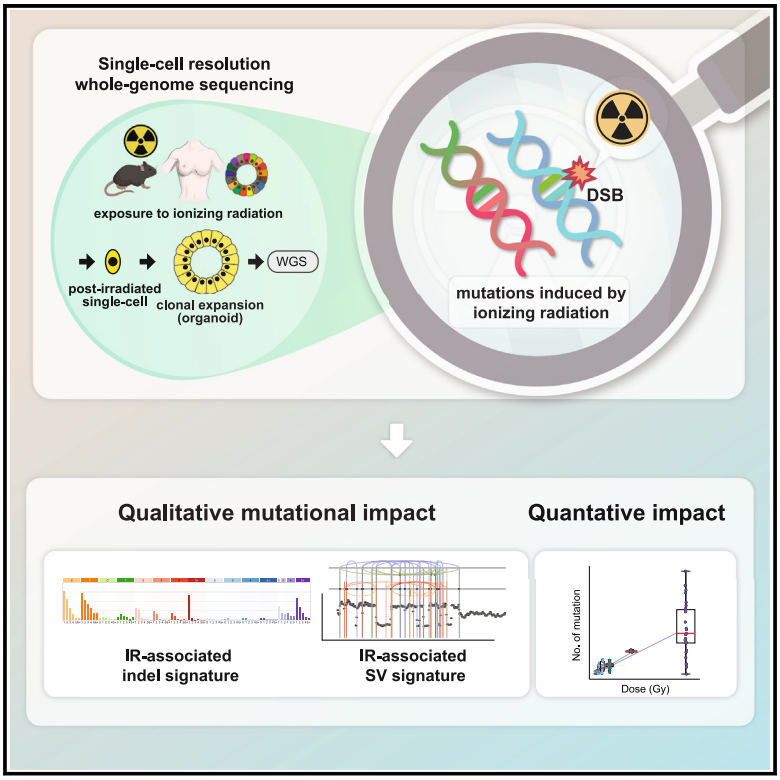
Cell Genomics

Article

Quantitative and qualitative mutational impact of ionizing radiation on normal cells

Graphical abstract



Highlights

- Single-cell-resolution DNA sequencing reveals IR-associated mutational signatures
- IR-associated mutations include short deletions and simple and complex rearrangements
- Simultaneous multiple DSBs cause chromothripsis, chromoplexy, and SVs by BFB cycles
- 1 Gy IR exposure generates ~2.33 mutations per Gb in mammalian cells

Authors

Jeonghwan Youk, Hyun Woo Kwon, Joonoh Lim, ..., Kyung Su Kim, Tae Gen Son, Young Seok Ju

Correspondence

jh.chang@snu.ac.kr (J.H.C.), kskim.cirt@snu.ac.kr (K.S.K.), tgson@dirams.re.kr (T.G.S.), ysju@kaist.ac.kr (Y.S.J.)

In brief

Youk, Kwon, Lim, Kim, Kim et al. used single-cell-resolution DNA sequencing to investigate the quantitative and qualitative impact of ionizing radiation (IR) on the mammalian genome. They found characteristic mutational signatures, comprised of specific short deletions and simple and complex structural variations. Complex genomic rearrangements, prevalent in cancer, were induced in post-irradiated cells, suggesting their role in tumor development. The study reveals mutational rates proportional to IR dose and consistent across cell types, underscoring IR's universal mutagenic impact on mammalian cells.

Youk et al., 2024, Cell Genomics *4*, 100499
February 14, 2024
https://doi.org/10.1016/j.xgen.2024.100499

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Quantitative and qualitative mutational impact of ionizing radiation on normal cells

Jeonghwan Youk,[1,21] Hyun Woo Kwon,[2,3,21] Joonoh Lim,[1,4,21] Eunji Kim,[5,6,21] Taewoo Kim,[1,21] Ryul Kim,[1,4] Seongyeol Park,[1,4] Kijong Yi,[1,4] Chang Hyun Nam,[1] Sara Jeon,[7] Yohan An,[1] Jinwook Choi,[8,9] Hyelin Na,[10,20] Eon-Seok Lee,[11] Youngwon Cho,[2,12] Dong-Wook Min,[2,12] HyoJin Kim,[11] Yeong-Rok Kang,[11] Si Ho Choi,[11] Min Ji Bae,[11] Chang Geun Lee,[11] Joon-Goon Kim,[1,13] Young Seo Kim,[1,13] Tosol Yu,[6,14] Won-Chul Lee,[4] Jong-Yeon Shin,[4] Dong Soo Lee,[2,15] Tae-You Kim,[2,12,16] Taeyun Ku,[1,13] Su Yeon Kim,[1] Joo-Hyeon Lee,[8,9] Bon-Kyoung Koo,[10] Hyunsook Lee,[7] On Vox Yi,[17] Eon Chul Han,[18] Ji Hyun Chang,[6,19,*] Kyung Su Kim,[6,19,*] Tae Gen Son,[11,*] and Young Seok Ju[1,4,22,*]

[1]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[2]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul 08826, Republic of Korea
[3]Department of Nuclear Medicine, Korea University College of Medicine, Seoul, Republic of Korea
[4]Genome Insight, Inc., San Diego, CA 92121, USA
[5]Department of Radiation Oncology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, 20, Boramae-ro 5 Gil, Dongjak-gu, Seoul 07061, Republic of Korea
[6]Department of Radiation Oncology, Seoul National University College of Medicine, Seoul 03080, Republic of Korea
[7]Department of Biological Sciences & IMBG, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
[8]Wellcome - MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, CB2 0AW Cambridge, UK
[9]Department of Physiology, Development and Neuroscience, University of Cambridge, CB2 3EL Cambridge, UK
[10]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[11]Research Center, Dongnam Institute of Radiological and Medical Science, Busan, Republic of Korea
[12]Cancer Research Institute, Seoul National University, Seoul, Republic of Korea
[13]KI for Health Science and Technology, KAIST, Daejeon 34141, Republic of Korea
[14]Department of Radiation Oncology, Dongnam Institute of Radiological and Medical Science, Busan 46033, Republic of Korea
[15]Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul 03080, Republic of Korea
[16]Department of Internal Medicine, Seoul National University Hospital, Seoul 03080, Republic of Korea
[17]Department of Breast Surgery, Dongnam Institute of Radiological and Medical Science, Busan, Republic of Korea
[18]Department of Surgery, Dongnam Institute of Radiological and Medical Science, Busan, Republic of Korea
[19]Department of Radiation Oncology, Seoul National University Hospital, Seoul, Republic of Korea
[20]Present address: TONY Investment Co., Ltd., Seoul 06180, Republic of Korea
[21]These authors contributed equally
[22]Lead contact
*Correspondence: jh.chang@snu.ac.kr (J.H.C.), kskim.cirt@snu.ac.kr (K.S.K.), tgson@dirams.re.kr (T.G.S.), ysju@kaist.ac.kr (Y.S.J.)
https://doi.org/10.1016/j.xgen.2024.100499

SUMMARY

The comprehensive genomic impact of ionizing radiation (IR), a carcinogen, on healthy somatic cells remains unclear. Using large-scale whole-genome sequencing (WGS) of clones expanded from irradiated murine and human single cells, we revealed that IR induces a characteristic spectrum of short insertions or deletions (indels) and structural variations (SVs), including balanced inversions, translocations, composite SVs (deletion-insertion, deletion-inversion, and deletion-translocation composites), and complex genomic rearrangements (CGRs), including chromoplexy, chromothripsis, and SV by breakage-fusion-bridge cycles. Our findings suggest that 1 Gy IR exposure causes an average of 2.33 mutational events per Gb genome, comprising 2.15 indels, 0.17 SVs, and 0.01 CGRs, despite a high level of inter-cellular stochasticity. The mutational burden was dependent on total irradiation dose, regardless of dose rate or cell type. The findings were further validated in IR-induced secondary cancers and single cells without clonalization. Overall, our study highlights a comprehensive and clear picture of IR effects on normal mammalian genomes.

INTRODUCTION

Somatic mutations accumulate in cells throughout an individual's lifetime, often leading to diseases, such as cancer.[1] As these mutations are a joint result of specific DNA damage and repair processes in cells,[1] each mutational process leaves a distinct pattern of mutations known as the "mutational signature."[2] Recently, a sophisticated statistical deconvolution of population-scale cancer genome data defined approximately 100 mutational signatures (available at the COSMIC database: https://cancer.sanger.ac.uk/signatures/).[2,3] Some of the known mutational signatures are associated with well-known carcinogens. For example, lung, skin, and liver cancer genomes frequently exhibit single-base substitution (SBS) signatures associated with exposure to tobacco smoking,[4] ultraviolet light,[5] and aristolochic acids,[6] respectively. Additional mutational signatures have been identified under experimental conditions using genome sequencing of cells exposed to various physical, chemical, and biological genotoxins.[7–9] These approaches have provided insight into DNA damage mechanisms of numerous carcinogens and downstream cellular repair processes.

Ionizing radiation (IR) is a well-known potent carcinogen that causes direct and indirect DNA damage[10] and is commonly encountered in the environment, for example as radon gas, and in medical practice. Due to its genotoxicity, numerous studies have investigated its mutational impact, particularly in the germline. Genome sequencing in model organisms has revealed a positive correlation between acquired genomic mutations and irradiation dose in budding yeasts, worms, and mice.[8,11–14] In humans, survivors from atomic bombings and nuclear power plant accidents have been analyzed.[15–17] To explore the mutational impact of IR on somatic cells, IR-irradiated induced pluripotent stem cells and radiation-induced secondary tumors have been investigated.[7,18,19] However, the precise mutational signatures of IR, particularly in normal human cells, remain incomplete, as previous studies were limited by insufficient sample size, inattention to structural variations (SVs), or other confounding genetic and environmental factors impacting cancer cells.[7,18–20] Due to similar reasons and the limited contribution of IR-associated secondary cancers in the discovery dataset, large-scale statistical methods have not delineated IR-associated mutational signatures.[2]

In this study, we conducted an in-depth genomic investigation of post-irradiated somatic cells employing whole-genome sequencing (WGS) of clonally expanded post-irradiated cells (hereafter referred to as colonies)[21,22] to elucidate the comprehensive IR-induced mutational spectrum and burden. Murine colonies were subjected to 0–20 Gy gamma radiation under controlled experimental conditions, and human colonies were established from adjacent normal tissues after radiation therapy involving exposure to approximately 50 Gy radiation. We also investigated the genomes of 22 radiation-induced secondary tumors.[18,23] To explore potential selection bias from relying on viable clones in our clonal-expansion approach, we additionally examined post-irradiated cells using various single-cell sequencing technologies.

RESULTS

Clonalization for accurate single-genome sequencing

WGS of clonally expanded cells allows comprehensive and accurate detection of all types of genomic variants accumulated within a single cell[22,24] but cannot be used to investigate the genomes of dead or non-proliferating cells. To efficiently expand single cells, we used organoid culture techniques (Figures S1A and S1B), which allow culturing of non-neoplastic normal single cells.[25] We used three different but complementary experimental settings for IR exposure (Figure 1A), including (1) IR exposure on cultured cells (IR^{vitro}) using a ^{137}Cs source; (2) IR exposure on living tissues (IR^{vivo}) using mostly a ^{60}Co source (Figure S1C), combined with a ^{137}Cs source for long-term exposure experiments; and (3) post-radiotherapy ($IR^{post\text{-}RT}$) with gamma rays from the accelerator as the main IR source. Further, we analyzed 22 genomes of IR-associated secondary tumors to evaluate IR-associated mutations in human cancers (Figure 1A).

We investigated the genome sequences of 135 colonies, including 88 colonies derived from post-irradiated normal cells and 47 from non-irradiated normal cells (controls) (Tables S1 and S2). The variant allele fraction (VAF) of acquired mutations (Figures 1B and S2–S4) confirmed that all colonies were monoclonal (n = 80; 59.3%) or dominant clonal (n = 55; 40.7%), thus allowing detection of IR-induced mutations acquired in the predominant founder cell.

Impact of IR on cell viability

We found that the organoid-forming efficiency of single cells was negatively correlated with IR exposure level. Approximately 2 Gy irradiation reduced the proliferation potential of mouse pancreatic organoids by 50% (Figure 1C). RNA sequencing of post-irradiated organoids revealed upregulation of DNA damage response genes, such as *Brca1*, *Brca2* (homologous recombination DNA repair pathway), *Mdm2* (p53 pathway), and *Cdkn1a* (cell-cycle pathway), 2 h post-irradiation (Figure 1D). However, the transcriptional activity of these genes returned to approximately baseline within 24 h, suggesting a short-term early cellular response to IR exposure (Figure 1D).

Short insertion or deletion (indel) mutations induced by irradiation

Analysis of the genomes of the 135 colonies revealed an average of 1,020 SBSs and 282 short indels per colony (Figure 2A). The acquired alterations included IR-induced mutations, IR-unrelated mutations, and technical artifacts from cell culture and sequencing. To sort out IR-induced acquired mutations, we deconvoluted mutational signatures using these alterations. Although IR exposure induces intracellular reactive oxygen species,[10,26] which cause DNA damages and C:G>A:T base substitutions,[27] we did not identify any SBS signatures exclusive to or enriched in the IR exposure group (Figures 2A and S5A; Tables S3, S4, S5, and S6), indicating that IR exposure does not lead to a substantial number of SBSs fixed in the genome of proliferative cells.

In contrast, a specific signature of indel mutations (referred to as ID-A), which was newly delineated from the indel mutations detected in this study, was almost exclusively observed in the

CellPress
OPEN ACCESS

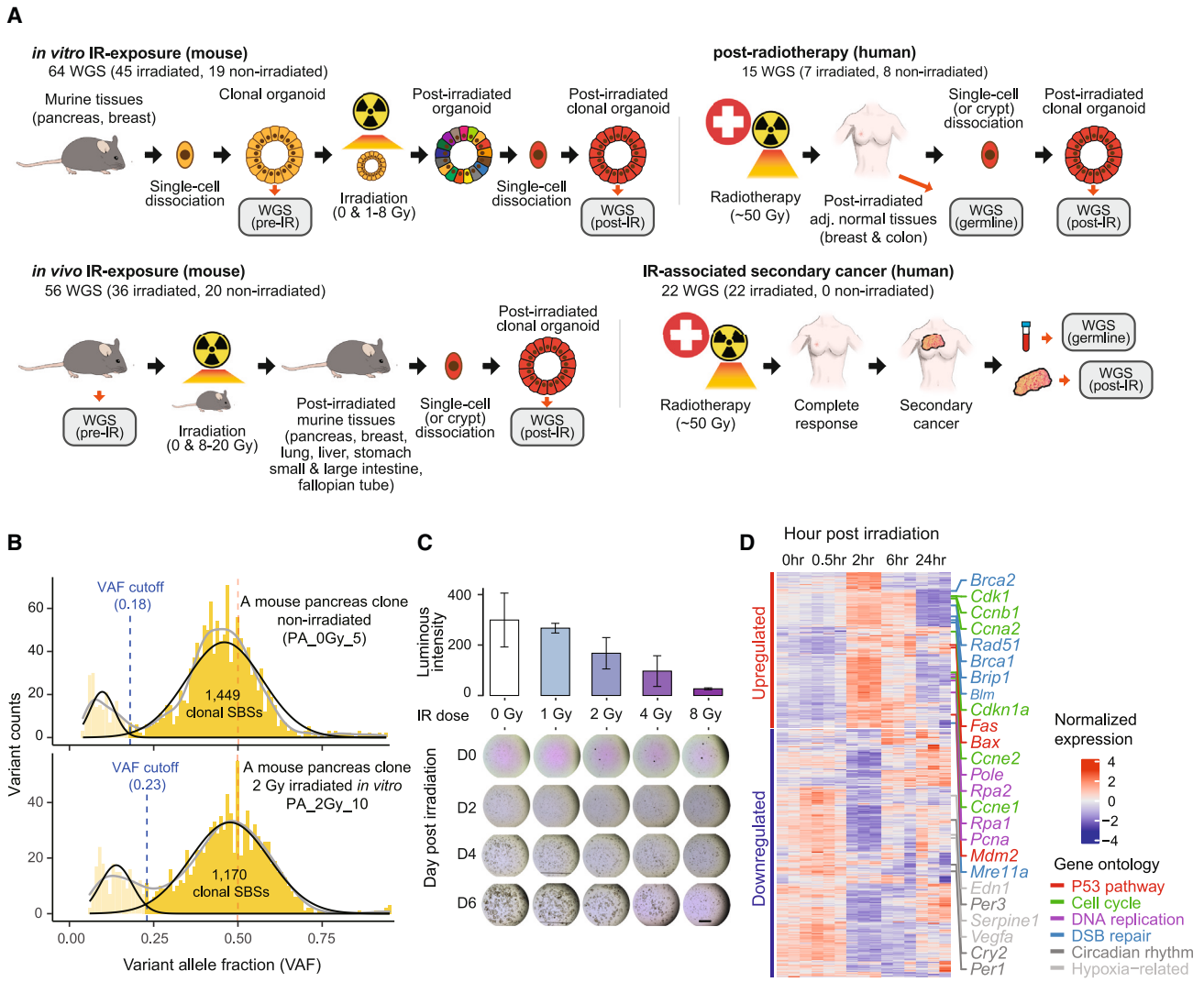


Figure 1. Overview of the experimental design

(A) Experimental design for detecting mutations induced by ionizing radiation (IR) in single cells using organoid culture technique. We collected cells irradiated in four different but complementary experimental settings, including (1) IR exposure on cultured cells (IR^{vitro}; top left), (2) IR exposure on living tissues (IR^{vivo}; bottom left), (3) post-radiotherapy ($IR^{post\text{-}RT}$; top right), and (4) IR-induced secondary cancers. The collected cells from the first three settings were clonally expanded into colonies using organoid techniques followed by WGS. IR-induced secondary malignancies were whole-genome sequenced without the clonalization step. For each experimental setting, we produced baseline (germline) genome sequences to sort out IR-associated somatic mutations from the colonies and cancer tissues.

(B) Variant allele fraction (VAF) distribution of acquired single-base substitutions (SBSs) in non-irradiated (top) and irradiated (bottom) mouse colonies (pancreas). The black solid lines are Gaussian curves fitted to the distribution, and the gray solid line is a density curve.

(C) Organoid-forming efficiencies of irradiated cells (bottom). Scale bar, 2.5 mm. Data are presented as mean ± SEM (the standard error of the mean) (top). n = 3.

(D) Gene expression changes in cells exposed to 2 Gy irradiation for 24 h. Genes in the p53, cell-cycle, DNA replication, and double-strand breakage pair pathways are colored red, green, purple, and blue, respectively.

