# Systematic mutagenesis assay promotes comprehension of the strand-bias laws for mutations induced by oxidative DNA damage

Hidehiko Kawai[1,2,*], Shungo Ebi[2], Ryusei Sugihara[2], Chiho Fujiwara[2], Yoshihiro Fujikawa[1,2], Shingo Kimura[3], and Hiroyuki Kamiya[1,2,*]

[1] Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan
[2] School of Pharmaceutical Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan
[3] Analysis Center of Life Science, Natural Science Center for Basic Research and Development, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan

* To whom correspondence should be addressed. Tel: +81-82-257-5301; Fax: +81-82-257-5334; Email: kawaih@hiroshima-u.ac.jp (H. Kawai) or hirokam@hiroshima-u.ac.jp (H. Kamiya).

## ABSTRACT

We have recently developed an efficient and sensitive method for analyzing mutations caused by various environmental and endogenous factors which utilizes nucleotide-barcoded *supF* shuttle vector libraries with a multiplexed NGS assay, referred to hereafter as *supF* NGS assay. Ionizing-radiation-induced cancer is known to be difficult to distinguish from spontaneous cancer, especially in the case of low-dose and low-dose-rate exposure, and discerning the underlying mechanisms of ionizing-radiation-induced cancer, especially the relationship between mutagenesis and carcinogenesis, is likely to be an arduous task. In the present study, we have attempted to address the mutations characteristic for exposure to low levels of ionizing radiation by using the *supF* NGS assay. As a result, a significant increase in mutations was detected at cytosines and guanines within 5'-TC-3':5'-GA-3' sites following chronic gamma-irradiation at a dose-rate of 1 Gy per day for the duration of 2 days. Since the number of detected mutations exceeded the expectations based on the quantity of DNA-damage induced by irradiation, we proceeded to explore the possibilities that a single DNA-lesion induced by irradiation may cause amplification of mutations. For this purpose, we utilized shuttle vector libraries with a single 8-oxo-7,8-dihydroguanine (8-oxo-G)-damaged residue introduced at different sites via an *in vitro* enzymatic method. Through a set of experiments, we revealed that a single 8-oxo-G-damaged residue can become a trigger for peripheral mutagenesis; intense generation of strand-biased mutations occured at 5'-TC-3':5'-GA-3' sites with specific localization in the secondary structures of single-stranded DNA, more frequently than not at sites different from the 8-oxo-G-damaged sites. Thus, this study provides a novel prospect for the role of DNA-lesions induced by environmentally or endogenously generated ROS in additional mutations. The high-performance mutagenesis assay presented in this study will advance research aimed at uncovering the mechanisms of mutagenesis and the intricacies relevant to carcinogenesis.

## Introduction

The progressive co-development of NGS technologies and cancer genomics databases have enabled the identification of distinct mutational signatures in various cancers, and so far about a half of the identified mutational signatures have underlying processes been assigned to them (1-4). Mutational signatures originate from the effects and interactions of a vast range of environmental and intrinsic sources, such as exposure to chemicals or physical factors like ultraviolet radiation among others, endogenously generated reactive oxygen species (ROS), errors in DNA replication, or defects in specific DNA repair mechanisms (5,6). Elucidating the specific vulnerabilities of cancer cells to different drugs and therapies can facilitate the development of more effective and personalized cancer treatments (7-10). Therefore, in addition to recognizing specific mutational signatures, it is also crucial to understand the processes responsible for them. Continuous research is expected to lead to full understanding of the individual mutational processes in sufficient detail to maximize their potential use in biomedical applications.

Ionizing radiation has been known to be mutagenic ever since Muller demonstrated that X-rays induce mutations in Drosophila sperm (11). The extensive studies on cancer incidence and mortality in atomic-bomb survivors from Hiroshima and Nagasaki have established exposure to ionizing radiation as an independent risk factor for many different types of cancer (12-14). It is now clear beyond doubt that high doses of ionizing radiation from medical radiation therapy or nuclear accidents can increase the risk of developing cancers such as leukemia, thyroid cancer and breast cancer (12-15). In addition, a large body of animal studies has also demonstrated that exposure to ionizing radiation can increase the risk of developing cancer in a dose-dependent manner (16-19). Numerous studies using murine germlines and *Arabidopsis thaliana* cells have demonstrated that ionizing radiation can cause various types of mutations, among which enrichment of indels (20-24). It is important to note that the risk of developing radiation-induced cancer depends on a complex multitude of factors, such as radiation dose and the dose-rate, duration of exposure, age at exposure, the cell types in the tissue and the level of their differentiation, as well as the individual's genetic susceptibility. Ionizing radiation can damage DNA through direct or indirect ionization mechanisms, resulting in various types of DNA lesions including damage to single bases, single-stranded and double-stranded DNA breaks (SSBs and DSBs) (25). The type of mutations induced in the process of mutagenesis is thought to have a decisive role in cancer risk. DSBs and clustered DNA lesions induced by ionizing radiation are considered to be the most severe types of DNA damage, having the highest toxicity to cells and mutagenicity (26). Notably, the repair of DSBs and cluster damage can be substantially delayed or compromised, which determines the higher likelihood of mutations and further genomic instability (27,28). Overall, comprehending the mutational signatures resulting from exposure to ionizing radiation is a formidable undertaking, which despite an ample amount of existing studies, including whole genome sequencing, is likely to continue in the foreseeable future.

Radiation exposure can be detected and measured down to very low levels, and according to the linear non-threshold (LNT) hypothesis, radiation exposure even at low doses is associated with a finite risk of cancer induction (29). Unlike other mutagens, radiation can produce complex DNA damage through both the direct ionization of DNA molecules or by indirect oxidation via reactive oxygen species (ROS), which are also continuously generated in living cells under physiological conditions (30,31). As a product of normal oxygen metabolism, ROS are a major source of endogenous DNA lesions *in vivo*, and their levels can substantially increase even further by exposure to various environmental stressors and inflammation (32). It has been postulated that one-third of radiation-induced DNA lesions result from direct radiation absorption and two-thirds result from the ionization of water molecules and the subsequent production of ROS (33). The chemical nature of the DNA lesions generated by ROS is indistinguishable between spontaneously generated ROS and indirect radiation effects. The rate of production of endogenous DNA lesions has been estimated for SSBs and apurinic/apyrimidinic sites (AP-sites or abasic sites) at more than 2,000 and 10,000, respectively; and for 8-oxo-7,8-dihydroguanine (8-oxo-G, also known as 8-hydroxyguanine) at 1,000 to 100,000 per cell per day (31,34-38). These amounts are equivalent to those induced by gamma-radiation delivered at dose rates of at least 1 to 10 Gy per day (31). In particular, 8-oxo-G is known as a major DNA lesion generated by ROS, and it is well known that 8-oxo-G on the replicating strand leads to misincorporations of adenine opposite the lesion at a rate of 10-75 %, subsequently resulting in G:C to T:A (C:G to A:T) transversion mutations (37,39,40).

Present-day large-scale integrated genomic analyses are able to distinguish the radiation-induced events associated with specific types of cancer (41). However, the fact that indirect ionization via ROS is the predominant mechanism by which radiation induces multiple types of lesions in genomic DNA is the reason why it is exceedingly difficult to distinguish between most radiation-induced cancers and spontaneous cancers. In recent years, whole-genome sequencing studies of clonal lines from somatic cells in various tissues of mice and humans have identified spontaneous somatic mutations (42,43). In these studies, C:G to T:A transitions and C:G to A:T transversions at non-CpG sites were observed as the majority of spontaneous mutations in addition to C:G to T:A transitions at CpG dinucleotides that occur from the deamination of 5-methylcytosine to thymine, which may reflect cell division (2). Furthermore, a recent study of somatic mutations using clonal cell populations from hematopoietic stem cells of whole-body X-ray-irradiated mice demonstrated that radiation exposure significantly increased small deletions and insertions, and also both C:G to T:A transitions and C:G to A:T transversions at non-CpG sites in a dose dependent manner (44). Moreover, experiments on the *gpt* delta mouse system with potassium bromate used as a ROS-inducing agent have also suggested that the same mutations at non-CpG sites can be induced by ROS (45). The mechanisms behind the mutations induced by ROS seem to be well explained with the damage or modification of either G or C bases, and their repair processes and errors. However, the prominent mutation hotspots are expected to be influenced by many factors in addition to the sequence context and gene's function, such as the DNA repair process, chromatin structure, cell cycle, and more.

Recently, we reported the development of an advanced mutagenesis assay using a random-DNA-barcoded *supF* shuttle vector library with non-SOS-induced *E. coli*, which we called "*supF* NGS assay" (46). In the study, we demonstrated that the *supF* NGS assay allowed us to quickly accumulate a very substantial amount of data and to contribute some new findings regarding UV-induced mutation signature and mutational processes in mammalian cells, which was surprising given that UV-associated mutagenesis have been studied in great depth and is considered well-known. Thus, we demonstrated that the *supF* NGS assay has advantages in sensitivity and throughput; it is also especially useful as a damage-specific mutagenesis assay, since different DNA lesions of interest can be introduced at different sites and sequences and in different numbers in shuttle vectors using *in vitro* construction. Various experiments can be designed based on environmental mutagen exposure and/or genetic perturbations, which will help to elucidate the mechanisms behind mutational signatures. In a previous study, we have demonstrated that chronic Cs-137 gamma-irradiation at a low dose-rate of 1 Gy per day efficiently induces senescence via ROS-mediated activation of an ATM-P53-P21 pathway in human fibroblast cells, but not in the many other types of cell lines that we tested (47,48). Unlike fibroblasts, other cell lines showed continuous growth with little mitotic cell death under conditions of chronic irradiation. Despite this different outcome, radiation-induced DNA damage detected by gamma-H2A.X and 53BP1 were similarly accumulated in a dose-rate dependent manner in all cell lines. In this study, using our recently developed *supF* NGS assay under experimental conditions of chronic Cs-137 gamma-irradiation, we have attempted to address the mutational signatures characteristic for continuous exposure to low levels of radiation.

## RESULTS AND DISCUSSION

### Mutation frequencies were increased in response to chronic gamma-irradiation at a dose-rate of 1 or 2 Gy per day.

In order to test whether we can investigate the mutational signatures associated with exposure to chronic gamma-irradiation at a dose-rate as low as 1 Gy per day in mammalian cells, at first the *supF* mutant frequencies were analyzed by using the RF01 *supF*-mutation indicator *E. coli* strain (49) and a conventional *supF* mutagenesis assay, which compares the number of colonies on selection plates versus titer plates. Briefly, the random 12-nucleotide barcoded ($N_{12}$-BC) pNGS2-K2 shuttle vector libraries containing the *supF* gene were constructed to be used with the NGS mutagenesis assay (Figures 1A and 1B). The constructed libraries were transfected into the human osteosarcoma U2OS cells, allowing the libraries to generate mutations with their replication in the cells. The cells were cultured for 2 days in a 5% $CO_2$, 37°C incubator at a dose-rate of 0, 1 or 2 Gy per day using a Cs-137 gamma-ray source (Figure 1C) (47). The shuttle vector libraries replicated in the cells during the 2 days of incubation were purified by digestion of unreplicated vectors with $N^6$-methyladenine-dependent restriction endonuclease *Dpn* I, and then were electroporated into RF01 and seeded on either titer or selection plates in appropriate dilutions. The *supF* mutant frequencies were obtained by counting colonies grown on the plates. The results from three independent transfection experiments

indicate that chronic gamma-irradiation at a dose-rate of 1 to 2 Gy per day can produce *supF* mutants with frequencies 0.5 to 1.5 × $10^{-3}$ higher than the spontaneous one (Figure 1D).

**Chronic gamma-irradiation induced strand-biased mutations at C:G sites were observed in the mutation spectra from colonies grown on *supF*-selection plates.**

Next, the mutation spectra were analyzed using the *supF* NGS assay, the procedure for which was described in detail in our previous report (46). Briefly, to obtain a substantial amount of data, including data on spontaneous mutations that occurred endogenously in cells, roughly two thousand colonies of RF01 grown on *supF*-selection plates and containing the $N_{12}$-BC shuttle vectors with a variety of mutations in the *supF* gene were harvested for each sample. Using the libraries purified from the harvested colonies as PCR templates, DNA fragments including the *supF* gene (Figure 1B) were amplified by two independent PCRs with the index-tagged primers and analyzed by the MGI DNBSEQ platform. The sequencing data containing the mutation spectra were classified and error-corrected by using the $N_{12}$-BCs and the variant frequencies (VFs). The VFs are the frequencies of identical variant call(s) found in a unique $N_{12}$-BC. Variant sequences exceeding 0.4 VF can be assumed to be true mutations and not errors that have occurred in the process of NGS analysis. In addition, to eliminate PCR errors from the data, only the identical $N_{12}$-BC with recurring mutations between two independent PCR samples were extracted and used for analysis. The number of mutations at each position in the sequence was plotted and shown in Figure 2A. The data from three independent transfections (TF1, TF2, and TF3) appear to have good reproducibility, and the mutations were located in the region of the *supF* gene (positions from 56 to145) and its promoter region (position -15 is the first nucleotide of the promoter), consistent with the results previously reported by many groups, including us (46). The positional bias for the number of detected mutations was caused by the *supF* selection process, which relies on the destruction of the function of transfer RNAs. This positional bias was observed in irradiated samples as well, as chronic gamma-irradiation seems to induce mutations at the same positions and some other specific positions (Figure 2A). In order to compare appropriately the mutation spectra and frequencies across samples, one hundred $N_{12}$-BCs with at least one mutation from the group with the highest number of reads, or the maximum number of available $N_{12}$-BCs with mutations were subjected to further analysis. These approximately one hundred unique $N_{12}$-BCs for each sample contain 176-211 mutations, indicating that more than one mutations were present in some of the $N_{12}$-BCs (Figure 2B). Single nucleotide substitutions (SNSs) were the predominant type of mutation, and the proportion of SNSs was not significantly altered by irradiation (Figure 2C). It is important to note that SNSs had occurred mainly from C:G (Non-transcribed:Transcribed strand, respectively) or G:C, where C and G were substituted into various bases (Figure 2D). Even though not significant, it was observed consistently that chronic gamma-irradiation induced mutations mainly from C:G in a dose-rate dependent manner, suggesting that due to some underlying reason, radiation-induced mutations were strand-biased. The mutational spectra and positions for SNSs were depicted in Figure 2E, and it was confirmed that mutations at

spontaneous hotspot positions were increased by irradiation, and C:G mutations were observed more frequently than G:C mutations.

**The mutation spectra of C:G substitutions differ between single mutations per $N_{12}$-BC and multiple mutations per $N_{12}$-BC.**

In order to address the mechanisms underlying the C:G strand-biased mutations induced by irradiation, the individual SNSs were analyzed in more detail. As shown in Figure 2E, the majority of mutations from C or G were detected at specific positions that are considered required for the *supF*-selection. The detected mutations were classified as single mutations in a unique $N_{12}$-BC only when they were selected as a standalone single mutation, while the multiple mutations per $N_{12}$-BC included those that were co-detected along with the aforementioned single mutations. It should be noted that the multiple mutations were able to be detected spontaneously and the frequencies of mutational events were as low as $10^{-3}$. These facts implied indirectly that most multiple mutations were produced in a single process, thereby, it was anticipated that some of the single mutations in the *supF*-responsible region could be distinguished as mutations directly produced from DNA damage induced by irradiation. Therefore, the mutations detected at different radiation dose rates were presented separately for the single or more than one mutation per $N_{12}$-BC (Figure 3). The percentages of $N_{12}$-BCs with either single or multiple mutations were not significantly altered by irradiation (Figures 3A, 3B, and 3C), suggesting that the strand-biased mutation induction by irradiation took place in both cases. For an in-depth look, the mutations were classified by number per $N_{12}$-BC (1, 2, 3, and 4~ mutations per $N_{12}$-BC) and depicted according to their positions (Figure 3-figure supplement 1). It was anticipated that single mutations would be exclusively located at specific positions in either the *supF*-tRNA cloverleaf or promoter region, multiple mutations would also include other regions, and C:G mutations would be dominant throughout all regions. However, contrary to expectations, a significant difference was observed between single and multiple mutations in the proportion of mutated bases at C:G/G:C sites (Figure 3D). Based on the data, C:G to G:C transversions were major for single mutations, while C:G to T:A transitions were major for multiple mutations. It appears that exposure to radiation may increase the frequency of C:G to A:T transversions as single mutations and C:G to T:A transitions as multiple mutations, while decreasing the frequency of G:C to A:T transitions in multiple mutations (Figure 3E). There is a remote possibility that by increasing C:G to A:T transversions, which are a characteristic feature of 8-oxo-G, exposure to radiation may become a trigger for the strand-biased C:G to T:A cluster mutations. In previous reports by us and others, the majority of the spontaneous mutations were located at 5'-T<u>C</u>-3':5'-<u>G</u>A-3' sites in both DNA strands, which is well known as COSMIC APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) signatures SBS2/SBS13 (7,50,51). In order to investigate in more detail the SNS mutation spectrum, all possible 192 trinucleotide contexts of mutations were compared between the complementary DNA strands, and it became clear that the SNSs occurred mainly at 5'-T<u>C</u>N-3':5'-N<u>G</u>A-3' sites in both DNA strands in all samples (Figure 3-figure supplement 2), without a noticeable difference between non-

irradiated and irradiated samples. From this data analysis it became clear that a deeper exploration of these complex mutation spectra was required to address the mutations induced at lower frequency.

**Mutational spectra of single versus multiple mutations provide evidence for the *supF*-selection bias.**

Next, the mutational positions within the sequence were ranked based on the number of detected mutations, and the mutational bases and positions with their flanking sequences were shown in Figure 4A. The ranks of single mutations were not significantly changed by irradiation. It proved challenging to identify any discernible features of irradiation, but consistent with the above analysis, in irradiated samples the number of multiple mutations located at 5'-TC-3':5'-GA-3' sites was increased more than those at 5'-GA-3':5'-TC-3' sites. It became clear that mutations at specific positions such as 111 and 125 can be detected only as multiple mutations, even though they are located in the *supF*-tRNA cloverleaf region (Figure 4A and Figure 4-figure supplement 1A) (52). The method of *supF*-selection has an advantage over other drug selection markers in the range of mutations it allows us to identify; however, there exists a significant bias in the analysis of mutation spectra due to its reliance on the *supF*-tRNA properties. The spectra of single mutations are depicted with the *supF*-tRNA cloverleaf structure in Figure 4B. As described in our previous report, the potential hairpin loop of the quasi-palindromic sequence between positions -8 and 12, GGCGCGTCATTTGATATGATGCGCC, in the promoter region of the *supF* gene may lead to template switching and give rise to complementary cluster mutations (Figure 4-figure supplement 1B). It is obvious that the mutations were distributed unequally at different positions, even at the same 5'-TC-3':5'-GA-3' sequence in the cloverleaf-region, suggesting that other factors in addition to the functional bias of the *supF*-gene may be involved in the mutation bias, such as the secondary structure of single-stranded DNA during the replication, repair, or transcription process, etc. Therefore, we predicted the single-stranded DNA secondary structures of the analyzed region by Mfold web server (Figure 4-figure supplements 2 and 3) (53,54), and the number and substituted bases of the identified mutations at each position in the analyzed sequence were depicted separately for single or multiple mutations per $N_{12}$-BC (Figure 4-figure supplements 4-9). However, it proved difficult to find any correlation between mutations and sequences or their structures, which may be due to the *supF*-selection bias.

**The NGS mutation analysis without *supF*-selection provides mutation frequency and non-biased mutation spectra.**

The process of *supF*-selection is effective to obtain a substantial amount of data on mutation spectra through concentrating the mutants at a magnitude of thousands. An appropriate *supF*-selection process provides accurate mutant frequency and trends. However, the mutation spectra obtained by *supF*-selection exhibit a strong bias in favor of certain positions and substituted bases that are associated with the function of the *supF* gene. On the other hand, without the selection it is difficult to

investigate thoroughly mutation spectra, especially at very low frequencies where it is hard to differentiate from background mutations. We hypothesized that our newly-developed NGS analysis can be applied to samples directly prepared from purified libraries digested by *Dpn* I without transformation into the indicator *E.coli*. Henceforth, the *Dpn* I-digested libraries purified from the cells were directly analyzed by using DNB-seq, following the PCR for multiplex tagging and amplification. The threshold for VF was set to 0.2. As expected, the regions between -67 and -20, where $N_{12}$-BC sequence oligonucleotides were inserted, display very similar pattern of variant sequences with high VFs close to 1.0, reproducible between libraries and independent transfection experiments (Figures 5A and 5B). These sequences turned out to be a suitable internal standard for the NGS analysis. The types of mutations were clearly different between the constructed $N_{12}$-BC region, positions from -67 to -20, and the *supF* gene, positions from -19 to 214 (Figure 5A). The mutational positions were spread relatively evenly in the sequence compared to the data of *supF*-selection shown in Figure 2A, and seemed to have a degree of similarity between samples. As expected, a similar number of $N_{12}$-BCs were classified across all samples, but the number of sequences with variants were significantly different and appeared to be increased by irradiation (Figure 5B). Interestingly, only $N_{12}$-BCs bearing a variant sequence with lower VFs were increased at positions from -19 to 214, and even from -67 to -20. Therefore, we obtained the mutant frequencies separately for the two regions (-67 to -20 and -19 to 214) from the number of reads with a variant sequence (only the highest number of reads with a variant sequence was selected if the unique $N_{12}$-BCs contains multiple variant sequences) divided by the number of total reads (Figure 5C). The results shown that the mutant frequencies were slightly but statistically significantly increased by chronic irradiation, which very closely resembles the results from the *supF* colony assay shown in Figure 1D. The proportions of mutation types and the sequences of SNSs at positions from -19 to 214 remained very similar between non-irradiated and irradiated samples (Figures 5D and 5E), indicating that the strand-biased mutations from C:G observed in the *supF*-selection condition were not present here. However, when the mutational positions were examined in detail, positions 5 and 27 emerged as significant hotspots in all samples, and their prominence among other hotspots was more evident compared to *supF*-selection. It appears that the mutations from G:C at positions at 5 and 27 counteracted the strand-biased mutations from C:G at multiple positions detected in *supF*-selection. When all possible 192 trinucleotide contexts of mutations were compared between the complementary DNA strands (Figure 5-figure supplement 1), it became clear that the results could be a reflection of positional biases. The data set demonstrates that chronic exposure to radiation leads to an increase in the frequency of SNSs at 5'-T<u>C</u>N-3':3'-N<u>G</u>A-5' sites in both DNA strands, along with multiple other mutations that exhibit a position-dependent strand bias. The positions where SNSs occurred were essentially the same between non-irradiated and irradiated samples. When the mutational sequences were ranked and compared between NGS analysis without selection (w/o-*supF*-selection) and *supF*-selection (Figure 5-figure supplements 2A and 2B), novel mutational hotspots were revealed, including positions 5, 27, 111, and 125, among others. These mutations could not be identified from *supF*-selection where they were masked by other position-biased mutations, although they were detected as multiple mutations per $N_{12}$-BC as shown in Figure 4A.

