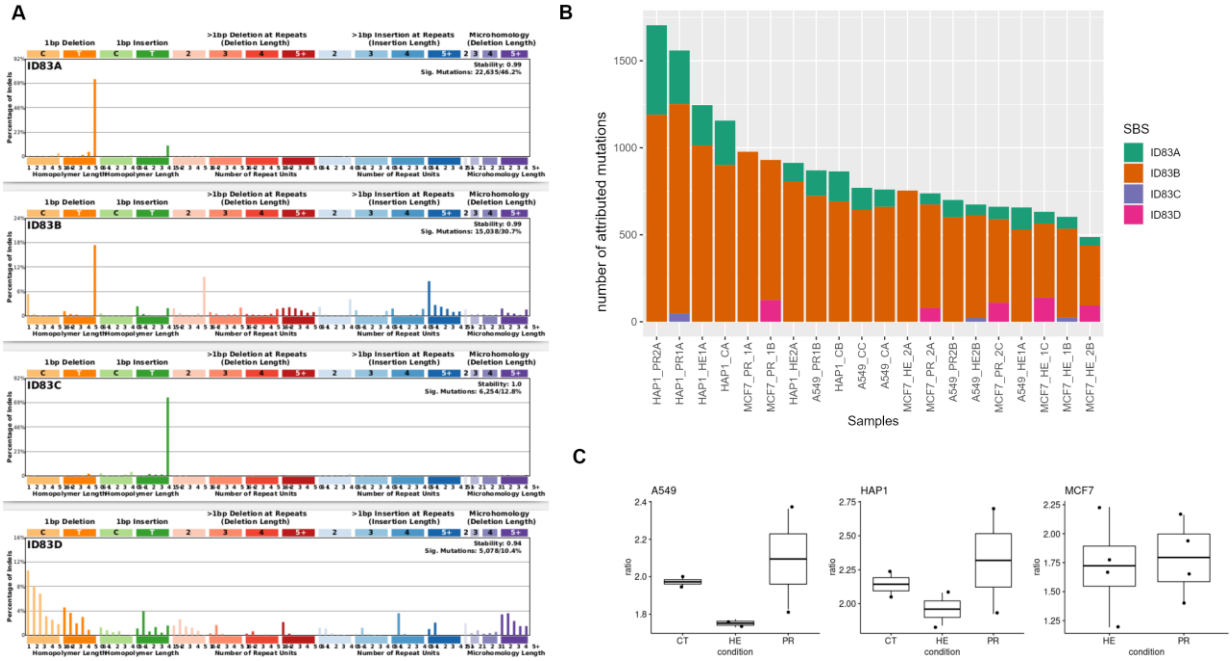


Figure 4. Mutational signatures of small indels in irradiated human cell lines. A. Mutational profiles of the 4 indel signatures extracted from pooled data (from our data and from the two other cell line studies that reported DNA repair deficient background and mutagenic chemical exposed cells[31,37]) using the SigProfiler algorithm. Indels were classified into 83 distinct categories following the procedure used in Alexandrov *et al.*[15]. B. Number of indels attributed to each of the 7 different extracted signatures, for each sample. ID83A and ID83B, characterized by DNA replication errors are found in all the clones. ID83C and ID83D are unique to radiated samples. C. Ratio of number of deletions over number of insertions.

In order to estimate the potential indel signatures as footprints of the helium and alpha radiation treatments, we ran the SigProfiler extractor tool on the generated indel count matrices. The matrices were from the pooled dataset, consisting as above of genomes from our irradiated MCF7, A549 and HAP1 cell lines, as well as genomes of mutagen-treated and DNA repair deficient cell lines from previous work[31,37] (see Methods for details on samples).

Four indel signatures were identified: ID83A, ID83B, ID83C and ID83D (Figure 4A), as well as the absolute activity of each in our experimental cell line samples (Figure 4B). To learn about the potential underlying mechanisms, we also estimated the signatures' activity in WGS of previous DNA repair knockouts and the mutagen-treated samples[31,37], as for SNV signatures above. ID83A was linked to DNA mismatch repair deficiency (*MSH2*, *MLH1* and *MSH6*). ID83B was found in DNA polymerase knockout samples (*POLB*, *POLM* and *POLE4* [37], see Supp. Figure 1B). The ID83C may be linked to various mechanisms (Supp. Figure 1B), and its spectrum resembles the PCAWG ID1, suggested to be related to slippage during DNA replication[15].

Finally, the ID83D signature, consisting mostly of 1-bp deletions and ≥5bp deletions at micro-homology regions, was found to be active in a cell line knocked out in *XRCC4*[37], a core

factor of the non-homologous end joining (NHEJ) pathway for DSB repair (Supp. Figure 1B). A loss of this gene (whose full name is “X-Ray Repair Cross Complementing 4”) is well-appreciated to increase sensitivity to ionizing radiation[46–48]. Moreover, this indel signature is similar to the indel mutational signature provoked by ionizing photon radiation, reported in a recent preprint, where clonal organoids from mouse and human cells were irradiated[49].

ID83A and ID83B were found in majority of our cell lines/conditions including controls, suggesting they may represent errors intrinsic to DNA replication. These errors may be enhanced in case of MMR failures, and thus observed more abundantly in MMR mutants (see above).

The signatures ID83C and ID83D were found in only a subset of our samples and more interestingly they were absent in the controls. ID83C was active in HAP1/proton, A549/helium and MCF7/helium conditions (Figure 4B), suggesting that radiation may cause a mutational footprint resembling that of strand slippage during DNA replication. In our set of cell lines, this mutational effect does not appear not-tissue specific nor genetic background-specific, suggesting an universal indel mutational footprint of particle-based radiation. However, we do not rule out that different cell types might be differentially affected by different radiation types (protons *versus* helium) generating different indel signature burdens.

ID83D was found only in the MCF7 cell line and in both irradiated conditions, suggesting a possible tissue-specific mutagenic effect (it was not seen in blood or lung cells); this leaves similar footprints to those reported elsewhere for photon-based ionizing radiation[49].

Mutational signatures of structural variation

Large structural variants (SVs, also referred to as rearrangements) were grouped into four master classes: large deletion (DEL), large insertions (INS), tandem duplications (DUP) and inversions (INV). No particular SV category has been found to be the most prevalent across the cell lines and across the cell lines/conditions (Figure 5A).