See also Figures S1–S4 and S8.

irradiated group (shown in purple in Figures 2A and S5B; Tables S7, S8, S9, and S10). Intriguingly, most indels observed in IR-associated secondary cancers were attributed to the ID-A signature (Figure 2A). Two concordant versions of the ID-A signature were independently extracted from murine and human (hID-A) colonies (cosine similarity = 0.924; Figures 2B and 2C). The ID-A signature was characterized by (1) predominantly base deletions (number of deletions:number of insertions = 4.87:1), (2) a broad spectrum with no particular enrichment in a specific sequence context, and (3) a marginally higher level of sequence microhomology than randomly expected (1–4 bp; Figure 2D). Assuming that IR-induced DNA double-strand breaks (DSBs) are randomly distributed genome-wide, the ID-A signature suggests that DNA ligation through end resection is more efficient when two breaks share sequence microhomology (Figures 2D and 2E). Interestingly, a component of the ID-A

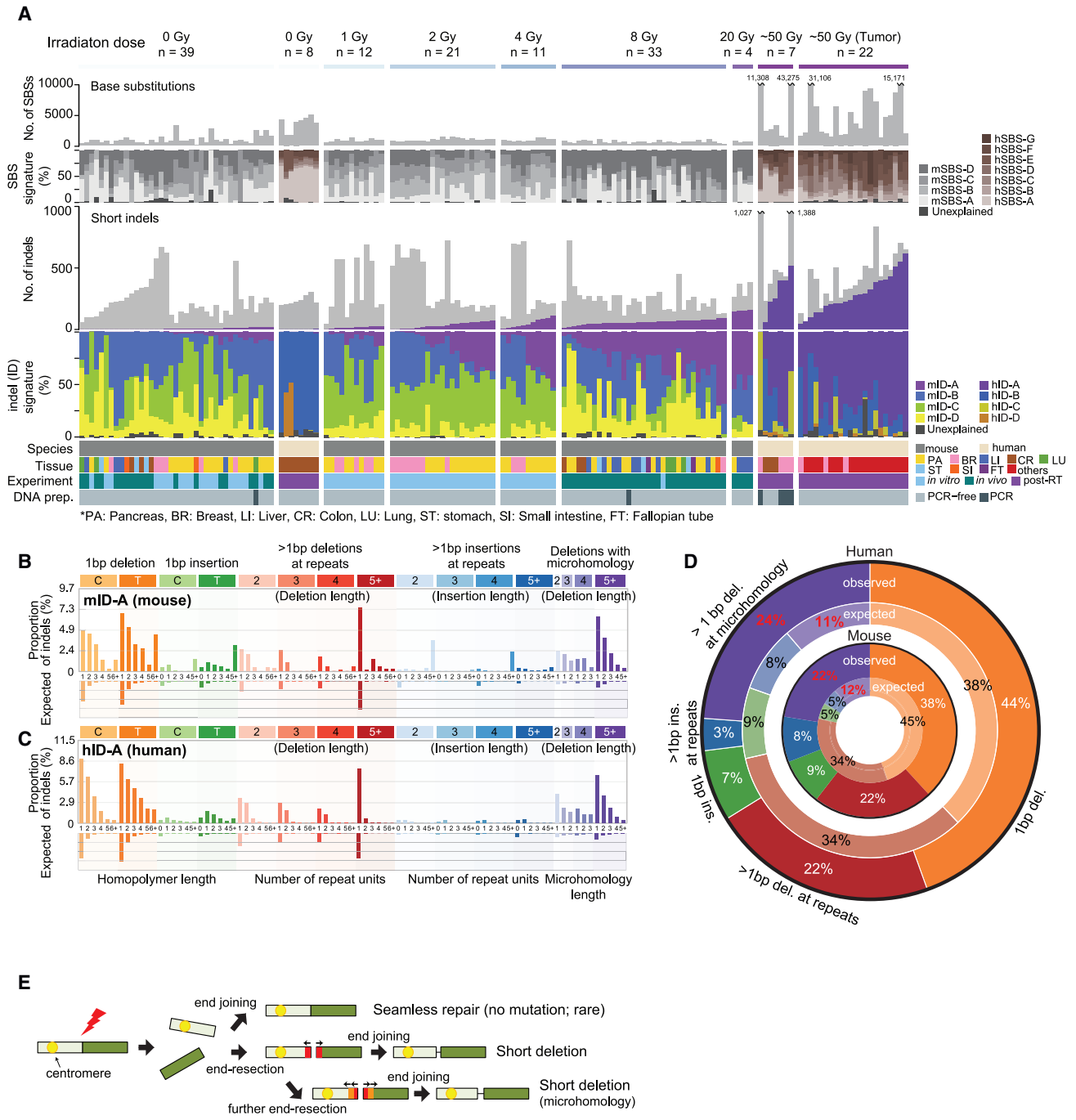


Figure 2. Landscape of SBSs and indels

(A) Summary of the acquired base substitution and indel mutations in the 135 colonies and 22 IR-induced secondary cancers. Stacked bar plots showing absolute number of SBS and indel mutations and the relative proportion of each SBS and indel signature. Full signatures delineated are shown in Figure S5. Annotation tracks include relevant information for each sample. The color codes in the figure legend are used for all subsequent figures.

(B) Mutational spectrum of IR-associated indels delineated in mice (mID-A) (top). Additional mutational signatures (IR unrelated) are shown in Figure S5B. Expected mutational spectrum assuming random mutations is shown with y axis flipped (bottom).

(C) Mutational spectrum of IR-associated indels delineated in humans (hID-A). Additional mutational signatures (IR unrelated) are shown in Figure S5B. Expected mutational spectrum assuming random mutations is shown with y axis flipped (bottom).

(D) Pie chart showing differences between observed and expected indel spectra in mice (inner circles) and humans (outer circles). The expected indel spectrum (lighter color in each circle) was calculated based on the reference genome sequence.

(E) Schematic representation of the formation of IR-induced DNA double-strand breaks (DSBs) and DNA repair processes for IRi-IDs.

See also Figure S5.

spectrum, particularly >1 bp deletions with microhomology (shown in purple in Figures 2B and 2C), appears highly similar to the COSMIC reference signature ID8, the proposed mechanism of which involves non-homologous end joining (NHEJ), a major repair mechanism for DNA DSBs.[2,28] Our spectrum further suggests that DSB repair pathways involving longer sequence homology, such as alternative end joining (2–20 bp microhomology), single-strand annealing (>20 bp homology), and homologous recombination (>100 bp homology), do not dominantly generate IR-induced indels.

SVs induced by irradiation

Analysis of copy-number changes and SVs revealed that chromosome and arm-level copy-number changes were 3.7 times more frequent in colonies from irradiated cells, although the difference was not statistically significant (17.8% vs. 4.9% in irradiated and non-irradiated groups, respectively; p = 0.14, Fisher's exact test; Figure 3A; Table S11). The 135 colonies collectively exhibited 2,596 SV breakpoints (Tables S12 and S13). Some SV classes, including medium-sized deletions (100 bp–1 Mb), tandem duplications, and templated insertions, were not considered IR associated, as they were also frequently present in the non-irradiated colonies (Figure 3A). Except for complex genomic rearrangements (CGRs; described in the following section), seven SV classes, whose number of breakpoints total 966, were substantially enriched in the irradiated colonies and were therefore defined as IR-inducible SVs (IRi-SVs) (Figure 3A). The majority of the IRi-SV breakpoints exhibited 0–4 bp microhomology (94.5%), implying that NHEJ is the primary underlying DNA repair mechanism leading to these rearrangements (Figure 3B).

Of the IRi-SV classes, balanced inversion was the most dominant type, present in approximately 60% of the colonies of the irradiated group. In contrast, colonies of the non-irradiated group did not harbor any inversion events (58.0% vs. 0% for the irradiated and non-irradiated groups, respectively; p = 2.0e−13, Fisher's exact test; Figure 3A). These results corroborate a previous study that reported frequent inversions in IR-mediated secondary malignancies.[18] Similarly, balanced translocation and long-deletion (≥1 Mb) classes were almost exclusive to the irradiated group (balanced translocation, 42.0% vs. 0% for the irradiated and non-irradiated groups, respectively; p = 6.3e−9; long-deletion, 30.7% vs. 2.1%, p = 3.3e−5, Fisher's exact test; Figure 3A). Despite low frequency, extrachromosomal DNA (ecDNA)[30,31] was also exclusively found in the irradiated group (2.3% vs. 0% for the irradiated and non-irradiated groups, respectively; p = 0.54, Fisher's exact test; Figure 3A). Topologically, the formation of these four SV classes involves at least two simultaneous DSBs (Figure 3C).

Colonies derived from irradiated cells exhibited three additional SV classes with more interlaced breakpoints involving at least three simultaneous DSBs (31% vs. 0% for the irradiated and non-irradiated groups, respectively; p < 2.2e−6, Fisher's exact test; Figure 3A). These included deletion-insertion (Del-Ins), deletion-inversion (Del-Inv), and deletion-translocation (Del-Tra) composites, which, to the best of our knowledge, have not previously been systematically defined (Figures 3D and S6A). We identified 17 such variants that were approximately 26-fold more frequent in the 22 IR-induced secondary malignancies than in IR-naive primary lung adenocarcinomas (Figure 3E; p < 2.2e−16, independent two-population proportions test), confirming that these SV classes are likely IR induced. We speculate that SVs involving multiple DSBs are prevalent in cells exposed to IR, as IR can induce multiple DSBs simultaneously in a cell.

CGRs induced by irradiation

In the 135 colonies, we identified 14 CGRs, comprising numerous breakpoints preferentially formed by a single catastrophic event.[32] These CGR events were exclusive to the irradiated group (14.8% vs. 0% for the irradiated and non-irradiated groups, respectively; p = 4.2e−3, Fisher's exact test), implying that these CGRs were induced by IR exposure (Figure 3A). We observed three typical CGR classes, including chromoplexy (n = 4), chromothripsis (n = 2), and SVs by breakage-fusion-bridge (BFB) cycles (BFB-SVs; n = 1).

Chromoplexy, characterized by "closed-chain" multi-chromosomal balanced translocations, is a major mechanism for the formation of oncogenic rearrangements in prostate, lung, bone, and soft-tissue cancers (Figure 4A).[29,33,34] Of note, DSB bursting in transcriptional hubs has been speculated to be an underlying mechanism of chromoplexy.[34] Our findings indicate that IR exposure may serve as a physical trigger for the induction of chromoplexy via the generation of simultaneous multi-chromosomal broken ends and subsequent erroneous rejoining of these ends (Figure 4B).

Two colonies in the irradiated group exhibited chromothripsis, characterized by extensive DNA fragmentation and reassembly localized on one or two chromosomes.[35] One chromothripsis event involved a 28-Mb-long region of chromosome 15, encompassing 47 breakpoints combined with copy-number oscillations between one and three copies in a colony derived from cells exposed to 4 Gy (Figure 4C). We speculate that IR exposure induced two DSBs in chromosome 15, which resulted in excision of the 28-Mb-long segment and the formation of a ring structure (Figure 4D). Due to the absence of a centromere and telomeres, the ring structure is unstable and therefore likely to be massively shattered in subsequent cell divisions.[36] The other chromothripsis event involved joint fragmentation and reassembly of chromosomes 12 and X (Figure 4E). We speculate that IR exposure contributed to disjoining the telomeres from the ends of the two chromosomes (Figure 4F), leading to the formation of a dicentric chromosome. Dicentric chromosomes are also structurally unstable, forming anaphase chromatin bridges during subsequent cell division, which can result in massive fragmentation and thus chromothripsis.[37] Consistent with the mechanism, we observed a localized hypermutation (kataegis) near the chromothripsis breakpoints (Figure S6B) that was acquired when the anaplastic DNA bridge was exposed to cytoplasmic *TREX1*.[37] Finally, we identified a BFB-SV event exhibiting typical patterns, including a stair-like increase in DNA copy number and enrichment of fold-back inversions (Figure 4G).[38] BFB-SVs can arise when two sister chromatids join end to end (Figure 4H). The genomes of IR-associated secondary cancers exhibited frequent CGRs (Figure S7). However, as CGRs also frequently occur in IR-naive primary cancer genomes,[39] it is unclear whether all the CGRs found in the secondary cancers were IR induced.

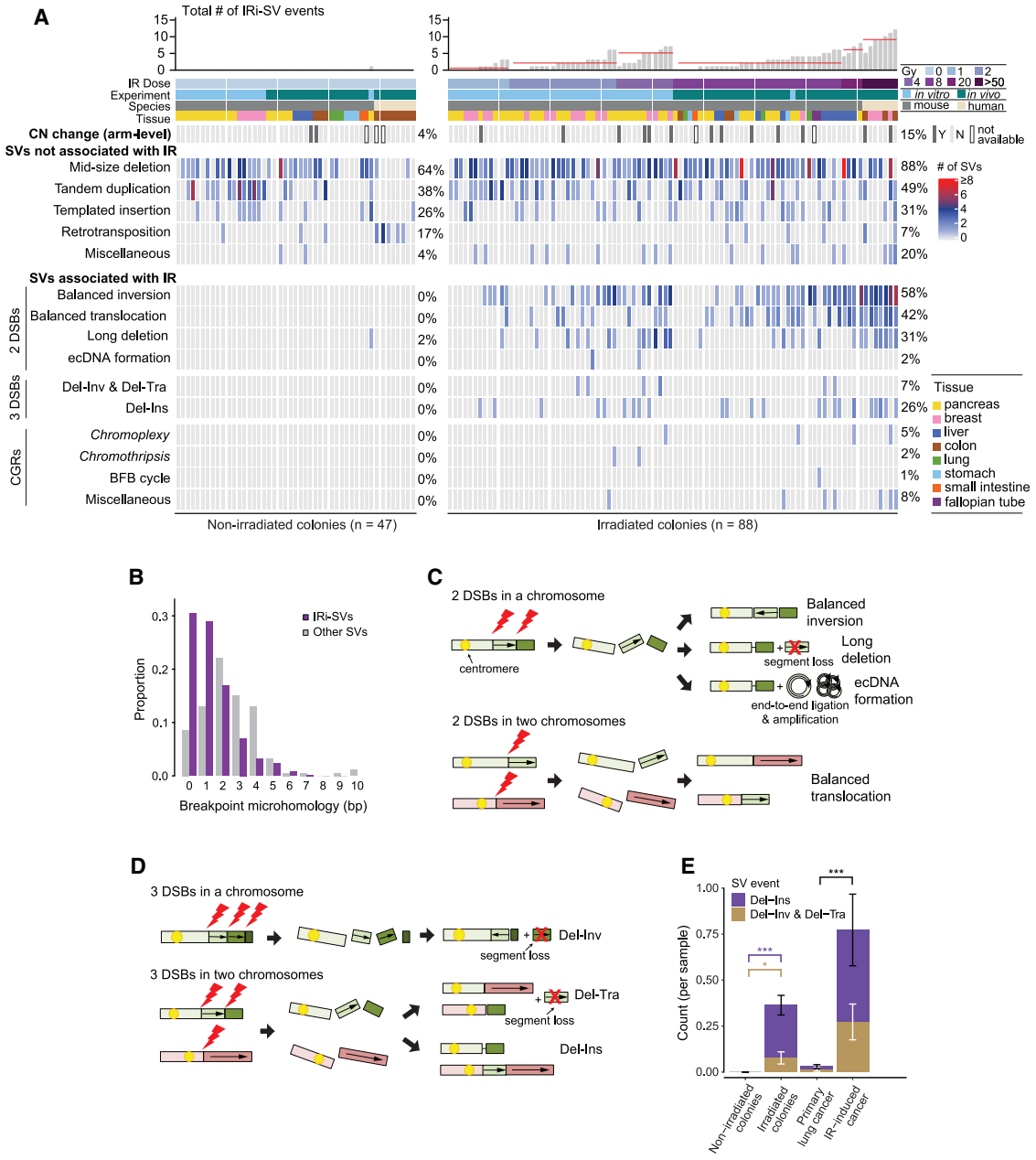


Figure 3. Landscape of structural variations (SVs)

(A) Summary of the acquired SVs in the 135 colonies. Each row indicates SV type, and each column represents the number of SVs observed in each sample. Colonies are classified into two groups: non-irradiated (n = 47) and irradiated (n = 88). SV types considered non-specific to the irradiated group are presented on the top. SV types specific to the irradiated organoids are presented on the bottom. Del-Inv, deletion-inversion composite; Del-Tra, deletion-translocation composite; Del-Ins, deletion-insertion composite.

(B) Size distribution of microhomology in IR-induced SVs (IRi-SVs; purple). Compared to non-IRi-SVs (gray), breakpoints of IRi-SVs harbor shorter microhomology ($p < 0.001$, Kolmogorov-Smirnov test), which supports that non-homologous end joining is a predominant repair process for IR-induced DSBs.

(C) Schematic representation of the consequences of IR-induced double DSBs in one chromosome (top) and in two chromosomes (bottom).

(D) Schematic representation of the consequences of IR-induced triple DSBs in one chromosome (top) and in two chromosomes (bottom).

(E) Frequency of Del-Ins and Del-Inv-and-Del-Tra events in non-irradiated organoids (n = 47), irradiated organoids (n = 88), primary lung cancer (n = 138; data from Lee et al.[29]), and IR-induced secondary cancer (n = 22). ***$p < 0.001$ for Del-Ins and *$p < 0.05$ for Del-Inv-and-Del-Tra, independent two-population proportions test. Data are presented as mean ± SEM.

See also Figure S6.

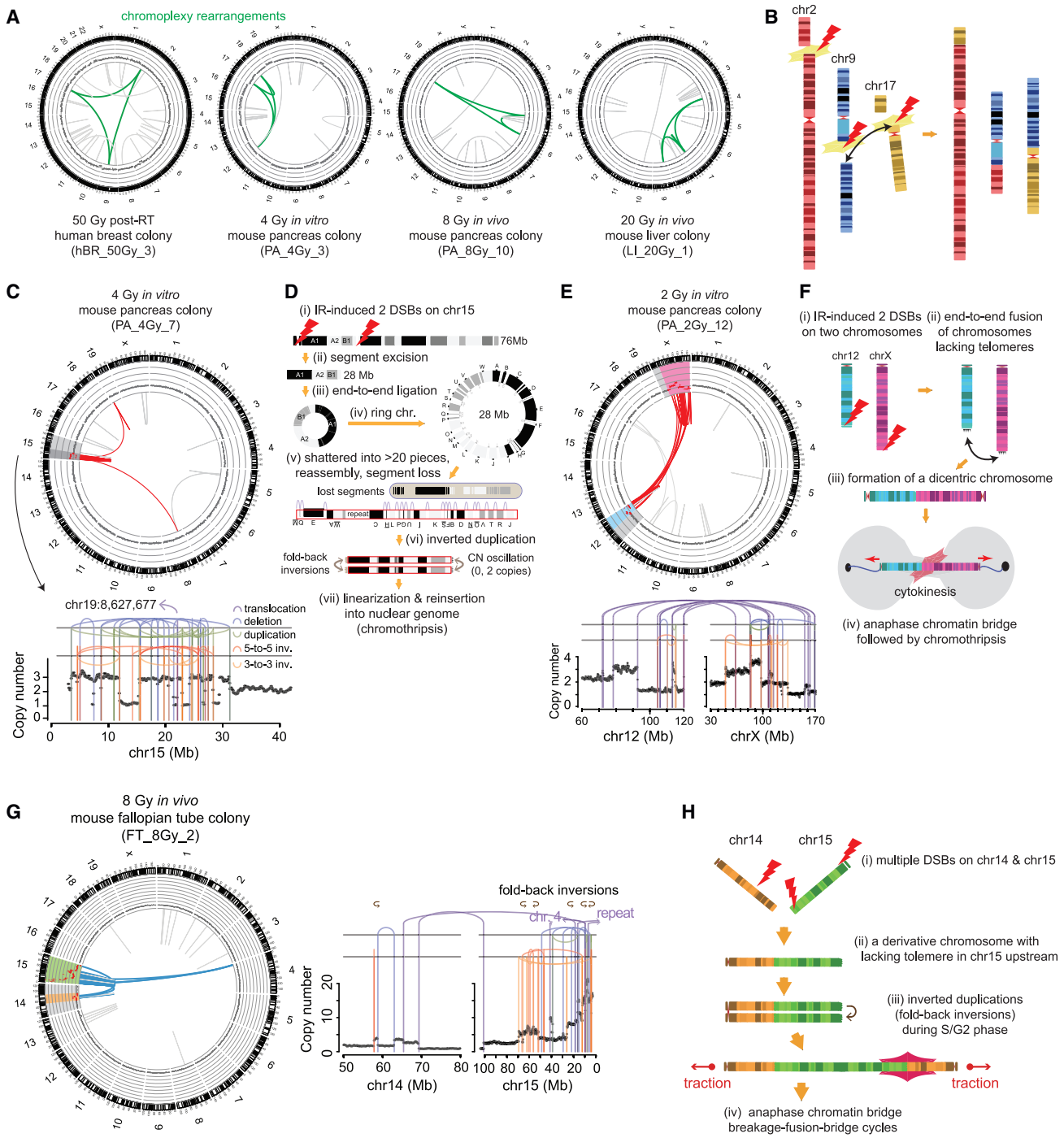


Figure 4. Complex genomic rearrangements

(A) Circos plots of four post-irradiated colonies exhibiting chromoplexy.