**Positional bias contributes to the discrepancy of the data on strand-biased mutations between *supF*-selection and without *supF*-selection.**

The NGS data from the w/o-*supF*-selection experiments was analyzed by the number of mutations per $N_{12}$-BC (Figure 6A). About 20 to 30 percent of $N_{12}$-BCs contained multiple mutations, and those seemed to be slightly increased by irradiation, although the limitations of the $N_{12}$-BC-variant sequence classification in the w/o-*supF*-selection condition bring a degree of uncertainty. In fact, both single and multiple mutations per $N_{12}$-BC appeared to be increased by irradiation (Figure 6B), and both were strand-biased; however, the radiation-induced increase occurred from G:C for single mutations, and from C:G for multiple mutations per $N_{12}$-BC (Figures 6C and 6D). The mutations were then depicted within the sequence and with the predicted secondary structure in Figure 6-figure supplements 2 and 3, analogous to Figure 4-figure supplements 4 and 5. For obvious reasons, the number of detected mutations at a single position was relatively less in the w/o-*supF*-selection condition compared to *supF*-selection; on the other hand, the w/o-*supF*-selection analysis could detect mutations at a same number of positions as the *supF*-selection, with some discrepancies. The data between *supF*-selection and w/o-*supF*-selection were compared in Figure 7. It may appear that the data regarding the effect of irradiation is in disagreement between *supF*-selection and w/o-*supF*-selection, but in reality the data were very consistent and highly compelling if strand bias and positional bias for spontaneous mutational hotspots at 5'-T<u>C</u>-3':5'-<u>G</u>A-3' sites are considered. The SNSs are the major mutations at positions from -19 to 214 which are generated in cells through replication. The positions of SNSs are relatively evenly spread in the sequence for all samples, but there exist hotspot sites that are shared with background mutations. So far, from the data two noteworthy points have emerged. First, chronic gamma-irradiation for 2 days at a dose-rate as low as 1 Gy per day induces a significant amount of mutations, which are very similar to the spontaneously induced mutations. This may indicate that chronic gamma-irradiation generates the same initial events that are responsible also for spontaneous mutations, that these events are the root cause of the strand bias associated with the 5'-T<u>C</u>-3':5'-<u>G</u>A-3' sequence in both strands, and are therefore an essential factor that influences the mutational frequencies and the substituted bases. This can be addressed by the experiments without *supF*-selection because, as already discussed, the strong bias associated with *supF*-selection obstructs the elucidation of these matters. As a second point, the estimated numbers of DNA lesions expected to be induced by gamma-irradiation, including the direct and indirect effects, are almost equal to the number of detected mutations. This may indicate that a DNA lesion directly causes a mutation without repair, or that a single DNA lesion can trigger mutagenesis and lead to amplification of mutations.

**A single 8-oxo-G efficiently induces mutations apart from the damaged base, referred to as action-at-a-distance mutations.**

The 8-oxo-G is known as a major DNA lesion generated by ROS produced endogenously or by gamma-irradiation (37,38). Previously, by using the conventional *supF* assay, we have reported that an 8-oxo-G can serve as a trigger for mutations at 5'-TC-3':5'-GA-3' sites, referred to as action-at-a-distance mutations, although the underlying mechanism still remains unclear (51,55). Here, to investigate the effects of 8-oxo-G in inducing SNSs and its potential contribution to radiation-induced mutagenesis, we introduced a single 8-oxo-G outside the region of the *supF* gene, at positions 176 or 177 (refer to Figure 1B), which has no effect on the tRNA function of the *supF*-gene, in a series of pNGS2 shuttle vectors. The schematic diagram of eight different shuttle vectors constructed by the *in vitro* enzymatic method (56) − pNGS2-K1, -K2, -K3, and -K4, containing either non-damaged guanine (G) or 8-oxo-G is depicted in Figure 8A. These vectors contain an M13 intergenic region with opposite orientations relative to the *supF* gene, which allows us to incorporate 8-oxo-G at specific sites in the opposite strands of the shuttle vector library; the vectors also contain a SV40 replication origin on opposite sides relative to the *supF* gene, which changes the direction of replication forks and/or transcription (46). The two different $N_{12}$-BC libraries (with inserted G or 8-oxo-G) for pNGS2-K1 to -K4 were independently prepared, and it was confirmed that no mutation could be detected by NGS analysis. In the preparation process for NGS samples, KOD-one polymerase synthesized a strand complementary to both the damaged and non-damaged sites evenly, which fortuitously allowed for the insertion of 8-oxo-G at positions 176 and 177 in the *in vitro* synthesized library to be confirmed by its detection as thymine instead of guanine in approximately half of the reads (Figure 8-figure supplement 1A). The *supF* mutant frequencies for each library containing either G or 8-oxo-G were analyzed following transfection into cells by using the conventional *supF* assay. As shown in Figure 8B, the *supF* mutant frequencies were increased by the insertion of 8-oxo-G in the region outside the *supF* gene, although the effect was statistically significant only in pNGS2-K3 and -K4. In parallel, the libraries after being purified from the cells and digested by *Dpn* I, were analyzed by NGS, as in w/o-*supF*-selection experiments. The nucleotide positions and types of mutations based on the combined data from three transfections are shown in Figure 8C. Consistent with the data on mutant frequencies obtained from the conventional *supF* assay, more mutations were detected after insertion of 8-oxo-G compared to the control insertion of G for all libraries (Figure 8C). Referring back to what was mentioned in Figure 5A, the region between -67 and -20 showed very similar pattern of variant sequences between G and 8-oxo-G. The number of mutations in the region from -19 to 214, excluding positions 176 and 177 with the introduced G or 8-oxo-G, were clearly increased. Notably, the patterns of mutations following insertion of G were similar in all libraries, while it appears that upon insertion of 8-oxo-G two distinct patterns were observed – one in pNGS2-K1 and -K4, and another in pNGS2-K2 and -K3. The number of $N_{12}$-BCs with at least one mutation for regions -67 to -20 and -19 to 214 indicate that a single 8-oxo-G caused the generation of other SNSs in the region (Figure 8D). The frequency of mutations induced at positions 176 and 177 with inserted 8-oxo-G was only in the order of $10^{-3}$ reads, suggesting that the inserted in the library 8-oxo-G is efficiently repaired in the cells (Figure 8-figure supplement 1B). The mutant frequencies calculated as a ratio between the number of reads with mutations and the number of total reads for positions from -19 to 214, except for positions 176 and 177, were relatively lower compared to those obtained from the conventional *supF* assay but

revealed very similar trends (Figures 8E versus 8B). This is consistent with the irradiation experiments using the conventional assay versus NGS, shown in Figures 1D and 5C. It is possible that the loss of data comes from reads containing true mutations with VFs below 0.2, which were excluded in order to avoid false positives from NGS errors.

**A set of pNGS2 libraries revealed that a single 8-oxo-G induces strand-biased mutations exclusively from G:C but not from C:G sites.**

The comparison of the types of mutations and the nucleotide bases of SNSs are shown in Figures 9A and 9B, respectively. The distribution of mutation types was not significantly altered between the G and 8-oxo-G inserted libraries, with SNSs being the dominant type of mutations observed (Figure 9A). On the other hand, the SNS patterns significantly varied depending on the library and the DNA-strand in which 8-oxo-G was introduced: while in pNGS2-K1 and -K4 substitutions were predominantly from C, in pNGS2-K2 and -K3 substitutions were from G (Figure 9B). The trinucleotide signatures of SNSs in both groups exhibited a bias towards 5'-TCN-3':5'-NGA-3' sites (Figure 9-figure supplement 1); however, this is not conclusive, again due to the strong positional biases as shown in graphs depicting the region (Figure 9C and Figure 9-figure supplements 2 and 3). In pNGS2-K1 and -K4, which have the 8-oxo-G inserted in the top strand of the *supF* gene region, the C to N mutations were uniformly distributed in specific positions throughout the region from 80 to 200. On the other hand, in pNGS2-K2 and -K3, which have the 8-oxo-G inserted in the bottom strand of the *supF* gene region, the G to N mutations were distributed in a wider region but in fewer highly-specific positions, such as 5 and 27. The positions 5 and 27 were previously mentioned as hotspots for spontaneous mutations that were further increased by irradiation. The results clearly demonstrate that the presence of a single 8-oxo-G can induce mutations at sites apart from the original damage-site, and the 5'-TC-3':5'-GA-3' site on the same strand emerges as the mutational signature associated with this phenomenon. Also, there is no significant difference between pNGS2-K1 and -K4 versus pNGS2-K2 and -K3, suggesting that the orientation of the replication/transcriptions are not a crucial factor for the strand bias of the mutations. However, the data also indicates that, in addition to the trinucleotide signatures, there is a strong positional bias, since there are 5'-TCN-3':5'-NGA-3' sites where mutations cannot be observed, such as positions 24 and 34. The incomplete classifications of the $N_{12}$-BC-variant sequences showed that the insertion of 8-oxo-G led to a slight increase in the proportion of multiple mutations per $N_{12}$-BC (Figure 10A). Nevertheless, in addition to the incomplete classification, it is also difficult to make conclusion due to the relatively low number of mutations; both single and multiple mutations per $N_{12}$-BC show a strand bias, and multiple mutations may even further reinforce this trend (Figure 10B and Figure 10-figure supplement 1).

**The NGS data from *supF*-selection complement the data from w/o-*supF* selection experiments.**

The NGS analysis in the w/o-*supF*-selection experiments provides less biased data compared to that of *supF*-selection; however, smaller amounts of mutation spectra data could be obtained. Based on the experiments we conducted with chronic irradiation, we can expect that the combined data from both w/o-*supF*-selection and *supF*-selection would provide valuable information if the bias from the selection is taken into account. Therefore, to delve deeper into the mutational spectra at the sites apart from the introduced G or 8-oxo-G in a series of libraries (referred to as G-library or 8-oxo-G library, respectively), the $N_{12}$-BC libraries from roughly 2,000 RF01 colonies grown on *supF* selection plates were analyzed. The $N_{12}$-BCs-variants, excluding positions 176 and 177, were extracted using VF 0.4. The detected number of $N_{12}$-BCs with variant sequence(s) ranged from 76 to 247, and between 120 and 390 mutations were detected at positions -19 to 214 (Figure 11-figure supplement 1A). In order to ensure a fair comparison between samples, the numbers of $N_{12}$-BCs with mutations were limited to about one hundred, as already explained about the data in Figure 2, and the results from TF1, TF2, and TF3 were combined (Figure 11–figure supplement 1B). The positions and types of mutations were very similar among G-libraries, as well as between pNGS2-K1 and -K4, and pNGS2-K2 and -K3 for 8-oxo-G-libraries (Figure 11A). Here the results exhibited a trend similar to the data from the w/o-*supF*-selection experiments shown in Figure 8C. However, there are a few distinct differences, such as the proportion of deletions or deletion-insertions (Figure 11-figure supplement 1 versus Figure 9A). The differences may be due to the *supF*-selection bias, which is often observed especially at lower mutation frequencies. Although the proportion of base substitutions were different, the proportions of individual SNSs seemed to be very similar between the data of *supF*-selection and w/o-*supF*-selection (Figure 11B versus Figure 9B). The proportions of SNSs showed significant difference depending on the 8-oxo-G-library: while in pNGS2-K1 and -K4 substitutions primarily occurred form C, in pNGS2-K2 and -K3 they occurred from G. As expected, the mutations were predominantly located in the tRNA-cloverleaf and promoter regions of the *supF* gene, which causes a noticeable bias on not only positions but also substituted bases. The G-libraries share more similarities in mutations, suggesting that background mutations can be obtained from *supF*-selection, which could help mitigate experimental bias. As described in our previous report, our novel *supF* NGS assay is at least ten-times higher in efficiency than the conventional *supF* assay, and is able to identify practically the whole range of SNSs in the *supF*-selection. Therefore, the quality of the libraries is essential for accurate analysis in the investigation of mutations induced by a variety of mutagens, environmental factors, or deficiency of genes. As already mentioned, all libraries were synthesized by *in vitro* enzymatic reactions and purified following treatment with T5 exonuclease. It has been observed that the frequencies of endogenous mutations in the *supF* gene are associated with the presence of nicks in shuttle vectors (46,57). However, it is impossible to completely eliminate vectors with poor quality, and this is why, in order to ensure the quality of libraries, it is important to thoroughly detect the mutations in the control libraries and treatments. The G-libraries shared similarities between pNGS2-K1 and -K3, with more mutations at positions -3, 5, and 12 compared to the pNGS2-K2 and -K4 libraries (Figure 11C and Figure 11-figure supplement 3). These mutations may have occurred via a template switching process (Figure 4-figure supplement 1B), and there is a possibility that the distance from the SV40 replication origin or the opposite replication strand may

have contributed to their frequency. In contrast to G-libraries, 8-oxo-G-libraries exhibited similar patterns for mutations in pNGS2-K1 and -K4, as well as in pNGS2-K2 and -K3. In pNGS2-K1 and -K4 the mutations were detected at the same positions in G-libraries and 8-oxo-G-libraries, but more mutations were observed at C in 8-oxo-G-libraries (Figure 11-figure supplement 2). In pNGS2-K2 and -K3, although the mutations were observed predominantly at G, mutations at positions outside of the *supF* cloverleaf and promoter regions were significantly decreased compared to the data from w/o-*supF*-selection. A strong strand bias was observed in the trinucleotide mutational signatures at 5'-TCN-3':5'-NGA-3' sites (Figure 11-figure supplement 3), which is similar to the data from w/o-*supF*-selection upon initial observation (Figure 9-figure supplement 1). In *supF*-selection experiments, the insertion of 8-oxo-G led to a slight increase in the proportion of multiple mutations per $N_{12}$-BC (Figure 11-figure supplement 4A), that is consistent with the data from w/o-*supF*-selection shown in Figure 10A. When the mutations detected by *supF*-selection were classified as single or multiple mutations per $N_{12}$-BC (Figure 11-figure supplement 4B), differences in the mutational spectra were revealed between single and multiple mutations per $N_{12}$-BC, and the multiple mutations were very similar to the case of w/o-*supF*-selection. The same trends were shared with the data from the experiments with gamma-irradiation using pNGS2-K2, shown in Figure 3. These data indicate that single mutations per $N_{12}$-BC may exhibit strong *supF*-selection bias due to the reliance on the loss of function of the *supF* gene as an amber suppressor tRNA.

The comparison between the positions and spectra of all mutations from *supF*-selection and w/o-*supF*-selection is shown in Figure 12 (Figure 9-figure supplements 2 and 3, and Figure 11-figure supplements 5-10). Relatively non-biased data on the mutations induced by 8-oxo-G were able to be obtained from the combination of w/o-*supF*-selection and multiple mutations per $N_{12}$-BC from *supF*-selection. The following is a summary of the observations: SNSs almost exclusively occurred at 5'-TC-3':5'-GA-3' sites throughout the analyzed region in all libraries; The mutations at 5'-TC-3':5'-GA-3' sites were generated spontaneously in all G-libraries without strand bias; The insertion of a single 8-oxo-G introduced strongly strand-biased mutations on the same strand at 5'-GA-3':5'-TC-3' sites. Furthermore, there were also apparent positional biases even without *supF*-selection. As mentioned above, positions between -8 and 17 represent a distinct and exclusive region for template switching which produced particular mutations: SNSs at -3 (G>A), 5 (G>C), 12 (T>C), and Del-Ins at 3-7 (TTGAT>ATCAA); these base substitutions were almost invariably biased and predetermined by fixed bases for a template strand in a quasi-palindromic sequence. Although the underlying molecular mechanisms of template switching are still unclear, it was hypothesized that the 8-oxo-G incidentally caused the replication fork to stall or collapse at a position separate from the 8-oxo-G, followed by disengagement of the lagging strand and invasion of an adjacent replication fork through the annealing to partial homologies in the other strand. Except for this region, the substituted bases in SNSs at 5'-TC-3':5'-GA-3' sequences were not predetermined by fixed bases (Figure 12), suggesting these SNSs may have occurred via abasic sites at 5'-TC-3':5'-GA-3' sites, which can be assumed with a high probability to have emerged from the uracil generated by APOBEC3 cytosine deaminase targeting the cytosine at 5'-TC-3' in single-stranded DNA (51). There exists a position-dependent bias for the substituted bases in all the libraries due to unknown reasons. Among the possible factors

involved are the efficiency of removal of uracil, the different translesion DNA synthesis (TLS) polymerases, etc. In contrast, it is clear that mutations were never detected at positions -1, 10, 24, and 34 (coldspots), even where the sequences at those positions were 5'-TC-3':5'-GA-3', and significantly more were detected at positions 5 and 27 (hotspots), especially in the libraries with an inserted 8-oxo-G.

**The secondary structure of single-stranded DNA is a crucial factor determining the positions of mutations and possibly the substitutions.**

In cells, single-stranded DNA is formed during DNA replication, recombination, and transcription. Single-stranded DNA is promptly coated by Replication Protein A (RPA) complexes, which serves as protection to prevent the induction of cluster mutations found in various cancer genomes (1,58,59). Based on the results from the NGS assay, it can be inferred that both template switching and mutational spectra at 5'-TC-3':5'-GA-3' sites are important for the mutational signature. It seems highly probable that the single-strand state of DNA may be involved in the mutational process induced by the introduction of 8-oxo-G and irradiation, or even induced spontaneously. Thus, to explore the factors contributing to the strand-biased and position-biased mutations, analysis of the secondary structure of single-stranded DNA was conducted using Mfold web server for the region between positions -20 and 60, which includes six 5'-TC-3':5'-GA-3 (5'-GA-3':5'-TC-3) sites with two mutational hotspots (5 and 27) and four coldspots (-1, 10, 24, and 34). The region was divided into overlapping smaller regions of 41 bases by applying a 5-base sliding window (Figure 13). As it was to a certain degree expected, the mutational hotspot positions 5 and 27 were located in the loops of two independent stem-loop structures, regardless of the exact configuration. These stem-loop structures shared the same strand between positions 13 to 26. Also, the above mentioned mutational cold spots, positions -1, 10, 24, and 34, were all located in the stem of the stem-loop structures; position -1 interacted with 10 by G:C base pairing, and the same was observed for 24 and 34. The secondary structures of regions other than -20 to 60 were also predicted by Mfold web server, but these stem-loop structures had a shorter stem, suggesting that they were less fastened, and therefore, it was difficult to find any correlations between the structures and the pattern of mutations (Figure 13-figure supplements 1 and 2).

**The presence of a single 8-oxo-G lesion increases the frequency of strand-biased multiple mutations at 5'-TC-3':5'-GA-3' sites.**

Finally, to answer the question whether 8-oxo-G influences the induction of mutations by chronic gamma-irradiation exposure, the same set of experiments conducted on pNGS-K2 were carried out using pNGS2-K3 and -K4 libraries with an inserted non-damaged G or an 8-oxo-G. A series of libraries was introduced into cells, and the cells were maintained for 2 days under non-irradiated conditions or chronic gamma-irradiation at a dose-rate of 2 Gy per day. The combined data from two

independently prepared libraries and three independent transfection experiments are presented in a set of figures analogous to those shown thus far. The mutations from the w/o *supF*-selection experiments in both libraries pNGS2-K3 and -K4 were increased in response to chronic irradiation as in pNGS2-K2 (Figures 5A and 5B), and interestingly, more mutations were observed in the pNGS2-K3 and -K4 libraries with inserted 8-oxo-G (Figures 14A and 14B). The proportions of mutation types were not significantly altered by irradiation, and SNSs represent the majority of mutations in both libraries (Figure 14-figure supplement 1A); also, the strand bias for the G:C mutations induced by 8-oxo-G was slightly intensified by irradiation (Figure 14-figure supplement 1B). The SNSs seemed to be increased at exactly the same positions at which they were induced by only insertion of an 8-oxo-G. In pNGS2-K3, SNSs were increased at positions 5, 27, 91, 162, 167, and elsewhere at 5'-GA-3':5'-TC-3' sites, while in pNGS2-K4 they were increased at positions 95, 111, 187, or 190, and other 5'-TC-3':5'-GA-3' sites (Figure 14-figure supplements 1C and 2-6). The increase of both the mutation frequencies obtained from the NGS assay (Figure 14C) and the absolute number of mutations (Figure 14-figure supplement 7A), excluding the mutations at positions 176 and 177, were synergistically greater than the additive effects of 8-oxo-G insertion and irradiation. As expected from the results in figures 6D and 10B, the proportions of SNSs were significantly changed by the insertion of an 8-oxo-G, while a similar effect of radiation could not be clearly inferred for either single or multiple mutations per $N_{12}$-BC (Figure 14-figure supplement 7B). Regarding the 8-oxo-G insertion position 176 for pNGS2-K3 or 177 for pNGS2-K4, the rates of insertion of an A base instead of C were also slightly increased by irradiation for both 8-oxo-G-inserted libraries (Figure 14D). This may suggest that irradiation somehow increases the frequency of incorporation of A by replicative DNA polymerases, or interferes with DNA repair pathways like OGG1-initiated BER for 8-oxo-G:C or MutYH-initiated BER for 8-oxoG:A, where the mechanisms involve many processes, such as damaged base excision or gap filling. The 8-oxo-G:C to T:A mutations were primarily single mutation per $N_{12}$-BC, suggesting that increased frequency of insertion of A is not directly required for the additional mutations (Figure 14-figure supplement 8). Although it has been reported that the mismatch repair pathway and/or OGG1 may be involved in the induction of mutations at positions apart from the damaged base, further studies are necessary to uncover the detailed mechanisms (50,60).

**The bias towards hotspot mutations at 5'-TC-3':5'-GA-3' sites related to the strand of insertion of 8-oxo-G and irradiation is a result from the process of *supF*-selection**

The subtle mutation spectra obtained from *supF*-selection experiments can provide complementary data sets to help us understand the data from the w/o-*supF*-selection experiments. Following the previous analyses of the *supF*-selection data, the mutational spectra for all samples were adjusted depending on the number of reads per $N_{12}$-BC (Figure 15-figure supplement 1A and 1B). As expected from all data sets so far, the major type of mutations was SNSs, and their proportions were slightly increased by both the insertion of an 8-oxo-G and irradiation (Figure 15-figure supplement 1C). A series of mutational biases associated with the insertion of an 8-oxo-G and irradiation were also observed as predicted: the positional bias specific for the *supF*-selection, causing the mutations to be

exclusively detected in the promoter and cloverleaf regions of the *supF* gene (Figure 15A, with reference to Figure 2A and Figure 11A); the strand bias manifested as mutations from C:G in opposite directions following the insertion of an 8-oxo-G, which depends on the damaged base inserted strand, and which is slightly intensified by irradiation (Figure 15B, refers to Figure 11B); the bias for hotspots, which were found at 5'-TC-3':5'-GA-3' sites (Figure 15C, with reference to Figures 2F, 11C, and 15-figure supplement 2); the multiple mutations per $N_{12}$-BC were increased by the insertion of an 8-oxo-G and slightly increased by irradiation (Figure 15-figure supplement 3A, with reference to Figures 3A and 11-figure supplement 4A) and the proportions of SNSs in single mutation per $N_{12}$-BC were significantly different from the w/o-*supF*-selection data, due to the hotspot bias and preferential substituted bases (Figure 16-figure supplement 3 versus Figure 15-figure supplement 7, with reference to Figure 3D versus Figure 6D, and Figure 10 versus Figure 11-figure supplement 4B).

