PNAS

Supplementary Information for

A quantification method of somatic mutations in normal tissues and their accumulation in pediatric patients with chemotherapy

Sho Ueda, Satoshi Yamashita, Miho Nakajima, Tadashi Kumamoto, Chitose Ogawa, Yu-Yu Liu, Harumi Yamada, Emi Kubo, Naoko Hattori, Hideyuki Takeshima, Mika Wakabayashi, Naoko Iida, Yuichi Shiraishi, Masayuki Noguchi, Yukio Sato, Toshikazu Ushijima

Toshikazu Ushijima
Email: tushijim@ncc.go.jp

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Materials and methods

Cell culture, clinical sample, and DNA preparation

The human TK6 lymphoblast cell line was obtained from the American Type Culture Collection (Manassas, VA). The human HPDE-4/E6E7 immortalized pancreatic ductal epithelial cell line was obtained from MS Tsao's laboratory. The human 293FT embryo kidney cell line was obtained from Thermo Fisher Scientific (Waltham, MA). Genomic DNA (gDNA) of TK6 and HPDE-4 cells were extracted by the phenol/chloroform method. gDNA of 293FT cells was extracted using a QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany).

DNA from peripheral blood cells was collected from serum-removed blood including leukocytes using a QIAamp DNA Mini Kit (QIAGEN), or FlexiGene DNA Kit (QIAGEN) and stored in the National Cancer Center Biobank (Tokyo, Japan). This study was approved by the Institutional Review Board of the National Cancer Center (approval No. 2018-024), and all the specimens were obtained with written informed consents.

All gDNAs were measured for nucleic acid concentration and double-stranded DNA (dsDNA) concentration. Nucleic acid concentration was measured using a NanoDrop 2000c (Thermo Fisher Scientific), which measures optical density at wavelengths of 260 and 280 nm. dsDNA concentration was measured using a Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific), which measures fluorescence derived from a dye that intercalates into dsDNA.

Construction of model DNA samples with artificially prepared mutations

The whole-genome sequencing (WGS) was performed for gDNAs of TK6 and HPDE-4 cell lines to detect SNPs. Sequence libraries were prepared using a TruSeq DNA PCR-Free kit (Illumina, San Diego, CA), and sequenced using NovaSeq 6000 (Illumina) to achieve 300M PE reads respectively at 150-bp PE sequencing. WGS reads were trimmed at their adaptor sequences using Fastp, aligned to the hg38 reference genome using BWA-MEM, and PCR duplicates were removed using Samtools. SNPs present only in HPDE-4 cells were identified using GenomonFisher (comparison option) with the following criteria: sequencing depth of ≥20 both in HPDE-4 and TK6 cells, variant allele frequency of ≥0.2 in HPDE-4 cells, variant allele frequency of ≤0.02 in TK6 cells, and Fisher p-value of ≤0.001.

Model DNA samples were prepared by mixing a small amount of HPDE-4 gDNA into TK6 gDNA at mixing ratios of 1, 0.3, 0.1, 0.03, and 0.01%, and the above SNPs present only in HPDE-4 cells were considered as artificial rare 'mutations'. An expected 'mutation' allele frequency was calculated as half of the mixing ratio. An expected 'mutation' frequency was then calculated by the expected 'mutation' allele frequency and the number of 'mutations' in a whole genomic region (analyzed regions by WGS) (Fig. 2*A*). After analysis of these model DNA samples using EcoSeq, a measured 'mutation' frequency was calculated by dividing the number of detected 'mutations' by the total number of analyzed base pairs (bp) (Fig. 2*A*).

Mutagen treatment of 293FT cells

Newly-cloned 293FT cells were treated with a mutagen, 4-nitroquinoline 1-oxide (4-NQO, Sigma-Aldrich) (*SI Appendix*, Fig. S3*A*). The cells were seeded at a density of 1×10^5 cells/10cm plate on day 0 and treated with 4-NQO on day 1 for three doses (0.1, 0.3, and 1 μg/mL). PBS treatment was used as a mock-treatment. The medium was changed to a normal medium on day 2 and 5, and the cells were harvested on day 8 to secure sufficient time for DNA lesions to be fixed as mutations.