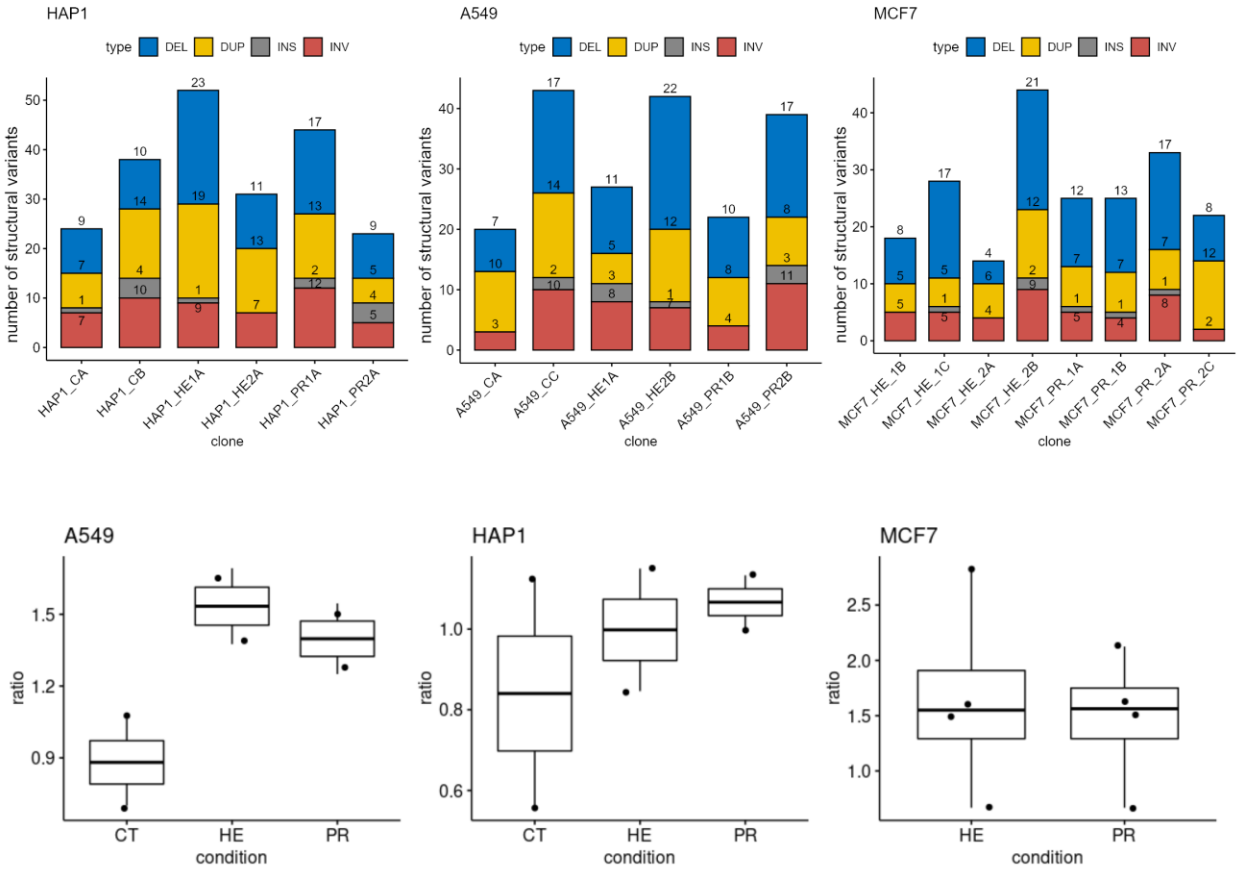


Figure 5. Structural variation in irradiated human cell lines. A. Distribution of the number of structural variants (SVs), decomposed into large deletions, large insertions, duplications and inversions, per clone inside each of the three analysed cell lines. Large insertion was the less-represented class of the SVs detected in the clones. B. Ratio of the number of large deletions over number of large insertions.

The rarest event type – insertions – is similarly observed across the cell lines and clones; their low frequency may however stem from technical difficulties in calling insertion variants from short-read WGS, resulting in many undetected variants. More generally, in our experiments, we observed globally a low number of SVs of any type, suggesting that the impact of the proton and alpha ray treatment on rearrangement rates in human cells may not be remarkable, at least in the treatment regimen applied here. Irradiated clones are not enriched in SVs when compared to control cells, except in the case of HAP1 cell line/alpha radiation, where we observed approximately two times more large deletions (23 in the most mutated clone) compared to the controls (10 in the most mutated clone).

Globally, helium (alpha) treatment generated the higher total number of SVs in both HAP1 and MCF7 cell lines, compared to proton treatment and to untreated (for HAP1). Also in the A549 lung cell line, results were generally consistent with this trend of propensity of alpha radiation to generate SVs, however less pronounced since both the non-treated clone and the helium-treated clone exhibited a high number of SVs.

We also investigated another quantitative measure of the identified variations: the distributions of the SV sizes, as commonly done in cancer genome analysis[27,29]. Again, we asked if there is a difference between two types of radiation treatments, also comparing them with the non-irradiated controls. Because insertions were rarely detected, here we considered only deletions, duplications and inversions. While the distribution of SV lengths was not significantly different across the conditions, nevertheless, we observed a trend for deletions that is consistent across all the cell lines. In particular, proton radiation generated the smallest deletion events, followed by alpha, and finally longest deletions were observed in untreated controls (Supp. Figure 3). Inversions behaved similarly: we also observed that treated clones, whatever the radiation type, contains inversions smaller than those in non-treated samples. Moreover, there is also a trend toward longer inversions resulting from helium compared to protons (Supp. Figure 3). Overall, proton radiation generated SVs of shorter lengths than alpha rays or than the default mutagenesis in untreated cells, suggesting a different mode of DNA damage and/or repair thereof following proton exposure in various human cell types.

Finally, we investigated the enrichment of deletion SVs compared to insertion and duplication SV (the latter two were considered to be similar and merged for this analysis). As previously reported for ionizing radiation-associated second malignancies[14], the ratio of deletions to insertions increases in our particle-irradiated human cell lines. We did not find this bias towards deletions to be tissue-specific (Figure 5B). Overall, different types of ionizing radiation, either X-rays or gamma as commonly employed for cancer therapy[13,14], or protons and helium/alpha particles as tested here, exhibit an overall similar spectrum of increase in deletion SVs relative to insertion/duplication SVs.

Regional mutation rate with respect to replication time and gene activity

DNA replication timing (RT) is correlated with regional variability in somatic mutation rates in human tissues. There is an increased mutation rate in late-replicating heterochromatic regions of the genome, because DNA repair mechanisms preferentially protect the early-replicating euchromatic, gene-rich regions[50,51]. We were interested if there is an enrichment of radiation-associated mutation rates in regions with a specific RT. Although the correlation between the SNV density and the RT bins are not significantly different between untreated cells and irradiated cells (and between conditions), we did observe a trend for radiation-associated SVs (Supp. Figure 4). In particular, alpha radiation, compared to untreated clones, generated more SV mutations in early-replicating regions, compared to late-replicating regions; this was observed in A549 and MCF7 cell lines. In proton radiation this trend is observed more strongly in the HAP1 cell line (Supp Figure 4B). Of note, the mutation enrichment in early-replicating DNA was not observed for point mutations (Supp Figure 4A), suggesting that this global “redistribution” of mutation rates[52] upon irradiation is specific to SV mutagenesis. Early-replicating regions have a higher gene density and are enriched with active genes. Therefore the increased rate of SV formation in early-replicating DNA upon particle radiation may have higher potential for functional impact, such as tumor suppressor gene inactivation.

Additionally, we considered mutation rate association in relation to presence of DNA methylation, which can be mutagenic e.g. it is thought to underlie generation of the very common mutational signature SBS1[53,54]. Also, gene body DNA methylation tends to be positively associated with gene expression levels, thus DNA methylation-associated mutational processes are expected to be more impactful[55]. We did not observe a clear association signal of mutation rates with DNA methylation for all the three cell lines (Supp Figure 4C). However, there is a trend in HAP1 blood cells for both SNVs and SVs, with fewer alterations in lower-methylated regions in irradiated cells compared to untreated cells (Supp Figure 4D). Prompted by this, we also investigated the distribution of mutations across regions harbouring various levels of gene expression. We observed that HAP1 proton condition generated more SV in highly expressed genes (Supp. Figure 4F) compared to the control non-radiated cell lines. This is consistent with the above RT analysis, suggesting that the exposures to particle radiation may increase SV mutation rates preferentially in active-chromatin regions, and thus would be expected to result in impactful alterations. However this effect may be cell type specific and/or genetic background specific.

Clustered mutation processes due to irradiation

Next, we considered patterns of mutation clusters, which can be particularly informative about various mutational processes in human cells[28], most prominently with regard to APOBEC enzyme DNA damaging activity[53,55,56] and also error-prone DNA polymerase usage[28]. Operationally, we defined a mutation cluster as a set of point mutations at a pairwise intermutational distance lower than 1Kb; given the overall modest mutation burden in our genomes, this threshold is unlikely to result in many artefactual (false-positive) mutation clusters. In the A549 cell line, 61 mutations were clustered in control clones, 109 in helium-irradiated clones and 93 in proton-irradiated samples. In the HAP1 cell lines, a total of 57 mutations were clustered in control clones, 262 in helium-irradiated clones and 181 in proton-irradiated samples. Finally, in the MCF7 cell line, we detected 344 clustered mutations in the helium-irradiated clones and 373 in the proton-irradiated clones (Figure 6A); note that untreated control DNA WGS is not available for MCF7.