**The hotspot sites for mutations induced by the insertion of 8-oxo-G and irradiation in addition to 5'-TC-3':5'-GA-3' sites**

Furthermore, abundant mutational spectra from all analyses revealed the common mutation sites that were detected at even higher frequency than some 5'-TC-3':5'-GA-3' sites among all the libraries pNGS2-K1 to -K4, at positions 71 (5'-CCC-3':5'-GGG-3'), 85 (5'-GGG-3':5'-CCC-3'), and 131 (5'-CCT:AGG-3'), (Figures 7, 12, and 15-figure supplements 4, 5). All these positions were C:G sites and were positioned next to 5'-TC-3':5'-GA-3' sites, namely 70 (5'-TCC-3':5'-GGA-3'), 86 (5'-GGA-3':5'-TCC-3'), and 130 (5'-TCC-3':5'-GGA-3'). As one possible mechanism, the C:G to T:A transition due to the deamination of C to U at 5'-TC-3':5'-GA-3' sites may result in the generation of secondary 5'-TTC-3':5'-GAA-3' sites for deamination. In fact, at all these positions the double nucleotide substitutions (DNSs), CC to TT/TA/AT/TG, were also detected with similar frequencies to the single mutations at the secondary sites. However, since the single mutations at the secondary sites were detected independently from the primary mutations, it is more likely that the deamination occurred at those two sites simultaneously rather than sequentially; another possibility is polymerase errors by TLS polymerases following the primary uracil or abasic sites, although this is unlikely because the sites of mutations were only C:G.

In addition to the series of mutational biases described so far, positional biases for the substituted bases at both 5'-TC-3':5'-GA-3' and 5'-GA-3':5'-TC-3' sites, although difficult to discern, were certainly present. The C:G to T:A transitions were the major SNSs detected at the sites in w/o-*supF*-selection and in multiple mutations from *supF*-selection experiments (Figures 3D, 6D, 10B, 11-figure supplement 4B, and 14-figure supplement 7B). The C to T transitions at 5'-TCN-3' are associated with COSMIC SBS2 and are a type of mutation observed in various cancers. These mutations are thought to arise from during DNA replication from uracil residues generated by the deamination of cytosine by AID/APOBEC enzymes (61,62). There seem to be some positional biases affecting their proportions, but it proved difficult to make conclusions from the limited number of mutations at each position. In addition to the transition mutations, the C:G to G:C and A:T transversions, which are associated with

COSMIC SBS13, were observed at the same sites and are believed to arise from the glycosylation of uracil, leading to abasic sites, and subsequent error-prone replication by TLS polymerases (63-65). The positional bias affecting the proportions of C:G to G:C and C:G to A:T transversions could also be observed (Figures 5F, 9C, 14-figure supplement 1B). The C:G to G:C transversions were predominantly observed at positions 95 and 125 for 5'-TC-3':5'-GA-3' sites, and 5, 91, 126, and 167 for 5'-GA-3':5'-TC-3' sites and were affected by both the insertion of 8-oxo-G and irradiation (Figure 14-figure supplements 3 and 5). Notably, the proportions of the mutations were not affected by any sequences other than the 5'-TC-3':5'-GA-3' sequence, prompting us to consider different factors that might be discerned from the data. We entertained the possibility that one such factor might be the secondary DNA structure, and therefore, the secondary structure predictions for single-stranded DNA were further explored. During replication, repair, and recombination, RPA is required for shielding exposed single-stranded DNA in order to prevent the formation of secondary structures, which plays a crucial role in maintaining genome stability (66). The interaction between RPA and single-stranded DNA needs to be highly dynamic, which aids in destabilizing secondary structures and hairpin structures of single-stranded DNA (67,68). This process creates small single-stranded regions that are essential for optimal DNA synthesis and repair. The structure of these single-stranded DNA regions is expected to exhibit high dynamics, and alterations are likely to occur around positions that have undergone deamination. Therefore, we conducted further predictions of the secondary structures of single-stranded DNA, as described in Figure 13, but the sequence at 5'-TC-3' sites was changed to T/U instead of C with a focus on the 5'-TC-3':5'-GA-3' sequence. Position 125 at 5'-TC-3':5'-GA-3' sites, and positions 126 and 167 at 5'-GA-3':5'-TC-3' sites, were located in the loop of the newly formed stem-loop structures (Figure 16). The underlying mechanisms are not known but it is possible that the stem-loop structure, which may contain a C, U or an abasic site either in single-stranded DNA or in the adjacent base-paired helical stem-structure, makes a difference for the DNA replication mechanisms of abasic sites. Abasic sites are endogenous DNA lesions that are highly frequent in cells, and their repair pathways and TLS are crucial for cells to maintain genomic stability and cellular survival (65). Recently, there has been a significant focus on understanding the difference between the abasic sites that occur in double-stranded DNA and those that occur in single-stranded DNA. Rev1 (REV1 DNA Directed Polymerase) and Pol ζ (DNA Polymerase zeta) are involved in TLS of abasic sites by filling in dCMP, resulting in C:G to G:C transversion mutations during DNA replication (69-71). Pol η (DNA polymerase eta), another TLS polymerase, is known to contribute to both C:G to G:C and A:T transversions when filling in abasic sites (72-74). In addition, PrimPol (Primase-Polymerase) has been reported to prevent the transversion mutations at abasic sites (75,76). Furthermore, HMCES (5-hydroxymethylcytosine binding, ES cell specific) protein interacts with abasic sites in single-stranded DNA exposed at DNA replication forks and blocks TLS (77-79).

Exposure to ROS is one of the major sources for endogenous DNA-lesions in genomic DNA, and 8-oxo-G is known as one of the most frequently produced oxidative base-damages. We have previously reported that chronic gamma-irradiation at a dose-rate of 1 Gy per day introduced cellular senescence and inflammatory reactions to the normal human fibroblast cells more efficiently than acute gamma-irradiation at a dose-rate of 1 Gy per minute, and produce ROS via mitochondrial damage-response

(47,48). In this study, a set of *supF*-NGS assays were carried out for the mutations induced by chronic gamma-irradiation and an 8-oxo-G, and an integrative perspective from the data of *supF*-selection and w/o-*supF*-selection provided much less biased mutation spectra comparing to an accumulated number of mutagenesis studies using various types of selective marker gene. As a consequence, the chronic gamma-irradiation induced base-substitutions from C:G at 5'-TC-3':5'-GA-3' sites, which were supposed to be increased at the same sites for spontaneous mutations rather than at the DNA-lesions directly induced by the radiation. A single 8-oxo-G produced by *in vitro* enzymatic reactions significantly increased the same mutations at the 5'-TC-3':5'-GA-3' sites apart from the introduced 8-oxo-G, and the mutations were highly strand-biased depending on a strand introduced with an 8-oxo-G. And, the strand-biased mutations were significantly further increased by chronic gamma-irradiation via the mechanism that remains undetermined yet. It is possible that it is because of the induction of APOBECs-expressions by exposure to the radiation and DNA-damages (79,80). It has been reported that the analysis for mutational signatures implicated APOBEC-mutagenesis following radiotherapy for patients with glioma (81). The frequencies and the substituted bases of mutations at each position, especially 5'-TC-3' sites, are strongly associated with the predicted secondary structure of single-stranded DNA. The underlying mechanisms for the achievement of the mutations are intricate and even stochastic, which is dependent on the balance of a number of physiological pathways influenced by many intra- and extracellular factors. The NGS analysis and large public data resources such as The Cancer Genome Atlas (TCGA) (82) and the Genomic Data Commons (GDC) (83), will keep providing results that are factually certain. However, in order to fully utilize the available data, the underlying individual and combined mechanisms must be addressed and confirmed by reproducible methods in an ordinary research laboratory setting without specialized equipment or techniques. The high-throughput *supF*-NGS assay provided in this study, that is developed as a more versatile and simple method than that presented in the previous study, can provide more prospect for various designed experiments in any laboratories. Especially, the data from w/o-*supF*-selection could provide less-biased mutational spectra, and the results indicate any nucleotide sequences, even a non-coding gene or artificially designed sequences, can be used for this assay. Therefore, it is promising that the further development of the sequence for the analysis or applying to long read sequencing, such as PacBio Revio system, will provide us more informative data. Accumulating data from future research will reveal the intricacies of the mutagenesis, and will hopefully contribute to machine learning for the discovering the patterns for the mutations induced by the complexed mechanisms, or even to improve base editing techniques for the treating genetic diseases.

## Materials and methods

### Construction of pNGS2-$N_{12}$-BC libraries

The vector maps of four series of *supF* shuttle vectors – pNGS2-K1, -K2, -K3, and -K4, and the construction method for the pNGS2-$N_{12}$-libraries were described in detail in our previous paper (46). Briefly, pNGS2 was digested with *Eco*RV-HF restriction enzyme (New England Biolabs, USA) and

purified by agarose gel electrophoresis followed by extraction using FastGene Gel/PCR Extraction kit (NIPPON Genetics, Japan). The purified *Eco*RV-HF-digested pNGS2 was combined with single-stranded randomized $N_{12}$-oligonucleotides (5'-dGGCCTCAGCGAATTGCAAGCTTCTAGAAGGCGATNNNNNNNNNNNNATCGAATTCGGATCCTTTCTCAACGTAA-3', where N = A, T, G or C) and assembled using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The randomized $N_{12}$-oligonucleotides were commercially synthesized and purified on a reversed-phase column (FASMAC, Japan). The *E. coli* JM109 electro-competent cells were transformed with either the assembly reaction mixture or non-assembly control mixture by electroporation. The transformants were serially diluted, seeded on LB agar plates, and incubated at 37°C overnight. The number of colonies on each plate was counted, and a higher than 99% efficiency of the assembly reaction was confirmed. Approximately $10^4$ colonies were harvested from the plates by scraping. The pNGS2-$N_{12}$-BC libraries were prepared by purification from the harvested cells by using either GenElute Plasmid Miniprep (Merck, Germany) or NucleoBond Xtra Midi (TaKaRa Bio Inc., Japan). Each prepared library was interrogated for the presence of potential *supF* mutants by using the indicator non-SOS induced *E. coli* RF01 strain (49), and a mutation frequency of less than $10^{-4}$ was considered mutant free.

**Construction of pNGS2-$N_{12}$-BC libraries containing either an 8-Oxo-7,8-dihydroguanine (8-oxo-G) or lesion-free (G)**

The procedures for construction of pNGS2 libraries containing either an 8-oxo-7,8-dihydroguanine (8-oxo-G) or lesion-free (G) were described in detail previously (56). The 5'-phosphorylated oligonucleotides carrying 8-oxo-G or G (5'-P-GAGGTGCTGCG*CCGCCGGATC-3' for pNGS2-K1 and -K4, and 5'-P-GATCCGGCGG*CGCAGCACCTC-3' for pNGS2-K2 and -K3; P=phosphate, G*=8-oxo-G or lesion-free G) were synthesized and purified by two sequential independent purifications (Hokkaido System Sciences, Japan). Briefly, *E. coli* HB101 bearing VCSM13ΔPS ($P_{BAD}$-pII) were transformed with each individual pNGS2-$N_{12}$-BC library and seeded on LB agar plates. At least $10^4$ colonies were harvested in LB medium by scraping, and the optical density of the cell suspensions was measured at 610 nm ($OD_{610}$). The cells were collected by centrifugation, resuspended in 15% glycerol/LB medium (adjusted to $OD_{610}$ = 50-250), and stored at -80 °C. For culturing, a batch from the cryopreserved stock was added to 2×□YT medium and incubated at 37□°C with shaking until $OD_{610}$ = 0.5. Arabinose at a final concentration of 0.2% was added, and the culture was further incubated overnight (37□°C with shaking). The supernatant containing the phage particles was collected by two consecutive centrifugations. The phage particles were precipitated by adding a one-tenth volume of 20% PEG-6000/2.5□M NaCl solution to the supernatant. The phage particles were then collected by centrifugation and resuspended in 10□mM Tris-HCl (pH□8.0). The suspension was subjected to enzymatic treatment with DNase I and RNase A in order to digest any contaminating bacterial nucleic acids. Then, the circular single-stranded DNA for the pNGS2-$BC_{12}$-libraries was extracted from the phage particles by treatment with proteinase K and SDS, followed by purification using QIAGEN-tip20 (QIAGEN, Venlo, Netherlands). To construct the circular double-stranded DNA

for pNGS2-$BC_{12}$-libraries containing either an 8-oxo-G or lesion-free, a five-fold molar excess of oligonucleotides bearing either 8-oxo-G or G was mixed with the phage-extracted circular single-stranded DNA in 1 × Phusion HF Buffer (New England BioLabs). After denaturation at 90 °C for 2 min followed by rapid cooling, the solutions were heated to 70 °C for 5 min, and then allowed to gradually cool to room temperature. The polymerase and ligase reactions were simultaneously carried out in 1 × Phusion HF Buffer supplemented with 1 mM dithiothreitol, 0.2 mM dNTPs, and 1 mM $NAD^+$, using Phusion High-Fidelity DNA polymerase and Taq DNA ligase (New England BioLabs). The reaction took place at 50 °C for 30 min, and then at 65 °C for 1 h. The sample was then vigorously vortexed to inactivate the enzymes, and after addition of dam methyltransferase and T5 exonuclease was incubated at 37 °C for 3 h to methylate the A in the 5'-GATC-3' sequence and digest the DNAs other than closed circular double-stranded DNA. The sample was purified with a PureLink PCR Purification Kit (Thermo Fisher Scientific, USA) and precipitated with ethanol. The quality of the circular single-stranded DNA for pNGS2-$BC_{12}$-libraries was checked by gel electrophoresis using agarose gels with GelRed (Biotium, USA).

**Cell culture, transfection, gamma-irradiation, and *supF* mutation analyses**

After their construction in the above-outlined manner, pNGS2-$N_{12}$-BC libraries were transfected into the human osteosarcoma U2OS cell line (HTB-96™, American Type Culture Collection, USA) using Lipofectamine 2000 (Thermo Fisher Scientific), as we have previously described (46,49,50). Briefly, trypsinized cells were seeded into either 6-well plates or 35-mm dishes at $3 \times 10^5$ cells/well the day before transfection. After culturing for 24 h, 400 ng of pNGS2-$BC_{12}$-library were transfected according to the manufacturer's instructions. Transfected cells were maintained at 37°C in a humidified, 5% $CO_2$ atmosphere for 48 hours. To investigate the effects of chronic gamma-irradiation at different dose-rates (1,000 or 2,000 mGy per 24 hours) on mutations, a specially designed radiation facility was used, where cell culture incubators were placed at different distances from a $^{137}$Cs-radiation-device (1.11 TBq), as described in our previous paper (47). The dose-rate at each location of sample placement inside the incubators was measured using a GD-302M glass dosimeter (AGC Techno Glass, Japan). At 48 hours after transfection the library was extracted from the cells as previously described (46), To remove unreplicated shuttle vectors, the library was digested with the restriction enzyme *Dpn* I (New England BioLabs), which possesses specificity for methylated 5'-$G^mATC$-3' sequences. The library was then precipitated by ethanol and dissolved in 15 μl of sterile deionized water (Nacalai, Japan). The *Dpn* I-treated libraries were used for both the conventional *supF* mutagenesis assay and the preparation of NGS samples (w/o-*supF*-selection). For analyzing the *supF* mutant frequency, the *Dpn* I-digested library was transformed into the indicator *E. coli* strain RF01. Serial 5-fold dilutions of transformed RF01 samples were then seeded on titer and selection LB agar plates. Titer plates contained kanamycin and chloramphenicol, while selection plates contained streptomycin and nalidixic acid in addition to kanamycin and chloramphenicol. The *supF* mutant frequencies were estimated based on the number of colonies grown on the individual plates. The presumed number of 2,000 colonies was harvested from selection plates as a cell pellet by scraping,

and libraries were extracted using GenElute plasmid miniprep kit for preparation of NGS samples (*supF*-selection).

**Preparation of samples with unique indexes for multiplexed NGS**

Each sample with a unique index for multiplexed NGS was prepared by PCR using KOD One PCR Master Mix -Blue- (TOYOBO, Japan). For the standard protocol, either 10 ng of the library extracted from the RF01 colonies grown on selection plates (*supF*-selection) or 1 μl of *Dpn* I-treated library (w/o-*supF*-selection), as described above, was used for a 50 μL PCR reaction (30 cycles of amplification at 98°C for 10 s, 60°C for 5 s, and 68°C for 1 s) with a series of primer sets containing a pre-designed 6-nucleotide sequence ($N_6$) as an index sequence for distinguishing each sample by multiplexed NGS data analysis (46). PCR products were purified using NucleoSpin Gel and PCR Clean-up (TaKara Bio Inc.), and then analyzed and quantified by Nanodrop spectrometer (Thermo Fisher Scientific) and agarose gel electrophoresis. The uniquely indexed samples were combined together in the proper proportion according to the expected number of NGS reads for each sample. More than ten reads per $N_{12}$-BC sequence were obtained for all samples in this study (approximately 16 Mb and 400 Mb of sequencing data for the selection plates and *Dpn* I-treated libraries, respectively).

**DNBSeq sequencing and sequencing processing**

Preparation of NGS libraries and sequencing were conducted by Seibutsu Giken inc. (Japan), as previously described (46). The DNA nanoballs were prepared by using the MGIEasy Circularization Kit (MGI Tech). The samples were subjected to multiplex{Canet, 2022 #15}ed deep sequencing in 200-bp paired-end mode on the NGS using a BGISEQ-G400 platform (MGI Tech, China).

**NGS data analysis**

The process of NGS data analysis was described in our previous paper (46). Briefly, the FASTQ files for each $N_6$-indexed data set in an NGS run (sorting was performed by Seibutsu Giken inc.) were used in the following manner: 1) The reverse-complement sequences were obtained using SeqKit tools (85); 2) The paired-end reads were merged using CASPER (86); 3) The merged sequences were mapped to a reference sequence using minimap2 (87); 4) The error corrections were performed based on a whitelist by UMI-tools (88). For the $N_{12}$-BC extraction, only one-base differences between $N_{12}$-BC sequences were allowed, and corrections were implemented using the following regular expression: (".*(GGCGAT){s<=1}(?P<cell_1>.{12})(?P<umi_1>.{0})(ATCGAA){s<=1}.*"); 5) Using VarDict, the sequence data for all $N_{12}$-BC sequences with more than 9 reads per $N_{12}$-BC was employed for the alignment to the reference sequence, variant calling and calculation of allele

frequencies (89); 7) Spotfire version 7.11.1 (TIBCO, USA) was used for data organization, filtering, and visualization.

## DATA AVAILABILITY

NGS raw data has been uploaded to DDBJ database (http://www.ddbj.nig. ac.jp) under the BioProject accession number PRJDB17510 and BioSample accession number SAMD00737992-SAMD00737998. See Supplementary table S1 for list of nucleotide sequences of primer sets used in multiplex sample preparation for NGS in the respective figures.

## SUPPLEMENTARY DATA

Figure 3 − figure supplements 1-2; Figure 4 − figure supplements 1-9; Figure 5 − figure supplements 1-2; Figure 6 − figure supplements 1-2; Figure 8 − figure supplement S1; Figure 9 − -figure supplement 1-3; Figure 10 − figure supplement 1; Figure 11 − figure supplements 1-10; Figure 13 − figure supplements 1-2; Figure 14 − figure supplements 1-8; Figure 15 − figure supplements 1-12; Supplementary table S1.

## ACKNOWLEDGEMENTS

A part of this study was carried out at the Joint Usage/Research Center Program of the Research Institute for Radiation Biology and Medicine (RIRBM), Hiroshima University.

## FUNDING

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI grant numbers JP 22K12375 (H. Kawai) and 22H03751 (H. Kamiya).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

1. Nik-Zainal, S., Alexandrov, Ludmil B., Wedge, David C., Van Loo, P., Greenman, Christopher D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, Lucy A. et al. (2012) Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell, 149, 979-993.

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

3. Degasperi, A., Zou, X., Dias Amarante, T., 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. Martínez-Jiménez, F., Movasati, A., Brunner, S.R., Nguyen, L., Priestley, P., Cuppen, E. and Van Hoeck, A. (2023) Pan-cancer whole-genome comparison of primary and metastatic solid tumours. Nature, 618, 333-341.

5. Sherwood, K., Ward, J.C., Soriano, I., Martin, L., Campbell, A., Rahbari, R., Kafetzopoulos, I., Sproul, D., Green, A., Sampson, J.R. et al. (2023) Germline de novo mutations in families with Mendelian cancer syndromes caused by defects in DNA repair. Nature Communications, 14, 3636.

6. Yurchenko, A.A., Rajabi, F., Braz-Petta, T., Fassihi, H., Lehmann, A., Nishigori, C., Wang, J., Padioleau, I., Gunbin, K., Panunzi, L. et al. (2023) Genomic mutation landscape of skin cancers from DNA repair-deficient xeroderma pigmentosum patients. Nature Communications, 14, 2561.

7. 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.

8. Mendelaar, P.A.J., Smid, M., van Riet, J., Angus, L., Labots, M., Steeghs, N., Hendriks, M.P., Cirkel, G.A., van Rooijen, J.M., Ten Tije, A.J. et al. (2021) Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. Nature Communications, 12, 574.

9. Kocakavuk, E., Anderson, K.J., Varn, F.S., Johnson, K.C., Amin, S.B., Sulman, E.P., Lolkema, M.P., Barthel, F.P. and Verhaak, R.G.W. (2021) Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. Nature Genetics, 53, 1088-1096.

10. Levatić, J., Salvadores, M., Fuster-Tormo, F. and Supek, F. (2022) Mutational signatures are markers of drug sensitivity of cancer cells. Nature Communications, 13, 2926.

11. Muller, H.J. (1927) Artificial Transmutation of the Gene. Science, 66, 84-87.

12. Ozasa, K., Shimizu, Y., Suyama, A., Kasagi, F., Soda, M., Grant, E.J., Sakata, R., Sugiyama, H. and Kodama, K. (2012) Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: an overview of cancer and noncancer diseases. Radiat Res, 177, 229-243.

13. Grant, E.J., Brenner, A., Sugiyama, H., Sakata, R., Sadakane, A., Utada, M., Cahoon, E.K., Milder, C.M., Soda, M., Cullings, H.M. et al. (2017) Solid Cancer Incidence among the Life Span Study of Atomic Bomb Survivors: 1958–2009. Radiation Research, 187, 513-537.

14. Ozasa, K., Cullings, H.M., Ohishi, W., Hida, A. and Grant, E.J. (2019) Epidemiological studies of atomic bomb radiation at the Radiation Effects Research Foundation. Int J Radiat Biol, 95, 879-891.

15. Rühm, W., Laurier, D. and Wakeford, R. (2022) Cancer risk following low doses of ionising radiation – Current epidemiological evidence and implications for radiological protection. Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 873, 503436.

16. Tanaka, I.B., Tanaka, S., Ichinohe, K., Matsushita, S., Matsumoto, T., Otsu, H., Oghiso, Y. and Sato, F. (2007) Cause of Death and Neoplasia in Mice Continuously Exposed to Very Low Dose Rates of Gamma Rays. Radiation Research, 167, 417-437, 421.

17. Tang, F.R., Loke, W.K. and Khoo, B.C. (2016) Low-dose or low-dose-rate ionizing radiation–induced bioeffects in animal models. Journal of Radiation Research, 58, 165-182.

18. Sasaki, S. and Fukuda, N. (2005) Temporal Variation of Excess Mortality Rate from Solid Tumors in Mice Irradiated at Various Ages with Gamma Rays. Journal of Radiation Research, 46, 1-19.

19. Kemp, C.J., Wheldon, T. and Balmain, A. (1994) p53-deficient mice are extremely susceptible to radiation-induced tumorigenesis. Nature Genetics, 8, 66-69.

20. 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. Nature Communications, 6, 6684.

21. Satoh, Y., Asakawa, J.-i., Nishimura, M., Kuo, T., Shinkai, N., Cullings, H.M., Minakuchi, Y., Sese, J., Toyoda, A., Shimada, Y. et al. (2020) Characteristics of induced mutations in offspring derived from irradiated mouse spermatogonia and mature oocytes. Scientific Reports, 10, 37.

22. Kitamura, S., Satoh, K. and Oono, Y. (2022) Detection and characterization of genome-wide mutations in M1 vegetative cells of gamma-irradiated Arabidopsis. PLOS Genetics, 18, e1009979.