(B) Schematic representation of the mechanism of chromoplexy observed in a human breast colony (hBR_50Gy_3).

(C) Circos plot showing chromothripsis (red lines) on chromosome 15 in a 4 Gy irradiated organoid (PA_4Gy_7; top). The copy-number state in the catastrophic segment oscillates between three and one (bottom).

(D) Schematic representation of a possible mechanism of chromothripsis observed in the 4 Gy irradiated organoid (PA_4Gy_7).

(E) Circos plot showing chromothripsis (red lines) localized in chromosomes 12 and X found in a 2 Gy irradiated organoid (PA_2Gy_12; top) and patterns of rearrangements and copy-number states (bottom).

(legend continued on next page)

Quantitative impact of IR exposure in colonies

The number of IRi indels (IRi-IDs), attributable to the ID-A signature, and IRi-SVs was positively correlated with IR dose (Figures 5A and 5B). Although we observed a huge level of intercellular variation, colonies derived from IR^{vitro} experiments exhibited 2.15 IRi-IDs per Gy per Gb of the genome (Figure 5A). In comparison, colonies derived from tissues exposed *in vivo* (IR^{vivo} and $IR^{post\text{-}RT}$) and secondary tumor samples exhibited slightly lower IRi-ID mutation rates (1.44 and 1.23 IRi-IDs per Gy per Gb, respectively), which may be attributed to stronger negative selection in these experiments. We speculate that cells with an excessive number of DSBs were less viable, making them not sequenceable due to decreased cell viability after IR exposure (Figure 1C). The negative selection of hypermutator cells in tissues was observed in a previous study, which reported the removal of hypermutated cells from the bronchial epithelium in ex-smokers.[40]

DNA DSB repair may result in either a seamless repair leaving no mutations or a mutation-forming repair. To estimate the rate of seamless repair, we investigated breakpoints in IRi-SVs and found predominantly end resections with gaps (approximately 87%; 1–20 bp) but some ligations with base insertions (approximately 7%; 1–9 bp) or without gaps (repair without end resection; approximately 6%) (Figure 5C). If the seamless repair rate of approximately 6% applies to indel formation, the rate of seamless DSB repair would be 0.15 per Gy per Gb for colonies derived from IR^{vitro} experiments.

The IRi-SV rate in colonies derived from IR^{vitro} experiments was 0.177 per Gy per Gb (Figure 5B), approximately 8% of the rate of IRi-IDs (Figure 5D). This lower rate of SVs may be attributed to the more complex mechanism involved in SV formation than indel mutations, requiring two or more simultaneous DSBs and their end joining in an erroneous configuration. Interestingly, the relative frequency of IRi-SVs and IRi-IDs was consistent within colonies derived from the IR^{vitro} experiments within the IR dose range of 1–4 Gy (Figure 5D). However, the ratio was reduced in colonies derived from IR^{vivo} and $IR^{post\text{-}RT}$, which were exposed to a higher IR dose (8–50 Gy) (Figure 5D). As speculated in the IR-induced indel rate, it may be attributed to negative selection that weeds out cells harboring a large number of DSBs. Cells that acquired many simultaneous DSBs, which are more likely to generate SVs, will be less viable, as SVs are generally more deleterious than indels.

There were no significant differences in IR-associated mutational burden of the different cell types, although colon-derived colonies exhibited slightly lower mutational burdens for both IRi-IDs and IRi-SVs (Figure 5E). There results imply that most cell types have similar DNA repair capacities in response to IR exposure, despite different background transcriptional and epigenetic profiles.

To investigate whether the irradiation dose rate can change the mutational impact, we established four colonies from mice using 8 Gy gamma irradiation at a very low IR rate (approximately 10e−4-fold; with approximately 5.5e−5 Gy/min continuously for 100 days). We found that the number of IRi-IDs and IRi-SVs in these colonies was not significantly different from those observed in the IR^{vivo} experiments using 0.714 Gy/min (Figure 5F), suggesting that the total dose of IR exposure, rather than the dose rate, is the determining factor for the mutational burden, at least within the given range of dose rates.

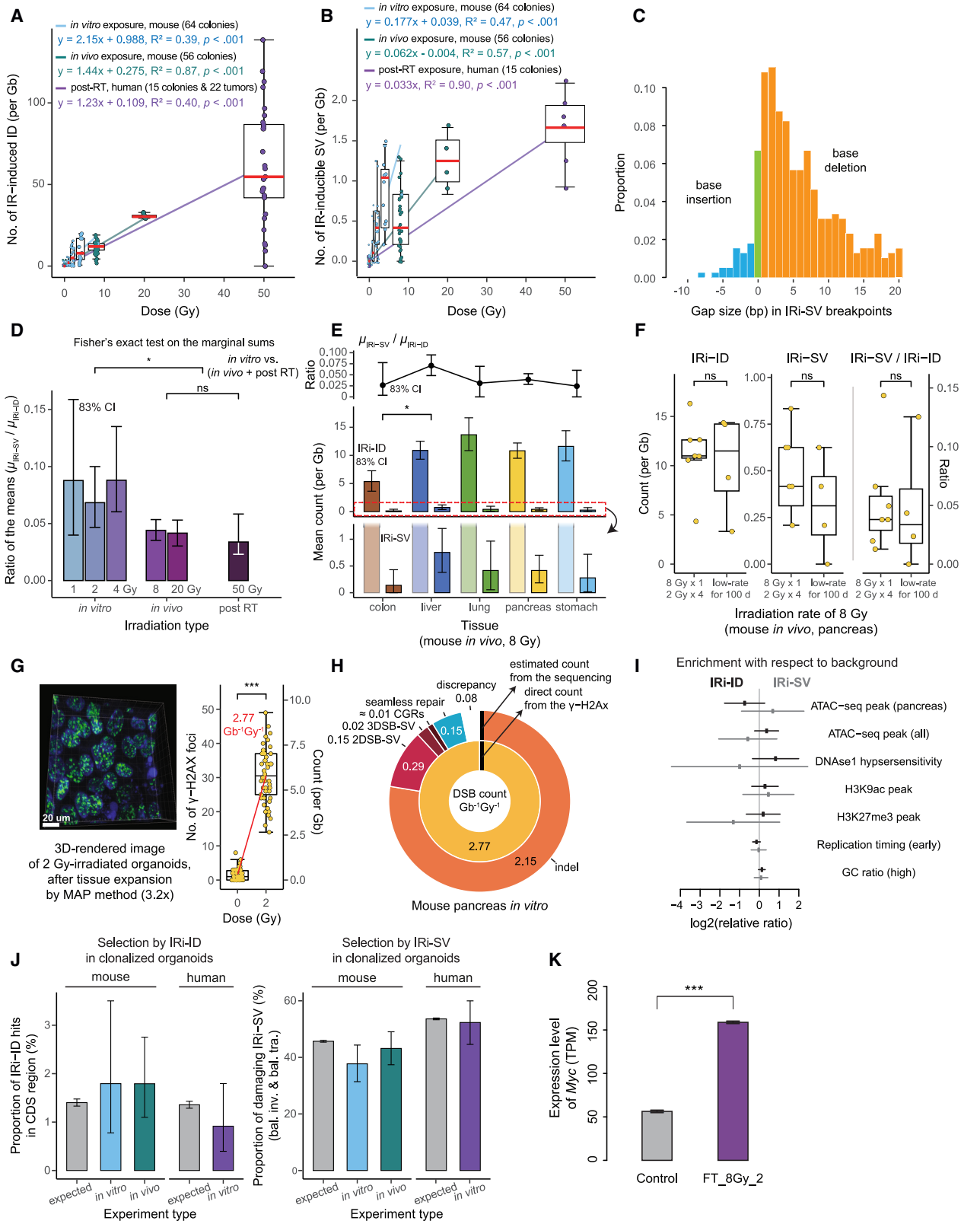
Finally, we compared the number of fixed IR-associated mutations with the number of acquired DSBs in the post-irradiated cells. To accurately measure the number of DSBs induced by IR irradiation, we counted gamma-H2AX foci in 50 cells exposed to 2 Gy *in vitro* using three-dimensional super-resolution imaging after physical expansion of cells.[42] After adjusting for background levels, we found that 1 Gy IR generated 2.77 DSBs per Gb in mouse cells (Figures 5G and 5H). The DSB burden was comparable to the rate of observed IRi mutations in this study, including 2.15 indels, 0.15 double-DSB-induced SVs (inversion, translocation, long deletion, and ecDNA), 0.02 triple-DSB-induced SVs (Del-Ins, Del-Inv, and Del-Tra), 0.01 CGR events, and 0.15 seamless repairs (Figure 5H), indicating that our clonalization-based approaches allowed sensitive detection of IR-associated mutations and accurate estimation of the mutation rate.

Genomic distribution and selection of IR-induced mutations in colonies

The IRi-IDs and IRi-SVs detected in the colonies were widely distributed throughout the genome and were not strongly associated with open/closed chromatin, specific histone markers, replication timing, or local GC content (Figure 5I). The functional impact of IR-associated mutations was close to random expectation. For example, approximately 1.8% and 0.9% of the indels involved protein-coding sequences in mice and humans, respectively, which was not significantly different from the background distribution (Figure 5J). Similarly, the proportions of gene-truncating balanced inversion and translocation were also close to random expectation (Figure 5J). These results indicate that negative selection of post-irradiated cells is not operative on pathogenic mutations after DNA repair but on the number of DSBs prior to the completion of DNA repair.

Finally, we examined whether any IR-associated mutations might have conferred selective advantage during clonal expansions. The 88 colonies from post-irradiated cells exhibited few gain-of-function mutations in known oncogenes and loss-of-function mutations in tumor-suppressor genes. There was only one event in which the *Myc* oncogene was amplified via an IR-induced BFB-SV event (Figure 4G), which resulted in transcriptional upregulation from 56 to 159 transcripts per million ($p < 0.001$, two-sample t test; Figure 5K). These findings suggest that IR-induced mutations in the colonies are not particularly biased during our technical clonalization steps.

(F) Schematic representation of the mechanism of chromothripsis observed in the 2 Gy irradiated organoid (PA_2Gy_12).
(G) Circos plot showing a BFB-SV identified in an 8 Gy irradiated human organoid (FT_8Gy_2; left) and patterns of rearrangements and copy-number states (right).
(H) Schematic representation of the mechanism of the BFB cycle observed in the 8 Gy irradiated human organoid (FT_8Gy_2).
See also Figure S7.



(legend on next page)

Direct sequencing to reduce bias in clonalization

To further evaluate potential selection bias during clonalization, we used various direct single-cell genome sequencing strategies, although such sequencing strategies generally exhibit a low signal-to-noise ratio and therefore limited sensitivity and specificity for detecting *de novo* somatic mutations.[43] First, 53 single cells were analyzed using whole-genome amplification (WGA) with multiple displacement amplification[44] (MDA; 6 single cells), WGA using primary-template-directed amplification[45] (PTA; 12 single cells), and Strand-seq[46] (35 single cells) (Figure 6A). The WGA approaches focused on indels and SVs, whereas Strand-seq targeted large-scale SVs. Second, we implemented two duplex sequencing strategies, including bottleneck sequencing[47] (10 bulk organoids) and Concatenating Original Duplex for Error Correction[48] (CODEC; 6 bulk organoids), to accurately detect IR-induced indels in post-irradiated bulk tissues without clonalization (Figure 6A). For IR exposure for the experiments, we used the ^{137}Cs source.

These techniques have variable capabilities for detection of IR-induced mutations. WGA using MDA of six single cells (three non-irradiated and three irradiated with 2 Gy) led to numerous amplification artifacts (11,603 SBS calls per single cell, 95% confidence interval [CI] [9,150–14,056], which is approximately 11.4-fold higher than SBSs detected from a colony, 95% CI [8.93–14.4]), unstable DNA copy numbers, and noisy distribution of the VAF of mutation calls (Figures 1B and S8A). Although stringent filter conditions to reduce erroneous calls in MDA provided an overall higher burden of breakpoint calls in post-irradiated cells (Figures 6B, S8B, and S8C), precise evaluation of the mutation rate and characterization of the mutational signature were not feasible due to the high background noise in non-irradiated cells.

Compared to MDA, WGA using PTA of 12 single cells (five non-irradiated and seven irradiated with 4 Gy) achieved more stable mutation calls (630 SBS calls per single cell, 95% CI [334–974], which was comparable to the number observed in the colonies; Figures S9A and S9B). The indel calls allowed deconvolution of an indel signature highly enriched in irradiated cells (PTA-hID-A), which was similar to the ID-A signatures extracted from the colonies (Figures 6C and S9C). Considering the average indel dropout rate, 4 Gy irradiated single cells acquired a median of 44 more hID-A indel calls (95% CI [6–130]) (Figures 6D and S9D), with the estimated rate (1.81 per Gy per Gb, 95% CI [0.15–5.23]) not significantly different from the rate observed in the colonies. The seven post-irradiated single cells analyzed using PTA exhibited a median of 7.1 more SVs per genome (approximately 0.4 SVs per Gy per Gb, 95% CI [0–1.25]; Figure 6E), also not meaningfully different from the colonies. Of note, one single cell had approximately 80 adjusted SV breakpoints genome-wide, implying that this was a hypermutator cell with numerous IR-induced DSBs.

Single-genome data from Strand-seq of 35 single cells (14 non-irradiated and 21 irradiated with 4 Gy) provided more robust insight into SV formation in non-clonalized single cells (Figure 6F). We found that post-irradiated cells had a higher SV burden than controls, including 2.47 long deletions (>8 Mb; 0.1/Gy/Gb), 1.38 inversions (>10 Mb; 0.05/Gy/Gb), and 2.26 sister-chromatid-exchange-like patterns (SCE-like patterns; 0.1/Gy/Gb) (Figure 6G). Except for SCE-like patterns, which can be challenging to interpret, as they may not represent real SVs (Figure S10), Strand-seq detected approximately 0.16 SVs (95% CI [0.11–0.20]), which is closely aligned with the rate observed in the colonies. In contrast to PTA, no cell with an unusually high number of SVs was observed in the Strand-seq data.

Figure 5. Quantitative analysis of radiation-associated indels and SVs

(A) Number of IR-induced indels per Gb with respect to radiation dose (Gy). Linear regression was used to estimate dose-response relationships. The data were divided into three subsets (experiments from IR^{vitro}, IR^{vivo}, and $IR^{post\text{-}RT}$). Boxplot shows median (midline; red), interquartile range (IQR) (box), and whiskers indicating minimum or maximum value within 1.5×IQR from the quartiles.

(B) Number of IRi-SVs per Gb with respect to radiation dose (Gy). Linear regression was used to estimate dose-response relationships. The data were divided into three subsets as shown above. Boxplot whiskers extend up to 1.5×IQR from the quartiles.

(C) Size distribution of breakpoint gap in IRi-SVs. IR-induced DSBs are usually accompanied by some nucleotide deletions.

(D) Ratio of two means ($\mu_{IRi\text{-}SV}/\mu_{IRi\text{-}ID}$) with respect to irradiation type (IR^{vitro}, IR^{vivo}, and $IR^{post\text{-}RT}$). Marginal sums were compared between (1) IR^{vitro} vs. IR^{vivo} groups and (2) IR^{vitro} vs. IR^{vivo} and $IR^{post\text{-}RT}$ groups using Fisher's exact test. The data are shown as the ratio of means with an 83% confidence interval (CI) calculated by Fieller's method[41]; *$p < 0.05$, not significant (ns) > 0.1.

(E) IRi-ID (middle), IRi-SV (bottom), and the ratio of the means ($\mu_{IRi\text{-}SV}/\mu_{IRi\text{-}ID}$; top) of *in vivo*, 8 Gy irradiated organoids with respect to tissue types. Only tissues with more than two data points were included. Fisher's exact test on the sum of IRi-IDs and IRi-SVs shows a significant difference between the colon and liver; *$p < 0.05$. Error bars indicate 83% CI.

(F) IRi-ID, IRi-SV, and the ratio (IRi-SV/IRi-ID) between high dose rate (one or four short exposures) and low dose rate (continuous exposure for 100 days) with a total irradiation of 8 Gy. The high rate involved irradiating a single 8 Gy or four 2 Gy doses, and the low rate involved approximately 0.08 Gy per day for 100 days; ns > 0.1. Boxplot whiskers extend up to 1.5×IQR from the quartiles.

(G) Representative image of gamma-H2AX immunofluorescence staining (green) in 2 Gy irradiated mouse pancreas organoids following 3.2× tissue expansion using magnified analysis of proteome method (left). Number of gamma-H2AX foci per cell in control and 2 Gy irradiated samples was counted; ***$p < 0.001$, two-sample t test. Boxplot whiskers extend up to 1.5×IQR from the quartiles.

(H) Pie chart showing direct DSB count from the gamma-H2AX experiment (inner circle) and estimated DSB count from the sequencing data (outer circle). We estimated DSB counts from the sequencing data using the number of variants and their minimum number of DSBs necessary (for example, 1 DSB for an indel, 2 DSBs for an inversion, or 3 DSBs for composite SV). The number of variants was based on the coefficient of the linear model for the IR^{vitro} experiment (Figures 5A and 5B). The number of seamless indels was estimated from the proportion of the SV without a gap and the number of indels (Figure 5C). The counts were adjusted to account for inaccessible and repetitive genomic regions.

(I) Enrichments of IRi-IDs and IRi-SVs in relation to genomic contexts. For an enrichment in the ATAC-seq (assay for transposase-accessible chromatin with sequencing) from the pancreas, SVs acquired from the pancreas colonies were used. Error bars indicate 95% CI.