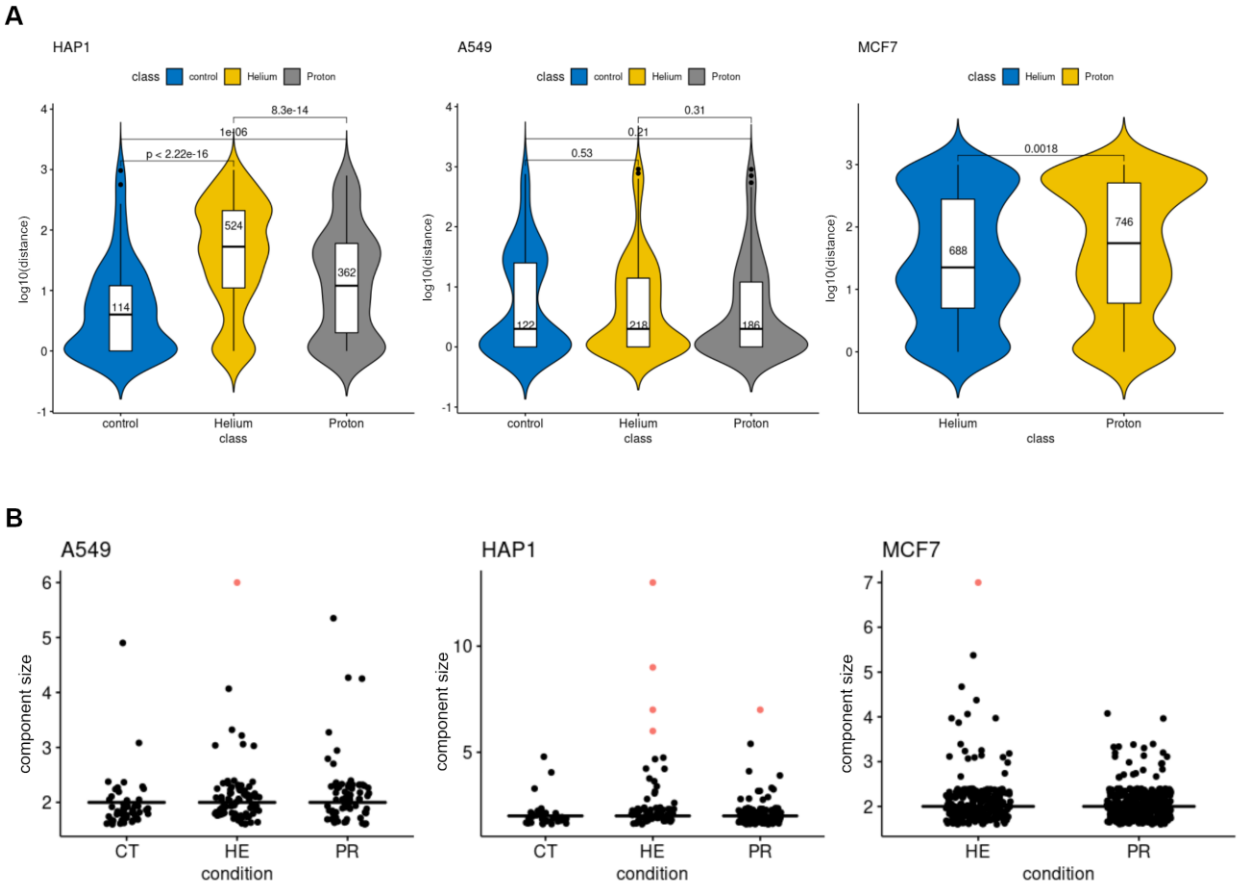


Figure 6. Clustered point mutations associated with irradiation. A. Distribution of the genomic distance between point mutations that were identified as being clustered together (see methods for the definition of a mutational cluster). *P*-values shown were computed using the *ggpubr* R package (t-test mean comparison). B. Distribution of the size of each individual cluster (number of mutations), defined as a graph component (see methods for details).

We detected more clustered mutations in radiation conditions compared to non-radiation conditions for both cell lines that could be analyzed, indicating that alpha and proton radiation generates mutation clusters in human cells. This pattern of mutagenesis is consistent with many reports of ionizing radiation damaging DNA in a clustered manner[57–59]. With respect to types of radiation, in the HAP1 and A549 cell lines (but not MCF7 cells) more clustered mutations are found in helium clones compared to proton clones.

We next asked if the clustered mutations, depending on the cell line and the radiation type, were found at closer distances. In HAP1 cells, clustered mutations were spaced closer together for both helium and proton radiations compared to the control clones. In that cell line, helium radiation generated more distant clustered mutations compared to proton radiation, but this was found to be the opposite in the MCF7 cell line (Figure 6A).

We were then interested in the repartition of the clustered mutations, *i.e.* do they form a lot

of small clusters or do they form a few large (multi-mutation) clusters, the latter case corresponding to a kataegis-like phenomenon[53] and the former to an omikli-like phenomenon[55]. We applied a graph-theory approach to quantify this (Methods). Considering large clusters as connected components of size higher than 5, we identified one large cluster in A549/helium, four in HAP1/helium, one in MCF7/helium and one in HAP1/proton. However there were no large clusters in the untreated cells, suggesting mutation showers can be generated by ionizing radiation (Figure 6B). Of note, the majority of clustered mutations, both in the irradiated and in the untreated conditions, were in smaller, *omikli*-like clusters..

Overall, this suggests that the clustering of mutations often results from exposure to particle radiation, that it may be tissue and/or genetic background specific, and that it may occasionally generate multi event, kataegis-like mutation clusters,.

Discussion

In this study, we analysed genomes of three human cancer cell lines irradiated with two different types of GCR – proton and helium fluxes – in order to characterize the spectra of DNA alterations that GCR would produce in different human tissues. Overall burden of SNV mutations, indels and SV events was not markedly different between treated and untreated cells, which suggests that GCR exposure – at least in the regime applied in our experiments – is not grossly mutagenic (see below for further considerations).

However, there were multiple differences in mutation spectra and distributions between GCR-treated and untreated cells. For instance, in agreement with a recent study analysing ionizing radiation-associated tumors[14], we observed an enrichment of deletions compared to insertions, and this was even more the case for large (SV) events (Figure 4C, Figure 5B). This phenomenon was also observed in a recent study that reported a 3.6-fold increase of small deletion burden when comparing radiotherapy treated glioma patients to untreated glioma patients[13]. We note that tumor radiotherapy usually employs X rays or gamma rays, rather than the protons and alpha particles tested here, and so this enrichment of deletion mutations appears broadly independent of the type of IR applied to cells.

We found that GCR generates point mutations that are more often clustered than those found in non-irradiated samples, moreover the distance between mutations in clusters and size of the clusters may vary between tissues and conditions. This is in line with recent work suggesting clustered mutations to be more specific indicators of mutagenic processes[28,55,56] than the general, genome-wide mutation signatures. The types of clusters were also variable: proton treatment generated clustered mutations with shorter inter-mutational distances compared to helium treatment in some cell lines. Helium treatment generates larger, multi-mutational clusters compared to proton treatment. Those two observations, overall, suggest that the clustered mutational distributions generated by helium and proton radiation exposure may differ, and potentially exhibit tissue-specificity and/or genetic background-specificity; the latter point would be clarified with additional experiments in future work.