23. Hirao, A.S., Watanabe, Y., Hasegawa, Y., Takagi, T., Ueno, S. and Kaneko, S. (2022) Mutational effects of chronic gamma radiation throughout the life cycle of Arabidopsis thaliana: Insight into radiosensitivity in the reproductive stage. Science of The Total Environment, 838, 156224.

24. 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. Nature Communications, 7, 12605.

25. Ward, J.F. (1994) The Complexity of DNA Damage: Relevance to Biological Consequences. International Journal of Radiation Biology, 66, 427-432.

26. Sage, E. and Harrison, L. (2011) Clustered DNA lesion repair in eukaryotes: Relevance to mutagenesis and cell survival. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 711, 123-133.

27. Hoeijmakers, J.H.J. (2009) DNA Damage, Aging, and Cancer. New England Journal of Medicine, 361, 1475-1485.

28. Eccles, L.J., O'Neill, P. and Lomax, M.E. (2011) Delayed repair of radiation induced clustered DNA damage: Friend or foe? Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 711, 134-141.

29. Strom, D.J., Cameron, J.R. and Cohen, B.L. (1998) The LNT model is appropriate for the estimation of risk from low-level (less than 100 mSv/year) radiation, and low levels of radon in homes should be considered harmful to health. Medical physics, 25, 273-278.

30. Vilenchik, M.M. and Knudson, A.G. (2000) Inverse radiation dose-rate effects on somatic and germ-line mutations and DNA damage rates. Proceedings of the National Academy of Sciences, 97, 5381-5386.

31. Vilenchik, M.M. and Knudson, A.G. (2003) Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences, 100, 12871-12876.

32. Hussain, S.P., Hofseth, L.J. and Harris, C.C. (2003) Radical causes of cancer. Nature Reviews Cancer, 3, 276-285.

33. Ward, J.F. (1985) Biochemistry of DNA Lesions. Radiation Research, 104, S103-S111.

34. Lindahl, T. and Barnes, D.E. (2000) Repair of endogenous DNA damage. Cold Spring Harbor symposia on quantitative biology, 65, 127-133.

35. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. Nature, 362, 709-715.

36. Burrows, C.J. and Muller, J.G. (1998) Oxidative Nucleobase Modifications Leading to Strand Scission. Chemical Reviews, 98, 1109-1152.

37. van Loon, B., Markkanen, E. and Hübscher, U. (2010) Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. DNA Repair, 9, 604-616.

38. Tubbs, A. and Nussenzweig, A. (2017) Endogenous DNA damage as a source of genomic instability in cancer. Cell, 168, 644-656.

39. Wallace, S.S. (2002) Biological consequences of free radical-damaged DNA bases. Free Radical Biology and Medicine, 33, 1-14.

40. Kamiya, H., Miura, K., Ishikawa, H., Inoue, H., Nishimura, S. and Ohtsuka, E. (1992) c-Ha-ras Containing 8-Hydroxyguanine at Codon 12 Induces Point Mutations at the Modified and Adjacent Positions1. Cancer Research, 52, 3483-3485.

41. Morton, L.M., Karyadi, D.M., Stewart, C., Bogdanova, T.I., Dawson, E.T., Steinberg, M.K., Dai, J., Hartley, S.W., Schonfeld, S.J., Sampson, J.N. et al. (2021) Radiation-related genomic profile of papillary thyroid carcinoma after the Chernobyl accident. Science, 372, eabg2538.

42. Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D.C., Tamuri, A.U., Martincorena, I., Petljak, M., Alexandrov, L.B., Gundem, G. et al. (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature, 513, 422-425.

43. 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.

44. Matsuda, Y., Uchimura, A., Satoh, Y., Kato, N., Toshishige, M., Kajimura, J., Hamasaki, K., Yoshida, K., Hayashi, T., Noda, A. et al. (2023) Spectra and characteristics of somatic mutations induced by ionizing radiation in hematopoietic stem cells. Proceedings of the National Academy of Sciences, 120, e2216550120.

45. Aoki, Y., Taniguchi, Y., Matsumoto, M., Matsumoto, M., Ohno, M., Masumura, K., Sasaki, S., Tsuzuki, T., Yamamoto, M. and Nohmi, T. (2020) Oxidative-stress-driven mutagenesis in the small intestine of the gpt delta mouse induced by oral administration of potassium bromate. Mutation research. Genetic toxicology and environmental mutagenesis, 850-851, 503136.

46. Kawai, H., Iwata, R., Ebi, S., Sugihara, R., Masuda, S., Fujiwara, C., Kimura, S. and Kamiya, H. (2022) Development of a versatile high-throughput mutagenesis assay with multiplexed short-read NGS using DNA-barcoded supF shuttle vector library amplified in E. coli. eLife, 11, e83780.

47. Cao, L., Kawai, H., Sasatani, M., Iizuka, D., Masuda, Y., Inaba, T., Suzuki, K., Ootsuyama, A., Umata, T., Kamiya, K. et al. (2014) A novel ATM/TP53/p21-mediated checkpoint only activated by chronic γ-irradiation. PloS one, 9, e104279.

48. Shimura, T., Sasatani, M., Kawai, H., Kamiya, K., Kobayashi, J., Komatsu, K. and Kunugita, N. (2017) ATM-mediated mitochondrial damage response triggered by nuclear DNA damage in normal human lung fibroblasts. Cell Cycle, 16, 2345-2354.

49. Fukushima, R., Suzuki, T. and Kamiya, H. (2020) New indicator Escherichia coli strain for rapid and accurate detection of supF mutations. Genes and Environment, 42, 28.

50. Suzuki, T., Zaima, Y., Fujikawa, Y., Fukushima, R. and Kamiya, H. (2022) Paradoxical role of the major DNA repair protein, OGG1, in action-at-a-distance mutation induction by 8-oxo-7,8-dihydroguanine. DNA Repair, 111, 103276.

51. Fukushima, R., Suzuki, T., Kobayakawa, A. and Kamiya, H. (2023) Action-at-a-distance mutations induced by 8-oxo-7,8-dihydroguanine is dependent on APOBEC3. Mutagenesis, gead023.

52. Kraemer, K.H. and Seidman, M.M. (1989) Use of supF, an Escherichia coli tyrosine suppressor tRNA gene, as a mutagenic target in shuttle-vector plasmids. Mutation research, 220, 61-72.

53. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31, 3406-3415.

54. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences, 95, 1460-1465.

55. Kamiya, H., Yamazaki, D., Nakamura, E., Makino, T., Kobayashi, M., Matsuoka, I. and Harashima, H. (2015) Action-at-a-Distance Mutagenesis Induced by Oxidized Guanine in Werner Syndrome Protein-Reduced Human Cells. Chemical Research in Toxicology, 28, 621-628.

56. Suzuki, T. and Kamiya, H. (2022) Easily-controllable, helper phage-free single-stranded phagemid production system. Genes and Environment, 44, 25.

57. Seidman, M.M., Bredberg, A., Seetharam, S. and Kraemer, K.H. (1987) Multiple point mutations in a shuttle vector propagated in human cells: evidence for an error-prone DNA polymerase activity. Proceedings of the National Academy of Sciences, 84, 4944-4948.

58. Roberts, Steven A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, Leszek J., Kryukov, Gregory V., Malc, E., Mieczkowski, Piotr A. et al. (2012) Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. Molecular Cell, 46, 424-435.

59. Wong, L., Sami, A. and Chelico, L. (2022) Competition for DNA binding between the genome protector replication protein A and the genome modifying APOBEC3 single-stranded DNA deaminases. Nucleic Acids Research, 50, 12039-12057.

60. Chen, J., Miller, B.F. and Furano, A.V. (2014) Repair of naturally occurring mismatches can induce mutations in flanking DNA. Elife, 3, e02001.

61. Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G. et al. (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet, 45, 970-976.

62. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.A., Hagen, L., Krokan, H.E. and Slupphaug, G. (2002) hUNG2 Is the Major Repair Enzyme for Removal of Uracil from U:A Matches, U:G Mismatches, and U in Single-stranded DNA, with hSMUG1 as a Broad Specificity Backup*. Journal of Biological Chemistry, 277, 39926-39936.

63. Koh, G., Degasperi, A., Zou, X., Momen, S. and Nik-Zainal, S. (2021) Mutational signatures: emerging concepts, caveats and clinical applications. Nature Reviews Cancer, 21, 619-637.

64. Wang, L., Zhang, H., Chen, W., Chen, H., Xiao, J. and Chen, X. (2022) Recent advances in DNA glycosylase assays. Chinese Chemical Letters, 33, 3603-3612.

65. Thompson, P.S. and Cortez, D. (2020) New insights into abasic site repair and tolerance. DNA Repair, 90, 102866.

66. Toledo, Luis I., Altmeyer, M., Rask, M.-B., Lukas, C., Larsen, Dorthe H., Povlsen, Lou K., Bekker-Jensen, S., Mailand, N., Bartek, J. and Lukas, J. (2013) ATR Prohibits Replication Catastrophe by Preventing Global Exhaustion of RPA. Cell, 155, 1088-1103.

67. Chen, R. and Wold, M.S. (2014) Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. BioEssays: news and reviews in molecular, cellular and developmental biology, 36, 1156-1161.

68. Pokhrel, N., Caldwell, C.C., Corless, E.I., Tillison, E.A., Tibbs, J., Jocic, N., Tabei, S.M.A., Wold, M.S., Spies, M. and Antony, E. (2019) Dynamics and selective remodeling of the DNA-binding domains of RPA. Nature structural & molecular biology, 26, 129-136.

69. Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H. and de Wind, N. (2006) Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. The Journal of experimental medicine, 203, 319-323.

70. Haracska, L., Unk, I., Johnson, R.E., Johansson, E., Burgers, P.M., Prakash, S. and Prakash, L. (2001) Roles of yeast DNA polymerases δ and ζ and of Rev1 in the bypass of abasic sites. Genes & development, 15, 945-954.

71. Chan, K., Resnick, M.A. and Gordenin, D.A. (2013) The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. DNA Repair, 12, 878-889.

72. Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S. and Prakash, L. (2000) Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. Nature, 406, 1015-1019.

73. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2009) DNA Synthesis across an Abasic Lesion by Human DNA Polymerase ι. Structure, 17, 530-537.

74. Choi, J.-Y., Lim, S., Kim, E.-J., Jo, A. and Guengerich, F.P. (2010) Translesion Synthesis across Abasic Lesions by Human B-Family and Y-Family DNA Polymerases α, δ, η, ι, κ, and REV1. Journal of Molecular Biology, 404, 34-44.

75. García-Gómez, S., Reyes, A., Martínez-Jiménez, M.I., Chocrón, E.S., Mourón, S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I.J. et al. (2013) PrimPol, an archaic primase/polymerase operating in human cells. Mol Cell, 52, 541-553.

76. Pilzecker, B., Buoninfante, O.A., Pritchard, C., Blomberg, O.S., Huijbers, I.J., van den Berk, P.C.M. and Jacobs, H. (2016) PrimPol prevents APOBEC/AID family mediated DNA mutagenesis. Nucleic Acids Research, 44, 4734-4744.

77. Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P. and Cortez, D. (2019) HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. Cell, 176, 144-153. e113.

78. Yaneva, D., Sparks, J.L., Donsbach, M., Zhao, S., Weickert, P., Bezalel-Buch, R., Stingele, J. and Walter, J.C. (2023) The FANCJ helicase unfolds DNA-protein crosslinks to promote their repair. Mol Cell, 83, 43-56.e10.

79. Sugimoto, Y., Masuda, Y., Iwai, S., Miyake, Y., Kanao, R. and Masutani, C. (2023) Novel mechanisms for the removal of strong replication-blocking HMCES-and thiazolidine-DNA adducts in humans. Nucleic Acids Research, 51, 4959-4981.

80. Oh, S., Bournique, E., Bowen, D., Jalili, P., Sanchez, A., Ward, I., Dananberg, A., Manjunath, L., Tran, G.P., Semler, B.L. et al. (2021) Genotoxic stress and viral infection induce transient expression of APOBEC3A and pro-inflammatory genes through two distinct pathways. Nature Communications, 12, 4917.

81. Isozaki, H., Sakhtemani, R., Abbasi, A., Nikpour, N., Stanzione, M., Oh, S., Langenbucher, A., Monroe, S., Su, W., Cabanos, H.F. et al. (2023) Therapy-induced APOBEC3A drives evolution of persistent cancer cells. Nature, 620, 393-401.

82. Kocakavuk, E., Anderson, K.J., Varn, F.S., Johnson, K.C., Amin, S.B., Sulman, E.P., Lolkema, M.P., Barthel, F.P. and Verhaak, R.G.W. (2021) Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. Nat Genet, 53, 1088-1096.

83. Chin, L., Hahn, W.C., Getz, G. and Meyerson, M. (2011) Making sense of cancer genomic data. Genes & development, 25, 534-555.

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

85. Shen, W., Le, S., Li, Y. and Hu, F. (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PloS one, 11, e0163962.

86. Kwon, S., Lee, B. and Yoon, S. (2014) CASPER: context-aware scheme for paired-end reads from high-throughput amplicon sequencing. BMC Bioinformatics, 15, S10.

87. Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34, 3094-3100.

88. Smith, T., Heger, A. and Sudbery, I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome research, 27, 491-499.

89. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C. and Dry, J.R. (2016) VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Research, 44, e108-e108.

## FIGURE LEGENDS

### Figure 1. Increase of *supF* mutation frequencies in response to chronic gamma-irradiation

(A) Schematic map of the *supF* shuttle vector $N_{12}$-BC library (pNGS2-K2). The pNGS2 encodes the amber suppressor tRNA (*supF*) gene, the TP53-/Rb-binding-deficient mutant SV40 large T antigen (SV40 T antigen E107K/D402E) gene, the SV40 replication origin (SV40 replication *ori*), the pBR327 origin of replication (pBR327 *ori*), the M13 intergenic region (M13 IG), the kanamycin-resistance gene ($km^r$), and the randomized 12-nucleotide barcode DNA sequence ($N_{12}$-BC). The $N_{12}$-BC is inserted before the sequence of the *supF* gene. (B) The nucleotide sequence for NGS analysis encoded in pNGS2-$N_{12}$-BC libraries (335 bp +12 bp for the $N_{12}$-BC). The underlined letters indicate the primer set for PCR amplification. Primer sets contain 0, 1, or 2 random nucleotides at the 5' end to ensure proper signal detection for amplicon sequencing. A pre-designed 6-nucleotide sequence ($N_6$) in the forward primers serves as an index sequence for multiplexed NGS. The sequence with green background represents commercially synthesized oligonucleotides used for the insertion of $N_{12}$-BC sequences ($N_{12}$) into the pNGS2. The sequence with red background indicates the *supF* gene starting at position 1. (C) A diagram of the low-dose-rate gamma-radiation facility at the Radiation Research Center for Frontier Science (RIRBM, Hiroshima University) equipped with a 1.11 TBq $^{137}$Cs source. Cell cultures were maintained in conventional incubators at 37 °C, 5% $CO_2$ under gamma-irradiation conditions at the indicated dose-rates (1 or 2 Gy/day). (D) Box plot of the *supF*-mutant frequency in response to gamma-irradiation. The *supF* mutant frequency was determined by the conventional *supF* assay by dividing the number of colonies grown on selection plates by the number of colonies grown on titer plates (n=3).

**Figure 2. Analysis by the *supF* NGS assay of the mutations induced by chronic gamma-irradiation in cells from 2,000 RF01 colonies grown on *supF*-selection plates (*supF*-selection, VF 0.4~1.0)**

(A) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.4 according to their nucleotide positions (Pos.) − data from three independent transfection experiments (TF1, TF2, and TF3) and different dose-rates of gamma-irradiation (0, 1, and 2 Gy/day for 2 days). In the left side graph, the colors reflect the variant frequency, as indicated by the heatmap on top. On the right side graph, the variant types are shown in different colors as indicated in the legend. The variant calls are categorized into six types: one-nucleotide deletion (1-Del), deletions larger than one nucleotide (Del), deletions with insertions (Del-in), dinucleotide substitutions (DNS), insertions (INS), and single nucleotide substitutions (SNS). (B) Stack bar graph representing the number of $N_{12}$-BC sequences with mutations from each sample used for further data analysis. Included are the one hundred mutation-bearing $N_{12}$-BCs with the highest number of reads per unique $N_{12}$-BC or the maximum number of available $N_{12}$-BC with mutations. The number in parentheses refers to the number of unique $N_{12}$-BC sequences with mutations. (C) Pie charts of the proportions of mutation types shown in different colors as indicated in the legend on top. The percentage and number (in parentheses) of SNSs are provided in each pie chart. (D) Pie charts of different base substitutions as proportions of SNSs. The individual base substitutions are shown in different colors as indicated in the legend on top. The percentage and number (in parentheses) of substitutions from guanine (G>) or cytosine (C>) are provided in each pie chart. (E) Number of SNSs according to their nucleotide position (Pos.) at different dose-rates of gamma-irradiation (0, 1, and 2 Gy/day for 2 days). The numbers represent data combined from three transfection experiments. The individual base substitutions are shown in different colors as indicated in the legend on top.

**Figure 3. Analysis of base-substitutions in single and multiple *supF* mutations per $N_{12}$-BC (*supF*-selection, VF 0.4~1.0)**

(A) Proportion of the number of $N_{12}$-BC sequences with single (1) or multiple (2~8) mutations from three independent transfection experiments (TF1, TF2, and TF3) at different dose-rates of gamma-irradiation (0, 1, and 2 Gy/day). The percentage of $N_{12}$-BCs with different number of mutations is indicated inside each bar, and the total number of mutations are indicated outside each bar in parentheses. (B) Number of $N_{12}$-BC sequences, combined data from three independent transfection experiments. The number of all SNSs, single mutations per $N_{12}$-BC, and multiple mutations (more than one mutation) per $N_{12}$-BC are separately presented. The total number of SNSs are shown above the bars in parentheses. The proportion of individual base substitutions is shown in different colors as indicated in the legend on top. (C) Stacked bar graph representing the relative proportion of the four bases with reference to the SNSs plotted in (B). The percentage and number (in parentheses) of SNSs from either C or G is provided inside each segment. The four bases are shown in different colors as indicated in the legend on top. (D) Stacked bar graph representing the relative proportion of

SNSs, data set identical to (B). The percentage of each SNSs from either C or G is provided inside each segment. The individual base substitutions are shown in different colors as indicated in the legend on top. (E) Box plot of the number of individual SNSs in response to different gamma-irradiation conditions (n=3). The horizontal line in each box plot represents the median. The number of all SNSs is presented in the left-side graph. Single-mutations per $N_{12}$-BC and multiple-mutations per $N_{12}$-BC are separately presented in the right-side graph. The asterisks indicate statistical significance determined by paired student's *t* test (*$P < 0.05$).

**Figure 4. Summary of SNSs induced in cells – data from *supF*-selection**

(A) SNSs listed by rank according to the frequency of their detection (number of $N_{12}$-BCs) from the combined data of three transfections for *supF*-selection. The ranks of the SNSs (shown in full black circles) are sorted separately for single mutation or multiple mutations per $N_{12}$-BC and for each radiation dose-rate (0, 1, and 2 Gy/day). Each SNS is denoted as an underlined nucleotide in a pentanucleotide sequence (including two nucleotides upstream and downstream of the mutation, and the sequence of the non-transcribed/transcribed strand). The 5'-TC-3' sites are shown in green color, and the 5'-GA-3' sites are shown in red color. The number next to the rank is the position of each mutation in the nucleotide sequence. The number in parentheses represents the number of detected $N_{12}$-BCs with the mutation. (B) The number of single mutations per $N_{12}$-BC according to their positions in the cloverleaf secondary structure of the *supF*-tRNA (positions from 56 to 145). The bar graphs in each frame provide the number of single mutations per $N_{12}$-BC for different radiation dose rates (the numbers are denoted above each bar), and their positions are indicated by dotted lines. The proportion of individual base substitutions is shown in different colors as indicated in the legend on top.

**Figure 5. Analysis by *supF* NGS assay of the *Dpn* I-treated libraries extracted from cells (w/o-*supF*-selection, VF 0.2~1.0)**

The layout of the figure is essentially analogous to Figure 2, only this time the NGS samples were prepared from the *Dpn* I treated libraries (used for the experiments in Figures 2 and 3) before transformation of RF01 (w/o-*supF*-selection). (A) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.2 according to their nucleotide position (refer to the legend in Figure 2A). (B) Stack bar graph representing the number of $N_{12}$-BC sequences with mutations from each sample (refer to the legend in Figure 2B). The data is separated into two graphs: the graph on the left side represents positions from -67 to -20, and the graph on the right side – positions from -19 to 214 containing the *supF* gene. (C) Box plot of the mutant frequency under different irradiation conditions calculated from NGS data. The mutant frequency was determined by the number of reads harboring mutation(s) vs the total number of reads, as described in the text. The data is separated into two graphs: positions from -67 to -20 (left side), and positions from -19 to 214 containing the *supF* gene (right side). The

results were obtained from three independent transfection experiments. The asterisks indicate statistical significance determined by paired student's *t* test (*$P < 0.05$). (D) Pie charts of the proportions of mutation types (refer to the legend in Figure 2C). (E) Pie charts of the proportions of SNSs represented by different base substitutions (refer to the legend in Figure 2D). (F) Number of SNSs according to their nucleotide position (refer to the legend in Figure 2E).

**Figure 6. Analysis of base-substitutions in single and multiple *supF* mutations per $N_{12}$-BC in response to chronic gamma-irradiation (w/o-*supF*-selection, VF 0.2~1.0)**

The layout of the figure is essentially analogous to Figure 3, only this time the NGS samples were prepared from the *Dpn* I-treated libraries (used for the experiments in Figures 2 and 3) before transformation into RF01. (A) $N_{12}$-BC sequences with single (1) or multiple (2~8) mutations from three independent transfection-experiments (TF1, TF2, and TF3) at different dose-rates of gamma-irradiation (0, 1, and 2 Gy/day). The percentage and number (in parentheses) of $N_{12}$-BCs with a single mutation per $N_{12}$-BC are presented inside each bar, and the total number of mutations is indicated next to the bars in parentheses. (B) Bar graph of the combined number of $N_{12}$-BC sequences from three independent transfection-experiments (refer to the legend in Figure 3B). (C) Stacked bar graph representing the relative proportion of the four bases in SNSs (refer to the legend in Figure 3C). (D) Stacked bar graph representing the relative proportion of SNSs (refer to the legend in Figure 3D).

**Figure 7. Analysis of SNSs by base substitutions according to positions in the sequence and dose-rates of irradiation (*supF*-selection versus w/o-*supF*-selection)**

Bar graphs depicting the number of SNSs at the indicated positions on the analyzed sequence (positions from -19 to 214) − combined data from three transfection-experiments (TF1, TF2, and TF3). Each frame depicts the number of SNSs for *supF*-selection (bars on top) and w/o-*supF*-selection (bars on bottom). The colors of the bars and segments of stack bars reflect the type of SNSs as indicated in the legend on the top-left side of the figure. The 5'-TC-3' and 5'-GA-3' sites in the sequence and their positions are shown in green and red, respectively.