Synthesis of the Duplex-loop adaptor

A Duplex-loop adaptor was constructed from a 94-mer (v0 adaptor) or 95-mer (v1 adaptor) oligonucleotide including one deoxyuridine (dU) base and 12-nt randomized sequence as a unique molecular identifier (UMI) (*SI Appendix*, Fig. S7*A*, *B*, and Table S9). This oligonucleotide was annealed to its intrinsic complementary sequence by heating at 95°C for 5 min and cooling slowly for 2 hours to 4°C to form a loop structure of NEBNext Adaptor (Illumina) and UMI. The single-strand UMI in an annealed oligonucleotide was converted to double-strand UMI using Klenow Fragment (3'->5' exo-) (New England Biolabs, Ipswich, UK) at 37°C for 90 min. After purification by ethanol precipitation and re-annealing, the end of this adaptor was digested by HpyCH4III (v0

adaptor) or Hpy188III (v1 adaptor) restriction enzyme at 37°C for 12-16 hours. After purification by ethanol precipitation and re-annealing once more, a 3'-dT-tailed (v0 adaptor) or 5'-TC-tailed (v1 adaptor) Duplex-loop adaptors, which could ligate 3'-dA-tailed (v0 adaptor) and 5'-GA-tailed (v1 adaptor) DNA fragments, respectively, were synthesized.

Collection of myelocytes and lymphocytes

Myelocytes were collected from 3 mL of peripheral blood using an EasySep HLA Chimerism Whole Blood Myeloid Positive Selection Kit on an EasySep Magnet (Stemcell Technologies, Vancouver, BC, Canada). The isolated cells were stained with anti-CD66b-PE antibody (clone G10F5; BioLegend, San Diego, CA) and enrichment was assessed using an Attune NxT Flow Cytometer (Thermo Fisher Scientific). Lymphocytes were collected from 3 mL of peripheral blood using an EasySep HLA Chimerism Whole Blood Lymphoid Positive Selection Kit on an EasySep Magnet (Stemcell Technologies). Enrichment of lymphocytes was confirmed using an Attune NxT Flow Cytometer after staining with anti-CD2-PE (clone RPA-2.10; BioLegend) and anti-CD20-FITC antibodies (clone 2H7; BioLegend).

EcoSeq library preparation and sequencing

500 ng (v0 adaptor) or 100 ng (v1 adaptor) double-strand DNA (measured by PicoGreen assay) was digested by BamHI restriction enzyme at 37 °C for 12-16 hours. The digested DNA fragments in a 50 µl volume were purified and size-selected using AMPure XP (Beckman Coulter, Brea, CA) to collect DNA fragments of 100 to 700-bp. Specifically, AMPure XP beads (0.7×, 35 µl) were added to the digestion sample and settled to the magnet. After collecting the supernatant, AMPure XP beads (1.1×, 55 µl) were added to the supernatant and settled to the magnet again. After discarding the supernatant, size-selected DNA binding beads were washed with 85% ethanol and eluted to Tris-HCl buffer (15 µl). A small peak between 100-700-bp in an electrophoretic image was confirmed, and an approximate number of moles was measured using Agilent 4200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, Santa Clara, CA). At this time, a large amount of DNA fragments of >1,500-bp remained even after size-selection, but they could not be amplified by PCR and could be ignored.

In the case of the v0 adaptor, ends of size-selected DNA fragments were repaired, and 3'-dA-tailed using ERAT enzyme in a GenNext® NGS Library Prep Kit (TOYOBO, Oosaka, Japan) at 30°C for 40 min and 65°C for 5 min. The 3'-dA-tailed DNA fragments were ligated to the 3'-dT-tailed v0 adaptor in 3-5-fold molar excess using ligation solution in a GenNext® NGS Library Prep Kit (TOYOBO) at 20°C for 20 min. In the case of the v1 adaptor, ends of size-selected DNA fragments were trimmed by partial filling-in, and 5'-GA-tailed using dATP, dGTP, and Klenow fragment (3'->5' exo-) (0.01U) (New England Biolabs) at 37°C for 5 min and 65°C for 5 min. The 5'-GA-tailed DNA fragments were ligated to the 50 fmol 5'-TC-tailed v1 adaptor using ligation solution in a GenNext® NGS Library Prep Kit (TOYOBO) at 20°C for 20 min. To excise dU-base of adaptors, adaptor-ligated DNA fragments were treated with USER enzyme (New England Biolabs) at 37°C for 30 min. dU-excised DNA fragments were purified using AMPure XP (0.7×), and eluted to low-TE buffer (20 µl).