We were also interested in the SNV trinucleotide mutational signatures that helium and

proton GCR can generate, and whether known biological mechanisms can be linked to those footprints. While most observed SNV signatures were consistent across all treated and untreated samples, e.g. reactive oxygen species mutagenesis was noted universally, some samples exhibited signatures consistent with MMR deficiency (this was however unlikely to have resulted from, or have been selected by irradiation, because we detected it also in a control sample). Interestingly, we observed a SNV signature in both irradiation conditions of the MCF7 cell line, but not in the controls we analysed (note these controls were in other cell lines apart from MCF7, see Methods): an T>A rich SBS signature, resembling previous signatures of aristolochic acid and some chemotherapy treatments[53,54], suggesting, speculatively, that particle radiation may generate DNA damage resembling bulky adducts in some cell types.

Similar to SNVs, some indel signatures were radiation-associated: ID83C and ID83D. ID83C was found in the three cell lines, in proton and helium exposed samples, thus suggesting a likely general effect over tissues for this particular mechanism. The second GCR-associated indel signature in our experiment, ID83D, was also highly active in *XRCC4*-knockout sample from an earlier study[37], suggesting that ID83D represents the mutational signature of a backup DSB repair pathway employed in absence of XRCC4 mediated non-homologous end joining repair. While ionizing radiation is expected to generate DSBs in all cell types similarly, the ID83D signature was observed only in MCF7 cells, suggesting this signature might derive from radiation applied to cells with a particular genetic background, potentially a DNA repair deficient one.

The solar radiation SNV signatures identified previously[15,31] (presumably UV generated) were observed at low intensities in our irradiated samples, likely stemming from imprecisions during statistical analysis. This is consistent with prior knowledge that ionizing cosmic radiation affects the DNA differently than the non-ionizing UV radiation[60].

Even though overall mutation rates were not increased in irradiated cells, we did note a relative increase of some mutation types, prominently SVs, in the active regions of the genome, which are early-replicating and contain highly expressed genes. This type of redistribution of mutation rates across the genome can increase the likelihood of impactful mutations occurring even in the absence of overall increased mutation burdens[52,55]. Such phenomena were previously linked with failures in DNA repair[50] and with activity of the APOBEC3A cytosine deaminase enzyme[55], and here we show that to a certain extent this happens also with ionizing radiation. Therefore, mutational impact on active genes may be increased by ionizing radiation exposures via directing SVs towards active chromatin.

Our study also has several limitations. For instance, some components of GCR were not considered in this study, in particular the heavy atom nuclei (often referred to as “HZE ions”), highly energetic particles likely originating from supernova explosions. Despite their small proportion in cosmic rays, their biological impact might be large, and their mutagenic effects on genome stability remain to be studied.

Next, we considered only a single dose of radiation that resulted in a ~40% to 50%

reduction of cell viability assessed by the colony formation assay, and a fractionation regime of four fractions (given every four days) exposures of this dose. The dose studied here had rather strong biological effects on cell viability, and such dosages would probably be well sufficient to observe the mutagenic effects of GCR. Since a gross increase in mutation rates was not observed, we infer that GCR exposures are not highly mutagenic. However, we cannot rule out the (less parsimonious) scenario where lower doses of GCR than those employed here might be more mutagenic e.g. by failing to trigger cell cycle checkpoints, thus 'slipping under the radar' of the mechanisms protecting genome integrity and introducing mutations.

Moreover, given that for several mutation patterns we observed tissue-specific and/or genetic background specific responses (these scenarios cannot be distinguished), we emphasize more experiments would be needed on different cell lines or other experimental models to ascertain tissue-specific responses to radiation types.

Finally, our data overall suggests that components of GCR are not overtly mutagenic to human cells, with some effects on SV distributions, and on indel and clustered mutation spectra However, the intermittent exposure regime that we employed for reasons of feasibility is different from longer-term, chronic exposures to GCR. Even modest increases in mutation rates under chronic GCR exposures e.g. during space flight might have detrimental effects on cancer rates, and possibly neurodegeneration and other aging phenotypes. Importantly, germline cells might be affected, with potential effects on reproductive health and on genetic disease incidence in the progeny. Therefore we highlight the necessity of further experimental work on cell and animal models, using longer-term exposures to components of galactic cosmic rays to ascertain their effects on the stability of the human genome.

Methods

Cell lines

We used three different cell lines: A549, HAP1 and MCF7. The lung adenocarcinoma cell line A549 (CCL-185) and the breast adenocarcinoma cell line MCF7 (HTB-22) were purchased from ATCC (the American Type Culture Collection, Manassas, VA). The near-haploid cell line derived from the KBM-7 cell line HAP1 (C859) was purchased from Horizon (Carle Place, NY). All cell lines were cultured and maintained according to recommended protocols. All the cell lines were authenticated by STR profiling by the corresponding repositories.
All the cells were grown at 37°C in a 5% CO_2 humidified incubator; the A549 cells were grown in Eagle's Minimum Essential Medium (MEM), MCF7 in Dulbecco modified MEM, and HAP-1 cells in Iscove's modified Dulbecco's medium (IMDM). All the media were supplemented with 10% heat-inactivated (56ºC, 30 min) fetal bovine serum (FBS), and 100 U/ml penicillin and 100 mg/ml streptomycin (all from Sigma–Aldrich Corp).

Cell irradiation, culture and clones isolation

A clonogenic assay[61] pilot experiment was conducted to assess the dose that resulted in 40-

50% lethality (i.e., 50-60% survival) to 11 MeV helium ions (LET = 65 keV/µm) or 5.4 MeV protons (LET = 10 keV/µm), which was 0.5 and 1 Gy, respectively (Figure 1A,B). Cells for the experiment were then exposed to four fractions delivered every four days. Cells were collected four days after the last fraction. For each sample, half of the cells were frozen at -80 C while the other half was re-plated until confluency. The cells were then collected, separated by vigorously passing through a 21 G syringe and 3 ul/ml of DAPI was added to the solution for cell sorting at the BD Influx cell sorter. For each cell line, single cells were sorted in two 96-well plates and incubated in 250 ul of medium supplemented with 20% FBS. Once colonies were visible, we picked ~30 colonies and re-plated each one in an individual well of a 12-well plate in the case of A549 and MCF7 and in 48-well plate in the case of HAP1 cells up to confluence. The cells were then collected and half frozen at -80C while the other half was replated in 6-well plates and then in 100-mm Petri dish until confluence. DNA was extracted with PureLink Genomic DNA kit (Invitrogen), frozen and shipped to the sequencing centre.

Cell irradiations at the track segment irradiation platform

The 5.5 MV Singletron accelerator at the Radiological Research Accelerator Facility (RARAF) served as source of energetic protons and alpha particles. The accelerator was operated at the maximum terminal voltage generating beams with nominal energies of 5.5 MeV and 11 MeV for protons and alphas, respectively. Cells were irradiated at the so called “track segment” irradiation platform whose name indicates the traversal of thin samples (typically cell monolayers) by a short segment of the ion’s trajectory resulting in a small variation of the linear energy transfer (LET) throughout the sample. It is valid therefore to assume that radiation doses at the track segment platform are delivered by mono-LET beams. The LET values of the applied beams were 10 keV/µm for protons and 65 keV/µm for alphas. Detailed description of the track segment irradiation platform and its operational principles can be found elsewhere[1,2,3,4]. We will state here only the main features of the irradiation protocol. The protocol can be generally divided in two parts. First, beam characterization and dosimetry were performed with the use of different detectors mounted on the metal wheel that rotates over the beam exit aperture. The aims of this step were to verify the energy and the LET of the beam, to check the uniformity of the irradiation field covering the 6 mm x 20 mm beam exit aperture (2.9 µm thick havar foil) and, finally, to calibrate the online beam monitor in terms of the absorbed dose delivered to the sample. In the second step, the dosimetry wheel was replaced with the sample-carrying wheel that can accommodate up to 20 custom made dishes. The dishes were manufactured by gluing 6 µm thick mylar foil over metal rings having a diameter of 5 cm. The mylar foil serves as the bottom of the dish on which the cells are attached, allowing thus the ions to penetrate through and reach the cells without losing much of their energy. The in-house developed software drives the stepper motor and controls the movement of sample dishes over the beam. A desired dose was delivered to the sample by exposing it to the beam until the appropriate number of monitor counts was reached according to dosimetry measurements performed during the first part of the irradiation protocol.