**Figure 8. The presence of a single 8-oxo-G lesion induces action-at-a-distance mutations (w/o-*supF*-selection, VF 0.2~1.0)**

(A) Schematic map of the $N_{12}$-BC libraries (pNGS2-K1~K4) with an inserted 8-Oxo-7,8-dihydroguanine (8-oxo-G, $G^O$) or a lesion-free G (G). An 8-oxo-G or G was inserted outside the *supF* gene at position 176 for pNGS2-K2 and pNGS2-K3 (represented as $^{176}$G:C for a lesion-free G-inserted and $^{176}G^O$:C for an 8-oxoG-inserted), or at position 177 for pNGS2-K1 and pNGS2-K4 (represented as $^{177}$C:G for a lesion-free G-inserted and $^{177}C:G^O$ for an 8-oxo-G-inserted). The

schematic diagrams of the relative localizations of the SV40 replication origin, *supF* gene, and 8-oxo-G or G -inserted position/strand are shown on the right side of figure. (B) Box plot of the *supF*-mutant frequency in transfected cells for G or 8-oxo-G inserted libraries. The *supF* mutant frequency were determined by the conventional *supF* assay, and statistical significance was determined by paired student's *t* test (n=3, $*P < 0.05$, $***P < 0.0005$). (C) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.2 according to their nucleotide position (refer to the legend in Figure 2A). The mutations at positions 176 and 177 were excluded. (D) Stack bar graph representing the number of $N_{12}$-BC sequences with mutations from each sample used for further data analysis (refer to the legend in Figure 5B). (E) Box plot of the mutation frequency calculated from NGS data as described in Figure 5C. The asterisks indicate statistical significance determined by paired student's *t* test (n=3, $*P < 0.05$, $***P < 0.005$).

**Figure 9. A single 8-oxo-G lesion induces strand-biased single nucleotide substitutions at G:C sites (w/o-*supF*-selection, VF 0.2~1.0)**

Analysis of mutations in w/o-*supF*-selection experiments ($^{176}$G:C or $^{176}$G$^{O}$:C for pNGS2 -K2 and -K3; $^{177}$C:G or $^{177}$C:G$^{O}$ for pNGS2 -K1 and -K4). (A) Pie charts of the proportions of different mutation types for pNGS2-K1~K4 with an inserted G or 8-oxo-G (refer to the legend in Figure 2C). (B) Pie charts of the proportions of different base substitutions in SNSs (refer to the legend in Figure 2D). (C) Number of SNSs according to their nucleotide position (refer to the legend in Figure 2E).

**Figure 10. Analysis of 8-oxo-G insertion induced base substitutions in single and multiple *supF* mutations per $N_{12}$-BC (w/o-*supF*-selection, VF 0.2~1.0).**

(A) Proportion of $N_{12}$-BC sequences with single (1) or multiple (2~7) mutations for pNGS2-K1~K4 with an inserted G or 8-oxo-G. The combined data from three independent transfection-experiments (TF1, TF2, and TF3); mutations at positions 176 and 177 are excluded. The percentage and number of $N_{12}$-BCs with a single mutation are denoted inside each bar, and the total number of $N_{12}$-BCs are indicated outside the bars in the parentheses. (B) Top-row bar graphs represent the number of SNSs in pNGS2-K1~K4 with an inserted G or 8-oxo-G − combined data from three independent transfection experiments without mutations at positions 176 and 177, for either all mutations (left side), single-mutation per $N_{12}$-BC (middle), or multiple-mutations per $N_{12}$-BC (right side). Middle-row bar graphs represent the relative proportions of the four bases in SNSs. Bottom-row bar graphs represent a detailed breakdown of SNSs according to base substitution, as indicated in the color legend on the right. Throughout panel (B) the numbers inside or above the bars indicate the percentage of SNSs, and the number of detected SNSs is denoted in parentheses.

**Figure 11. Analysis by *supF* NGS assay of the mutations induced by the insertion of an 8-oxo-G (*supF*-selection, VF 0.4~1.0).**

(A) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.4 according to their nucleotide position (Pos.) – combined data from three independent transfection experiments (TF1, TF2, and TF3) for pNGS2-K1~K4 with an inserted G or 8-oxo-G (refer to the legend of Figure 2A). (B) Pie charts depicting individual base substitutions (in different colors, in the legend on top) as proportions of SNSs. The percentage (and number, in parentheses) of total substitutions from guanine (G>) or cytosine (C>) are indicated in the pie charts. (C) Number of SNSs according to their nucleotide position (Pos.) – combined data from three transfection experiments. The individual base substitutions are shown in different colors as indicated in the legend on top.

**Figure 12. Analysis of SNSs by proportions of substituted bases and mutational positions for pNGS2 libraries -K1 ~ -K4 with an inserted G or 8-oxo-G (*supF*-selection versus w/o-*supF*-selection)**

Bar graphs depicting the number of SNSs at the indicated positions on the analyzed sequence, (positions from -19 to 214) − combined data from three transfection experiments (TF1, TF2, and TF3). Each frame depicts the number of SNSs for *supF*-selection (bars on top) and w/o-*supF*-selection (bottom bars). The colors of the bars and segments of stack bars reflect the type of SNSs as indicated in the legend on the top-left side of the figure. The 5'-TC-3' and 5'-GA-3' sites in and their positions are shown in green and red, respectively.

**Figure 13. Localization of each 5’-TC-3’:5’-GA-3’ site and the predicted secondary structure of single-stranded DNA.**

The secondary structure of the single-stranded non-transcribed (top) or transcribed strand (bottom) from position -21 to position 175 of the *supF*-analyzed sequence were predicted by Mfold web server. The analyzed sequence was divided into 27 overlapping regions of 41-bases with a sliding window of 5-bases; the first and last positions of the regions are shown in the top corners of each frame (Fig. 13-figure supplements 1 and 2). The 5'-TC-3' sites are shown in a green-outline box, and the 5'-GA-3' sites are shown in a red-outline box for each strand.

**Figure 14. Effects of chronic gamma-irradiation on 8-oxo-G induced action-at-a-distance mutations in cells (pNGS2-K3 and -K4 with an inserted non-damaged G or 8-oxo-G, dose-rate of irradiation 0 or 2 Gy/day for 2 days, w/o-*supF*-selection, VF 0.2~1.0)**

(A) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.2 according to their nucleotide position (refer to the legend in Figure 2A). The mutations at positions 176 and 177 were excluded. (B)

Stack bar graph representing the number of $N_{12}$-BC sequences with mutations in each sample used for further data analysis (refer to the legend in Figure 5B). The data is separated into two graphs: on the left side positions from -67 to -20, and on the right side – positions from -19 to 214 containing the *supF* gene. (C) Box plot of the mutation frequency calculated from NGS data as described in Figure 5C. The asterisks indicate statistical significance determined by paired student's *t* test (n=3, $*P < 0.05$). (D) Mutation frequencies for positions 176 and 177 with inserted G or 8-oxo-G − combined data from three transfection experiments. The different base-substitutions or types of mutations are shown in different colors as indicated on in the legend.

**Figure 15. Effects of chronic gamma-irradiation on 8-oxo-G induced action-at-a-distance mutations in cells (pNGS2-K3 and K4 with an inserted non-damaged G or 8-oxo-G, dose-rate of irradiation 0 or 2 Gy/day for 2 days, *supF*-selection, VF 0.4~1.0)**

(A) Number of $N_{12}$-BC sequences with a variant exceeding VF 0.4 according to their nucleotide position (refer to the legend in Figure 2A). The mutations at positions 176 and 177 were excluded. (B) Pie charts of the proportions of different base-substitutions in SNSs. The individual base-substitutions are shown in different colors as indicated in the legend on the right side. The percentage (and number, in parentheses) of the total substitutions from guanine (G>) or cytosine (C>) are provided in each pie chart. (C) Number of SNSs according to their nucleotide position (Pos.) – combined data from three transfection experiments. The individual base-substitutions are shown in different colors as indicated on the right side.

**Figure 16. Localization of each 5’-TC-3’:5’-GA-3’ site and the predicted secondary structure of single-stranded DNA.**

The secondary structure of the single-stranded non-transcribed (top) or transcribed strand (bottom) of the *supF*-analyzed sequence were predicted by Mfold web server. The analyzed sequence was divided into overlapping regions of 41-bases, with the 5'-TC-3':5’-GA-3’ sites (70, 95, 101, 108, 111, 117. 125, 130, 134, and 147 for the *supF*-non-transcribed strand; 73, 86. 91, 112, 118, 126, 162, and 167 for the *supF*-transcribed strand) located at the center of the sequence. The bottom structure in each frame is the predicted single-stranded DNA secondary structure in the case of the replacement of cytosine with uracil at the site. The first and last positions of the analyzed sequence are shown in the upper left and right corners of each frame. The positions of 5’-T<u>C</u>-3' sites in the non-transcribed strand are shown in a green-outline box, and the 5’-T<u>C</u>-3’ sites in the transcribed strand are shown in a purple-outline box for each region.

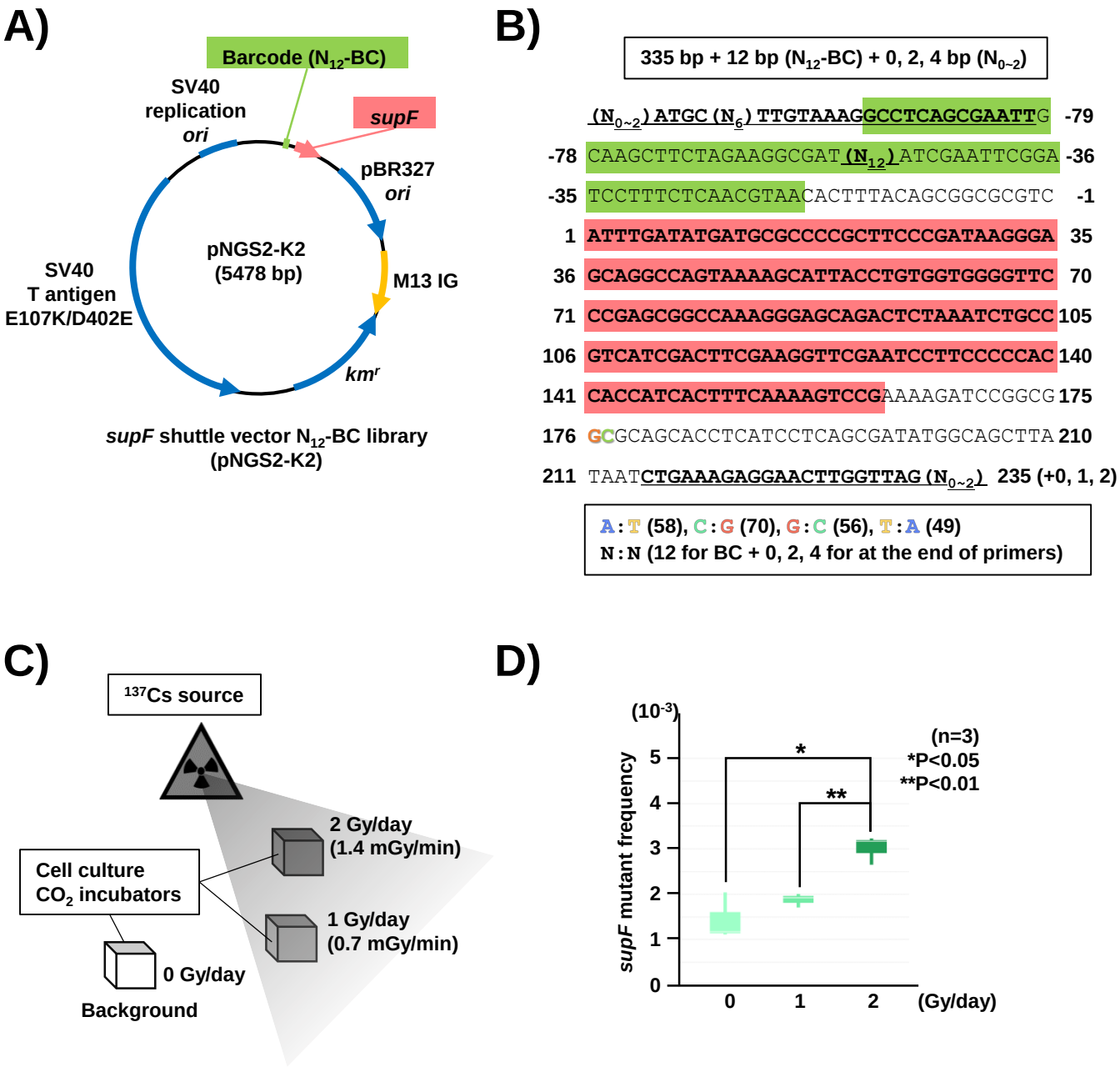

A)
Barcode ($N_{12}$-BC)
SV40 replication ori
supF
pBR327 ori
pNGS2-K2 (5478 bp)
M13 IG
SV40 T antigen E107K/D402E
$km^r$
supF shuttle vector $N_{12}$-BC library (pNGS2-K2)
B)
335 bp + 12 bp ($N_{12}$-BC) + 0, 2, 4 bp ($N_{0\sim2}$)
($N_{0\sim2}$) ATGC ($N_6$) TTGTAAAGGCCTCAGCGAATTG -79
-78 CAAGCTTCTAGAAGGCGAT ($N_{12}$) ATCGAATTCGGA -36
-35 TCCTTTCTCAACGTAACACTTTACAGCGGCGCGTC -1
1 ATTTGATATGATGCGCCCCGCTTCCCGATAAGGGA 35
36 GCAGGCCAGTAAAAGCATTACCTGTGGTGGGGTTC 70
71 CCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCC 105
106 GTCATCGACTTCGAAGGTTCGAATCCTTCCCCCAC 140
141 CACCATCACTTTCAAAAGTCCGAAAAGATCCGGCG 175
176 GCGCAGCACCTCATCCTCAGCGATATGGCAGCTTA 210
211 TAATCTGAAAGAGGAACTTGGTTAG ($N_{0\sim2}$) 235 (+0, 1, 2)
A : T (58), C : G (70), G : C (56), T : A (49)
N : N (12 for BC + 0, 2, 4 for at the end of primers)
C)
$^{137}$Cs source
2 Gy/day (1.4 mGy/min)
Cell culture $CO_2$ incubators
1 Gy/day (0.7 mGy/min)
0 Gy/day
Background
D)
($10^{-3}$)
supF mutant frequency
(n=3)
*P<0.05
**P<0.01
(Gy/day)

Fig. 1

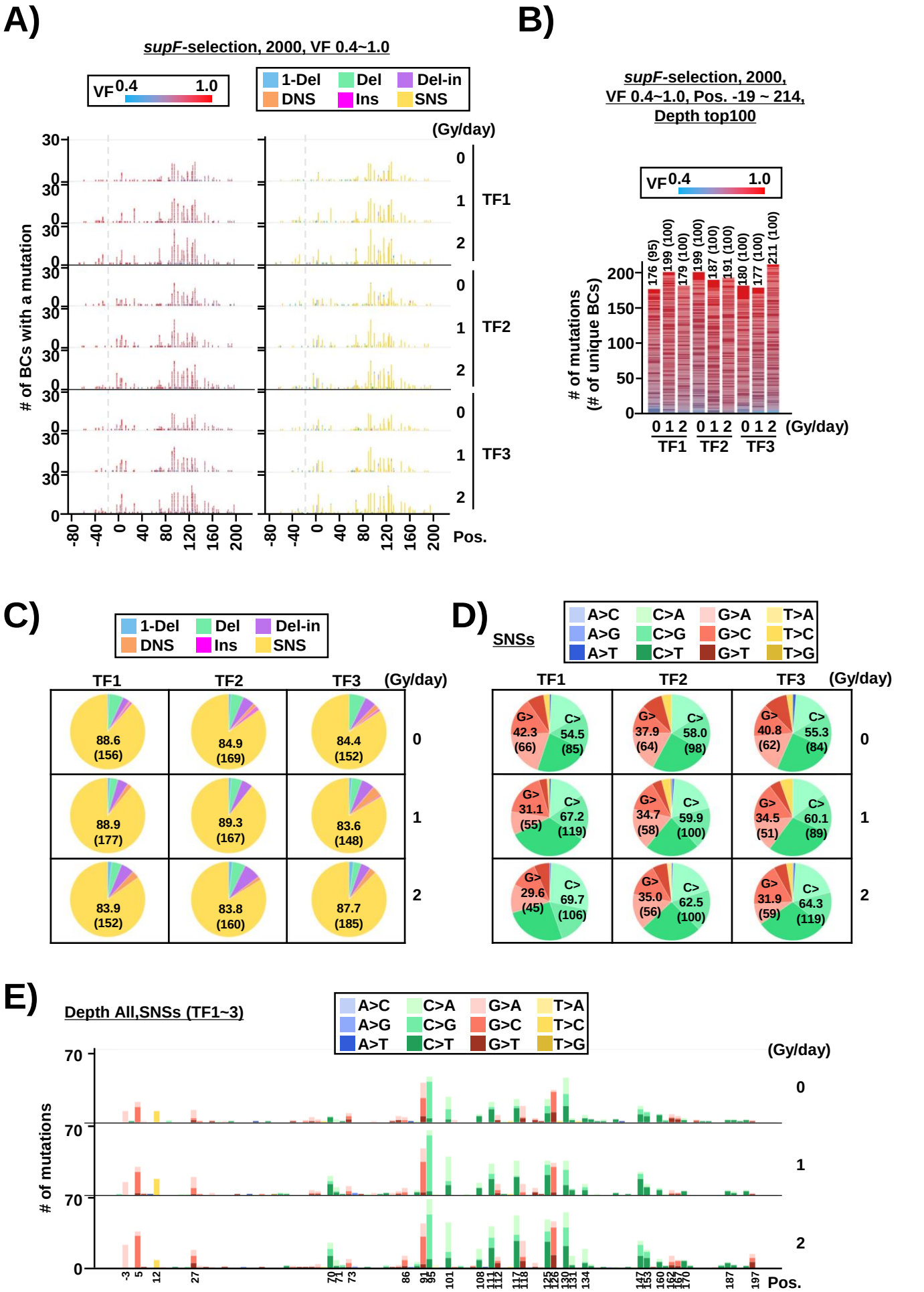

A)
supF-selection, 2000, VF 0.4~1.0
VF 0.4 1.0
1-Del Del Del-in
DNS Ins SNS
(Gy/day)
# of BCs with a mutation
TF1
TF2
TF3
Pos.
B)
supF-selection, 2000,
VF 0.4~1.0, Pos. -19 ~ 214,
Depth top100
VF 0.4 1.0
# of mutations
(# of unique BCs)
176 (95)
199 (100)
179 (100)
199 (100)
187 (100)
191 (100)
180 (100)
177 (100)
211 (100)
(Gy/day)
C)
1-Del Del Del-in
DNS Ins SNS
TF1 TF2 TF3 (Gy/day)
88.6 (156)
84.9 (169)
84.4 (152)
88.9 (177)
89.3 (167)
83.6 (148)
83.9 (152)
83.8 (160)
87.7 (185)
D)
SNSs
A>C C>A G>A T>A
A>G C>G G>C T>C
A>T C>T G>T T>G
TF1 TF2 TF3 (Gy/day)
G> 42.3 (66)
C> 54.5 (85)
G> 37.9 (64)
C> 58.0 (98)
G> 40.8 (62)
C> 55.3 (84)
G> 31.1 (55)
C> 67.2 (119)
G> 34.7 (58)
C> 59.9 (100)
G> 34.5 (51)
C> 60.1 (89)
G> 29.6 (45)
C> 69.7 (106)
G> 35.0 (56)
C> 62.5 (100)
G> 31.9 (59)
C> 64.3 (119)
E)
Depth All,SNSs (TF1~3)
A>C C>A G>A T>A
A>G C>G G>C T>C
A>T C>T G>T T>G
(Gy/day)
# of mutations
Pos.

Fig. 2

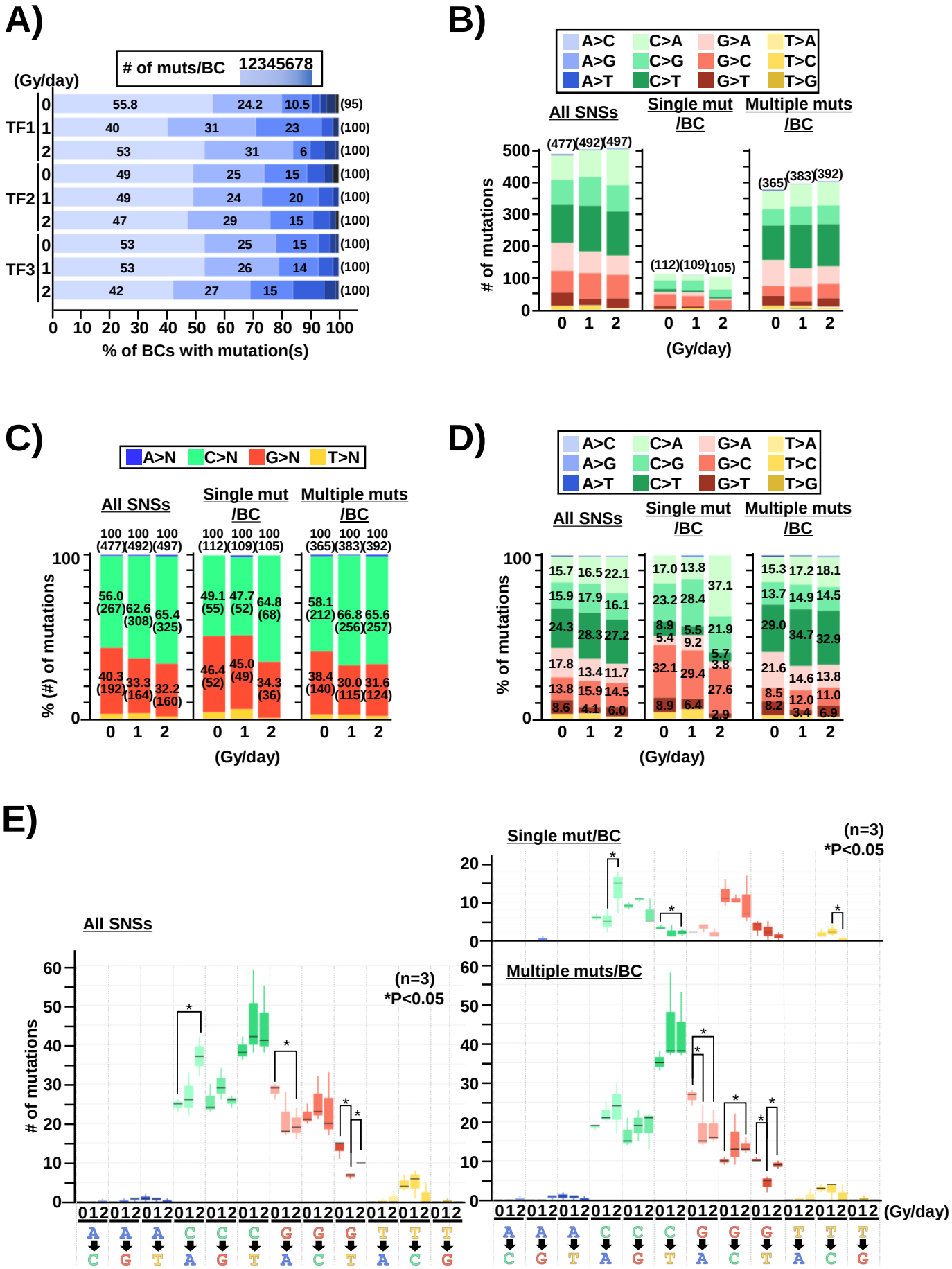

A)
# of muts/BC
12345678
(Gy/day)
TF1
TF2
TF3
% of BCs with mutation(s)
B)
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
All SNSs
Single mut /BC
Multiple muts /BC
# of mutations
(Gy/day)
C)
A>N
C>N
G>N
T>N
All SNSs
Single mut /BC
Multiple muts /BC
% (#) of mutations
(Gy/day)
D)
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
All SNSs
Single mut /BC
Multiple muts /BC
% of mutations
(Gy/day)
E)
All SNSs
(n=3)
*P<0.05
# of mutations
Single mut/BC
(n=3)
*P<0.05
Multiple muts/BC
(Gy/day)