(J) Functional consequences of IRi-IDs (left) and IRi-SVs (right) detected from colonies. Error bars indicate 95% CI.

(K) The difference of *Myc* expression in control and irradiated organoids (FT_8Gy_2). Data are presented as mean ± SEM. n = 3. TPM, transcripts per million.

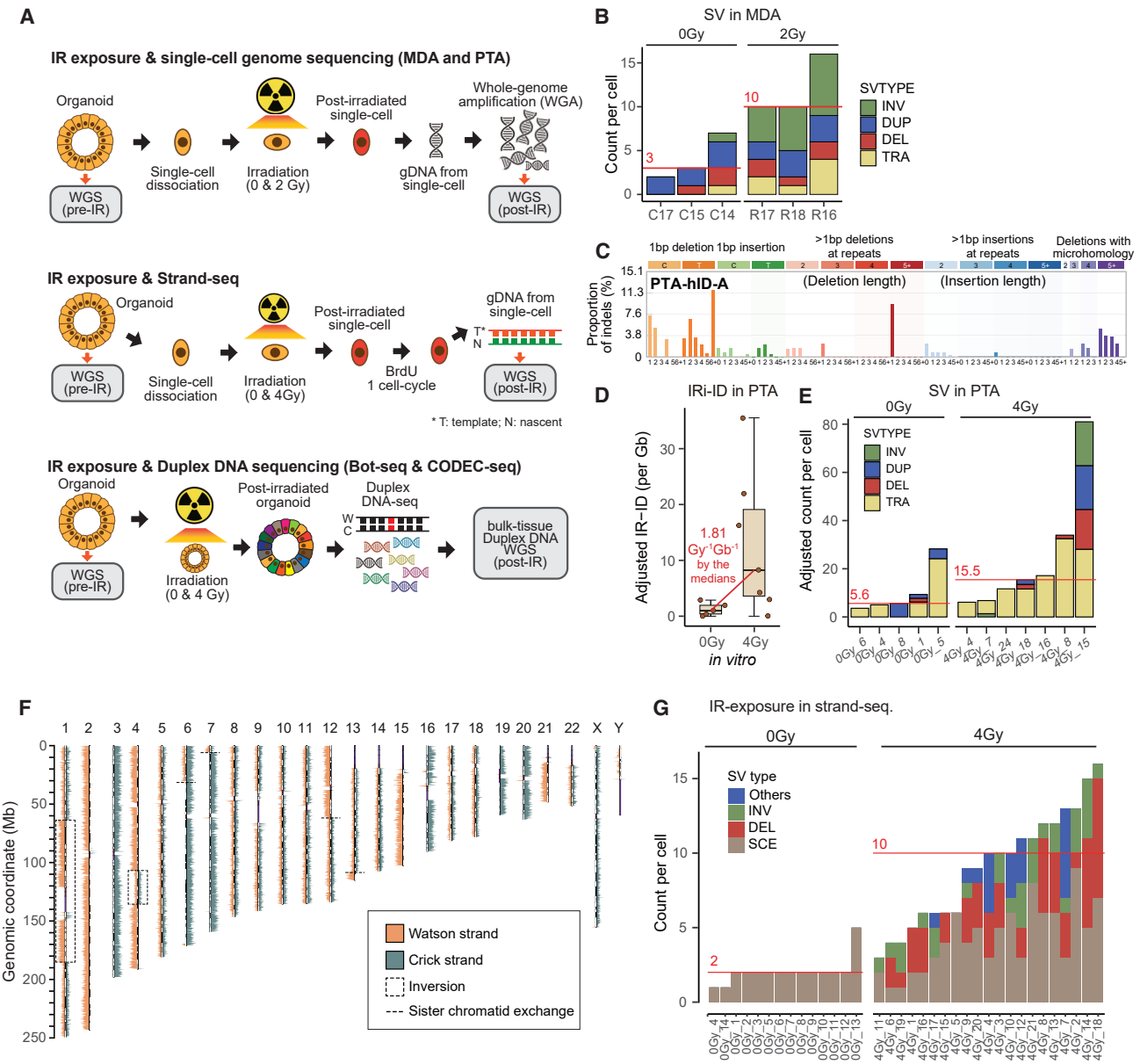


Figure 6. Direct single-cell genome sequencing

(A) Experimental design for detecting IR-induced mutations in single cells without clonal expansions, using WGS of the whole-genome-amplified single cells (MDA and PTA), Strand-seq, and duplex DNA sequencing (bottleneck sequencing and Concatenating Original Duplex for Error Correction) techniques.

(B) SVs detected in the MDA experiments.

(C) Mutational signature of IR-associated indels delineated from the PTA experiments. Full signatures are shown in Figure S9.

(D) Number of indels observed in the PTA experiments, adjusted for allelic dropout rates. Boxplot whiskers extend up to 1.5×IQR from the quartiles.

(E) Number of SVs observed in the PTA experiments, adjusted for allelic dropout rates.

(F) Example of Strand-seq result of a 4 Gy irradiated single cell.

(G) Number of SVs and sister-chromatid-exchange-like patterns observed in the Strand-seq experiments.

See also Figures S9–S11.

Finally, we applied the bottleneck sequencing and CODEC (both are duplex DNA WGS methods) to detect IR-induced point mutations in post-irradiated bulk organoids. Despite much lower sequencing error rates reported[47,48] (10e−6 to 10e−7 errors per bp), these methods still generated a significant amount of background noise, masking IR-induced point mutations (Figure S11). Of note, an IR dose of 4 Gy generated ~50 indels in the diploid genome, which is equivalent to 8 × 10e−9 indels per bp. Such an amount of indels may be too low to be precisely detected through the duplex DNA sequencing methods.

Collectively, our observations from the direct single-cell sequencing data suggested that (1) these techniques were not fully competent to comprehensively and precisely detect IR-induced mutations and (2) the best estimates of IR-induced mutation rates from these techniques were not substantially different from the rate observed in the colonies.

DISCUSSION

In this study, we investigated the mutational impact of IR on normal cells at nucleotide-level resolution. Despite well-established mutagenicity, comprehensive mutational signatures of IR in normal somatic cells have not been elucidated due to technical challenges posed by single-cell genomic analysis. Although WGA techniques present a technically simple, direct approach[7,9] and duplex DNA sequencing can be applied to bulk cells,[47,49] these techniques are insufficient for accurately capturing small numbers (10–100) of IR-induced somatic mutations acquired in post-irradiated cells, as evidenced by our single-cell and duplex genome sequencing results. Particularly, SVs, which are the main mutations induced by IR, are challenging to detect using these techniques. Strand-seq is also insufficient, as it only detects large-scale SVs without breakpoint sequences. Although our clonalization-based approach was labor intensive and only applicable to cells with sustained proliferation potential, it allowed accurate and precise detection of mutations in single cells, which is essential for obtaining accurate, comprehensive mutational signatures.

The acquisition of indels and inversions after IR exposure was previously observed in a genome-sequence study of IR-associated secondary cancers.[18] However, a clear picture of mutational signatures and rare classes of SVs were not elucidated due to challenges in identifying direct mutations induced by IR exposure amid coinciding, unrelated mutations in cancer genomes. Furthermore, although cytogenetic investigation of post-irradiated cells may provide evidence of some rare SVs, it inevitably overlooks information regarding SV breakpoints and lacks the resolution required for identifying indel mutations.[50] This study is therefore the first to report a comprehensive profile of IR mutational signatures in normal mammalian somatic cells.

The quantitative mutational impact of IR observed in the colonies requires cautious interpretation, as cells that have lost proliferation potential, and are more likely to harbor a higher mutational burden, could not be investigated. The mutation rate observed in this study may therefore underestimate the absolute mutational burden in post-irradiated cells or tissues. Indeed, results from the single-genome techniques indicated a slightly higher mutational burden, although the observed rate may not be precise due to high background noise. As severely damaged cells are negatively selected in our tissues after real-world IR exposure, our results from the clones likely reflect the biological effects of IR more accurately.

Compared to IRi-IDs, the IRi-SV burden in post-irradiated cells is more variable across cells, as CGRs are stochastically formed via unstable intermediate chromosomal structures during the repair of multiple DSBs. Surprisingly, approximately 15% of cells irradiated with $\geq$1 Gy acquired catastrophic CGRs, including chromoplexy (approximately 5%), chromothripsis (approximately 2%), and BFB-SV (approximately 1%). These CGRs have a greater functional impact than indels and simple SVs, contributing to the development of IR-associated secondary malignancy to a greater degree. Although it is quite challenging to accurately estimate the functional consequences of CGRs, if a substantial fraction of CGRs detected here activate cellular proliferation, then the frequency of CGRs in this study could be overestimated.

During typical RT, millions of normal cells in adjacent target tissues are exposed to at least 50 Gy IR. Our results from colonies established from post-RT cells imply that approximately 300 mutations are acquired in every cell under such circumstances. Some of these post-RT cells will acquire pathogenic variants, such as gain-of-function mutations in oncogenes or loss-of-function mutations in tumor-suppressor genes. Mutant cells harboring these pathogenic alterations may facilitate disease development, as well as silently persist until sufficient acquisition of additional mutations allow cellular transformation. Future studies tracing phenotypic changes in mutant cells carrying functionally important mutations *in vivo* will provide further insights into the tumorigenesis of IR-associated secondary malignancies.

Limitations of the study

This study had some limitations. Most IR-induced mutations observed in this study were identified in single-cell expanded colonies (organoids), therefore representing genomic alterations acquired in adult stem cells with proliferative potential. As mentioned in the results, post-irradiated cells that acquired an excessive number of DSBs may have been underrepresented in the study. In addition, mutations that activate or inactivate cell proliferations could be over- or under-detected, respectively. Furthermore, mutations acquired in terminally differentiated cells, such as neurons, could not be measured. A few direct genome sequencing techniques applied in the study exhibited insufficient competence to accurately detect IR-induced mutations. Exploring a larger number of genomes collected from diverse tissues through more advanced single-genome techniques will be necessary to achieve an unbiased assessment of the qualitative and quantitative mutagenic impact of IR exposure across various cell types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Experimental mice
  - Human normal samples
  - Human cancer samples
  - Pathology review

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100499.

ACKNOWLEDGMENTS

The authors thank Jinwoo Seong and Young-Yun Kong (Seoul National University) for mouse mammary gland acquisition and Myungsuk Choi (KAIST) for technical help. This work was supported by the Suh Kyungbae Foundation (SUHF-18010082 to Y.S.J.); Young Investigator Grants from the Human Frontier Science Program (RGY0071/2018; to Y.S.J. and B.-K.K.); the National Research Foundation of Korea (Leading Researcher Program NRF-2020R1A3B2078973 and NRF-2020M3A9E4039670 to Y.S.J.; for the Brain Pool Program NRF-2019H1D3A2A02061168 to Y.S.J. and S.Y.K.; Basic Science Research Program NRF-2020R1I1A1A01072873 to K.S.K.; and NRF-2017R1A2B4003535 to J.H.C. and NRF-2021R1G1A1009606 to H.W.K.); the National R&D Program funded by the Ministry of Science, ICT, and Future Planning through the Dongnam Institute of Radiological & Medical Sciences (DIRAMS) under grants 50491-2018 and 50491-2019 (to T.G.S.); and research grants from Korea University Anam Hospital, Seoul, Republic of Korea (grant nos. O1800741 and O2000681 to H.W.K.).

AUTHOR CONTRIBUTIONS

Y.S.J., T.G.S., K.S.K., J.Y., and H.W.K. designed the experiments; H.W.K., T.G.S., H.K., Y.-R.K., C.G.L., and M.J.B. conducted the *in vitro* and *in vivo* mouse irradiation experiments; K.S.K., O.V.Y., and E.C.H. provided adjacent normal human tissue samples; E.K. and J.H.C. provided secondary malignancy samples; K.S.K., E.K., and J.H.C. oversaw all clinical data collection and curation; J.Y., H.W.K., S.J., E.-S.L., T.Y., and T.G.S. performed organoid culture, clonal expansion, and DNA/RNA extraction, with S.H.C. and D.S.L. providing advice; S.K. (pancreas), J.C. (lung), H.N. (Wnt), J.-H.L. (lung), B.-K.K. (Wnt, colon), and H.L. (pancreas) provided training in organoid culture technologies; H.W.K. and Y.C. performed cell viability analysis and organoid staining; D.-W.M., T.-Y.K., H.W.K., J.-G.K., Y.S.K., and T. Ku performed organoid expansion and obtained high-resolution confocal images; T. Kim and C.H.N. conducted single-cell WGS analysis; Y.A., W.-C.L., and J.-Y.S. conducted duplex sequencing analysis; J.Y. performed most bioinformatic analyses, including alignment, mutation calling, and data curation, with S.P., K.Y., R.K., and Y.S.J. providing advice; J.L. and J.Y. performed downstream data analysis, refinement, and interpretation, with Y.S.J. providing advice; J.L. conducted most of the statistical and statistical learning analyses, including mutational signature analysis; S.Y.K. oversaw statistical analysis; and J.Y., H.W.K., J.L., K.S.K., T.G.S., and Y.S.J. wrote the manuscript. All authors finalized the manuscript. Y.S.J. supervised the project.

DECLARATION OF INTERESTS

Y.S.J. is a genomic co-founder and chairman of Genome Insight.

Received: December 6, 2022
Revised: October 23, 2023
Accepted: January 9, 2024
Published: February 14, 2024

REFERENCES

1. Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.
2. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101.
3. Degasperi, A., Zou, X., Amarante, T.D., Martinez-Martinez, A., Koh, G.C.C., Dias, J.M.L., Heskin, L., Chmelova, L., Rinaldi, G., Wang, V.Y.W., et al. (2022). Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. Science *376*, abl9283.
4. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. Science *354*, 618–622.
5. Lindberg, M., Boström, M., Elliott, K., and Larsson, E. (2019). Intragenomic variability and extended sequence patterns in the mutational signature of ultraviolet light. Proc. Natl. Acad. Sci. USA *116*, 20411–20417.
6. Hoang, M.L., Chen, C.H., Sidorenko, V.S., He, J., Dickman, K.G., Yun, B.H., Moriya, M., Niknafs, N., Douville, C., Karchin, R., et al. (2013). Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. Sci. Transl. Med. *5*, 197ra102.
7. Kucab, J.E., Zou, X., Morganella, S., Joel, M., Nanda, A.S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S.P., et al. (2019). A compendium of mutational signatures of environmental agents. Cell *177*, 821–836.e16.
8. Volkova, N.V., Meier, B., González-Huici, V., Bertolini, S., Gonzalez, S., Vöhringer, H., Abascal, F., Martincorena, I., Campbell, P.J., Gartner, A.,

and Gerstung, M. (2020). Mutational signatures are jointly shaped by DNA damage and repair. Nat. Commun. *11*, 2169.

9. Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H.M., Nomburg, J., Gurjao, C., Manders, F., Dalmasso, G., Stege, P.B., et al. (2020). Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature *580*, 269–273.
10. Lomax, M.E., Folkes, L.K., and O'Neill, P. (2013). Biological consequences of radiation-induced DNA damage: relevance to radiotherapy. Clin. Oncol. *25*, 578–585.
11. St Charles, J., Hazkani-Covo, E., Yin, Y., Andersen, S.L., Dietrich, F.S., Greenwell, P.W., Malc, E., Mieczkowski, P., and Petes, T.D. (2012). High-resolution genome-wide analysis of irradiated (UV and γ-rays) diploid yeast cells reveals a high frequency of genomic loss of heterozygosity (LOH) events. Genetics *190*, 1267–1284.
12. Meier, B., Volkova, N.V., Hong, Y., Bertolini, S., González-Huici, V., Petrova, T., Boulton, S., Campbell, P.J., Gerstung, M., and Gartner, A. (2021). Protection of the C. elegans germ cell genome depends on diverse DNA repair pathways during normal proliferation. PLoS One *16*, e0250291.
13. Wang, S., Meyer, D.H., and Schumacher, B. (2023). Inheritance of paternal DNA damage by histone-mediated repair restriction. Nature *613*, 365–374.
14. Adewoye, A.B., Lindsay, S.J., Dubrova, Y.E., and Hurles, M.E. (2015). The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. Nat. Commun. *6*, 6684–6688.
15. Yeager, M., Machiela, M.J., Kothiyal, P., Dean, M., Bodelon, C., Suman, S., Wang, M., Mirabello, L., Nelson, C.W., Zhou, W., et al. (2021). Lack of transgenerational effects of ionizing radiation exposure from the Chernobyl accident. Science *372*, 725–729.
16. Horai, M., Mishima, H., Hayashida, C., Kinoshita, A., Nakane, Y., Matsuo, T., Tsuruda, K., Yanagihara, K., Sato, S., Imanishi, D., et al. (2018). Detection of de novo single nucleotide variants in offspring of atomic-bomb survivors close to the hypocenter by whole-genome sequencing. J. Hum. Genet. *63*, 357–363.
17. Costa, E.O.A., Pinto, I.P., Gonçalves, M.W., da Silva, J.F., Oliveira, L.G., da Cruz, A.S., Silva, D.d.M.E., da Silva, C.C., Pereira, R.W., and da Cruz, A.D. (2018). Small de novo CNVs as biomarkers of parental exposure to low doses of ionizing radiation of caesium-137. Sci. Rep. *8*, 5914.
18. Behjati, S., Gundem, G., Wedge, D.C., Roberts, N.D., Tarpey, P.S., Cooke, S.L., Van Loo, P., Alexandrov, L.B., Ramakrishna, M., Davies, H., et al. (2016). Mutational signatures of ionizing radiation in second malignancies. Nat. Commun. *7*, 12605. http://dharmasastra.live.cf.private.springer.com/articles/ncomms12605#supplementary-information.
19. Rose Li, Y., Halliwill, K.D., Adams, C.J., Iyer, V., Riva, L., Mamunur, R., Jen, K.-Y., Del Rosario, R., Fredlund, E., Hirst, G., et al. (2020). Mutational signatures in tumours induced by high and low energy radiation in Trp53 deficient mice. Nat. Commun. *11*, 394.
20. Nakamura, N., Suyama, A., Noda, A., and Kodama, Y. (2013). Radiation effects on human heredity. Annu. Rev. Genet. *47*, 33–50.
21. Park, S., Mali, N.M., Kim, R., Choi, J.W., Lee, J., Lim, J., Park, J.M., Park, J.W., Kim, D., Kim, T., et al. (2021). Clonal dynamics in early human embryogenesis inferred from somatic mutation. Nature *597*, 393–397.
22. Youk, J., Kwon, H.W., Kim, R., and Ju, Y.S. (2021). Dissecting single-cell genomes through the clonal organoid technique. Exp. Mol. Med. *53*, 1503–1511.
23. Kim, E., Han, D.-J., Kim, B.H., Yoo, J., Kim, H.J., Wu, H.G., Kim, K.S., Kim, H.S., Han, I., Moon, K.C., et al. (2023). Whole-Genome Sequencing Reveals Mutational Signatures Related to Radiation-Induced Sarcomas and DNA-Damage-Repair Pathways. Mod. Pathol. *36*, 100004.
24. Olafsson, S., and Anderson, C.A. (2021). Somatic mutations provide important and unique insights into the biology of complex diseases. Trends Genet. *37*, 872–881.
25. Kim, J., Koo, B.-K., and Knoblich, J.A. (2020b). Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. *21*, 571–584.
26. Leach, J.K., Van Tuyle, G., Lin, P.S., Schmidt-Ullrich, R., and Mikkelsen, R.B. (2001). Ionizing radiation-induced, mitochondria-dependent generation of reactive oxygen/nitrogen. Cancer Res. *61*, 3894–3901.
27. Waris, G., and Ahsan, H. (2006). Reactive oxygen species: role in the development of cancer and various chronic conditions. J. Carcinog. *5*, 14.
28. Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat. Rev. Mol. Cell Biol. *18*, 495–506.
29. Lee, J.J.-K., Park, S., Park, H., Kim, S., Lee, J., Lee, J., Youk, J., Yi, K., An, Y., Park, I.K., et al. (2019). Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. Cell *177*, 1842–1857.e21.
30. Kim, H., Nguyen, N.P., Turner, K., Wu, S., Gujar, A.D., Luebeck, J., Liu, J., Deshpande, V., Rajkumar, U., Namburi, S., et al. (2020a). Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat. Genet. *52*, 891–897.
31. Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature *543*, 122–125.
32. Yi, K., and Ju, Y.S. (2018). Patterns and mechanisms of structural variations in human cancer. Exp. Mol. Med. *50*, 1–11.
33. Anderson, N.D., de Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., et al. (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science *361*, eaam8419.
34. Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.
35. Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.
36. Zhang, C.Z., Spektor, A., Cornils, H., Francis, J.M., Jackson, E.K., Liu, S., Meyerson, M., and Pellman, D. (2015). Chromothripsis from DNA damage in micronuclei. Nature *522*, 179–184.
37. Maciejowski, J., Li, Y., Bosco, N., Campbell, P.J., and De Lange, T. (2015). Chromothripsis and kataegis induced by telomere crisis. Cell *163*, 1641–1654.
38. Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L., et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature *467*, 1109–1113.
39. Cortés-Ciriano, I., Lee, J.J.-K., Xi, R., Jain, D., Jung, Y.L., Yang, L., Gordenin, D., Klimczak, L.J., Zhang, C.Z., Pellman, D.S., et al. (2020). Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat. Genet. *52*, 331–341.
40. Yoshida, K., Gowers, K.H.C., Lee-Six, H., Chandrasekharan, D.P., Coorens, T., Maughan, E.F., Beal, K., Menzies, A., Millar, F.R., Anderson, E., et al. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. Nature *578*, 266–272.
41. Fieller, E.C. (1954). Some problems in interval estimation. J. R. Stat. Soc. B (Methodol.). *16*, 175–185.
42. Ku, T., Swaney, J., Park, J.Y., Albanese, A., Murray, E., Cho, J.H., Park, Y.G., Mangena, V., Chen, J., and Chung, K. (2016). Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. Nat. Biotechnol. *34*, 973–981.
43. Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. *17*, 175–188.

44. Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2006). Whole-genome multiple displacement amplification from single cells. Nat. Protoc. *1*, 1965–1970.

45. Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., Shaner, B., Annu, K., Putnam, D., Chen, W., et al. (2021). Accurate genomic variant detection in single cells with primary template-directed amplification. Proc. Natl. Acad. Sci. USA *118*, e2024176118.

46. Sanders, A.D., Falconer, E., Hills, M., Spierings, D.C.J., and Lansdorp, P.M. (2017). Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. Nat. Protoc. *12*, 1151–1176.

47. Hoang, M.L., Kinde, I., Tomasetti, C., McMahon, K.W., Rosenquist, T.A., Grollman, A.P., Kinzler, K.W., Vogelstein, B., and Papadopoulos, N. (2016). Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. Proc. Natl. Acad. Sci. USA *113*, 9846–9851.

48. Bae, J.H., Liu, R., Roberts, E., Nguyen, E., Tabrizi, S., Rhoades, J., Blewett, T., Xiong, K., Gydush, G., Shea, D., et al. (2023). Single duplex DNA sequencing with CODEC detects mutations with high sensitivity. Nat. Genet. *55*, 871–879.

49. Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., Russell, A.J.C., Alcantara, R.E., Baez-Ortega, A., Wang, Y., et al. (2021). Somatic mutation landscapes at single-molecule resolution. Nature *593*, 405–410.

50. Wilkins, R.C., Romm, H., Kao, T.C., Awa, A.A., Yoshida, M.A., Livingston, G.K., Jenkins, M.S., Oestreicher, U., Pellmar, T.C., and Prasanna, P.G.S. (2008). Interlaboratory comparison of the dicentric chromosome assay for radiation biodosimetry in mass casualty events. Radiat. Res. *169*, 551–560.

51. Mihara, E., Hirai, H., Yamamoto, H., Tamura-Kawakami, K., Matano, M., Kikuchi, A., Sato, T., and Takagi, J. (2016). Active and water-soluble form of lipidated Wnt protein is maintained by a serum glycoprotein afamin/alpha-albumin. Elife *5*, e11621.

52. de Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S.W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M., et al. (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293–297.

53. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

54. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

55. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

56. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

57. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323.

58. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

59. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

60. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594.

61. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

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

63. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

64. Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D., et al. (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat. Genet. *52*, 306–319.

65. Favero, F., Joshi, T., Marquard, A.M., Birkbak, N.J., Krzystanek, M., Li, Q., Szallasi, Z., and Eklund, A.C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann. Oncol. *26*, 64–70.

66. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

67. Grün, B., and Leisch, F. (2008). FlexMix, version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. J. Stat. Softw. *28*, 1–35.

68. Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D. (2009). Mixtools: an R package for Analyzing Finite Mixture Models. J. Stat. Softw. *32*, 1–29.

69. Soetaert, K. (2015). rootSolve: Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations. software. R-Package Version 1.

70. Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a, in: a haystack: Detecting knee points in system behavior. In 2011 31st international conference on distributed computing systems workshops (IEEE Publications), pp. 166–171.

71. Young, D.S. (2010). Tolerance: an R package for Estimating Tolerance Intervals. J. Stat. Soft. *36*, 1–39.

72. Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593.

73. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.

74. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

75. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

76. Cahan, W.G., and Woodard, H.Q. (1948). Sarcoma arising in irradiated bone; report of 11 cases. Cancer *1*, 3–29. -7.

77. Lee, E.S., Won, Y.J., Kim, B.C., Park, D., Bae, J.H., Park, S.J., Noh, S.J., Kang, Y.R., Choi, S.H., Yoon, J.H., et al. (2016). Low-dose irradiation promotes Rad51 expression by down-regulating miR-193b-3p in hepatocytes. Sci. Rep. *6*, 25723.

78. Bartfeld, S., Bayram, T., van de Wetering, M., Huch, M., Begthel, H., Kujala, P., Vries, R., Peters, P.J., and Clevers, H. (2015). In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. Gastroenterology *148*, 126–136.e6.

79. Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

80. Fujii, M., Matano, M., Nanki, K., and Sato, T. (2015). Efficient genetic engineering of human intestinal organoids using electroporation. Nat. Protoc. *10*, 1474–1485.

81. Broutier, L., Andersson-Rolf, A., Hindley, C.J., Boj, S.F., Clevers, H., Koo, B.K., and Huch, M. (2016). Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. Nat. Protoc. *11*, 1724–1743.

82. Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A., Begthel, H., et al. (2018). A living biobank of breast cancer organoids captures disease heterogeneity. Cell *172*, 373–386.e10.

83. Kessler, M., Hoffmann, K., Brinkmann, V., Thieck, O., Jackisch, S., Toelle, B., Berger, H., Mollenkopf, H.J., Mangler, M., Sehouli, J., et al. (2015). The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. Nat. Commun. *6*, 8989.

84. Lee, J.H., Bhang, D.H., Beede, A., Huang, T.L., Stripp, B.R., Bloch, K.D., Wagers, A.J., Tseng, Y.H., Ryeom, S., and Kim, C.F. (2014). Lung stem cell differentiation in mice directed by endothelial cells via a BMP4-NFATc1-thrombospondin-1 axis. Cell *156*, 440–455.

85. Youk, J., Kim, T., Evans, K.V., Jeong, Y.-I., Hur, Y., Hong, S.P., Kim, J.H., Yi, K., Kim, S.Y., Na, K.J., et al. (2020). Three-dimensional human alveolar stem cell culture models reveal infection response to SARS-CoV-2. Cell Stem Cell *27*, 905–919.e10.

86. Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC−/− colon carcinoma. Science *275*, 1784–1787.

87. Jager, M., Blokzijl, F., Sasselli, V., Boymans, S., Janssen, R., Besselink, N., Clevers, H., van Boxtel, R., and Cuppen, E. (2018). Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. Nat. Protoc. *13*, 59–78.

88. Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. Nature *538*, 260–264.

89. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

90. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93.

91. Leisch, F. (2004). Flexmix: A General Framework for Finite Mixture Models and Latent Glass Regression in R. J. Stat. Softw. *11*, 1–18.

92. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. *47*, D941–D947.

93. Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

94. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering signatures of mutational processes operative in human cancer. Cell Rep. *3*, 246–259.

95. Brown, L.D., Cai, T.T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. Stat. Sci. *16*, 101–117.

96. Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. Ann. Math. Statist. *27*, 713–721.

97. Zakov, S., Kinsella, M., and Bafna, V. (2013). An algorithmic approach for breakage-fusion-bridge detection in tumor genomes. Proc. Natl. Acad. Sci. USA *110*, 5546–5551.

98. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *21*, 29.21–29.29.

99. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46*, D794–D801.

100. Ku, T., Guan, W., Evans, N.B., Sohn, C.H., Albanese, A., Kim, J.G., Frosch, M.P., and Chung, K. (2020). Elasticizing tissues for reversible shape transformation and accelerated molecular labeling. Nat. Methods *17*, 609–613.

101. Abramoff, M., Magalhães, P., and Ram, S.J. (2003). Image processing with ImageJ. Biophotonics Int *11*, 36–42.

102. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

103. Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. 2013–2015.

104. Glenn, T.C., Nilsen, R.A., Kieran, T.J., Sanders, J.G., Bayona-Vásquez, N.J., Finger, J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., et al. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ *7*, e7755.

105. Hills, M., O'Neill, K., Falconer, E., Brinkman, R., and Lansdorp, P.M. (2013). BAIT: organizing genomes and mapping rearrangements in single cells. Genome Med. *5*, 82.

106. Goldstein, H., and Healy, M.J.R. (1995). The graphical presentation of a collection of means. J. R. Stat. Soc. A. *158*, 175–177.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PE Rat anti-Mouse CD326	BD Biosciences	Cat#563477; RRID:AB_2738233
Rat anti-Mouse CD31	BD Biosciences	Cat#550274; RRID:AB_393571
APC Rat anti-Mouse CD45	BD Biosciences	Cat#561018; RRID:AB_10584326
APC-Cy7 Rat anti-Mouse Ly-6A/E	BD Biosciences	Cat#560654; RRID:AB_1727552
Phospho-Histone H2A.X (Ser139) Antibody	Cell Signaling Technology	Cat#2577; RRID:AB_2118010
Donkey anti-Rabbit IgG antibody-Alexa Fluor 488	Thermo Fisher Scientific	Cat#A21206; RRID:AB_2535792
Goat anti-Rabbit IgG antibody-Alexa Fluor 488	Thermo Fisher Scientific	Cat#A11008; RRID:AB_143165
Bacterial and virus strains		
DH5α Chemically Competent E. coli	Enzynomics	Cat#CP011
Biological samples		
Human colon, breast normal tissues and blood	This paper	N/A
Human sarcoma tissues	This paper	N/A
Chemicals, peptides, and recombinant proteins		
DAPI	Invitrogen	Cat#D3571
Acrylamide	Sigma-Aldrich	Cat#A9099
Sodium acrylate	AK scientific	Cat#R624
Bis-acrylamide	Biorad	Cat#1610156
VA-044	Wako Chemical	Cat#011-19365
4% Paraformaldehyde	Biosesang	Cat#PC2031-050-00
EDTA	Invitrogen	Cat#15575-020
HBSS	Gibco	Cat#14025092
Collagenase P	Roche	Cat#11213865001
DNAseI	Sigma-Aldrich	Cat#DN25
Collagenase I	Sigma-Aldrich	Cat#C0130
Dispase	Corning	Cat#354235
Collagenase/dispase	Sigma-Aldrich	Cat#10269638001
RBC lysis buffer	Sigma-Aldrich	Cat#R7757
Advanced DMEM/F12	Gibco	Cat#12634028
Opti-MEM reduced serum medium	Thermo Fisher Scientific	Cat#31985062
FBS	Gibco	Cat#16000-044
Lipofectamine 2000 reagent	Invitrogen	Cat#11668019
Y-27632	Sigma-Aldrich	Cat#Y0503
Corning Matrigel Basement Membrane Matrix	Corning	Cat#356235
Corning Matrigel Growth Factor Reduced Basement Membrane Matrix	Corning	Cat#354230
Cultrex Organoid Harvesting Solution	Trevigen	Cat#3700-100-01
Cell Recovery Solution	Corning	Cat#354253
TrypLE Express	Gibco	Cat#12604013
Accutase	STEMCELL Technologies	Cat#07920
HEPES	Gibco	Cat#15140-122
Glutamax	Gibco	Cat#35050-061
Penicillin/Streptomycin	Gibco	Cat#15630-080
Bovine Serum Albumin solution	Sigma-Aldrich	Cat#AB412-100ML
NP40	Sigma-Aldrich	Cat#98379

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
B27 supplement	Thermo Fisher Scientific	Cat#17504044
FGF7	PEPROTECH	Cat#100-19
FGF10	PEPROTECH	Cat#100-26
Noggin	PEPROTECH	Cat#120-10C
EGF	PEPROTECH	Cat#100-15
N-Acetylcysteine	Sigma-Aldrich	Cat#A9165
Nicotinamide	Sigma-Aldrich	Cat#N0636
SB431542	Calbiochem	Cat#616461
CHIR99021	TOCRIS	Cat#4423
SB202190	Sigma-Aldrich	Cat#S7067
A83-01	TOCRIS	Cat#2939
Primocin	Invivogen	Cat#Ant-pm-1
Triton X-100	Sigma-Aldrich	Cat#T9284-100ML
N-2 Supplement	Invitrogen	Cat#17502048
Neuregulin-1	PEPROTECH	Cat#100-03
Dexamethasone	Sigma-Aldrich	Cat#D4902
Forskolin	TOCRIS	Cat#1099
Critical commercial assays		
DNeasy Blood and Tissue Kit	Qiagen	Cat#69506
TaKaRa MiniBEST Universal Genomic DNA Extraction Kit	Takara	Cat#9765A
TruSeq DNA PCR-Free Library Prep Kit	Illumina	Cat#20015963
TruSeq Nano DNA Kit	Illumina	Cat#20015965
Accel-NGS 2S Plus DNA Library Kit	Swift	Cat#21024
RNeasy Mini Kit	Qiagen	Cat#74104
TruSeq Stranded Total RNA Gold Kit	Illumina	Cat#20020599
TruSeq RNA Library Prep Kit	Illumina	Cat#RS-122-2001
Illumina Tagment DNA Enzyme and Buffer Kit	Illumina	Cat#20034197
Nextera DNA Library Kit	Illumina	Cat#FC-121-1030
NEBNext High-Fidelity 2X PCR Master Mix	New England Labs	Cat#M0541L
Power SYBR Green PCR Master Mix	Thermo Fisher Scientific	Cat#4367659
PrimeSTAR HS Taq polymerase	Takara	Cat#R010A
Pfu DNA polymerase	Thermo Fisher Scientific	Cat#EP0501
Dual Luciferase reporter assay	Promega	Cat#E1910
CellTiter-Glo 3D Cell Viability Assay	Promega	Cat#G9681
Deposited data		
Mouse organoid whole-genome sequence data	This study	KoNA:KAP230688
Mouse organoid transcriptome data	This study	KoNA:KAP230688
Human organoid whole-genome sequence data	This study	KoNA:KAP230688
Human organoid transcriptome data	This study	KoNA:KAP230688
Experimental models: Cell lines		
Afamin-Wnt3a-producing HEK293	Mihara et al.[51]	N/A
Cultrex HA-R-Spondin 1-Fc 293T	Trevigen	Cat#3710-001-01
Experimental models: Organisms/strains		
Mouse: BALB/c (male, 6–7 weeks old)	Central Laboratory Animal Incorporated	N/A
Mouse: C57BL/6 (female, 8–9 weeks old)	Central Laboratory Animal Incorporated	N/A
Oligonucleotides		
PA_2Gy_11 deletionF, 5′-GCCTGTGTTC AAATTGGGGG-3′	BIONICS	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PA_2Gy_11 deletionR, 5′-AAAACCCATG CCCTCTGCTT-3′	BIONICS	N/A
PA_2Gy_11 insertionF, 5′-CCACATGGA ACCTTATGCTGC-3′	BIONICS	N/A
PA_2Gy_11 insertionR, 5′-TTCTGACTG CCTTGGCACAG-3′	BIONICS	N/A
Recombinant DNA		
TOPcloner Blunt core Kit	Enzynomics	Cat#EZ012S
TOPcloner TA core Kit	Enzynomics	Cat#EZ011S
M50 Super 8x TOPFlash	de Lau et al.[52]	Addgene plasmid #12456
M51 Super 8x FOPFlash	de Lau et al.[52]	Addgene plasmid #12457
ResolveDNA Whole Genome Amplification Kit	BioSkryb	Cat#PN100068
Software and algorithms		
R 3.6.0	The R Foundation	http://www.r-project.org/
Burrow-Wheeler Aligner	Li et al.[53]	https://github.com/lh3/bwa
Samtools	Li et al.[54]	http://htslib.org
Picard	McKenna et al.[55]	http://broadinstitute.github.io/picard
Genome Analysis Toolkit	McKenna et al.[55]	https://gatk.broadinstitute.org/hc/en-us
Cutadapt	Martin et al.[56]	https://github.com/marcelm/cutadapt
RSEM v1.3.1	Li et al.[57]	https://github.com/deweylab/RSEM
ComplexHeatmap R package	Gu et al.[58]	https://github.com/jokergoo/ComplexHeatmap
Enrichr	Kuleshov et al.[59]	https://maayanlab.cloud/Enrichr
Strelka2	Kim et al.[60]	https://github.com/Illumina/strelka
Varscan2	Koboldt et al.[61]	https://varscan.sourceforge.net
Python 2.7.5	Python Software Foundation	https://www.python.org
Pysam Python package	Li et al.[54]	https://github.com/pysam-developers/pysam
Delly 0.7.6	Rausch et al.[62]	https://github.com/dellytools/delly
Integrative Genomics Viewer	Robinson et al.[63]	https://software.broadinstitute.org/software/igv
TraFiC-mem	Rodriguez-Martin et al.[64]	https://github.com/cancerit/TraFiC
Sequenza R package	Favero et al.[65]	https://cran.r-project.org/package=sequenza
Circos	Krzywinski et al.[66]	http://circos.ca
flexmix R package	Grun et al.[67]	https://cran.r-project.org/package=flexmix
mixtools R package	Benaglia et al.[68]	https://cran.r-project.org/package=mixtools
rootSolve R package	Soetaert et al.[69]	https://cran.r-project.org/package=rootSolve
numDeriv R package	Satopaa et al.[70]	https://cran.r-project.org/package=numDeriv
Julia 1.2.0	The Julia Project	https://julialang.org/
tolerance R package	Young[71]	https://cran.r-project.org/package=tolerance
phangorn R package	Schliep et al.[72]	https://cran.r-project.org/package=phangorn
ape R package	Paradis et al.[73]	https://cran.r-project.org/package=ape
Bowtie2	Langmead et al.[74]	https://bowtie-bio.sourceforge.net/bowtie2
liftOver	Kent et al.[75]	https://genome.ucsc.edu/cgi-bin/hgLiftOver
ImageJ	ImageJ software	https://ImageJ.nih.gov/ij/L
mratios R package	Fieller et al.[41]	https://cran.r-project.org/package=mratios
Mutational signature analysis	This study	https://doi.org/10.5281/zenodo.7402806
Other		
Confocal microscope	Carl Zeiss	LSM 780
Cs-137 gamma irradiator	J.L. Shepherd and Associates	MK1-68
Cs-137 source	Chiyoda Technol Corp.	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Co-60 irradiator	Best Theratronics	GBX200
Radiochromic film	ISP Corporation	Gafchromic EBT3
Nanodrop 2000 spectrophotometer	Thermo Fisher Scientific	ND-2000
PCR Thermal Cycler	Applied Biosystems	Veriti Thermal Cycler
Luminescence Counter	Perkin Elmer	1420 Victor Light

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Young Seok Ju (ysju@kaist.ac.kr).