Irradiations of the cells for mutational analysis and clone isolations

The obtained survival curves were fitted with the linear quadratic (LQ) model. Doses of 1 Gy and 0.5 Gy were selected for investigating the mutational signatures of protons and alpha particles, respectively. According to the LQ fits, the selected doses resulted in preservation of between 50% and 60% clonogenic capacity for all cell lines. Four fractions of the same dose were delivered to the cells every 96 hours to amplify the number of mutations in surviving cells. Between the fractions, the cells were not removed from the mylar dishes. To prevent the cells to overgrow the dish size before the end of the irradiations, initial number of plated cells was < 1000.

Whole-genome sequencing

Total genomic DNA was extracted from pelleted cells using PureLink™ Genomic DNA Mini Kit (K182001, Thermo Fisher Scientific). DNA was sequenced on NovaSeq 6000, 150 nt paired-end mode. Reads were aligned to the reference human genome Hg38 (GRCh38.d1.vd1) using BWA version 0.7.17[62]. GATK Base Quality Score Recalibration was applied to adjust quality scores assigned during the sequencing process[63]. The average coverage of the various samples was 18.5X to 44.8X.

Variant detection

Strelka2 variant calling algorithm was used to detect point mutations and small insertions and deletions[64]. Strelka2 was launched in the joint genotyping mode independently across each sample and each cell line, *i.e.* using all the samples of one particular cell line as the input of the variant calling process, in one batch. Mutations were annotated using annovar tool[65] and the gnomAD genome database v2.1.1[66]. To call structural variants, i.e. large insertions and deletions, duplications and inversions, we used the Manta tool in the joint calling mode[67], in the same manner we used Strelka2. For inversions, Manta returns a list of breakpoints that described the inversions it detected. To transform those breakpoints into inversions, we applied the convertInversion.py python script provided by the Manta developers (https://github.com/Illumina/manta). In order to remove common germline variants, we removed all the mutations reported in the gnomAD database at a frequency higher than 0.1%. In order to remove the rest of germline variations that are shared mutations across the clones of the same cell lines, we removed all mutations present in more than one clone. From further analyses, we excluded the data from the MCF7 control (unirradiated) sample, that harboured an implausibly high number of mutations (116,971) possibly due to the low quality of the sequenced DNA, or other artefacts.

Mutation clustering

Point mutations were considered to be clustered together if they appear at positions with a genomic distance lower than 1kb; such mutations would be considered to fall into the omikli ("mutation fog") category[55.] To consider large, multi-mutation clusters resembling a kataegis (mutation shower) event[53,68], we used an approach based on graph theory. Roughly, we considered each mutation to be a node of the graph, and edges connect two nodes if the

distance between corresponding mutations is lower than 1kb. Then, we computed the components of the graph, that corresponds to a connected sub-graph, and subsequently the size of the components, *i.e.* how many nodes or mutations are present in the component.

Extraction of mutational signatures

We pooled our samples with two already published datasets in order to help NMF to converge and in order to compare our datasets to signatures of DNA damage and signatures of various mutagens. We first pooled our data with the mutations reported in Zou *et al.* study[37], which analysed 173 clones generated by CRISPR-Cas9 knockouts of 73 DNA repair/replicative pathway genes in human induced pluripotent stem cells. We then pooled the data with the mutations reported in Kucab *et al.* study[31], which reported 54 samples containing a mutational signature (note there were additional samples in the study without a prominent mutation pattern, which were not included here), together with our cosmic-radiation exposed samples of three cell lines. This pooled dataset of a total of 247 individual samples. In order to extract SNV signatures generated by our radiation treatments, we applied SigProfiler Extractor algorithm[15] with default settings on the set of detected mutations, classified into the 96 SNV categories. The suggested solution contained 12 extracted signatures. Nevertheless, SigProfiler also provided the summary of all tester set of signatures, and a set of 7 signatures was also found to be potentially plausible by SigProfiler (based on low reconstruction error and high reproducibility in mutational signatures, corresponding to an high signature stability, see refs.[54,69] for details). We observed that, when considering the set of 12 mutational signatures, some mechanisms were spread across multiple signatures suggesting that some could be merged. We thus considered the set of 7 mutational signatures for the rest of the analyses. Extracted signature were assigned to a known PCAWG “SBS” signature[15], if the cosine similarity between the two mutational profiles was at least 0.85. The SigsPack R package[70] was used to obtain the mutational profiles of PCAWG signatures.
For indel signatures, we used the SigProfiler Matrix Generator tool to correctly classify the observed indels into one of the 83 categories defined in Alexandrov *et al.*[15]. Then, the SigProfiler Extractor tool was launched in the a total set of 321 samples from Kucab *et al.*, Zou *et al.*, and our experiments, using those generated matrices. We considered the optimal solution given by the tool, resulting in a total set of 4 extracted indel signatures.

Regional enrichment analysis

We performed a regional enrichment analysis in order to estimate if the mutations detected in the irradiated clones were located in regions with a specific epigenetic pattern compared to the mutations detected in the non-treated clones. We tested three genomic features: the replication timing, the DNA repair histone mark H3K36me3 and the gene expression levels based on RNA-seq data. Replication timing bins, genomic bins extracted from chromatin mark H3K36me3, and regions based on variable gene expression were computed as in Supek and Lehner[28]. To detect significant association of mutations in specific genomic regions, we fitted a negative binomial regression using the glm.nb function from the MASS R package, on the counts of mutations per bins, controlling for the 3-nucleotide context and type of mutation in the case of SNV, as in

Supek and Lehner[28]. The regional enrichment analysis source code is implemented using Nextflow[71] and is available on the GitHub platform (https://github.com/tdelhomme/RegionalEnrichment-nf/).

Statistical analyses

All the statistical analyses were performed using R software version 3.6.0. VariantAnnotation[72] R package version 1.32 was used to read the VCF files. The R packages BSgenome.Hsapiens.UCSC.hg38 version 1.4.1 and BSgenome.Hsapiens.UCSC.hg19 version 1.4.0 were used to load the human genome inside R. MutationalPatterns[73] R package version 1.12 was used to generate the trinucleotide contexts. Ggraph version 2.0.5 and Igraph version 1.2.6 R packages was used to identify the cluster components.