Fig. 3

# A)

| SNSs | 0 Gy/day | 1 Gy/day | 2 Gy/day |
|---|---|---|---|
| Single mut/BC | ❶ 95 CTCTA/TAGAG (20)<br>❷ 130 ATCCT/AGGAT (16)<br>❸ 126 TCGAA/TTCGA (15)<br>❹ 91 CAGAC/GTCTG (12)<br>❺ 5 TTGAT/ATCAA (9)<br>❻ 101 ATCTG/CAGAT (8)<br>❼ 118 TCGAA/TTCGA (7)<br>❽ 12 GATGC/GCATC (4)<br>❽ 131 TCCTT/AAGGA (4)<br>❾ 85 AGGGA/TCCCT (3)<br>❾ 117 TTCGA/TCGAA (3)<br>❾ 121 AAGGT/ACCTT (3) | ❶ 95 CTCTA/TAGAG (24)<br>❷ 126 TCGAA/TTCGA (15)<br>❸ 91 CAGAC/GTCTG (11)<br>❹ 130 ATCCT/AGGAT (8)<br>❺ 5 TTGAT/ATCAA (7)<br>❺ 12 GATGC/GCATC (7)<br>❺ 118 TCGAA/TTCGA (7)<br>❽ 71 TCCCG/CGGGA (4)<br>❽ 101 ATCTG/CAGAT (4)<br>❽ 117 TTCGA/TCGAA (4)<br>❽ 134 TTCCC/GGGAA (4) | ❶ 95 CTCTA/TAGAG (18)<br>❷ 130 ATCCT/AGGAT (16)<br>❸ 126 TCGAA/TTCGA (10)<br>❹ 5 TTGAT/ATCAA (9)<br>❹ 101 ATCTG/CAGAT (9)<br>❻ 91 CAGAC/GTCTG (8)<br>❼ 131 TCCTT/AAGGA (7)<br>❼ 134 TTCCC/GGGAA (7)<br>❾ 71 TCCCG/CGGGA (5)<br>❿ 117 TTCGA/TCGAA (4)<br>❿ 118 TCGAA/TTCGA (4) |
| Multiple muts/BC | ❶ 130 ATCCT/AGGAT (28)<br>❷ 91 CAGAC/GTCTG (27)<br>❸ 95 CTCTA/TAGAG (23)<br>❹ 125 TTCGA/TCGAA (22)<br>❺ 111 ATCGA/TCGAT (19)<br>❻ 117 TTCGA/TCGAA (18)<br>❼ 126 TCGAA/TTCGA (17)<br>❽ 147 ATCAC/GTGAT (15)<br>❾ 101 ATCTG/CAGAT (13)<br>❿ -3 GCGTC/GACGC (10)<br>❿ 27 CCGAT/ATCGG (10)<br>❿ 153 TTCAA/TTGAA (10) | ❶ 95 CTCTA/TAGAG (29)<br>❶ 125 TTCGA/TCGAA (29)<br>❸ 117 TTCGA/TCGAA (26)<br>❹ 101 ATCTG/CAGAT (25)<br>❺ 111 ATCGA/TCGAT (24)<br>❻ 91 CAGAC/GTCTG (22)<br>❼ 130 ATCCT/AGGAT (20)<br>❼ 147 ATCAC/GTGAT (20)<br>❾ 70 TTCCC/GGGAA (16)<br>❿ 5 TTGAT/ATCAA (14)<br>❿ 27 CCGAT/ATCGG (14) | ❶ 125 TTCGA/TCGAA (32)<br>❷ 111 ATCGA/TCGAT (27)<br>❷ 117 TTCGA/TCGAA (27)<br>❹ 130 ATCCT/AGGAT (25)<br>❺ 91 CAGAC/GTCTG (22)<br>❺ 147 ATCAC/GTGAT (22)<br>❼ 95 CTCTA/TAGAG (21)<br>❽ 101 ATCTG/CAGAT (20)<br>❾ 126 TCGAA/TTCGA (14)<br>❿ -3 GCGTC/GACGC (13)<br>❿ 70 TTCCC/GGGAA (13) |

# B)

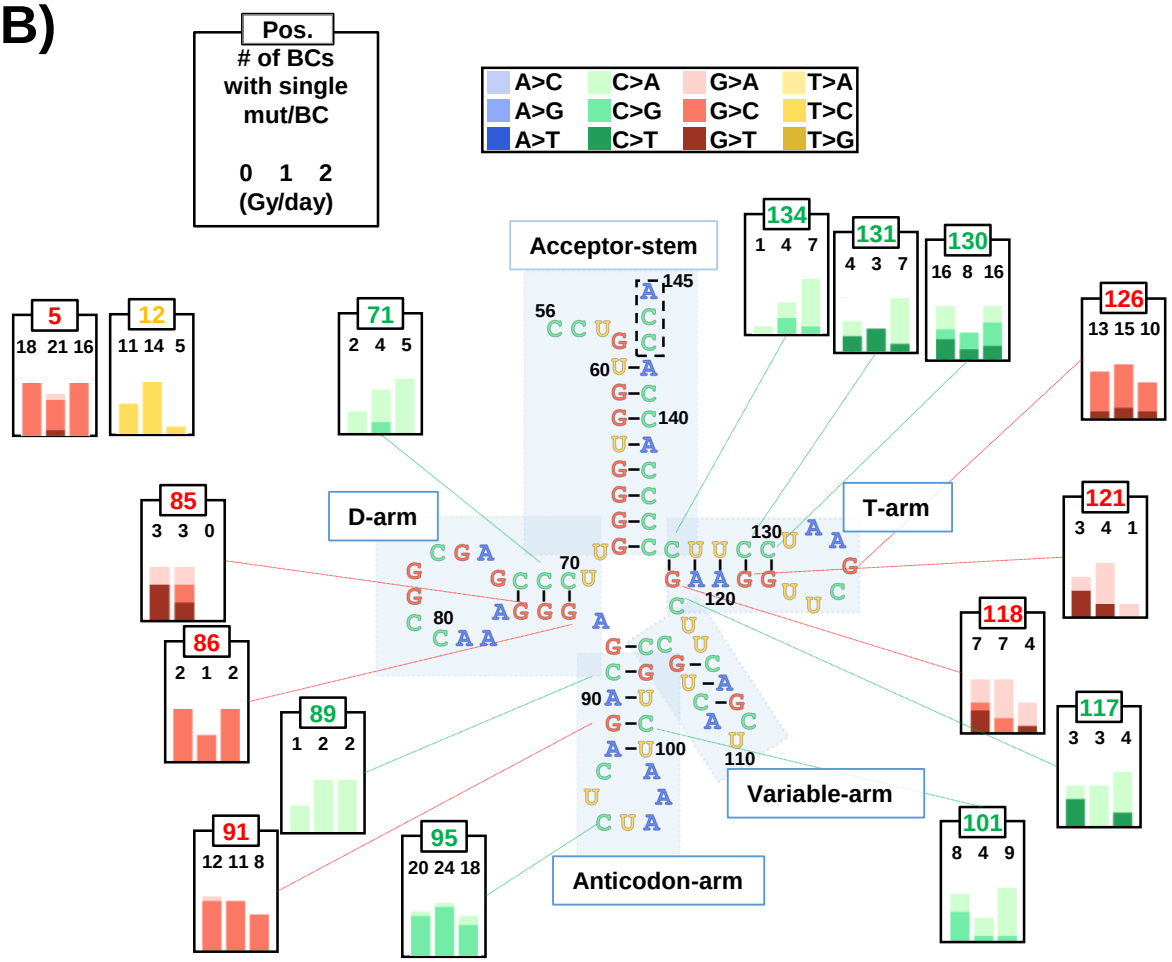



Fig. 4

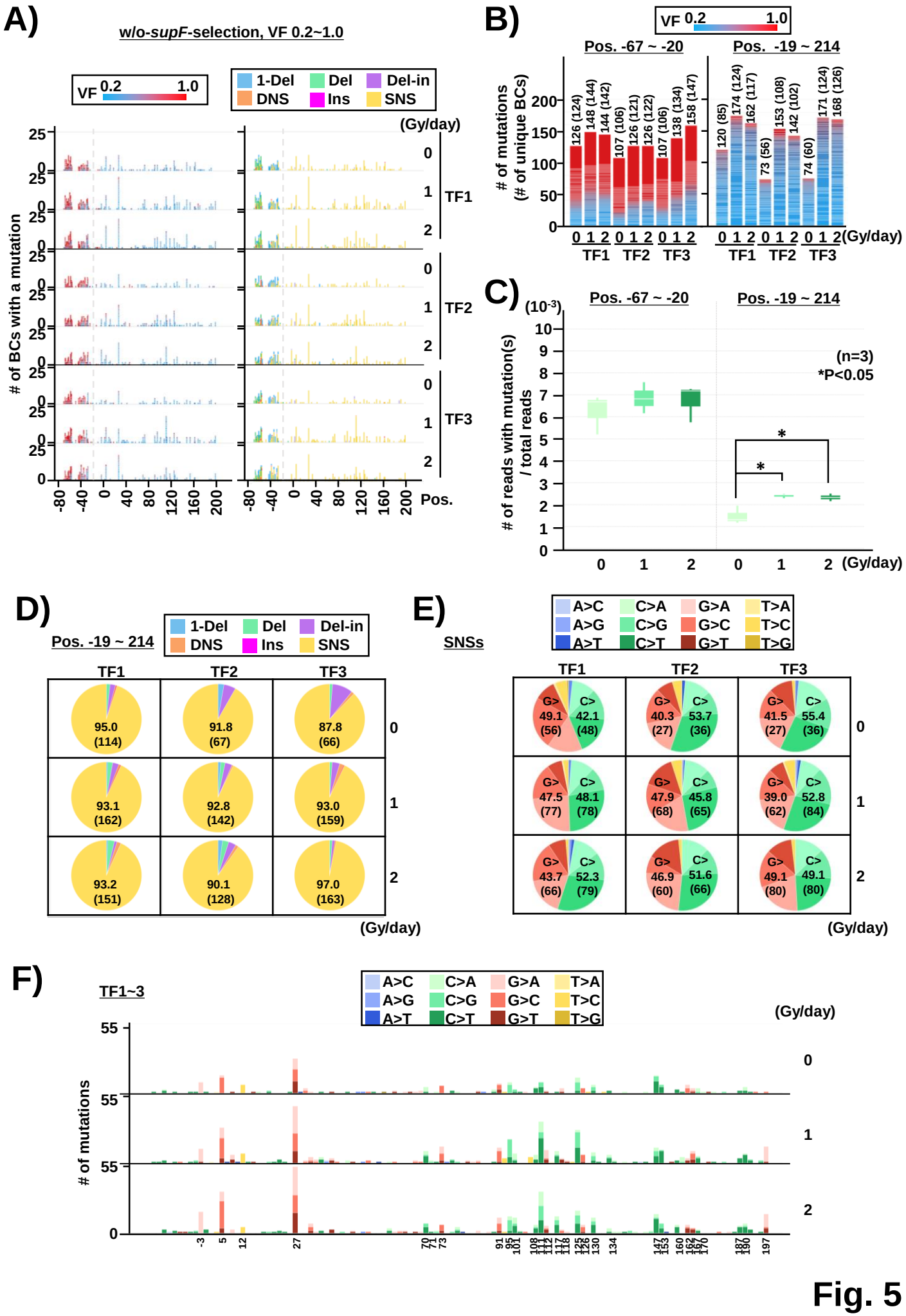

A)
w/o-supF-selection, VF 0.2~1.0
VF 0.2 1.0
1-Del
Del
Del-in
DNS
Ins
SNS
(Gy/day)
TF1
TF2
TF3
# of BCs with a mutation
Pos.
B)
VF 0.2 1.0
Pos. -67 ~ -20
Pos. -19 ~ 214
# of mutations (# of unique BCs)
126 (124)
148 (144)
144 (142)
107 (106)
126 (121)
126 (122)
107 (106)
138 (134)
158 (147)
120 (85)
174 (124)
162 (117)
73 (56)
153 (108)
142 (102)
74 (60)
171 (124)
168 (126)
(Gy/day)
C)
Pos. -67 ~ -20
Pos. -19 ~ 214
(10-3)
# of reads with mutation(s) / total reads
(n=3)
*P<0.05
(Gy/day)
D)
Pos. -19 ~ 214
1-Del
Del
Del-in
DNS
Ins
SNS
95.0 (114)
91.8 (67)
87.8 (66)
93.1 (162)
92.8 (142)
93.0 (159)
93.2 (151)
90.1 (128)
97.0 (163)
(Gy/day)
E)
SNSs
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
G> 49.1 (56)
C> 42.1 (48)
G> 40.3 (27)
C> 53.7 (36)
G> 41.5 (27)
C> 55.4 (36)
G> 47.5 (77)
C> 48.1 (78)
G> 47.9 (68)
C> 45.8 (65)
G> 39.0 (62)
C> 52.8 (84)
G> 43.7 (66)
C> 52.3 (79)
G> 46.9 (60)
C> 51.6 (66)
G> 49.1 (80)
C> 49.1 (80)
(Gy/day)
F)
TF1~3
# of mutations
(Gy/day)

Fig. 5

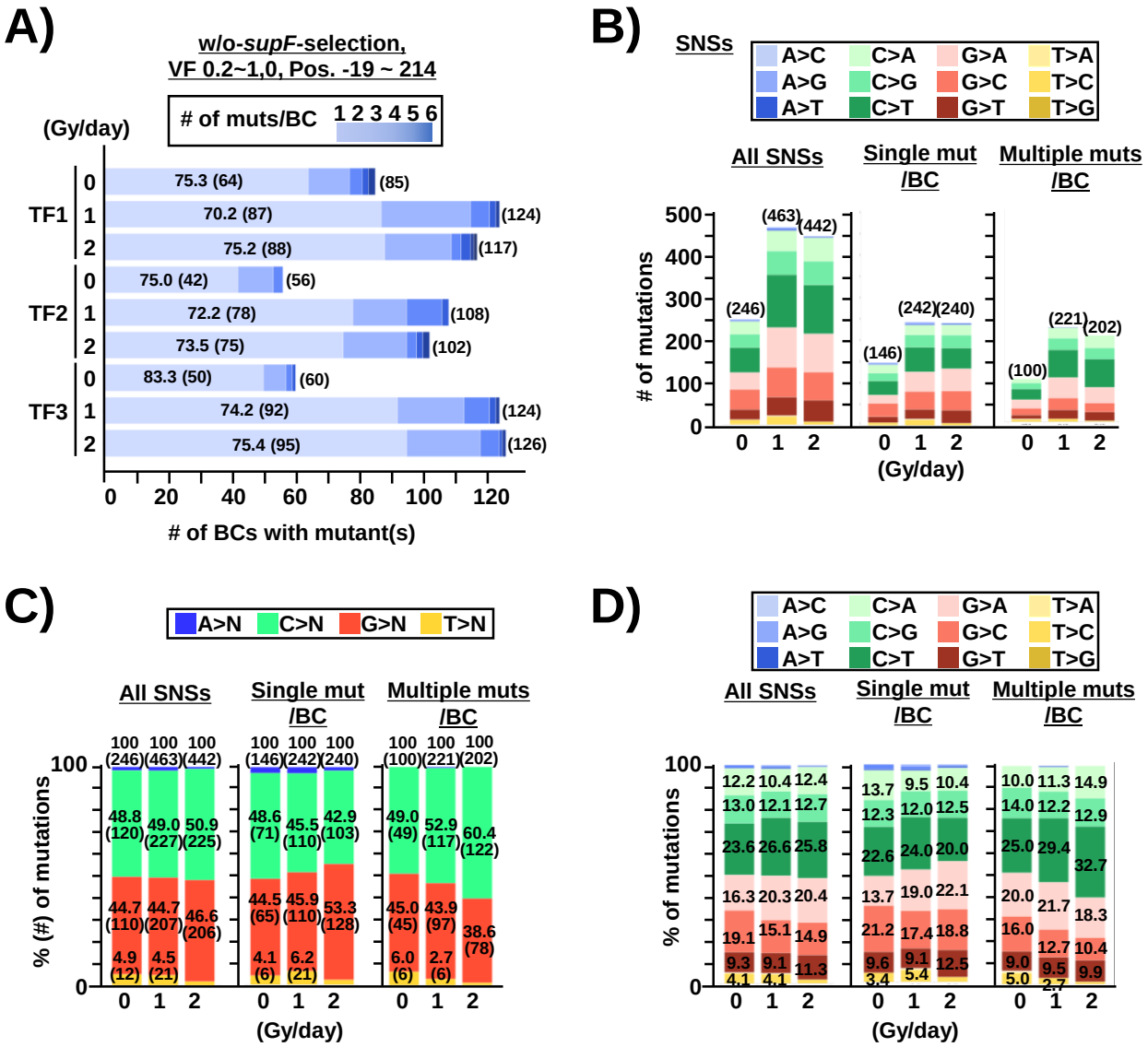

A)
w/o-supF-selection,
VF 0.2~1,0, Pos. -19 ~ 214
# of muts/BC 1 2 3 4 5 6
(Gy/day)
TF1
TF2
TF3
75.3 (64)
(85)
70.2 (87)
(124)
75.2 (88)
(117)
75.0 (42)
(56)
72.2 (78)
(108)
73.5 (75)
(102)
83.3 (50)
(60)
74.2 (92)
(124)
75.4 (95)
(126)
# of BCs with mutant(s)
B)
SNSs
A>C C>A G>A T>A
A>G C>G G>C T>C
A>T C>T G>T T>G
All SNSs
Single mut /BC
Multiple muts /BC
# of mutations
(Gy/day)
C)
A>N C>N G>N T>N
All SNSs
Single mut /BC
Multiple muts /BC
% (#) of mutations
(Gy/day)
D)
A>C C>A G>A T>A
A>G C>G G>C T>C
A>T C>T G>T T>G
All SNSs
Single mut /BC
Multiple muts /BC
% of mutations
(Gy/day)

Fig. 6

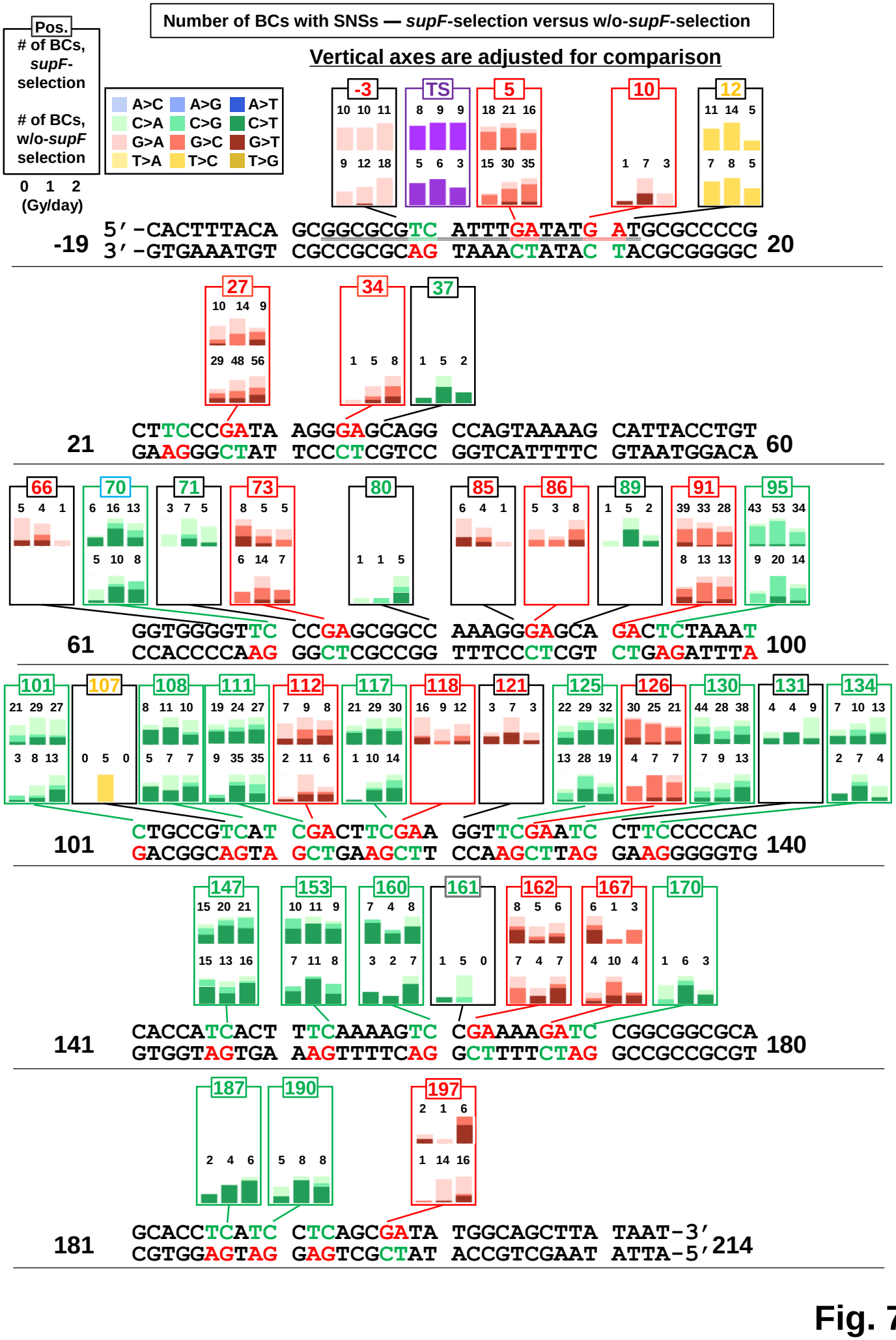

Number of BCs with SNSs — supF-selection versus w/o-supF-selection
Vertical axes are adjusted for comparison
Pos.
# of BCs, supF-selection
# of BCs, w/o-supF selection
0 1 2 (Gy/day)
A>C A>G A>T
C>A C>G C>T
G>A G>C G>T
T>A T>C T>G
-3 10 10 11 9 12 18
TS 8 9 9 5 6 3
5 18 21 16 15 30 35
10 1 7 3
12 11 14 5 7 8 5
-19 5′-CACTTTACA GCGGCGCGTC ATTTGATATG ATGCGCCCCG 20
3′-GTGAAATGT CGCCGCGCAG TAAACTATAC TACGCGGGGC
27 10 14 9 29 48 56
34 1 5 8
37 1 5 2
21 CTTCCCGATA AGGGAGCAGG CCAGTAAAAG CATTACCTGT 60
GAAGGGCTAT TCCCTCGTCC GGTCATTTTC GTAATGGACA
66 5 4 1
70 6 16 13 5 10 8
71 3 7 5
73 8 5 5 6 14 7
80 1 1 5
85 6 4 1
86 5 3 8
89 1 5 2
91 39 33 28 8 13 13
95 43 53 34 9 20 14
61 GGTGGGGTTC CCGAGCGGCC AAAGGGAGCA GACTCTAAAT 100
CCACCCCAAG GGCTCGCCGG TTTCCCTCGT CTGAGATTTA
101 21 29 27 3 8 13
107 0 5 0
108 8 11 10 5 7 7
111 19 24 27 9 35 35
112 7 9 8 2 11 6
117 21 29 30 1 10 14
118 16 9 12
121 3 7 3
125 22 29 32 13 28 19
126 30 25 21 4 7 7
130 44 28 38 7 9 13
131 4 4 9
134 7 10 13 2 7 4
101 CTGCCGTCAT CGACTTCGAA GGTTCGAATC CTTCCCCCAC 140
GACGGCAGTA GCTGAAGCTT CCAAGCTTAG GAAGGGGGTG
147 15 20 21 15 13 16
153 10 11 9 7 11 8
160 7 4 8 3 2 7
161 1 5 0
162 8 5 6 7 4 7
167 6 1 3 4 10 4
170 1 6 3
141 CACCATCACT TTCAAAAGTC CGAAAAGATC CGGCGGCGCA 180
GTGGTAGTGA AAGTTTTCAG GCTTTTCTAG GCCGCCGCGT
187 2 4 6
190 5 8 8
197 2 1 6 1 14 16
181 GCACCTCATC CTCAGCGATA TGGCAGCTTA TAAT-3′ 214
CGTGGAGTAG GAGTCGCTAT ACCGTCGAAT ATTA-5′