Based on the copy number of a pre-PCR library measured by real-time PCR using KOD SYBR® qPCR Mix (TOYOBO) and pre-PCR-specific primers (*SI Appendix*, Table S6), 1M copies were used as a template for PCR. PCR was performed using KAPA HiFi HotStart ReadyMix (KAPA BIOSYSTEMS) and NEBNext Universal and Index Primers (New England Biolabs), and individual samples were uniquely barcoded using primers with different index primer sequences. PCR was conducted in two steps with 98°C (20 seconds) and 68°C (5 seconds) for 18 cycles. The copy number of a post-PCR library was measured by real-time PCR using KOD SYBR® qPCR Mix (TOYOBO) and post-PCR-specific primers (*SI Appendix*, Table S6). The prepared libraries were pooled to equal volume and sequenced using HiSeq X Ten (Illumina) or NovaSeq 6000 (Illumina) to achieve 40M PE reads per sample at 150-bp PE sequencing.

Data processing for mapping

A flowchart of the data processing is shown in Fig. S8 (*SI Appendix*). The shell script and R script for the data processing are available at the GitHub repository (https://github.com/EpigeneticField/EcoSeq).

An intact sequence was expected to consist of a 12-nt UMI tag, a 5-nt spacer sequence, and an insert DNA sequence starting from a partial BamHI sequence (GATCC). From a pair of sequences (Read 1 and 2), the two UMIs were combined into a read header as a 24-nt tag. The spacer sequence was removed using a pre-existing script from Duplex sequencing: tag_to_header.py (1). The partial BamHI sequence at the 5'-end was also removed to eliminate errors using Fastp (2). The trimmed reads were then aligned to the hg38 reference genome using BWA-MEM, and multi-mapped reads were filtered out using grep (original sequence, OS). The OS reads were sorted and converted to a BAM file using Samtools.

Three or more OS reads containing identical tag sequences were merged into a single-strand consensus sequence (SSCS) using ConsensusMaker.py from Duplex sequencing (1). Two SSCSs from a single double-strand insert DNA were merged into a double-strand consensus sequence (Duplex consensus sequence, DCS) using DuplexMaker.py from Duplex sequencing (1). If a partial BamHI sequence (GGATC) appeared in the middle of DCS, the following sequence at its 3' side was removed using Fastp (2). DCS reads were aligned to the hg38 reference genome using BWA-MEM, and converted to the BAM file using Samtools.

Data processing for mutation calling by DCSs

To determine genomic regions analyzed by an EcoSeq analysis, sequencing depths of OS reads and DCS reads were counted from each BAM file using Samtools. The ENCODE Blacklist (3) and specific genomic regions obtained from the UCSC genome browser that contained simple repeat, duplication, and self-chain were filtered out using bedtools. The analyzed regions were determined using a threshold of sequencing depth of OS reads ≥20 and DCS reads ≥5 using R software. The total number of analyzed bp was calculated based upon the analyzed genomic regions and sequencing depth by DCS reads.

All variants were first identified from mapped DCS reads regardless of their depths using GenomonFisher (single option) (https://github.com/Genomon-Project). Then, only variants located in the analyzed genomic regions were selected by superimposing the variant list and the analyzed genomic regions. At the same time, SNPs were identified as variants that had an allele frequency of ≥10% in OS reads or a variant count of ≥2 in DCS reads. The SNPs were removed from the variants, and the remaining variants were considered as mutations. A common SNP database (NCBI dbSNP Build 153) was referred to in order to exclude common SNPs in the list of the identified mutations. Furthermore, positions of all detected mutations were confirmed using the Integrative Genomics Viewer (4), and mutations located in terminal regions of a DCS that had failed to be trimmed or those found in a DCS with three variants or more were removed as artifacts.

Mutation frequency was calculated by dividing the number of detected mutations by the total number of analyzed bp. The difference in mutation frequencies was assessed by the Student t-test, and a p-value of <0.05 was considered as statistically significant. Mutational profiles were assessed by the 96 mutation type classification (5), and the COSMIC v3 signatures (5) and environmental mutagen signatures (6) estimated to contribute to the profiles were analyzed using Signal (7).