Materials availability

Organoids established in this study are available under a material transfer agreement. To do so, please contact the lead author (ysju@kaist.ac.kr).

Data and code availability

Mouse and human whole-genome and transcriptome sequencing raw data produced in this study were deposited on the Korean Nucleotide Archive (KoNA; https://www.kobic.re.kr/kona/). The project accession ID is KAP230688. All data were provided for review purposes upon reviewers' request. Tumor genome sequences analyzed in this study were obtained from previous studies (EGAS00001000138, EGAS00001000147, EGAS00001000195; EGAD00001004162).[18,23] Essential in-house scripts used in this study are available on Zenodo (DOI: https://doi.org/10.5281/zenodo.7402806).

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Experimental mice

Male BALC/c (6–7 weeks old) and female C57BL/6 (8–9 weeks old) mice were purchased from Central Laboratory Animal Incorporated (Seoul, Korea). All experiment protocols were approved by Dongnam Institute of Radiological and Medical Sciences (DIRAMS AEC-2018-007 and DIRAMS AEC-2019-007) and Seoul National University (SNU-180101-1-3). Animal studies were conducted in accordance with guidelines established by the committee on Use and Care of Animals of Dongnam Institute of Radiological and Medical Sciences and Seoul National University Animal Care and Use Committee. Animals were treated humanely in accordance with the Ministry of Food and Drug Safety on the ethical use of animals. Mice were sacrificed with carbon dioxide inhalation euthanasia to get tissues for organoid cultures and spleens for matched germline controls.

Human normal samples

For all patients, tumor-adjacent normal tissues or matched blood were collected from irradiated and non-irradiated individuals (controls). The protocol of this study was approved by the institutional review board of Dongnam Institute of Radiological and Medical Sciences (D-1804-023-002, D-1810-007-002, and D-2206-021-002). Radiation-free human normal colon tissues were acquired from surgical specimens of two colorectal cancer patients (Table S1). The enrolled patients had no experience of previous chemotherapy or radiotherapy. Tumor adjacent normal colon tissues were cut out from a region >5 cm away from primary tumor mass. Irradiated normal colon samples were taken from a patient diagnosed with rectal cancer located 7cm away from the anal verge. The patient received preoperative radiotherapy (45 Gy/25 fx + 5.4 Gy/3 fx) with 5-fluorouracil and leucovorin to preserve the anus. After ultralow anterior resection surgery, a tumor-adjacent normal colon tissue in the radiation field was acquired from a region >5 cm away from primary tumor mass. Normal human breast samples were collected from adjacent normal tissues from a total mastectomy of recurred breast cancer at Dongnam Institute of Radiological and Medical Sciences (Table S1). Patients had breast-conserving surgery 1 (HBIR-1) and 17 (HBIR-2) years ago, followed by prophylactic adjuvant radiotherapy (50 Gy/25 fx + 10 Gy/5 fx as a boost for HBIR-1 and 50.4 Gy/28 fx + 9 Gy/5 fx as a boost for HBIR-2). The HBIR-1 patient received neoadjuvant chemotherapy with Adriamycin, cyclophosphamide and docetaxel at initial diagnosis. The HBIR-2 patient received adjuvant cyclophosphamide, methotrexate and 5-fluorouracil at initial diagnosis. The patient took tamoxifen for several years. After recurrence, the HBIR-2 patient had neoadjuvant chemotherapy with doxorubicin, cyclophosphamide, docetaxel and trastuzumab before total mastectomy.

Human cancer samples

After receiving an approval by the institutional review board of Seoul National University Hospital (H-1506-026-678), we collected patients who met the following modified Cahan's criteria for radiation-induced sarcoma between 2015 and 2019; (A) pathologically confirmed sarcoma which was histologically different from the primary cancer, (B) occurred in the field of radiation, and (3) developed at least 6 months after radiation therapy.[76] Sixteen patients were identified and surgically resected in our institution, and we finally obtained whole-genome sequences of ten matched tumor and normal samples. We also collected clinical information including initial tumor diagnosis, radiation dose, latency period, and pathologic diagnosis of radiation-induced sarcoma (Table S1). The median prescribed radiation dose was 50.4 Gy. Radiation-induced sarcomas were located in the extremities (n = 5), trunk (n = 4), and head & neck (n = 1). Radiation-induced sarcomas included 4 undifferentiated pleomorphic sarcomas, 3 undifferentiated spindle cell sarcomas, 2 osteosarcomas, and 1 undifferentiated epithelioid sarcoma.

Pathology review

Tumors had been fresh frozen (n = 8) or formalin-fixed/paraffin-embedded (FFPE, n = 2). Normal samples were obtained from blood (n = 7) or adjacent normal tissues of FFPE (n = 1) and fresh frozen specimens (n = 2). Two experienced pathologists reviewed specimens and performed microdissection to isolate the tumor and normal tissues.

METHOD DETAILS

Irradiation experiments

For *in vitro* organoids irradiation, ^{137}Cs gamma irradiator (MK1-68; J.L. Shepherd and Associates) was employed. Dose distribution in the floor of the irradiation cavity was mapped by a 24-well plate containing PBS-submerged radiochromic film (RCF, GAFCHROMIC EBT2, ISP Corporation). Film samples were digitized using a flatbed scanner (Expression 10000XL, Epson) with accompanying software (Silverfast Epson IT8). The optical density from RGB (red, green, and blue) uncompressed tagged film images was analyzed for radiation exposure. After the harvest and dissociation of organoids, the cell suspension was transferred to 1.5 mL Eppendorf tube with culture media. Single cells were irradiated at a dose rate of 31 mGy/s. After the irradiation, the cells were sorted and inoculated in Matrigel (Corning). For *in vivo* mice irradiation, a cobalt irradiator (GBX200, Best Theratronics) with ^{60}Co source (~5,000 Ci) was used. An ionization chamber was used to ensure accurate dosimetry. Ionization chambers for whole-mouse irradiation measured air kerma and an accurate dose was irradiated at an experimental radiation field. The field size of ^{60}Co was 20 cm × 15 cm, and the air kerma was measured by inserting the ionization chamber into a 50 mL conical tube with a holed cap, which was placed at a source-to-chamber distance (SCD) of 60 cm (Figure S1C). Whole-body γ-ray irradiations of 2 ± 0.098 Gy (0.4763 Gy/min) were carried out in an irradiation room equipped with a ^{60}Co source. The standard uncertainty of mouse irradiation using a cobalt irradiator is shown in Table S15. A total 8 and 20 Gy were irradiated cumulatively by irradiating 2 Gy per day with 4 times and 10 times, respectively. The animals were sacrificed 24 h and 2 weeks after irradiation for primary bulk organoid culture and primary single-cell/crypt culture, respectively. For low-rate long-term irradiation, 8 Gy radiation was irradiated to whole-body with a rate of 3.33 mGy/h using ^{137}Cs source (185 GBq) (Chiyoda Technol Corp.)[77] for 100 days, followed by sacrifice after three months from irradiation. Sham mice were also placed in the same tube without irradiation and then sacrificed on the same day as the other irradiated mice.

Organoid culture

All organoid establishment procedures and media compositions were adopted from previous literature with slight modifications (Table S14). For crypt isolation of the mouse stomach (antrum),[78] small intestine[79] and colon,[80] each tissue was dissected from euthanized mice. Each tissue was opened up longitudinally and washed in cold PBS 3 times. In the small intestine, villi were scrapped off using coverslips. Tissues were cut into small 2–4 mm tissues and transferred to 10 mM EDTA (Invitrogen) in 50 mL conical tubes, followed by incubation on an automatic rolling machine for 20–30 min at room temperature. After incubation, the tubes were shaken to extract crypts from the tissues. The supernatant was transferred to new tubes and centrifuged at 300 g for 5 min. The pellet was washed once with PBS. Isolated crypts were embedded in Matrigel and plated in a 12- or 24-well plate (TPP). The plates were transferred into an incubator at 37°C for 5-10 min to solidify Matrigel. Each well was overlaid with 0.5–1 mL of organoid culture media for a 24- or 12-well plate, respectively. For the mouse pancreas,[81] liver,[81] breast[82] and fallopian tubes,[83] each tissue was collected from euthanized mice and washed in cold HBSS (Gibco). Each tissue was placed in a 100-mm Petri dish and minced into small pieces using scalpel blades. For the pancreas, liver and breast, 10 mg/mL collagenase P (Roche) and 0.1 mg/mL DNAse (Sigma) in HBSS were prepared as digestion solutions. For fallopian tubes, 0.5 mg/mL collagenase I (Sigma) was prepared for digestion. The minced tissue was transferred to a 50 mL conical tube, and 10 mL of prewarmed digestion buffer was added, followed by incubation at 37°C with shaking at 230 rpm for 20–45 min. The digested tissues were washed with PBS two times and shaken in PBS. Because insufficiently digested tissues were frequently observed in the pancreas and liver, the tubes with digested tissues were allowed to stand for 20-30 s to remove the insufficiently digested tissues. Then, the supernatant was collected and centrifuged at 300 rcf for 3 min. For breast and fallopian tubes, the tubes with digested tissues were centrifuged at 300 rcf for 3 min. After centrifugation, pellets were embedded in Matrigel. In the case of the liver, isolation media was used for primary culture cases during the initial 3–4 days. Then, expansion media was added to culture the liver organoids. For mouse lung organoids,[84] a rib cage was opened up in euthanized mice. Lung perfusion was carried out through the right ventricle using 10 mL DPBS in a 10 mL syringe with a 26 G

needle. After removing the heart, the trachea was dissected and cut off at the most proximal portion. A 2 mL dispase (Corning) was infused through the proximal opening of the trachea until the expansion of all lung lobes. The whole lung was separated from the thoracic cage by lifting the trachea. Five lobes were dissected from the whole lung using scissors, followed by digestion in 3 mL PBS with 65 μL of collagenase/dispase (Sigma) in a 37°C shaking incubator with 230 rpm. After 35 min of digestion, 7.5 μL of DNase I (Sigma) was added to the digestion solution. With 10–20 min of additional incubation, the digested solution was filtrated using a 40 μm strainer (Falcon) to remove insufficiently digested materials. Then, filtrates were centrifuged at 400 rcf for 5 min. Because several RBCs were frequently observed in the pellet, the pellet was incubated in RBC lysis buffer (Sigma) for 1–1.5 min, followed by adding 10 mL of advanced DMEM/F12 and 500 μL of FBS (Gibco) at the bottom. After centrifuging at 400 rcf for 5 min, we discarded the supernatant. The pellet was resuspended in 200 μL of PF10 (10% FBS in DPBS). The resuspended solution was incubated with Anti-EpCAM PE antibody (BD Biosciences, 563477), anti-CD31 APC antibody (BD Biosciences, 550274), anti-CD45 APC antibody (BD Biosciences, 561018) and anti-Ly-6A/E APC-Cy7 antibody (BD Biosciences, 560654). After 30–60 min of incubation, 800 μL PF10 was added and centrifuged at 13,000 rpm for spin-down. The pellet was washed two times with PF10. Then, type 2 alveolar epithelial cells (EpCAM-positive, CD31-negative, CD45-negative and Ly-6A/E-negative) were isolated using a fluorescence-activated cell sorter (FACSAria II, BD Biosciences) (Figure S1B). The type 2 alveolar epithelial cells were embedded in a 1:1 mixture of complete lung organoid media[85] and Growth Factor Reduced Matrigel (Corning), and the droplets of organoids and Matrigel mixture were seeded in 6.5 mm transwells (Corning) in a 24-well plate. After 20 min of incubation, 500 μl complete lung media with 10 μM Y-27632 (Sigma) were added into each well. Plates were transferred to a humidified 37°C/5% CO_2 incubator, and the medium was changed every 2–3 days. Organoids were passaged by removing the medium and dissolving Matrigel by adding Cultrex organoid harvesting solution (Trevigen) or cell recovery solution (Corning) for 40–50 min on ice. After removing the dissolving solution, 0.5–1 mL prewarmed TrypLE Express (Gibco) or prewarmed Accutase (Stemcell Technology) was added to disrupt organoids, followed by centrifuging at 300 rcf for 3 min. In the case of mouse breast organoids, a 1cc syringe with a 26 G needle was used to dissociate before centrifugation. Pellet was resuspended with Matrigel and seeded in a new 12- or 24-well plate at ratio of 1:4 to 1:6. After polymerization of Matrigel in a 37°C incubator, 500–1000 μL complete media were added to each well. In the several complete organoid media, Wnt3a or R-Spondin-1 conditioned media was included. We prepared the conditioned media using Afamin-Wnt3a producing HEK293 cell line[51] and Cultrex HA-R-Spondin 1-Fc 293T cell line (Trevigen), respectively. The activity of the harvesting conditioned media was tested using TOP/FOP assay.[52,86]

Single-cell-derived clonal organoid acquisition

For the clonalization of single cells, we used two alternative strategies. First, bulk organoids were harvested and single-cell dissociated using prewarmed TrypLE Express or Accutase. After centrifuging, supernatants were removed, and organoids were resuspended using AdDF+++ (Advanced DMEM/F12 with 10mM HEPES, 1X Glutamax and 1% penicillin/streptomycin). Organoid suspensions were filtered through a 40 μm strainer (Falcon). Using a cell sorter (FACSAria II, BD Biosciences), single cells were sorted into a FACS tube with 2 mL of AdDF+++. Single cells were selected in the FACSDiva software based on forward- and side-scatter characteristics.[87] Sorted cells were seeded with Matrigel (500–1500/well). Alternatively, single whole-crypts were collected from the stomach, small intestine, and colon after 14 days from whole-body irradiation. Of note, a crypt is physiologically clonalized *in vivo* due to the competitive proliferation of single cells.[88] When mono-to oligo-clonal organoids were acquired right after primary culture, digested materials were sparsely seeded with Matrigel at P0. After growing organoids, a single organoid distant from other organoids was picked up using a 200 μL pipette under a microscope. Clonal organoid generation was evaluated by variant allele frequency of somatic point mutations.

DNA extraction

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen) or TaKaRa MiniBEST Universal Genomic DNA Extraction Kit Ver. 5.0 (TaKaRa Bio Inc.). At the last step, the DNA sample was eluted twice with 40 μL of elution buffer. The concentration of DNA was measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific, ND-2000) and stored at −20°C or −80°C before use.

Whole-genome sequencing and read alignment

DNA libraries for whole-genome sequencing (WGS) were generated using Truseq DNA PCR-Free Library Prep Kits (Illumina) from >500 ng of genomic DNA or TruSeq Nano DNA Kit (Illumina) from >100 ng of genomic DNA, according to the manufacturer's protocols. When genomic DNA was insufficient to use the PCR-Free Library Kits, Accel-NGS 2S Plus DNA Library Kit for Illumina (Swift) was alternatively used to construct DNA libraries (n = 10; Table S1). The libraries were sequenced with paired-end (2x151bp) using HiSeq 2500, HiSeq X10, or Novaseq 6000 platforms (Illumina) to generate a minimum of 30X depth for most samples except B3S100 (100X), radiation-induced sarcomas (90X), and tumor-matched normal controls (60X). The raw FASTQ files were aligned to the human reference genome, GRCh37/hg19, and the mouse reference genome, GRCm38/mm10, using Burrows-Wheeler Aligner mapping tools.[53] Samtools was used to convert sam to bam and to sort.[54] Duplicated reads were marked and removed using the Picard package (https://broadinstitute.github.io/picard). Then, local realignments and base recalibration were done using the Genome Analysis Toolkit.[55]

Whole transcriptome sequencing of organoids

Mouse pancreas organoids in Matrigel were irradiated (2 Gy). After irradiation, the irradiated organoids were retrieved from Matrigel at five different time points (0, 0.5 h, 2 h, 6 h, and 24 h; three biological replicates at each time point). After centrifugation at 300*g* for 3 min, total RNA was extracted from the organoid pellets using RNeasy Mini Kit (Qiagen, #74104). Total RNA sequencing library was constructed using Truseq Stranded Total RNA Gold kit (Illumina) according to the manufacturer's protocol. For validation of *Myc* amplification in mouse fallopian tube organoids, total RNA was extracted from FT_8Gy_1 (*Myc* non-amplified; three biological replicates) and FT_8Gy_2 (*Myc* amplified; three biological replicates). TruSeq RNA Library Prep Kit v2 (Illumina) was used to generate cDNA libraries. We performed high-throughput sequencing with 2 x 101 bp using Hiseq 2500. Adapter contamination was eliminated from the FASTQ files using Cutadapt.[56] Then, the trimmed FASTQ files were aligned to GRCh38/mm10 using STAR v2.6.1day,[89] and normalized counts of RNA expression were calculated using RSEM v1.3.1.[57] The top 2500 genes with the highest variations (standard deviations/mean of transcripts per million (TPM) values of each gene) in irradiated organoids at five time points (0, 0.5, 2, 6, and 24 h) were demonstrated by a heatmap using the R package "ComplexHeatmap".[58] Enriched pathways were obtained from the Enrichr website with the input of the top 2500 variable genes.[59]

Detection of somatic variants

Strelka2 and Varscan2 were used to find single base substitutions (SBSs) and small insertions and deletions (indels).[60,61] After taking a union of SBS and indel call sets passed in each caller, an in-house python script using pysam module was utilized to annotate read information and filter out false positive calls. To exclude sequencing artifacts, we used germline samples to calculate loci-specific background mutation rates. In the case of mouse samples, normal organoid samples from other mice of the same strains were also applied because the number of germline samples at each strain was not sufficient. This information was further utilized to rule out false positive mutations. Structural variations (SVs), or genomic rearrangements, were called using Delly 0.7.6.[62] Loci-specific background rearrangement rates were also calculated as described above. Accurate breakpoint positions and microhomology sequences were determined using the "SA tag" of clipped reads. Further, read information around breakpoints was annotated by in-house script with pysam module, including the number of wild-type read pairs spanning breakpoints with appropriate orientation and normal insert size ($\leq$1000 bp) and the number of variant read pairs spanning breakpoints with inappropriate orientation or large insert size (>1000 bp). Then, breakpoints suggesting false positive variants were removed by an in-house script. Briefly, variant calls that meet any of the following criteria were regarded as false positives and excluded from the downstream analysis: (i) breakpoints having high depth of read pairs in a matched germline sample (read pairs $\geq$200), (ii) $\geq$2 SA tag being the same with a somatic sample at a breakpoint in a matched germline sample, (iii) sequencing artifacts (rearrangements detected in $\geq$2 unmatched normal samples), or (iv) low supporting variants in somatic samples (less than ten discordant read pairs without supporting SA tag, less than three discordant read pairs with 1 supporting SA tag). Then, all rearrangements were visually inspected using Integrative Genomics Viewer (IGV)[63] to remove remaining false positive variants and to rescue false negative variants near breakpoints. Gains and losses of <100 bp sequences were classified as indels. Somatic LINE-1 retrotranspositions were called using TraFiC-mem[64] and manually investigated using IGV to filter false positive calls. Several false negative calls were rescued during visual inspection of structural variations. Segmented copy number alterations (CNAs) of clonal organoids were estimated by Sequenza, an R package, with modification for GRCm38/mm10 assembly.[65] Except for a few organoids with chromothripsis or breakage-fusion-bridge cycles, purity and ploidy were assigned as 1 and 2.0, respectively. When abrupt CNAs were without breakpoints at callable regions, we manually investigated the regions to find false negative breakpoints using IGV. To show SVs and CNAs in each sample, we used the Circos package.[66]

Publicly available datasets

In addition to the newly sequenced mouse and human organoids and radiation-induced sarcomas, we obtained additional WGS from previously published cancer datasets to validate mutational signatures of ionizing radiation (IR). For validation of IR-induced indel signatures, 12 radiation-induced cancer,[18] 64 bone sarcoma, and 31 soft tissue sarcoma data[90] were used. For comparison of deletion-insertion composite (Del-Ins) incidence, 138 lung cancer[29] and 369 diverse cancer data (liver, esophageal, prostate, bladder, stomach, biliary tract, rectal, and oral cancer, skin cutaneous melanoma) from the PCAWG datasets[90] were used. Most WGS datasets were re-processed from raw FASTQ files as described above. For 12 radiation-induced tumor samples, we used available variant call sets of SBSs, indels, and SVs for downstream analyses.