Fig. 7

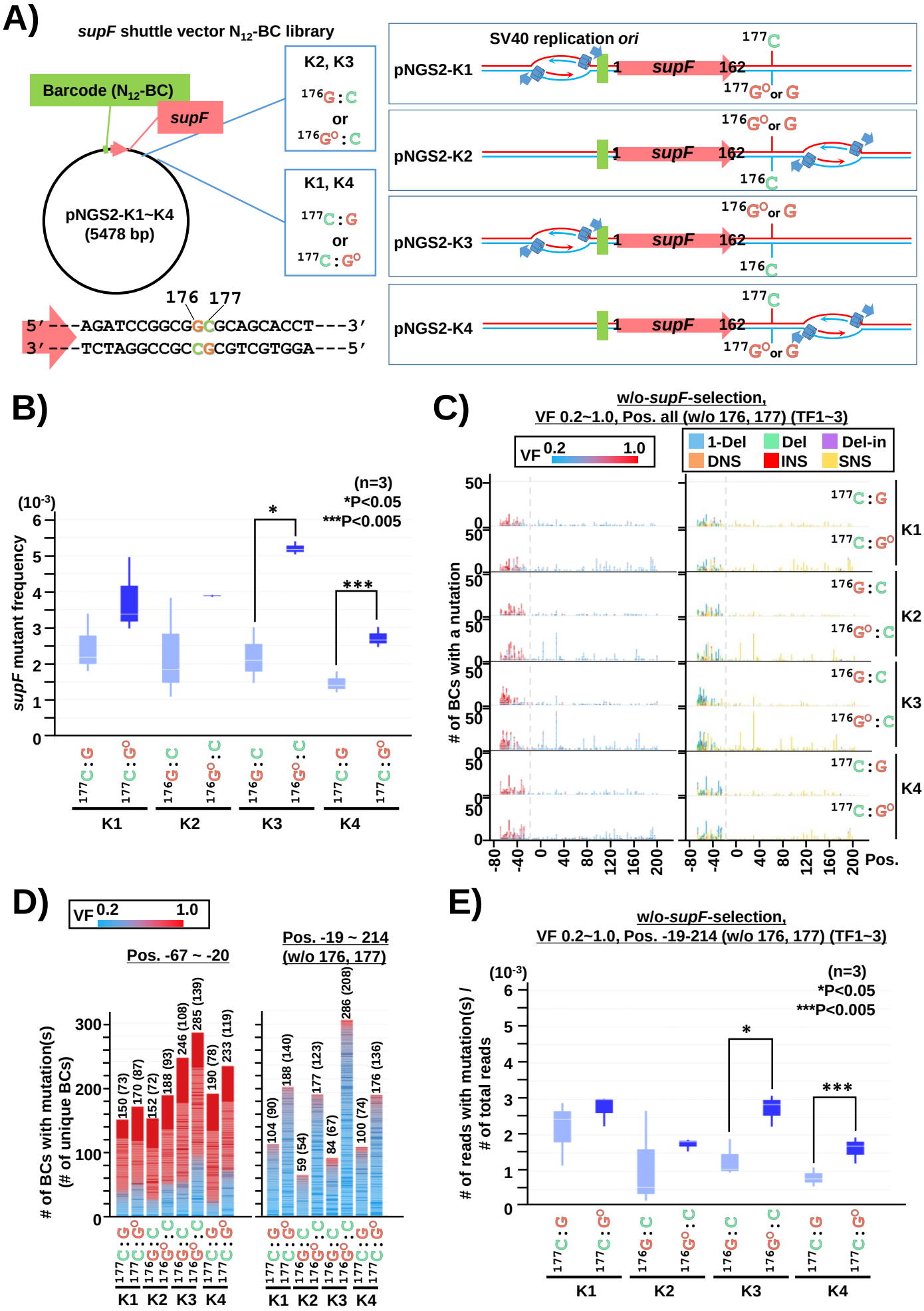

A)
supF shuttle vector N12-BC library
Barcode (N12-BC)
supF
pNGS2-K1~K4
(5478 bp)
K2, K3
176G : C
or
176G° : C
K1, K4
177C : G
or
177C : G°
176 177
5′---AGATCCGGCGGCGCAGCACCT---3′
3′---TCTAGGCCGCCGCGTCGTGGA---5′
SV40 replication ori
pNGS2-K1
pNGS2-K2
pNGS2-K3
pNGS2-K4
1
supF
162
177C
177G° or G
176G° or G
176C
B)
(n=3)
*P<0.05
***P<0.005
(10-3)
supF mutant frequency
K1
K2
K3
K4
C)
w/o-supF-selection,
VF 0.2~1.0, Pos. all (w/o 176, 177) (TF1~3)
VF 0.2 1.0
1-Del
Del
Del-in
DNS
INS
SNS
# of BCs with a nutation
Pos.
D)
VF 0.2 1.0
Pos. -67 ~ -20
Pos. -19 ~ 214
(w/o 176, 177)
# of BCs with mutation(s)
(# of unique BCs)
150 (73)
170 (87)
152 (72)
188 (93)
246 (108)
285 (139)
190 (78)
233 (119)
104 (90)
188 (140)
59 (54)
177 (123)
84 (67)
286 (208)
100 (74)
176 (136)
E)
w/o-supF-selection,
VF 0.2~1.0, Pos. -19-214 (w/o 176, 177) (TF1~3)
(n=3)
*P<0.05
***P<0.005
(10-3)
# of reads with mutation(s) /
# of total reads

Fig. 8

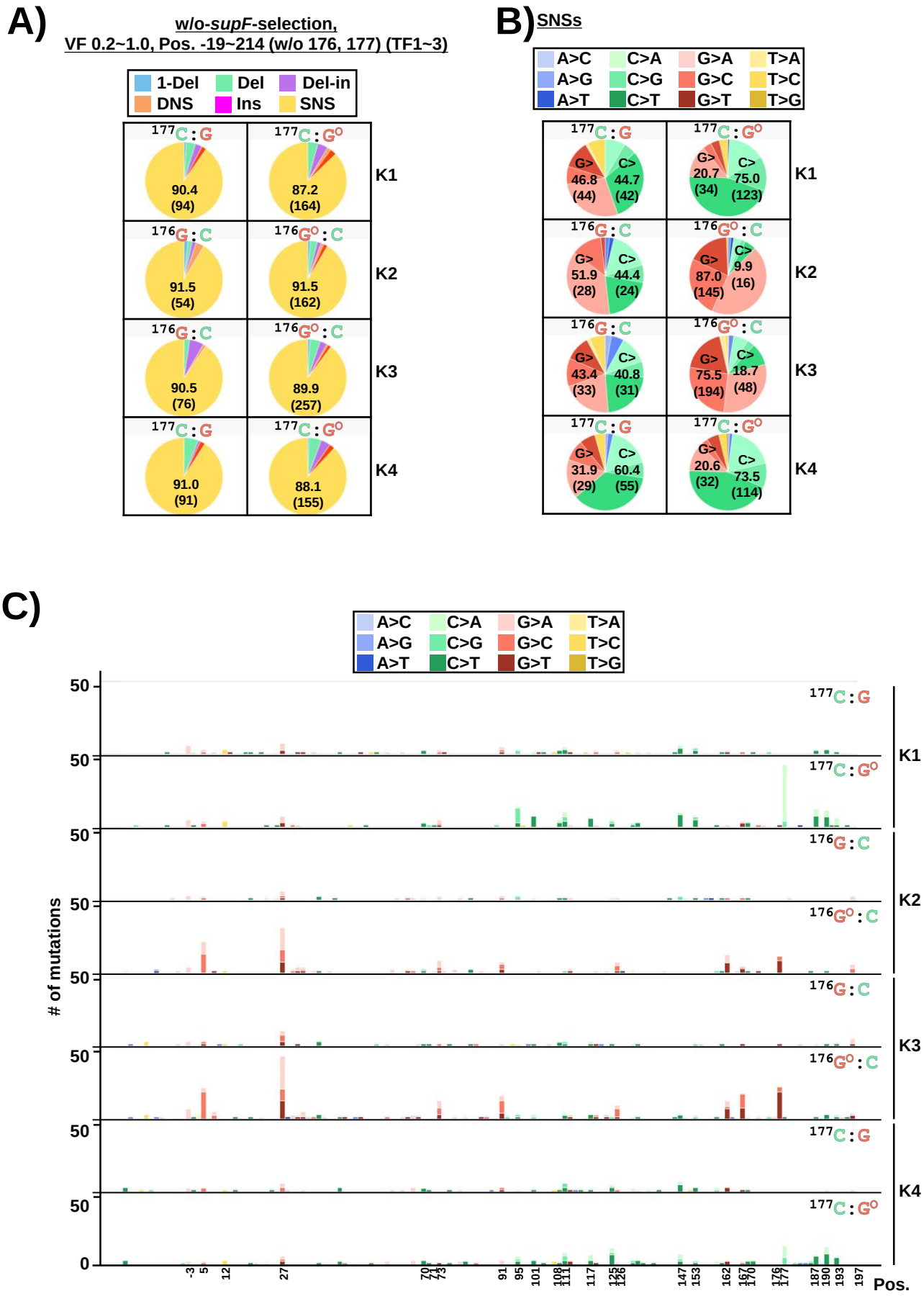

A)
w/o-supF-selection,
VF 0.2~1.0, Pos. -19~214 (w/o 176, 177) (TF1~3)
1-Del
Del
Del-in
DNS
Ins
SNS
177C : G
90.4 (94)
177C : G°
87.2 (164)
K1
176G : C
91.5 (54)
176G° : C
91.5 (162)
K2
176G : C
90.5 (76)
176G° : C
89.9 (257)
K3
177C : G
91.0 (91)
177C : G°
88.1 (155)
K4
B)
SNSs
A>C
C>A
G>A
T>A
A>G
C>G
G>C
T>C
A>T
C>T
G>T
T>G
177C : G
G> 46.8 (44)
C> 44.7 (42)
177C : G°
G> 20.7 (34)
C> 75.0 (123)
K1
176G : C
G> 51.9 (28)
C> 44.4 (24)
176G° : C
G> 87.0 (145)
C> 9.9 (16)
K2
176G : C
G> 43.4 (33)
C> 40.8 (31)
176G° : C
G> 75.5 (194)
C> 18.7 (48)
K3
177C : G
G> 31.9 (29)
C> 60.4 (55)
177C : G°
G> 20.6 (32)
C> 73.5 (114)
K4
C)
A>C
C>A
G>A
T>A
A>G
C>G
G>C
T>C
A>T
C>T
G>T
T>G
# of mutations
50
0
177C : G
177C : G°
176G : C
176G° : C
K1
K2
K3
K4
-3
5
12
27
70
71
73
91
95
101
108
111
117
125
126
147
153
162
167
170
176
177
187
190
193
197
Pos.

Fig. 9

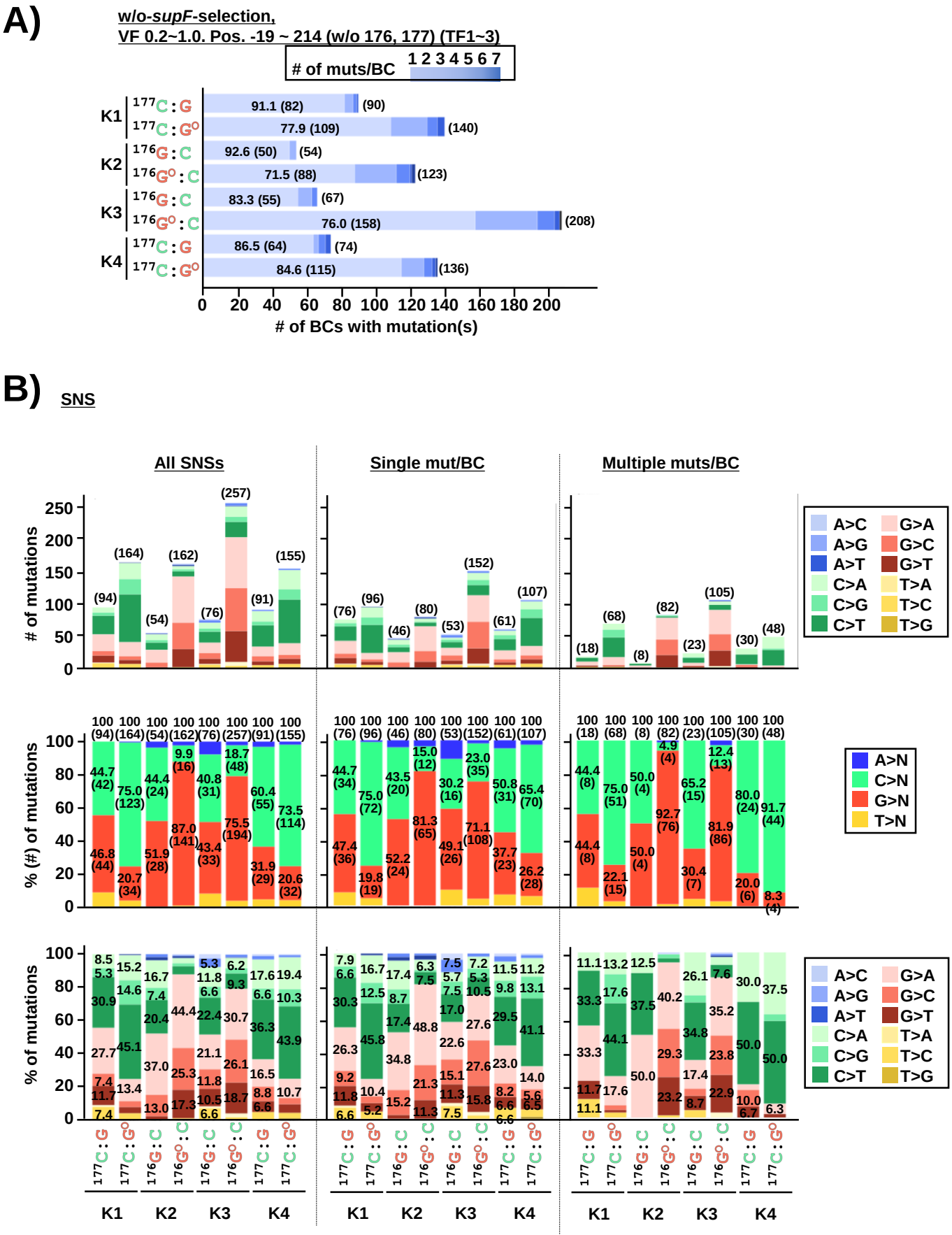

A)
w/o-supF-selection,
VF 0.2~1.0. Pos. -19 ~ 214 (w/o 176, 177) (TF1~3)
# of muts/BC
1 2 3 4 5 6 7
K1
K2
K3
K4
91.1 (82)
(90)
77.9 (109)
(140)
92.6 (50)
(54)
71.5 (88)
(123)
83.3 (55)
(67)
76.0 (158)
(208)
86.5 (64)
(74)
84.6 (115)
(136)
# of BCs with mutation(s)
B)
SNS
All SNSs
Single mut/BC
Multiple muts/BC
# of mutations
% (#) of mutations
% of mutations
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
A>N
C>N
G>N
T>N

Fig. 10

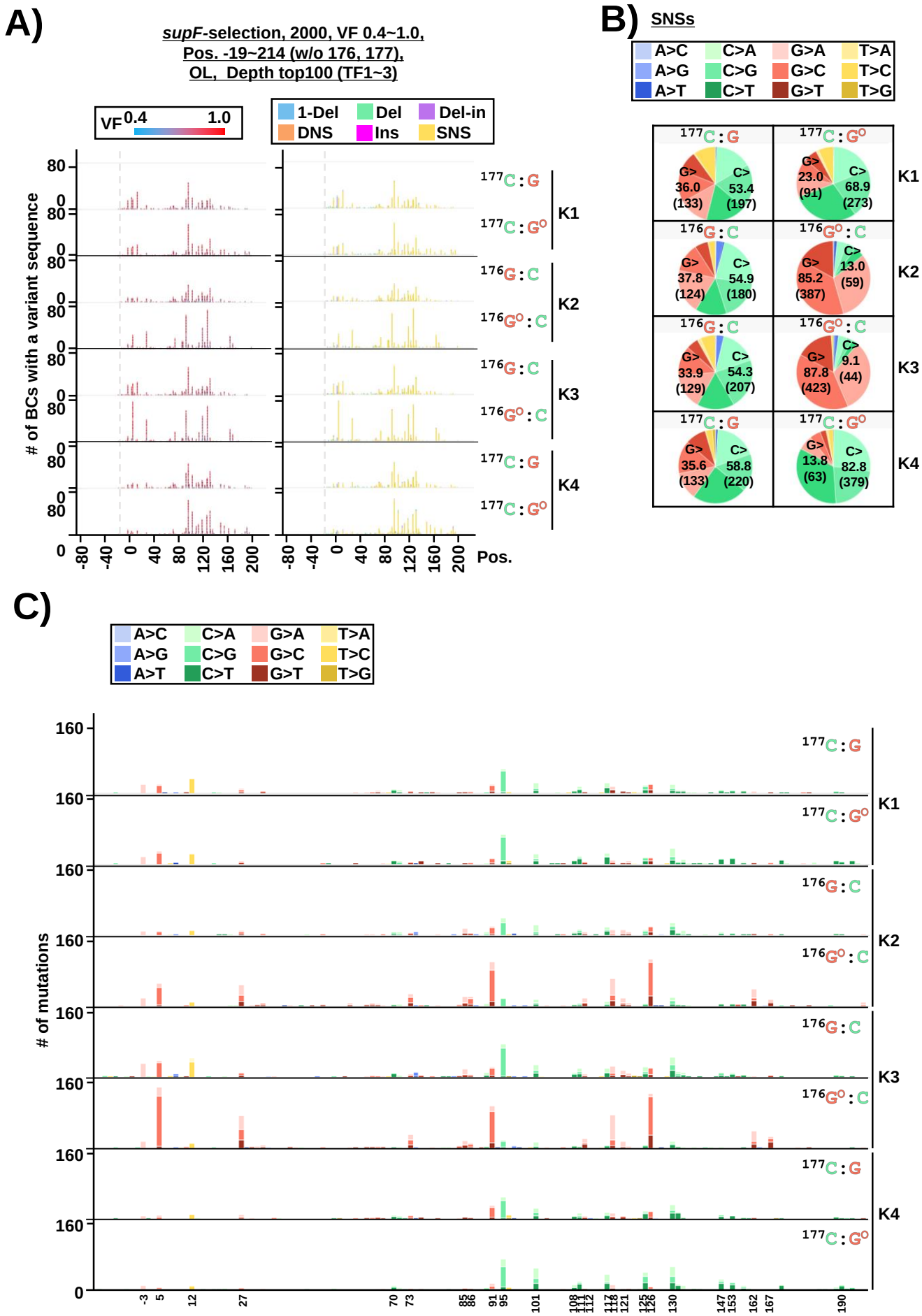

A)
supF-selection, 2000, VF 0.4~1.0,
Pos. -19~214 (w/o 176, 177),
OL, Depth top100 (TF1~3)
VF 0.4 1.0
1-Del
Del
Del-in
DNS
Ins
SNS
# of BCs with a variant sequence
177C : G
177C : G°
176G : C
176G° : C
K1
K2
K3
K4
Pos.
B)
SNSs
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
177C : G
G> 36.0 (133)
C> 53.4 (197)
177C : G°
G> 23.0 (91)
C> 68.9 (273)
176G : C
G> 37.8 (124)
C> 54.9 (180)
176G° : C
G> 85.2 (387)
C> 13.0 (59)
176G : C
G> 33.9 (129)
C> 54.3 (207)
176G° : C
G> 87.8 (423)
C> 9.1 (44)
177C : G
G> 35.6 (133)
C> 58.8 (220)
177C : G°
G> 13.8 (63)
C> 82.8 (379)
C)
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
# of mutations
177C : G
177C : G°
176G : C
176G° : C
K1
K2
K3
K4