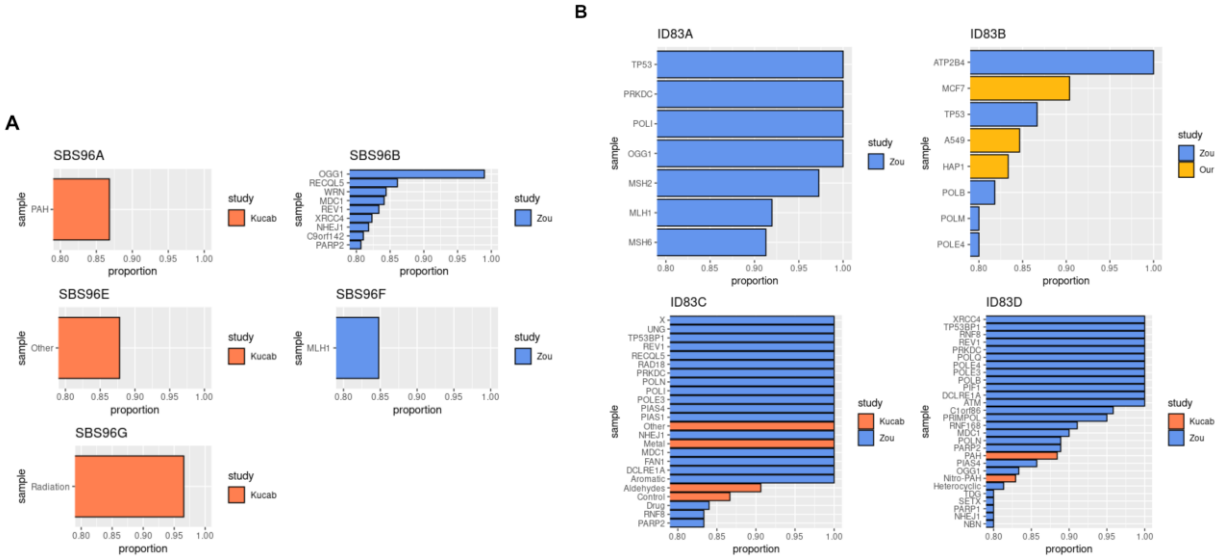
References

1. Larose, T. L. *Tumors in Space: Preparation for Spaceflight*. *Preparation of Space Experiments* (IntechOpen, 2020). doi:10.5772/intechopen.93465.
2. Jastrow, R. Definition of Air Space. in *First Colloquium on the Law of Outer Space* (eds. Haley, A. G. & Heinrich, W.) 82–82 (Springer, 1959). doi:10.1007/978-3-7091-4414-5_16.
3. 100km Altitude Boundary for Astronautics. https://www.fai.org/page/icare-boundary (2017).
4. McDowell, J. C. The edge of space: Revisiting the Karman Line. *Acta Astronaut.* **151**, 668–677 (2018).
5. Thirsk, R., Kuipers, A., Mukai, C. & Williams, D. The space-flight environment: the International Space Station and beyond. *CMAJ Can. Med. Assoc. J.* **180**, 1216–1220 (2009).
6. Panesar, S. S. & Ashkan, K. Surgery in space. *BJS Br. J. Surg.* **105**, 1234–1243 (2018).
7. Hellweg, C. E. & Baumstark-Khan, C. Getting ready for the manned mission to Mars: the astronauts' risk from space radiation. *Naturwissenschaften* **94**, 517–526 (2007).
8. Cucinotta, F. A. Space Radiation Risks for Astronauts on Multiple International Space Station Missions. *PLoS ONE* **9**, e96099 (2014).
9. Nikiforov, Y. E. Thyroid carcinoma: molecular pathways and therapeutic targets. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **21 Suppl 2**, S37-43 (2008).
10. Morton, L. M. *et al.* Radiation-related genomic profile of papillary thyroid carcinoma after the Chernobyl accident. *Science* **372**, eabg2538 (2021).
11. Brenner, A. V. *et al.* I-131 dose response for incident thyroid cancers in Ukraine related to the Chornobyl accident. *Environ. Health Perspect.* **119**, 933–939 (2011).
12. Ozasa, K. Epidemiological research on radiation-induced cancer in atomic bomb survivors. *J. Radiat. Res. (Tokyo)* **57**, i112–i117 (2016).
13. Kocakavuk, E. *et al.* Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat. Genet.* **53**, 1088–1096 (2021).
14. Behjati, S. *et al.* Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
15. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
16. Berrington de González, A. & Darby, S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. *Lancet Lond. Engl.* **363**, 345–351 (2004).
17. IARC. *Ionizing Radiation, Part 1: X- and Gamma (γ)-Radiation, and Neutrons*.
18. Hughes, J. R. & Parsons, J. L. FLASH Radiotherapy: Current Knowledge and Future Insights Using Proton-Beam Therapy. *Int. J. Mol. Sci.* **21**, E6492 (2020).
19. Yuan, T.-Z., Zhan, Z.-J. & Qian, C.-N. New frontiers in proton therapy: applications in cancers. *Cancer Commun.* **39**, 61 (2019).
20. Li, M. *et al.* Clinical Efficacy and Safety of Proton and Carbon Ion Radiotherapy for Prostate Cancer: A Systematic Review and Meta-Analysis. *Front. Oncol.* **11**, (2021).
21. Hu, M., Jiang, L., Cui, X., Zhang, J. & Yu, J. Proton beam therapy for cancer in the era of precision medicine. *J. Hematol. Oncol.J Hematol Oncol* **11**, 136 (2018).
22. Vogel, J. *et al.* Proton therapy for pediatric head and neck malignancies. *Pediatr. Blood Cancer* **65**, e26858 (2018).
23. Jain, V. *et al.* Predicted Secondary Malignancies following Proton versus Photon Radiation for Oropharyngeal Cancers. *Int. J. Part. Ther.* **6**, 1–10 (2020).
24. Huang, J. & Mehta, M. Can proton therapy reduce radiation-related lymphopenia in glioblastoma? *Neuro-Oncol.* **23**, 179–181 (2021).
25. Mohan, R. *et al.* Proton therapy reduces the likelihood of high-grade radiation-induced lymphopenia in glioblastoma patients: phase II randomized study of protons vs photons. *Neuro-Oncol.* **23**, 284–294 (2021).
26. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
27. Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
28. Supek, F. & Lehner, B. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* **170**, 534-547.e23 (2017).
29. Degasperi, A. *et al.* Substitution mutational signatures in whole-genome–sequenced cancers in