Monoclonal and polyclonal organoids

To extract precise mutational burden and signatures harbored in the original single-cell, we excluded polyclonal organoids lacking sufficient clonal mutations due to noisy subclonal mutations (i.e., culture-associated mutations). To filter out polyclonal organoids, we used the distribution of variant allele frequency (VAF) of somatic SBSs, following a similar procedure from a previous report.[88] In particular, mouse organoids with a VAF peak <0.4 or not FACS-sorted were deemed polyclonal (Table S1). Of note, all human colon organoid samples (n = 11) were monoclonal, and human breast organoids harboring IRi-SVs with VAF around 0.5 (n = 4) were regarded as monoclonal organoids. The rest of the organoids were polyclonal, but most of them had a clear, albeit not the most prominent, VAF peak between VAF 0.25 and 0.5 (mostly near 0.5), suggesting the presence of a dominant clone that contribute more than 50% of the cell pool with some minor clones (referred to as 'dominant-clonal' organoids). Among polyclonal organoids, organoids that lack dominant clones (n = 20) based on manual inspection of SBS VAF distribution were excluded from further analysis (not listed

in Table S1). Although 3 organoids (CR_0Gy_3, LI_8Gy_8, LI_8Gy_9) did not meet the criteria, they had dominant clonal indel VAF distribution and thus were reincluded. These are referred to as 'non-dominant-clonal' to be consistent with the naming based on SBS VAF (Table S1). The discrepancy between SBS and indel VAF distribution in these samples implies that a certain subclone gained substantial amount of SBSs for some reason while there is still only a handful of clones. The complete list of the excluded samples is: BR_0Gy_7, BR_0Gy_8, BR_0G_9, BR_1Gy_1, BR_8Gy_1, BR_8Gy_2, FT_0Gy_1, LI_0Gy_6, Li_8Gy_7, LU_0Gy_1, LU_0Gy_2, LU_8Gy_1, PA_0Gy_17, PA_0Gy_20, PA_4Gy_1, PA_4Gy_8, PA_8Gy_12, PA_8Gy_13, SI_0Gy_2, SI_8Gy_2. In addition, an extreme outlier, FT_0Gy_2, bearing ~2400 indels, was also excluded from further analysis. For SV VAF, many organoids had VAF between 0.15 and 0.5 (mostly the middle of the range), lower than expected. This is likely because many informative sequencing reads are missed due to the noisy nature of SV breakpoints, making it difficult to evaluate SV VAF accurately.

Determination of VAF cutoffs for dominant clonal mutations

To separate clonal mutations from subclonal mutations in the monoclonal and dominant clonal organoids, we determined VAF cutoffs for each organoid so that only the mutations having higher VAF than the cutoff were taken to be clonal mutations. Because of the aforementioned slight discrepancy between SBS and indel VAF distributions, VAF cutoffs for both variants were calculated separately. Briefly, each VAF distribution was modeled as a Gaussian mixture, and a VAF cutoff was equated to the intersection of the rightmost fitted Gaussian curve and the second rightmost curve. If a single Gaussian curve best explained the VAF distribution, the VAF cutoff was set to 0, such as in some monoclonal organoids. The detailed procedure is as follows. Supposing that k is the number of Gaussian components, we used an R package 'flexmix'[67,91] for $k = 1$ and 'mixtools'[68] for $k > 1$ to fit Gaussian curves to the distribution using expectation maximization. To promote accurate and stable solutions, we initialized μs as modes of the density curve fitted to the histogram (with bandwidth 0.03), σs as 0.1, and λs as $1/k$, and calculated for k from 1 to 5. To find the modes, we used an R package 'rootSolve' and 'numDeriv' to solve for the first and second derivatives and chose VAF values that meet $df(\mathrm{VAF}) = 0$, $d^2f(\mathrm{VAF}) < 0$, and $\mathrm{VAF} < 0.6$.[69] We computed Akaike information criteria for each k and used the Kneedle algorithm[70] to infer a value of k that best explains the VAF distribution. For each result, we manually examined whether the result from the chosen best k agreed well with the VAF distribution. Wherever an apparent discrepancy was seen, we refined the fit by fixing k to be an appropriate value, manually setting initial values of μs and σs, and re-ran the fitting procedure. We iterated the above procedure until all the solutions were stably reproducible. The final Gaussian curve fits on SBS VAF distributions, and VAF cutoffs are shown in Figures S2–S4. The numerical values of SBS and indel VAF cutoffs are listed in Table S1.

Mutational signature analysis of single base substitutions and indels

The representation of mutational signature was borrowed from the COSMIC database,[92] where SBS and indel are summarized by a mutation profile of 96 and 83 mutation features, respectively. The features are mutation types and sequence contexts.[2] Briefly, learning new mutational signatures here amounts to encoding the mutation profiles of a group of samples to a matrix and carrying out a mathematical technique called non-negative matrix factorization (NMF). Given an input matrix $V \in R^{m \times n}$ (i.e., m features and n samples), our objective is to solve $V = WH + \varepsilon$ for W and H where $W \in R_+^{m \times k}$, $H \in R_+^{k \times n}$, and an error term $\varepsilon \in R_+^{m \times n}$. We seek for an approximate solution by formulating it as an optimization problem:

$$\underset{W,\,H}{\operatorname{argmin}}\, D(V\|WH) \text{ where } D(V\|WH) = \sum_{ij}\left(V_{ij}\, log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}\right).$$

$D(V\|WH)$ is a generalized Kullback-Leibler divergence between V and WH.[93] To avoid overfitting and promote stability of a solution, a bootstrap aggregating is used, where K-means clustering is used to obtain an averaged solution.[94] For implementation, we referred to the original MATLAB code,[94] a predecessor of the SigProfiler, one of the widely used tools for signature extraction.[2] We followed mainly the same procedure, except for manually choosing some of the computation and model parameters. In particular, great care was taken to determine k, the number of presumed mutational processes, mainly through examining whether a set of signatures obtained from a chosen k leads to a coherent story without under or over-explaining the mutation history. We used a measure of stability and reconstruction error to get a reasonable initial guess for k.[94] For input data, 15 normal organoids and 22 radiation-induced cancers were included for human SBS and indel signature learning, and 67 monoclonal organoids were included for mouse SBS and indel signature learning. There were some peculiar signatures attributed uniquely to one or two samples. These were not considered solutions because they only introduced noises to the estimates of signature exposure in other samples without adding any meaningful information. Also, in the human indel signature solution set, we observed that one signature resembled ID18 from the COSMIC database, which has been linked to *E. coli* and colibactin exposure. Our solution contained more signals that should presumably have been attributed to other mutational signatures, a known phenomenon of the method. Therefore, we replaced it with ID18 from the COSMIC, which is a supposedly pure signal of *E. coli* and colibactin exposure.[9] Here, it is referred to as hID-D to be consistent with the naming of signatures (Figure S5). Of note, in the learning phase, instead of using the learned VAF cutoffs, we used a constant VAF cutoff 0.3 for all samples to minimize non-clonal signals and obtain a quality mutational signature set regarding radiation exposure. It was based on our observation that, in almost all cases except some monoclonal cases, it is inevitable to include non-clonal signals regardless of the VAF cutoff, for a VAF distribution is almost always composed of overlapping Gaussian curves (Figures S2–S4). Then, we refitted our signature solution sets using a non-negative least squares method to all 135 organoid and

22 tumor samples on the variants filtered by learned VAF cutoffs accordingly with respective species and variant types. To account for inflated proportions of unexplained mutations from ignoring the peculiar signatures and refitting the learned signatures, we defined "unexplained" counts to quantify to what extent our solution, W, cannot explain the observed mutations. For that, we cast each mutational signature spectrum to a multinomial distribution. Then, we regarded each mutation as a sample from probability distribution constructed from a convex combination of these distributions. The residual was then considered a deviation from the mean by random sampling. The rationale is that residuals that do not fall within a certain expected interval should be attributed to unknown mutational processes. More precisely, given that fitting on a sample resulted in the residual $r = v - \hat{v}$ where $\hat{v}$ is our reconstructed mutational spectrum, we defined an unexplained count for the sample as

$$\sum_{i}^{m} \max\left(0, 1_{r_i \geq 0}\left(|r_i| - \left|CI^{+}_{95\%}(\hat{v}_i)\right|\right) + 1_{r_i < 0}\left(|r_i| - \left|CI^{-}_{95\%}(\hat{v}_i)\right|\right)\right),$$

where m is the number of features (i.e., 96 for SBS and 83 for indel) and $1_C(x)$ is an indicator function that becomes x only when the condition C is met and 0 otherwise. $CI^{+}_{95\%}(\hat{v}_i)$ (or $CI^{-}_{95\%}(\hat{v}_i)$) is a value obtained by subtracting $\hat{v}_i$ from the upper (or lower) bound of 95% confidence interval of $\hat{v}_i$. We assumed that $\hat{v}_i$ is a reasonable mean estimate of the sum of binomial random variables and used the Jeffreys method to compute the binomial confidence interval using an R package 'tolerance'.[95,71] A caveat is that our mean estimate for the mutation count of each feature can be an underestimate of the actual count by the binomial sum variance inequality.[96] Also, ignoring counts that fall within the 95% CI of $\hat{v}_i$ acts against the unexplained count. Nevertheless, our definition is conservative in that the greater unexplained count strongly suggests the presence of potentially unknown processes.

Simulation of IR-inducible indel mutations

We examined the IR-associated indel signature (ID-A) to determine if there is any bias in the mutation types, hoping to understand possible mechanistic underpinnings of the indel process of IR. We ran an indel simulation to infer the expected signature spectrum when a random mutation is assumed for the IR-associated indel signature. First, the learned IR indel signatures were reduced to a probability vector (p) of deletions and insertions of length from 1 to 5+ by summing up the probabilities of the respective features. Then, we sampled random positions in the mouse or human genome without replacement. Subsequently, a deletion or insertion was introduced for each position depending on a random sample from p. We sampled an arbitrarily large number (e.g., 10000) of samples that is supposedly enough to accurately capture the true underlying distribution.

Reconstruction of genomic rearrangements

Given confident rearrangement call sets (n = 1,298) from 135 organoids, we reconstructed the rearrangement patterns and classified them into 16 types of SVs (Tables S12 and S13), including (i) medium-sized simple-deletion (<1 Mb), (ii) tandem duplication, (iii) templated insertion, (iv) retrotransposition, (v) miscellaneous type, (vi) balanced inversion, (vii) balanced translocation, (viii) long simple-deletion ($\geq$1 Mb), (ix) ecDNA formation, (x) deletion-inversion composite (Del-Inv), (xi) deletion-translocation composite (Del-Tra), (xii) deletion-insertion composite (Del-Ins), (xiii) chromoplexy, (xiv) chromothripsis, (xv) SV by Breakage-fusion-bridge cycle (BFB-SV), and (xvi) other complex genomic rearrangements (CGRs). All the balanced inversions and balanced translocations showed no overlapping segments or overlapping less than 10 bp, except in one case (225 bp in one 20 Gy irradiated liver organoid). If a pattern of SV suggests balanced inversion or balanced translocation combined with >500 bp copy number loss in one of the breakpoints, we considered those SVs as Del-Inv or Del-Tra. An SV was defined as ecDNA formation if one read of a deletion-type rearrangement and one read of a duplication-type rearrangement are located at one breakpoint and the other reads of the two rearrangements are located at another breakpoint together with copy number gain or loss. Del-Ins was an ectopic insertion of a short (typically 100–1000 bp) deleted segment like cut-and-paste type mobilization of DNA transposons. Chromoplexy was defined as a cluster of reciprocal rearrangements involving three or more chromosomes.[34] Chromothripsis is a cluster of localized (usually involving 1 or 2 chromosomes) rearrangements with massive breakpoints exhibiting copy number oscillation between two or three copy number states.[35] BFB-SV was defined as a type of SV best explained by repetitive BFB cycles, characterized by a cluster of rearrangements with multiple copy number amplifications and a foldback inversion pattern.[97] Other CGRs were characterized by rearrangements of three or more segments by a cut-and-paste mechanism. Some breakpoints (n = 21) were defined as miscellaneous if the rearrangement is not involved in the other 15 types.

Adjusted count per Gb

To convert the unit of count into per Gb, we divided the count accordingly by either the mouse or human diploid genome size (of autosomal and X chromosomes), which is 5,267,553,344 bp or 6,072,607,692 bp, respectively, and then multiplied it by 10^9 and a correction factor. The correction factor was used to account for missed calls in (1) genomic regions of insufficient read coverage and (2) repetitive regions where indel calls are more likely to be missed. The correction factor for indel counts was 1.122 for mice and 1.143 for humans, which take both (1) and (2) above into account. For SV counts, the correction factor was 1.097 for mice and 1.123 for humans. The correction factors were derived from read coverages of representative clones for mice and humans and the reference genome sequences.

Analysis of double-strand breaks (DSBs) induced by irradiation

DSBs must reflect the extent of initial damage induced by irradiation. Therefore, we used the number of DSBs to quantify the damage and consequences of irradiation (Figure 5H). DSB is not directly measurable from the sequencing data but can be estimated from the number of breakpoints associated with indel and SVs. However, some breakpoints may not be observable if they have been repaired seamlessly. To estimate the frequency of such seamless repairs, we looked into breakpoints in IR-induced SVs, except deletions, chromothripsis and BFB cycle (Figure 5C), examining traces of end-repairs occurred during junctional ligations, assuming that the proportion of seamless repair for total DSBs would be similar across all SV types. Our estimate for the seamless repair is 6.7% of all DSBs induced by irradiation. For observable breakpoints, each indel and SV was converted to DSB counts using the following conversion rules: (1) an indel is comparable to 1 DSB, (2) balanced inversion, balanced translocation, long deletion, ecDNA, and chromothripsis result from 2 DSBs, (3) Del-Inv, Del-Tra, Del-Ins, chromoplexy, and CGRs generally result from 3 DSBs. A few exceptions to this rule were three organoid samples: (1) Chromoplexy found in LI_20Gy_1 was accompanied by Del-Ins and can be explained by 4 initial DSB events (and so only 1 DSB was counted toward the Del-Ins), (2) CGRs found in both PA_2Gy_5 and LI_8Gy_1 can be explained by 4 DSBs. To validate the conversion approach to quantify radiation damage, we compared the estimated DSBs from the sequencing data with direct DSB count from the γ-H2AX immunostaining (the details of which will be followed; Figure 5G). We used these DSB counts to estimate the number and types of mutations induced by 1 Gy of irradiation (Figure 5H). The expected number of indel and SVs were taken from the coefficient of the regressors for *in vitro* samples, as shown in Figures 5A and 5B, respectively.

Analysis of IR-associated copy number variants

Manual inspection was carried out on the genome view results of Sequenza.[65] Primarily only arm-level copy number changes were considered CNVs, yet a few >10 Mb-sized copy number changes missing from SV calls were counted toward CNV. We observed that some CNVs, mostly of whole chromosome level, appear to be of the same origin across different samples. These CNVs could be ancestral variants, recurrent variants that occurred independently, or both. Another possibility being a germline variant was ruled out since not all the samples that are known to originate from the same mother cell shared the variant. To account for this potential source of bias and count CNVs that are only relevant to the IR exposure, we assessed whether these ''common'' CNVs are ancestral by examining a phylogenetic relationship of the samples. We inferred three phylogenetic trees, each from mouse pancreas organoids, mouse breast organoids, and human tumors, groups where common CNVs were found. In the trees, we looked for mutations attributed to the terminal branches, which correspond to the time points where the samples were irradiated. Accordingly, CNVs satisfying any of the following criteria were removed from the final CNV count: 1) a CNV shared across samples more than once in non-irradiated (0 Gy) samples, which is deemed culture-associated, 2) a CNV shared across samples that are tightly grouped in the phylogenetic tree, deemed an ancestral variant, 3) deletion or duplication (of a large size) that have been already included in structural variation calls, and 4) depth change at 0.5 unit, deemed 50:50 subclones (i.e., no dominant clones). To reconstruct the phylogenetic tree, we first merged somatic SBS variant-call-formatted files (VCFs) using samtools. We then calculated pairwise p-distance using an R package 'phangorn'[72] and built a tree using the neighbor-joining algorithm using an R package 'ape'.[73] The phylogeny was manually confirmed by examining the BAM files through the IGV at random positions among samples in the same lineage, whether the same variants are shared only in the samples. In addition, we also checked whether all the samples having the same common CNV had any shared mutation, which turned out there were no shared mutations.