Fig. 11

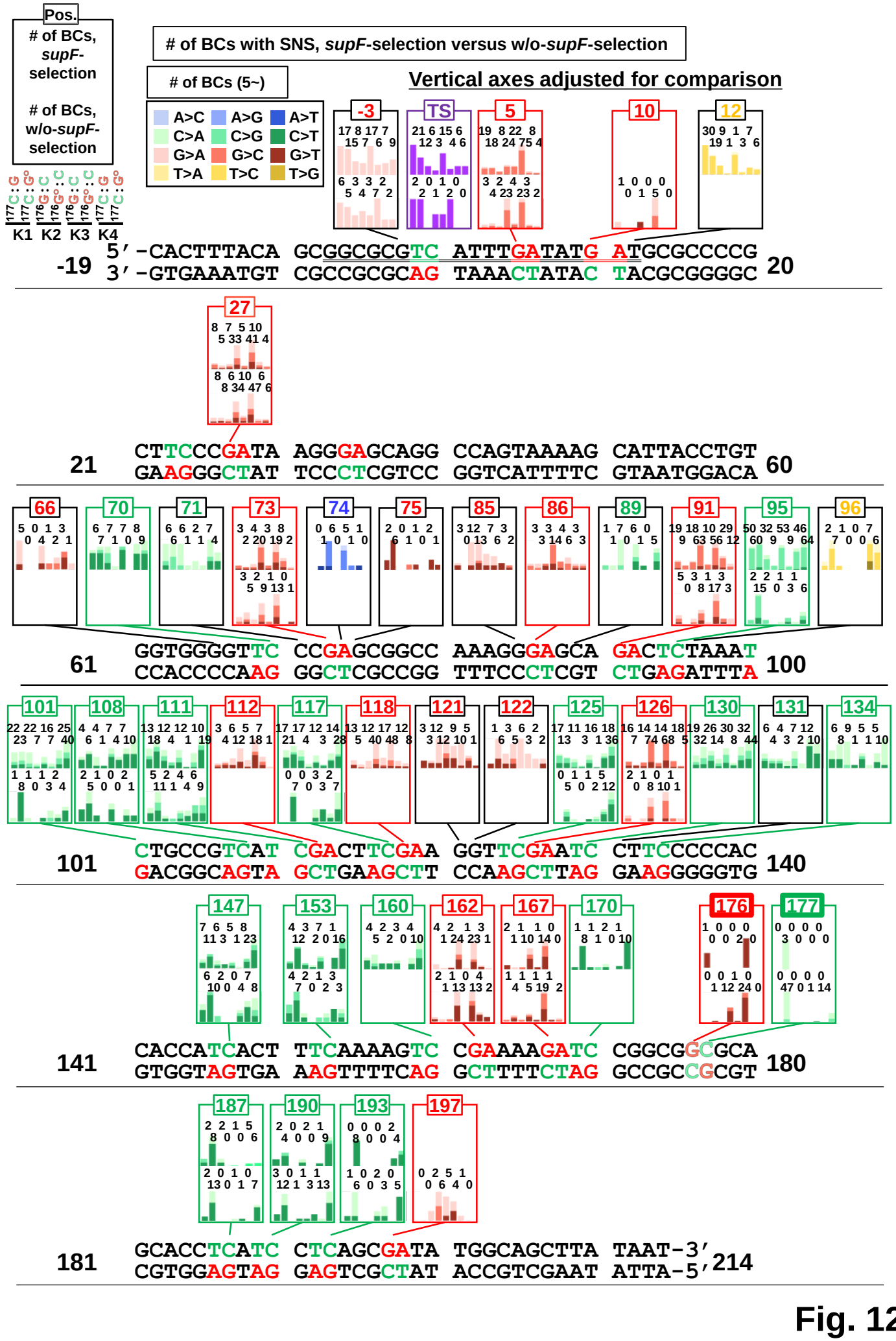

Pos.
# of BCs, supF-selection
# of BCs, w/o-supF-selection
# of BCs with SNS, supF-selection versus w/o-supF-selection
# of BCs (5~)
Vertical axes adjusted for comparison
A>C A>G A>T
C>A C>G C>T
G>A G>C G>T
T>A T>C T>G
K1 K2 K3 K4
-3 TS 5 10 12
-19 5′-CACTTTACA GCGGCGCGTC ATTTGATATG ATGCGCCCCG 20
3′-GTGAAATGT CGCCGCGCAG TAAACTATAC TACGCGGGGC
27
21 CTTCCCGATA AGGGAGCAGG CCAGTAAAAG CATTACCTGT 60
GAAGGGCTAT TCCCTCGTCC GGTCATTTTC GTAATGGACA
66 70 71 73 74 75 85 86 89 91 95 96
61 GGTGGGGTTC CCGAGCGGCC AAAGGGAGCA GACTCTAAAT 100
CCACCCCAAG GGCTCGCCGG TTTCCCTCGT CTGAGATTTA
101 108 111 112 117 118 121 122 125 126 130 131 134
101 CTGCCGTCAT CGACTTCGAA GGTTCGAATC CTTCCCCCAC 140
GACGGCAGTA GCTGAAGCTT CCAAGCTTAG GAAGGGGGTG
147 153 160 162 167 170 176 177
141 CACCATCACT TTCAAAAGTC CGAAAAGATC CGGCGGCGCA 180
GTGGTAGTGA AAGTTTTCAG GCTTTTCTAG GCCGCCGCGT
187 190 193 197
181 GCACCTCATC CTCAGCGATA TGGCAGCTTA TAAT-3′ 214
CGTGGAGTAG GAGTCGCTAT ACCGTCGAAT ATTA-5′

Fig. 12

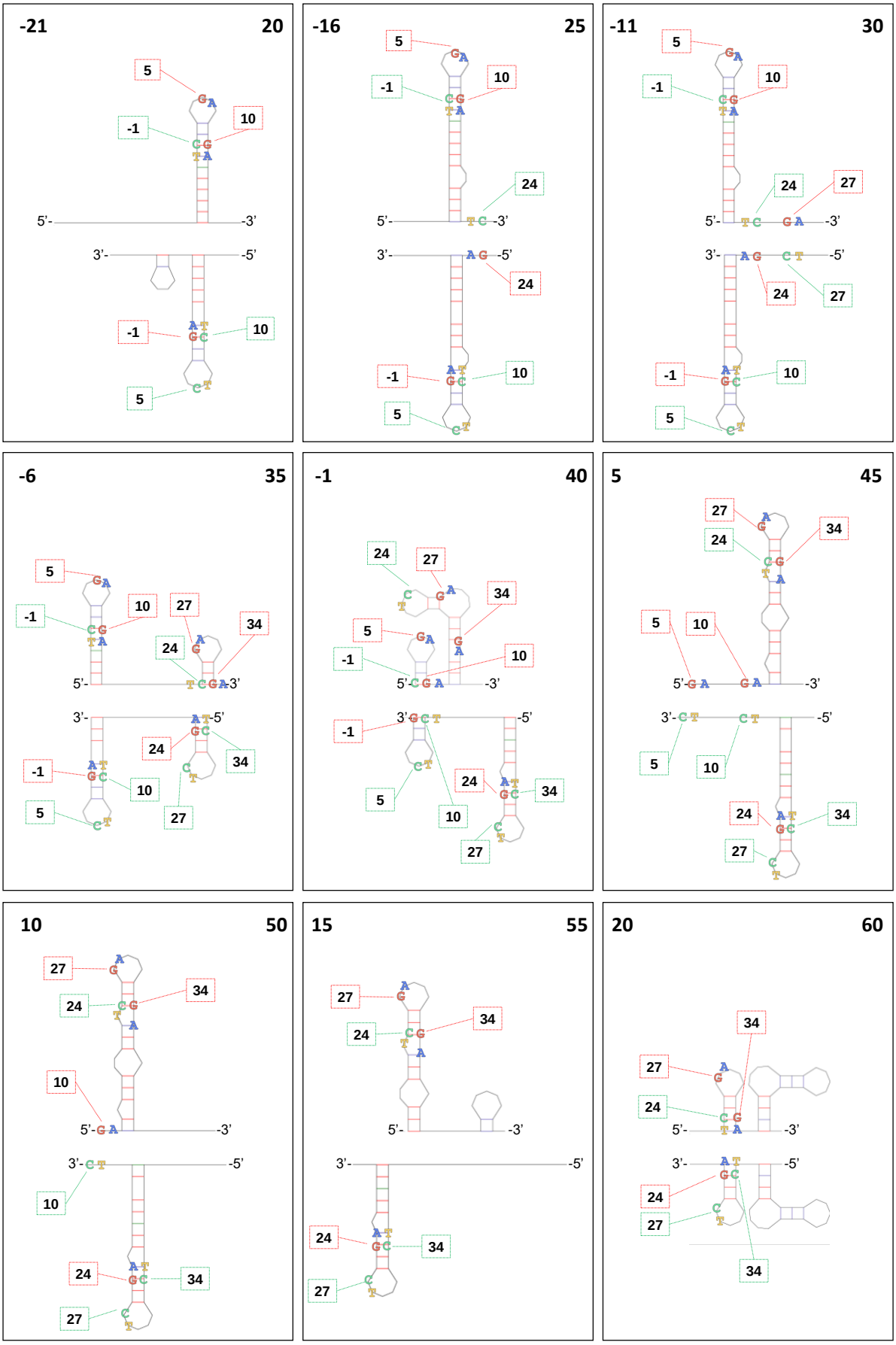



Fig. 13

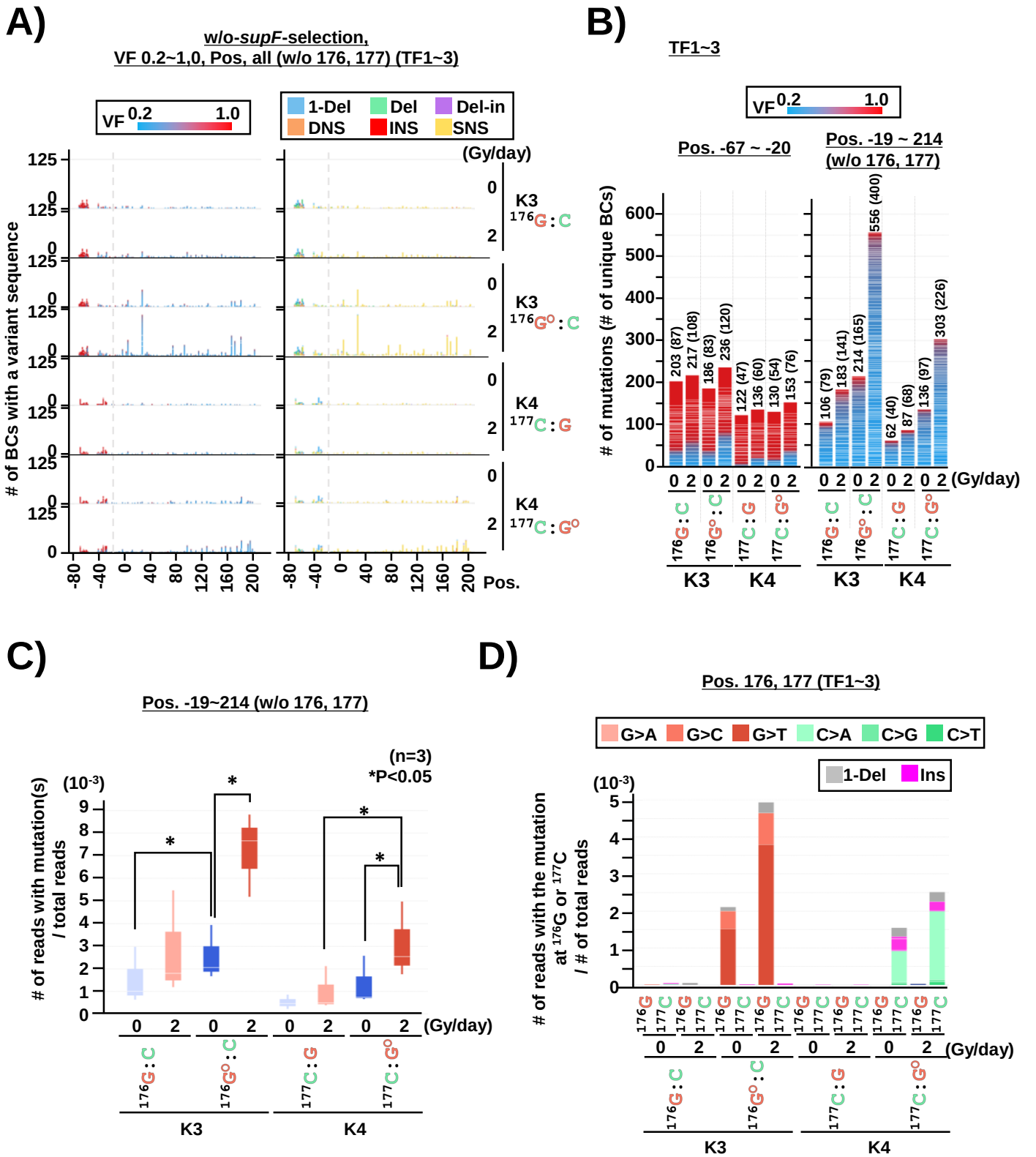

A)
w/o-supF-selection,
VF 0.2~1.0, Pos, all (w/o 176, 177) (TF1~3)
VF 0.2 1.0
1-Del
Del
Del-in
DNS
INS
SNS
(Gy/day)
# of BCs with a variant sequence
K3 176G : C
K3 176Go : C
K4 177C : G
K4 177C : Go
Pos.
B)
TF1~3
VF 0.2 1.0
Pos. -67 ~ -20
Pos. -19 ~ 214 (w/o 176, 177)
# of mutations (# of unique BCs)
203 (87)
217 (108)
186 (83)
236 (120)
122 (47)
136 (60)
130 (54)
153 (76)
106 (79)
183 (141)
214 (165)
556 (400)
62 (40)
87 (68)
136 (97)
303 (226)
(Gy/day)
K3
K4
C)
Pos. -19~214 (w/o 176, 177)
(n=3)
*P<0.05
# of reads with mutation(s) / total reads
(10-3)
(Gy/day)
K3
K4
D)
Pos. 176, 177 (TF1~3)
G>A
G>C
G>T
C>A
C>G
C>T
1-Del
Ins
# of reads with the mutation at 176G or 177C / # of total reads
(10-3)
(Gy/day)
K3
K4

Fig. 14

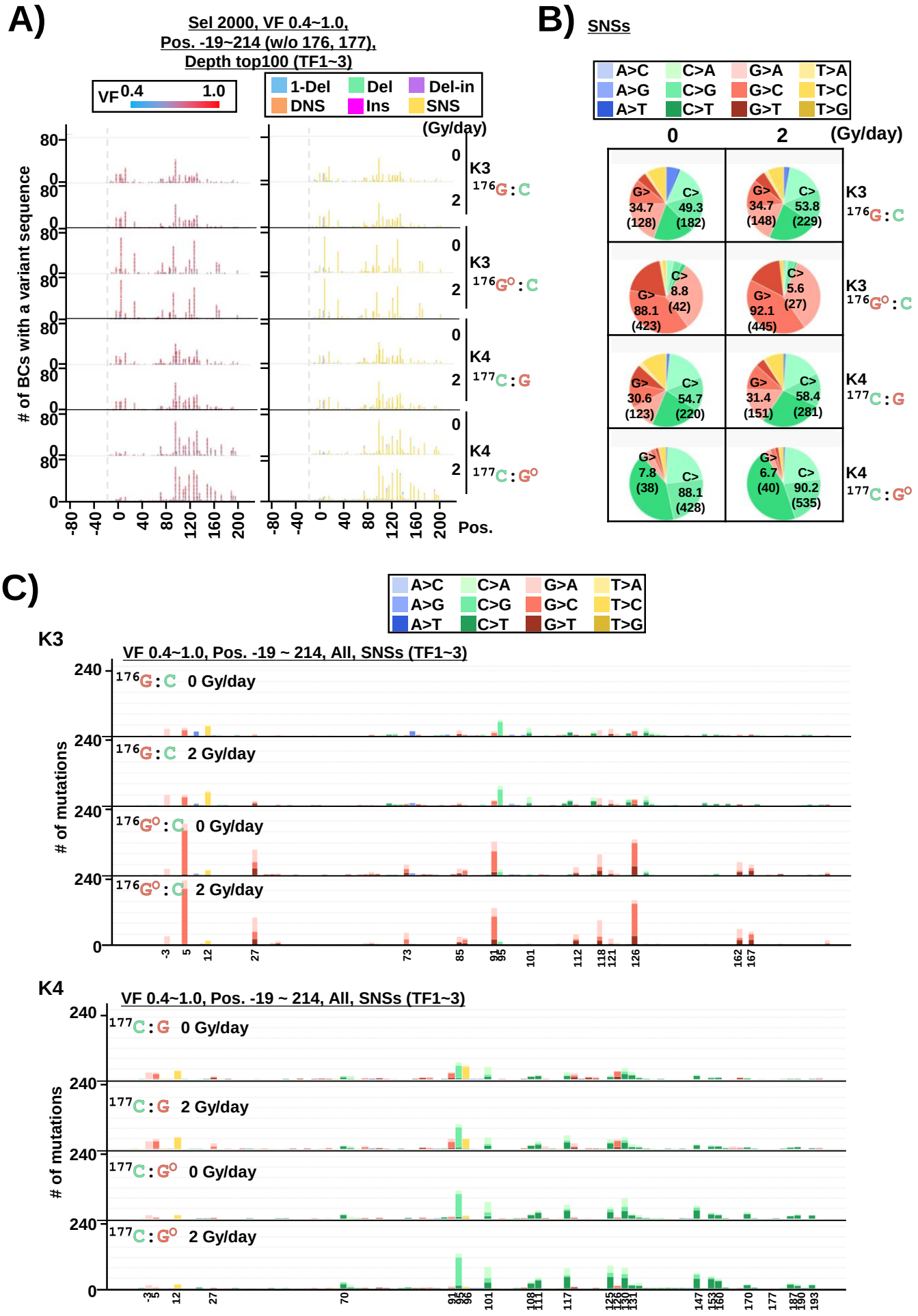
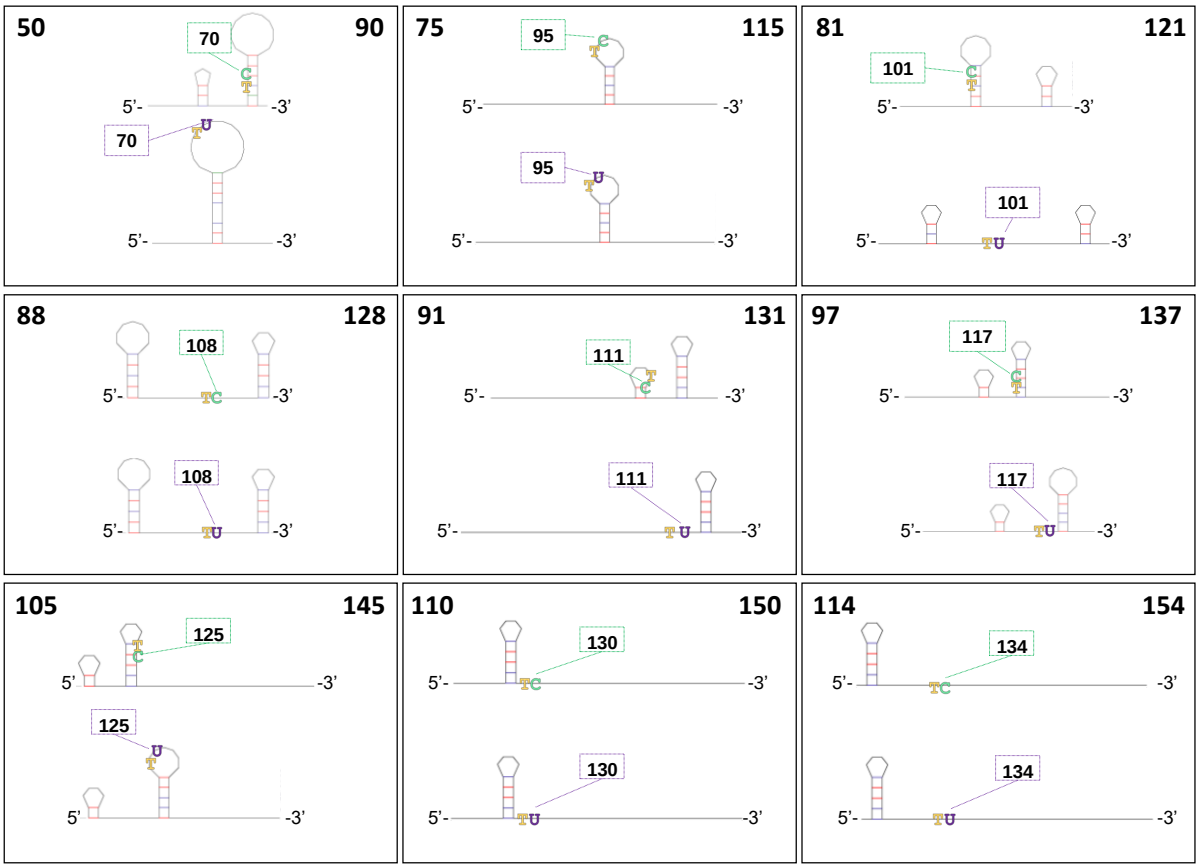

A)
Sel 2000, VF 0.4~1.0,
Pos. -19~214 (w/o 176, 177),
Depth top100 (TF1~3)
VF 0.4 1.0
1-Del
Del
Del-in
DNS
Ins
SNS
(Gy/day)
# of BCs with a variant sequence
K3 176G : C
K3 176G° : C
K4 177C : G
K4 177C : G°
Pos.
B) SNSs
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
0
2
(Gy/day)
G> 34.7 (128)
C> 49.3 (182)
G> 34.7 (148)
C> 53.8 (229)
G> 88.1 (423)
C> 8.8 (42)
G> 92.1 (445)
C> 5.6 (27)
G> 30.6 (123)
C> 54.7 (220)
G> 31.4 (151)
C> 58.4 (281)
G> 7.8 (38)
C> 88.1 (428)
G> 6.7 (40)
C> 90.2 (535)
C)
K3
VF 0.4~1.0, Pos. -19 ~ 214, All, SNSs (TF1~3)
176G : C 0 Gy/day
176G : C 2 Gy/day
176G° : C 0 Gy/day
176G° : C 2 Gy/day
# of mutations
K4
VF 0.4~1.0, Pos. -19 ~ 214, All, SNSs (TF1~3)
177C : G 0 Gy/day
177C : G 2 Gy/day
177C : G° 0 Gy/day
177C : G° 2 Gy/day
# of mutations

Fig. 15

## Non-transcribed strand of *supF* gene




## Transcribed strand of *supF* gene

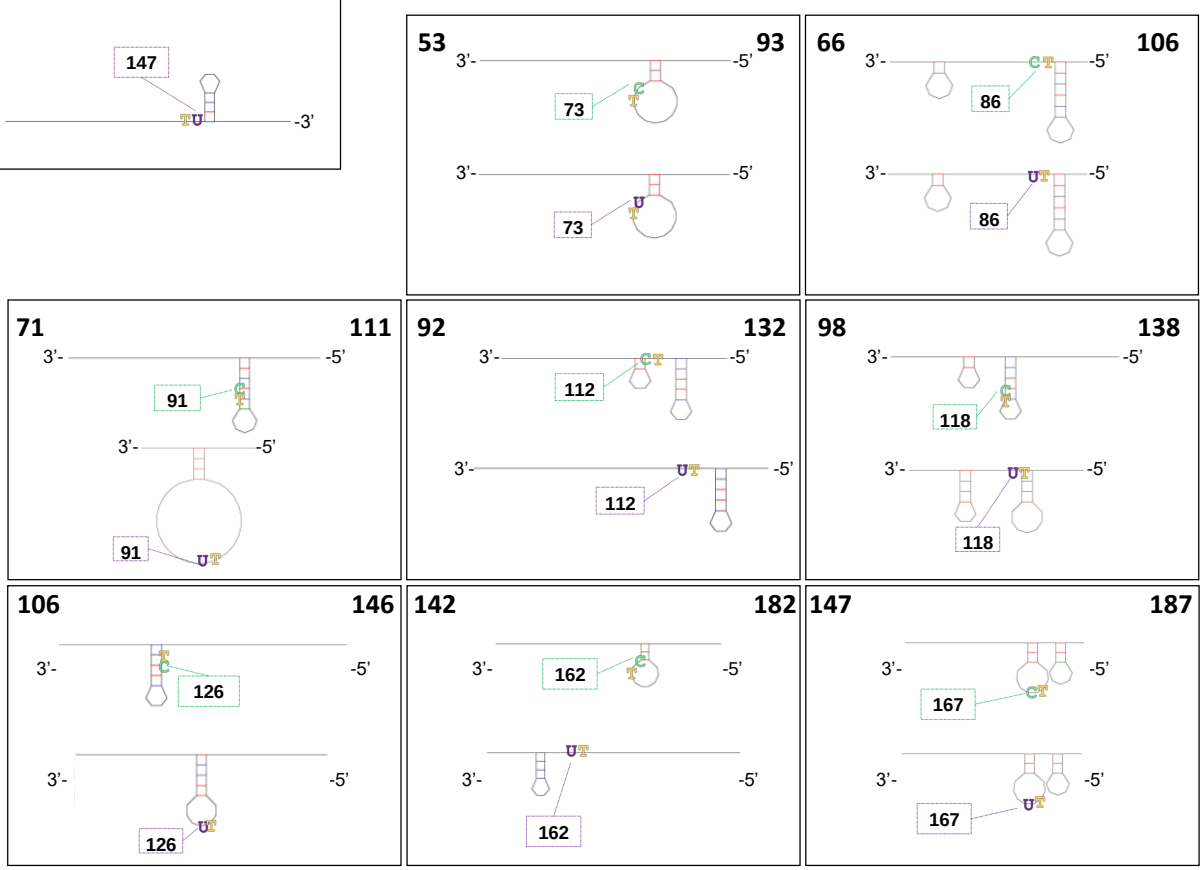



Fig. 16