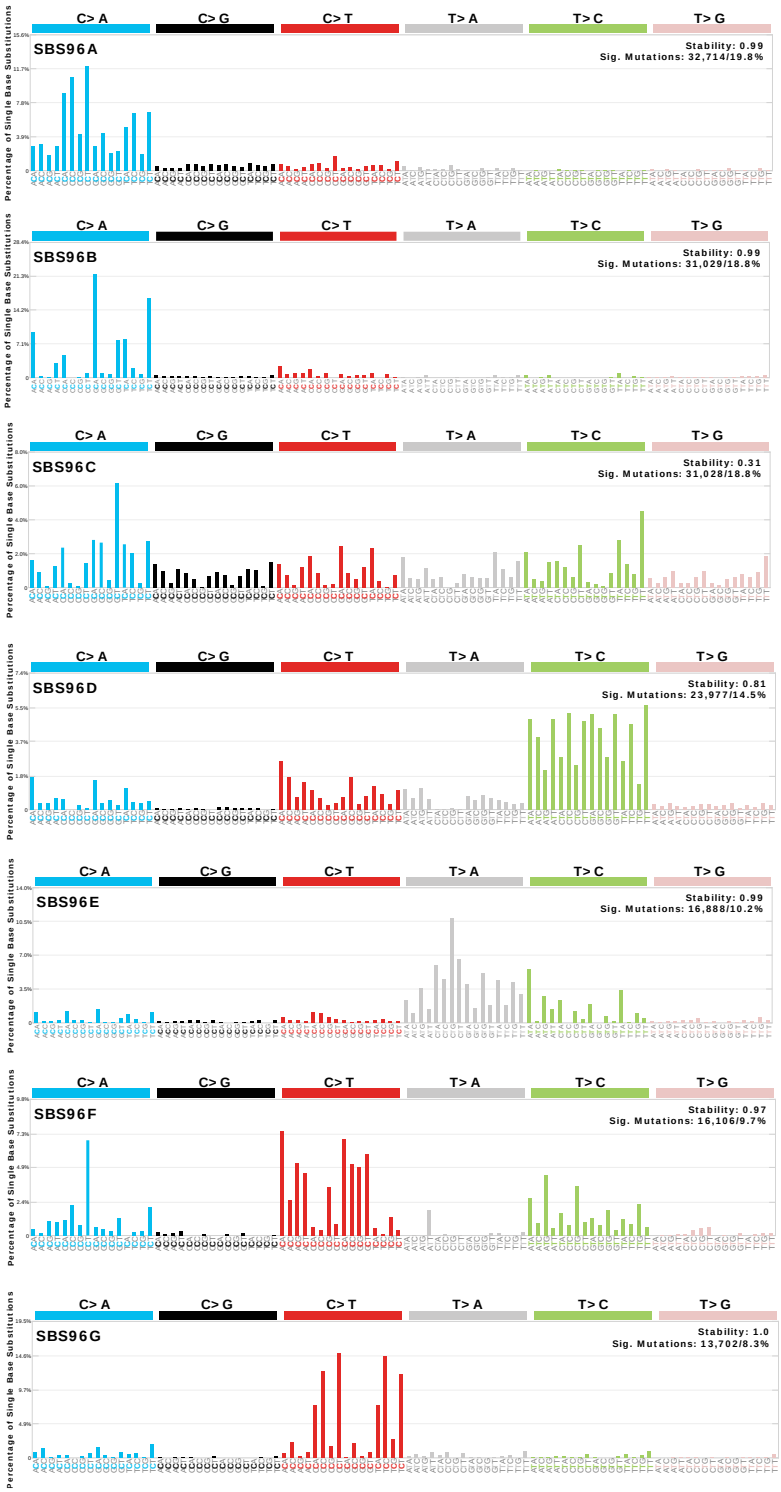
the UK population. *Science* **376**, abl9283 (2022).
30. Pleasance, E. *et al.* Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. *Nat. Cancer* **1**, 452–468 (2020).
31. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e16 (2019).
32. Saini, N. *et al.* The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. *PLOS Genet.* **12**, e1006385 (2016).
33. Franco, I. *et al.* Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol.* **20**, 285 (2019).
34. Davidson, P. R., Sherborne, A. L., Taylor, B., Nakamura, A. O. & Nakamura, J. L. A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies. *Sci. Rep.* **7**, 7645 (2017).
35. Sherborne, A. L. *et al.* Mutational Analysis of Ionizing Radiation Induced Neoplasms. *Cell Rep.* **12**, 1915–1926 (2015).
36. Volkova, N. V. *et al.* Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
37. Zou, X. *et al.* A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
38. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
39. Sekiguchi, M. & Tsuzuki, T. Oxidative nucleotide damage: consequences and prevention. *Oncogene* **21**, 8895–8904 (2002).
40. Kawanishi, S., Hiraku, Y. & Oikawa, S. Mechanism of guanine-specific DNA damage by oxidative stress and its role in carcinogenesis and aging. *Mutat. Res.* **488**, 65–76 (2001).
41. Wang, R. *et al.* OGG1-initiated base excision repair exacerbates oxidative stress-induced parthanatos. *Cell Death Dis.* **9**, 628 (2018).
42. Keijzers, G. *et al.* Human Exonuclease 1 (EXO1) Regulatory Functions in DNA Replication with Putative Roles in Cancer. *Int. J. Mol. Sci.* **20**, 74 (2018).
43. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
44. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
45. Mas-Ponte, D., McCullough, M. & Supek, F. Spectrum of DNA mismatch repair failures viewed through the lens of cancer genomics and implications for therapy. *Clin. Sci. Lond. Engl. 1979* **136**, 383–404 (2022).
46. Bassing, C. H. *et al.* Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8173–8178 (2002).
47. Delacôte, F., Han, M., Stamato, T. D., Jasin, M. & Lopez, B. S. An xrcc4 defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. *Nucleic Acids Res.* **30**, 3454–3463 (2002).
48. Wen, Y., Dai, G., Wang, L., Fu, K. & Zuo, S. Silencing of XRCC4 increases radiosensitivity of triple-negative breast cancer cells. *Biosci. Rep.* **39**, BSR20180893 (2019).
49. Youk, J. *et al.* Mutational impact and signature of ionizing radiation. 2021.01.12.426324 (2021) doi:10.1101/2021.01.12.426324.
50. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
51. Zheng, C. L. *et al.* Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep.* **9**, 1228–1234 (2014).
52. Supek, F. & Lehner, B. Scales and mechanisms of somatic mutation rate variation across the human genome. *DNA Repair* **81**, 102647 (2019).
53. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
54. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
55. Mas-Ponte, D. & Supek, F. DNA mismatch repair promotes APOBEC3-mediated diffuse

hypermutation in human cancers. *Nat. Genet.* **52**, 958–968 (2020).
56. Bergstrom, E. N. *et al.* Mapping clustered mutations in cancer reveals APOBEC3 mutagenesis of ecDNA. *Nature* **602**, 510–517 (2022).
57. Sage, E. & Shikazono, N. Radiation-induced clustered DNA lesions: Repair and mutagenesis. *Free Radic. Biol. Med.* **107**, 125–135 (2017).
58. Sutherland, B. M., Bennett, P. V., Weinert, E., Sidorkina, O. & Laval, J. Frequencies and relative levels of clustered damages in DNA exposed to gamma rays in radioquenching vs. nonradioquenching conditions. *Environ. Mol. Mutagen.* **38**, 159–165 (2001).
59. Hada, M. & Georgakilas, A. G. Formation of clustered DNA damage after high-LET irradiation: a review. *J. Radiat. Res. (Tokyo)* **49**, 203–210 (2008).
60. Ravanat, J.-L. & Douki, T. UV and ionizing radiations induced DNA damage, differences and similarities. *Radiat. Phys. Chem.* **128**, 92–102 (2016).
61. Puck, T. T. & Marcus, P. I. Action of x-rays on mammalian cells. *J. Exp. Med.* **103**, 653–666 (1956).
62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
63. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
64. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
65. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
66. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
67. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinforma. Oxf. Engl.* **32**, 1220–1222 (2016).
68. Roberts, S. A. *et al.* Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Mol. Cell* **46**, 424–435 (2012).
69. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. 2020.12.13.422570 (2022) doi:10.1101/2020.12.13.422570.
70. Schumann, F. *et al.* SigsPack, a package for cancer mutational signatures. *BMC Bioinformatics* **20**, 450 (2019).
71. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
72. Obenchain, V. *et al.* VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinforma. Oxf. Engl.* **30**, 2076–2078 (2014).
73. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).