Assay for transposase-Accessible chromatin using sequencing (ATAC-seq) analysis

To investigate whether chromatin status was associated with the formation of IRi-SVs, we performed ATAC-seq on control mouse pancreas organoids (n = 4, biological replicates). According to the published protocol,[98] nuclei isolation was initially performed as follows. Organoids (approximately 10k-100k cells) were retrieved from Matrigel using Cell Recovery Solution (Corning). After centrifugation at 300*g* for 3 min, 10k-100k cells were prepared and resuspended in 100 μL cold lysis buffer (Tris·HCl pH7.4 10 mM, NaCl 10 mM, MgCl2 3 mM, BSA 1%, Tween 20 0.1%, NP40 0.1% (Sigma), Digitonin 0.01% in distilled water). After 3 min incubation, 1 mL wash buffer (Tris·HCl pH7.4 10 mM, NaCl 10 mM, MgCl2 3 mM, BSA 1%, Tween 20 0.1% in distilled water) was added and centrifuged at 500 g for 5 min. The supernatant was discarded, and nuclei were resuspended in 95 μL DPBS. Approximately 10-20k nuclei were prepared and centrifuged at 500 g for 5 min. The supernatants were discarded, and the remaining nuclei were gently resuspended in 16.5 μL of DPBS. After the nuclei isolation, a transposition reaction mixture (16.5 μL nuclei in DPBS, 12.5 μL 2x TD buffer (Illumina), 0.5 μL 1% digitonin, 0.5 μL 10% tween 20, and 2.5 μL transposase (Illumina) in 5 μL distilled water) was prepared and incubated at 37°C for 45 min. A post-clean-up process was performed using MinElute PCR Purification Kit (Qiagen), and the transposed DNA was eluted in a 10 μl Elution Buffer. PCR amplification of the transposed DNA was performed with the following in PCR tubes (10 μL of transposed DNA, 2.5 μL Nextera index primer1 (Illumina), 2.5 μL Nextera index primer2, 25 μL NEBNext High-Fidelity 2x PCR Master Mix (New England Labs) in 10 μL distilled water). The thermal cycle was as follows: 72°C 5 min, 98°C 30 s, 5 cycles of (98°C 10 s, 63°C 30 s, 72°C 60 s), then 4°C hold. To determine the appropriate number of qPCR cycles, 5 μL of PCR amplified DNA was run with 5 μL NEBNext High-Fidelity 2x PCR Master Mix, 0.5 μL Nextera index primer1, 0.5 μL Nextera index primer2 in 3.85 μL distilled water +0.15 μL SYBR Green I (Invitrogen). A cycle was as follows: 98°C 30 s, 20 cycles of (98°C 10 s, 63°C 30 s, 72°C 1 min). The number of cycles corresponding to 1/3 of maximum fluorescent intensity was applied to qPCR for a 45 μL remnant of the PCR-amplified DNA. Lastly, we performed double library size selection using SPRIselect. Briefly, 22.5 μL (0.5x) of SPRIselect

beads were added, and the supernatant was transferred. Then, 58.5 μL (1.8x) of SPRIselect beads were added, and libraries attached to the beads were eluted in 40 μL of Elution Buffer. The final product (2x101 bp) was sequenced by Hiseq 4000 (Illumina). Using the raw fastq files of control organoids (n = 4), we performed primary data processing according to the ENCODE ATAC-seq Pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline), including alignment to GRCm38/mm10 by Bowtie2 aligner.[74] Then, peak calling was done by Model-based Analysis for CHIP-Seq (MACSv2, https://github.com/macs3-project/MACS). We used narrow peak calls in the downstream analysis.

Association of IRi-SV incidences with genomic features

To investigate whether IR-induced DSBs occurred more frequently in a certain genomic context, we first collected several genomic contexts information from our experiment (i.e., ATAC-seq in mouse pancreas organoids) as well as from Mouse ENCODE, including DNase-seq of the liver, H3K9ac CHIP-seq of the liver, H3K27me3 CHIP-seq of the liver (8 weeks old male C57BL/6), and Repli-chip of embryonic fibroblast (13.5 days male C57BL/6 embryo).[99] GC ratio in 50 bp windows was calculated using the mouse reference genome sequence. If genomic coordinates were from mm9, we converted them to mm10 using liftOver.[75] Using SVs identified from the irradiated mouse organoids (IRi-SVs) and SVs from controls (SVs from non-irradiated mouse organoids), we tested enrichments of IRi-SVs in euchromatin (open chromatin), early replicating regions, or genomic regions with high GC ratio, using the proportions of breakpoints around regions spanning peaked position ±20 bp. For ATAC-seq data, we additionally tested the enrichment of IRi-SVs detected in mouse pancreas organoids.

Sanger validation of deletion-insertion composite (Del-Ins)

For validation of Del-Ins, primers were designed to cover the candidate deletion and insertion area of the Del-Ins:

PA_2Gy_11 deletionF, 5′-GCCTGTGTTCAAATTGGGGG-3’; PA_2Gy_11 deletionR, 5′-AAAACCCATGCCCTCTGCTT-3’; PA_2Gy_11 insertionF, 5′-CCACATGGAACCTTATGCTGC-3’; PA_2Gy_11 insertionR, 5′-TTCTGACTGCCTTGGCACAG-3’; These regions were PCR-amplified using 500 ng genomic DNA, 5× buffer 10 μL, 0.25 mM of each dNTP, 0.2 μM primers and 0.5 unit of PrimeSTAR HS Taq polymerase (Takara) in a final volume of 50 μL or 500 ng genomic DNA, 10× buffer 5 μL, 0.2 mM of each dNTP, 0.2 μM primers and 1.25 unit of Pfu DNA polymerase (ThermoFisher) in a final volume of 50 μL at 95°C for 10 min; followed by 30 cycles of 98°C for 10 s, 60°C for 10 s, 72°C for 30 s; with a final extension at 72°C for 10 min (Taq polymerase) or 98°C for 5 min; followed by 98°C for 10 s, 60°C for 30 s, 72°C for 30 s; with a final extension at 72°C for 7 min (Pfu polymerase). The PCR reaction was run in the Veriti Thermal Cycler (Applied Biosystems) according to the manufacturer’s protocol. The PCR products were cloned into pTOP Blunt V2 vector system (Enzynomics), and Sanger sequencing was carried out using M13 universal primer.

Viability assay

Dissociated pancreas organoids were irradiated (0, 1, 2, 4, and 8 Gy; three biological replicates at each dose) and re-seeded in a 24-well plate (25,000/well). The growth of irradiated organoids was visually monitored for up to 6 days. On day 6 after plating, organoids were harvested for quantitative viability assay using CellTiter-Glo 3D kit (Promega). The total ATP amount in each well was measured by a luminometer (1420 Victor Light).

γ-H2AX immunostaining and 3D visualization

Mouse pancreas organoids on 12 well plates were irradiated (0 Gy and 2 Gy) and incubated for 1 h at 37°C. Then the organoids were harvested into 15 mL tube without dissociation procedure and centrifuged at 500 g for 3 min. Collected organoids were fixed in 4% formaldehyde for 20 min and permeabilized with 0.5% Triton X-100 in PBS for 20 min at room temperature. Blocking was performed with 4% bovine-serum albumin (BSA) in buffer (0.2% Triton X-100, 0.05% Tween 20 in PBS) for 1 h. The blocked samples were incubated with a primary antibody (phospho-histone H2A.X (Ser139) antibody, Cell signaling, #2577) in PBST (PBS with 0.1% (w/v) Triton X-100 and 0.02% (w/v) sodium azide) with a 1:300 dilution at 37°C for 4 h, followed by washing at 37°C for 2 h in PBST three times. The samples were then incubated with a secondary antibody (donkey anti-rabbit IgG antibody-Alexa Fluor Plus 488, Invitrogen, #A32790) in PBST with a 1:300 dilution at 37°C for 2 h, followed by washing at 37°C for 1 h in PBST three times. For nuclear staining, the samples were incubated in 1 μg/ml DAPI (Invitrogen) in PBST for 30 min followed by brief PBST washing three times. The samples were mounted on a slide glass with a spacer (iSpacer 0.15 mm, SUNJin Lab) filled with PBS and then covered with a coverslip. The mounted samples were imaged by confocal laser-scanning microscopy (Zeiss LSM 780) with a 63× objective (C-Apochromat 63x/1.20 W Korr M27).

Physical expansion of organoids for super-resolution imaging

For super-resolution microscopy of organoids, we used a modified magnified analysis of proteome (MAP) protocol.[42] Organoids were embedded in dense polyelectrolyte hydrogels via a physical tissue-gel hybridization approach,[100] for which a modified MAP solution with a decreased concentration of acrylamide was used (20% (w/v) acrylamide (Sigma), 10% (w/v) sodium acrylate (AK Scientific), 0.1% (w/v) bis-acrylamide (BIORAD), 0.03% (w/v) VA-044 (Wako Chemical) in PBS). Immunolabeled organoid samples were post-fixed with 4% PFA in PBS for 10 min followed by brief PBST washing three times. Post-fixed organoid samples were incubated in the modified MAP solution for 2 h. To locate organoids efficiently in 100-μm-thick gels in which organoids were embedded, we used a 3 mm punch to take out regions of interest containing organoids. After expanding the 3 mm punched gel with distilled water

three times, about a 3.2x linear expansion ratio was observed. Samples were mounted on a slide glass with distilled water using the Blu-Tack adhesive (Bostik) as a spacer, sealing with a coverslip on the top. The mounted samples were imaged using the confocal laser scanning microscope with the 63× objective.

γ-H2AX counting

ImageJ[101] was used for image handling. We divided the microscopic fields of view into 3-by-3 to select cells in organoid images. Fifty cells, recognized through the nuclear staining using DAPI, were randomly selected to count the numbers of γ-H2AX foci found in the Alexa Fluor Plus 488 channel. When selecting the cells, those with broad and high-intensity signals of γ-H2AX were avoided to minimize error. The number of DSBs in each selected cell was manually counted.

Selection pressure analysis

We conducted analysis of selection pressure exerted by IRi-ID and IRi-SV. First, we annotated all indel calls and breakpoints of SV calls in the 135 colonies, using the Ensembl Variant Effect Predictor (VEP).[102] For indel, we estimated the background distribution by randomly sampling 100,000 positions from the reference genome. The number of indels that fell into coding-sequence (CDS) regions were then compared between the observed and simulated data. If the proportion of IRi-ID hits in CDS regions was smaller than that in the simulated data, the observed data was considered to be under negative selective pressure. To isolate IRi-ID, or indels that were caused by IR damage, we calculated the likelihood of each variant being attributed to IR-associated indel signatures, or mID-A and hID-A, based on the estimated amounts of indel signatures present in the colony and distributions of indel features, or 83 indel sequence contexts, within these signatures. For SV, we considered only two SV classes, balanced inversions and balanced translocations, as these were the only classes that allowed for proper definitions of the background distribution. As in the indel case, we randomly sampled 100,000 positions from the reference genome. We compared breakpoints of the SVs located in exon, intron, and UTR regions between the observed and simulated data, categorizing them as potentially damaging variants.

Single-cell sequencing

Irradiation experiments for single cell sequencing were performed using 12 well plates. Organoids were seeded in neighboring 2 wells of 12 well plates. For irradiation, plates were placed with organoids centered in the irradiator room. After irradiation, organoids were incubated during 24–36 h at 37°C. Genomic libraries of single cells were obtained by direct single cell manipulation (see 'single-cell-derived clonal organoid acquisition') or respective library generation experiments.

For whole-genome amplification using MDA, 2Gy radiation was irradiated in mouse pancreas organoids. Single cells placed in 96 well plates were processed by REPLI-g Advanced DNA Single Cell Kit (Qiagen) using the manufacturer's protocol. Amplified gDNA were sequenced through Novaseq 151bp * 2 with 100 Gb. Sequenced fastq files were aligned on GRCm38/mm10 with BWA-MEM.[53] Structural variations were called using Delly 0.7.6[62] and filtered using an in-house script (SVAIR, https://github.com/phansol/gremlin). Then, all variants were visually inspected using IGV.[63] To exclude false positive calls generated during whole-genome amplification, only variants with (i) distances between breakpoints larger than 10kb for deletions and duplications and 100kb for inversions and (ii) breakpoints not overlapping with RepeatMasker[103] are included.

For primary template-directed amplification (PTA), single cells dissociated, 4Gy-irradiated human colon organoids were prepared following ResolveDNA Whole Genome Amplification Kit's instructions (BioSkryb cat # PN 100068). Samples were sequenced through Novaseq 151bp * 2 with 100 Gb. Sequenced fastq files were aligned on GRCh37 with BWA-MEM. SNV and indels were called by the union of Varscan and GATK Mutect2, and further filtered out with two additional modifications on previous filter criteria.[21] First, mutations where the number of 28 PTA samples with VAF $\geq$3% is larger than 2 were filtered out. Second, for indel, mutations, where the number of reads which harbor different indel mutations from called mutation is larger than 1, were filtered out. Among 28 PTA samples, 9 inefficiently amplified single cells and 7 multi-clonal cells were filtered out. To acquire single-cell mutations, mutations with VAF larger than 25% were used. To calculate the indel mutation drop-out rate, we used germline heterozygous SNPs from the bulk organoid. Total 10,000 heterozygous indel mutations with 40% < VAF <60% and depth $\geq$30 were listed, and calculate the number of mutations that pass our filter criteria for 12 single-cell PTA samples. We adjusted the number of mutations in each PTA sample using its own drop-out rate. For structural variations, we first called mutations from Delly and followed the in-house script (SVAIR, https://github.com/phansol/gremlin). Because of depth fluctuation, we considered only obvious breakpoints where the fraction of split alignment reads at both breakpoints was greater than 20% and both breakpoint depths were greater than 20. For short SV error, only indels longer than 7000 bp and duplication and deletion longer than 2000 bp were used.

For Strand-seq, to incorporate BrdU in human colon organoids, we added BrdU after a 4Gy irradiation experiment. A single nuclei isolation experiment was performed 36 h after irradiation using a standard nuclei isolation buffer.[46] After sorting G1 phase BrdU incorporated cells, we followed the research paper's instructions with three additional modifications. We used adapters and unique indexes for paired-end libraries,[104] 302nm UV light, and 2X beads–based short library removal. Pooled libraries were sequenced through Novaseq 151bp * 2 with average 3 Gb per sample. Sequenced libraries were aligned on GRCh37 with BWA-MEM. The number of Watson and Crick strand reads were calculated from the BAIT R package.[105] Samples with uneven coverage or coverage less than 1% were removed. Two researchers independently examined for structural variations and reached a consensus.

BotSeq and analysis

Mouse pancreas organoids were used for the BotSeq experiment. Samples from each well of the plate were considered as separate ones. Genomic DNA was extracted from 2Gy-irradiated sample using the DNeasy Blood & Tissue Kit (Qiagen).

DNA libraries were constructed with Truseq DNA PCR-Free Library Prep Kits (Illumina, San Diego, CA) following the manufacturer's protocol. Subsequent steps for BotSeq libraries were carried out as described in the previous literature.[47] First, quantification of DNA libraries was performed with the KAPA Library Quantification Kit ILLUMINA Platforms (Roche, KK4824) following the manufacturer's protocol. Next, each library was diluted to the target concentration using EB buffer (QIAGEN, 19086). The diluted libraries underwent 20 cycles of PCR amplification using KAPA HiFi HotStart ReadMix (Roche, KK2602) following the manufacturer's protocol, with custom primers from IDT (Coralville, IA) with a phosphorothioate bond(*) at the 3′ end.

Forward: 5′-AATGATACGGCGACCACCGAG*A-3‘

Reverse: 5′-CAAGCAGAAGACGGCATACGA*G-3′

The amplified libraries were cleaned-up with SPRIselect beads (BECKMAN COULTER life Sciences, B23317) following the manufacturer's protocols. The completed libraries were sequenced on the Novaseq platform with paired-end 151-bp reads, generating a total of 100Gb.

Sequenced fastq files were aligned on GRCm38/mm10 with BWA-MEM. Short indels were called by Varscan2, and we retained only variants that met the following criteria. (1) the number of F1R2 and F2R1 reads exceeds 4 each, with a VAF greater than 0.9 for both F1R2 and F2R1 reads. (2) no variant-supporting reads were detected in the mother organoids. (3) VAF in a panel of normals composed of mouse organoids was less than 0.1. (4) No additional mismatches are identified except within 5 bp of the variant in order to remove artifacts due to misalignment while retaining complex indels. (5) the distance from the read ends was greater than 5 bp. (6) the number of simple base repeats near the variant was fewer than 5. Additionally, all variants underwent visual inspection using Integrative Genomics Viewer (IGV)[63] to remove remaining false positive variants generated by misalignment.

CODEC and analysis

Human colon organoids were used for the CODEC experiment. Samples from each well of the plate were considered as separate ones. Genomic DNA was extracted from 4Gy-irradiated samples using Exgene Cell SV mini kit (GeneAll). All CODEC sequencing and bioinformatics analytic procedures were adopted from the previous literature.[48] Each library was prepared with 20 ng of input gDNA extracted from organoids. We produced ~150 Gb of genome sequences from each of the 12 libraries using 2 x 150 bp with the S4 flow cells in NovaSeq 6000 sequencers (Illumina Inc). Subsequently, they were mapped to the reference genome and the proportion of genomic regions covered by duplex DNA reads was measured for each sample using scripts in the previous literature (Figure S11D).[48] This information was used to calculate the adjust number of indels per diploid genome.

Confidence intervals of statistics

For IRi-SV and IRi-ID, we assumed Poisson distribution and computed $F^{-1}(p;\nu = 2n\mu) / 2n$ to derive the confidence interval of the means, where F^{-1} is the inverse cumulative distribution function of Chi-square distribution, ν is the degree of freedom, p is 0.915 for the upper bound and 0.085 for the lower bound, μ is the mean estimate, and n is the number of samples included in the analysis (Figure 5E). To estimate the confidence interval of the ratio of the means ($\mu_{\text{IRi-SV}}/\mu_{\text{IRi-ID}}$) for each dose group across different irradiation types (*in vitro*, *in vivo*, and post RT), we calculated Fieller's interval[41] using 'mratios' R package (Figure 5D). A caveat is that Fieller's method assumes normality for the underlying distributions. IRi-ID may well be approximated by normal distribution but IRi-SV is not because IRi-SV count is very small. This is in line with an observation that a dispersion relative to the mean, or the coefficient of standard deviation, is greater in lower doses (e.g., 1.33 vs. 0.52 for 1Gy vs. 4Gy). Therefore, comparisons between low doses (1Gy or 2Gy) and higher doses cannot be drawn confidently. To circumvent this, we used marginal sums for irradiation types and compared (1) *in vivo* vs. post-RT and (2) *in vitro* vs. *in vivo* plus post-RT (Figure 5D).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical calculations were performed using R version 3.6.0 (R Core Team, Vienna, Austria). A simple linear regression (using 'lm' function) was used for the relationship between two continuous variables, and one-sample t-test was used to evaluate p values for the learned coefficients. We used Fisher's exact test to assess a meaningful difference in the proportion of the number of events between two groups. To compare the means of a continuous variable between two groups, we used the Wilcoxon rank-sum test. p value < 0.05 was considered statistically significant. For the visual aid of assessing the meaningful difference between the mean estimates at a near 0.05 p value level, an 83% confidence interval was used where appropriate.[106] Otherwise, a 95% confidence interval was used. Detailed information for the calculations can be found in the figure legends. The "n" values appearing in the text and figure refer to the number of independent samples.