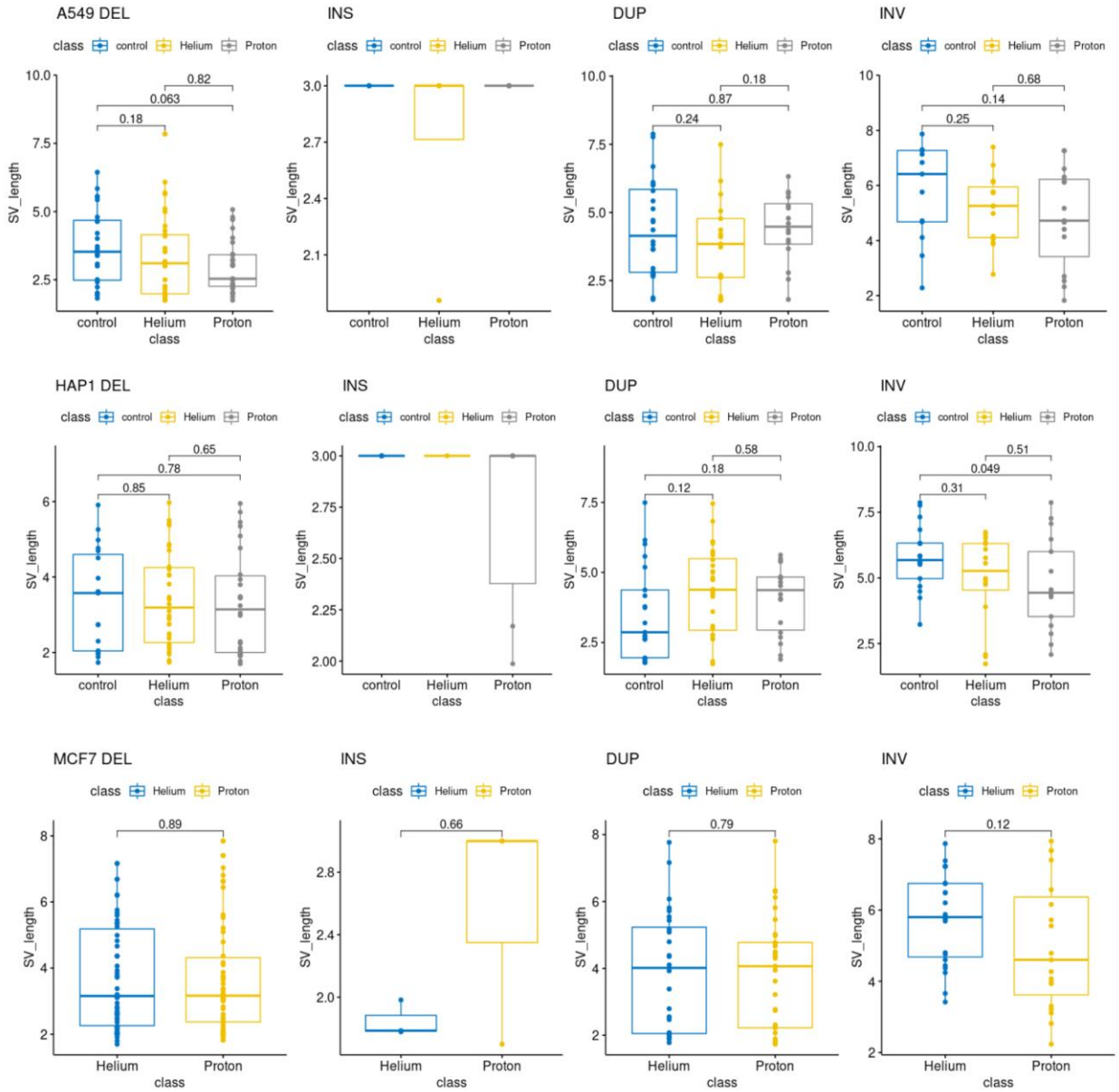
Supplementary figures



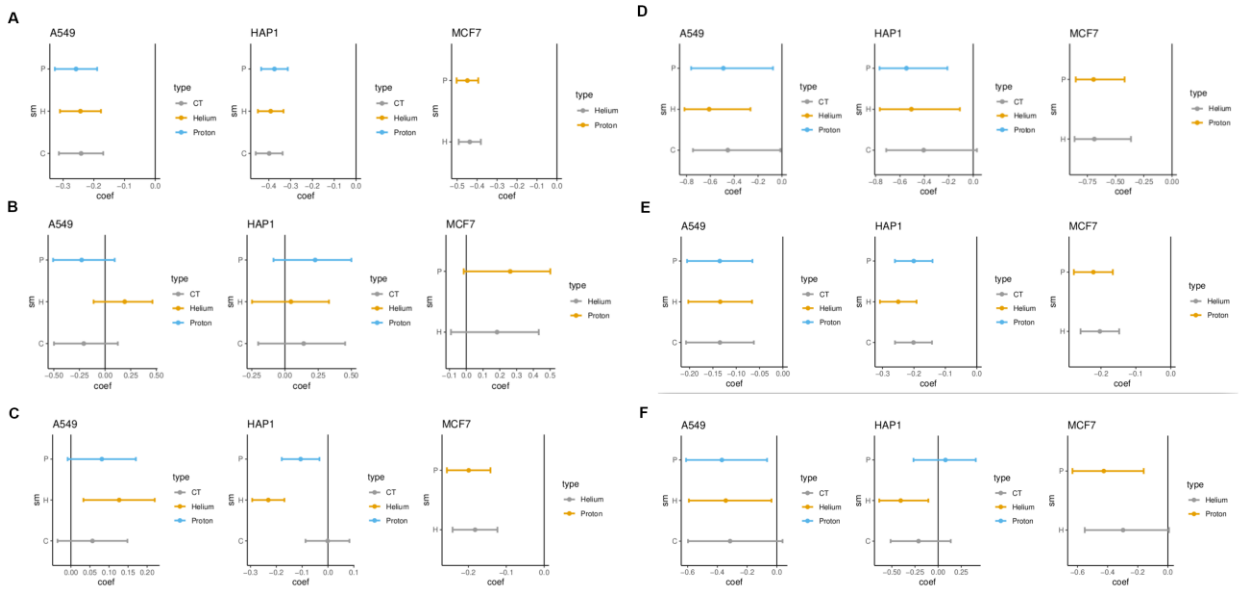
Supp. Figure 1. Sample attribution to a mutational signature. A signature was considered to be generated by a condition (basically a mutagen or a gene knock-out) if at least 80% of the total number of mutations detected in the condition sample are attributed to the signature. Note that this is different than the activity of a signature that can be detected in one particular sample. This was computed independently for point mutations (A) and for indels (B). Samples were coloured according to the corresponding study: blue are Zou *et al*. knock-out samples[37], orange are Kucab *et al.* mutagenesis samples[31] and yellow are the samples from this study.



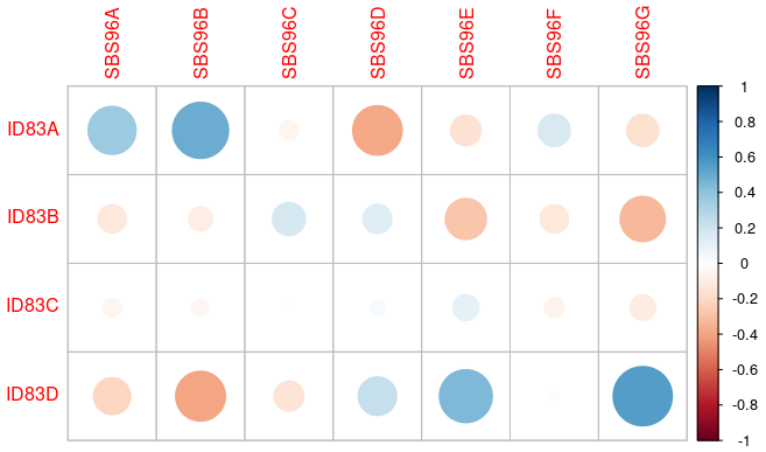
Supp. Figure 2. Extracted SNV spectra. Mutational spectrum of the 7 SNV signatures identified in the pooled-data, from the SigProfiler analysis. Each signature corresponds to the distribution of point mutations into one of the 96 possible substitution categories taking a trinucleotide context: 6 possible point mutations (taking into account base complementarity, *i.e.* A=T; G=C) at the target position and 4 at each surrounding genomic location = 6*4*4=96 combinations.



Supp. Figure 3. Structural variation lengths. Distribution of the length of structural variations, classified into the 4 main categories, *i.e.* large deletions, large insertions, duplications and inversions. Length values are shown in the $\log_{10}$ scale. Rows represented the three different cell lines, A549, HAP1 and MCF7. *P*-values shown in the plots were computed using the ggpubr R package (t-test mean comparison). We observed a particular trend: proton radiation provokes the smallest deletion and inversion events, followed by helium, and finally controls.



Supp. Figure 4. Regional mutation rates analysis, comparing negative binomial regression coefficient distribution. A negative binomial regression was computed for each cell line/condition/clone, in order to detect association between number of mutations (mutation rate) and genomic region, corrected by trinucleotide context in the case of SNVs. A similar approach than previously developed[55] was applied. The regression was applied to 3 different types of genomic components: replication time (A: SNVs, B: SVs), methylation (C: SNVs, D: SVs), and gene expression (E: SNVs, F: SVs). See methods for details on regional bins computations.



Supp. Figure 5. Correlation between SNV signatures and indel signature activities. For each sample we computed, for each SNV signature and each indel signature, the proportion of mutations attributed to that particular signature. The plot shows the correlation between each pair of indel signature activities and SNV signature activities. In other words, a high value would mean that the corresponding indel and SNV signatures are both highly active in the same samples, whereas a low value would mean they are both slightly active in the same samples.