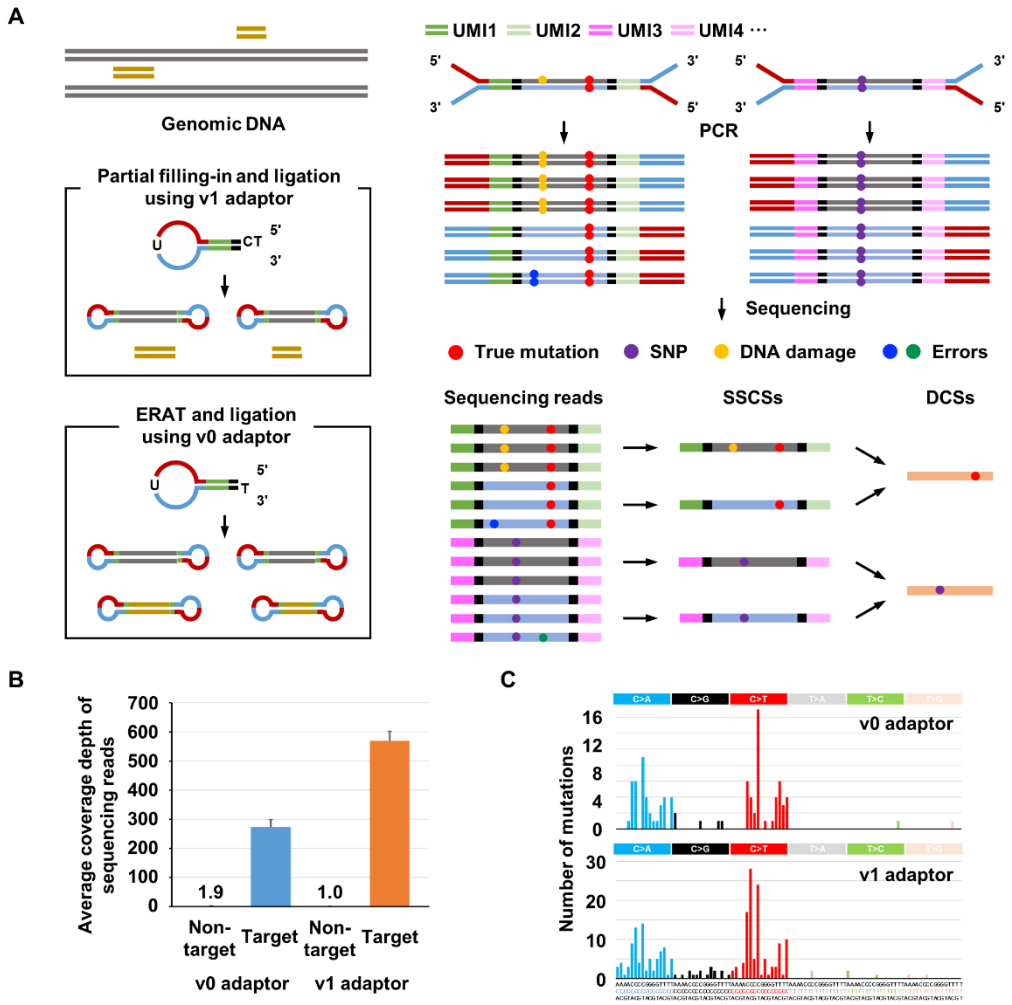

Fig. S1. EcoSeq using an improved adaptor (v1) and an original adaptor (v0). (A) Schema of EcoSeq library preparation, sequencing, and DCS assembly. Genomic regions are reduced by digestion using BamHI restriction enzyme and size selection. The improved adaptor (v1 adaptor) can be specifically and efficiently ligated to BamHI-digested fragments after partial filling-in, while the original adaptor (v0 adaptor) could be ligated to degraded short fragments (brown fragments) after blunting and A addition. After Uracil excision, adaptor-ligated fragments are amplified and sequenced by next generation sequencing. Single-strand DNA damage, PCR errors, and sequencing errors are excluded in the process of DCS assembly. Real mutations and SNPs can be detected by mapping DCSs to a reference genome. ERAT, end repair and adenine-tailing enzyme; UMI, unique molecular identifier; SSCS, single-strand consensus sequence; DCS, duplex consensus sequence. (B) Average coverage depth of sequencing reads in target and non-target regions. Genomic regions expected to be covered by BamHI-digested 100 to 700-bp fragments were defined as the target regions, and the other regions were defined as the non-target regions. For these two regions, average coverage depths of sequencing reads obtained from EcoSeq libraries prepared by the v0 adaptor (n=9) and the v1 adaptor (n=7) were analyzed. Almost all reads were mapped to the target regions, and the v1 adaptor showed higher depths than the v0 adaptor. Error bar represents standard deviation (SD). (C) Mutational profiles of EcoSeq libraries prepared by v0 adaptor (n=4) and v1 adaptor (n=3) from the same cell line sample. Similar profiles were shown between the two adaptors.

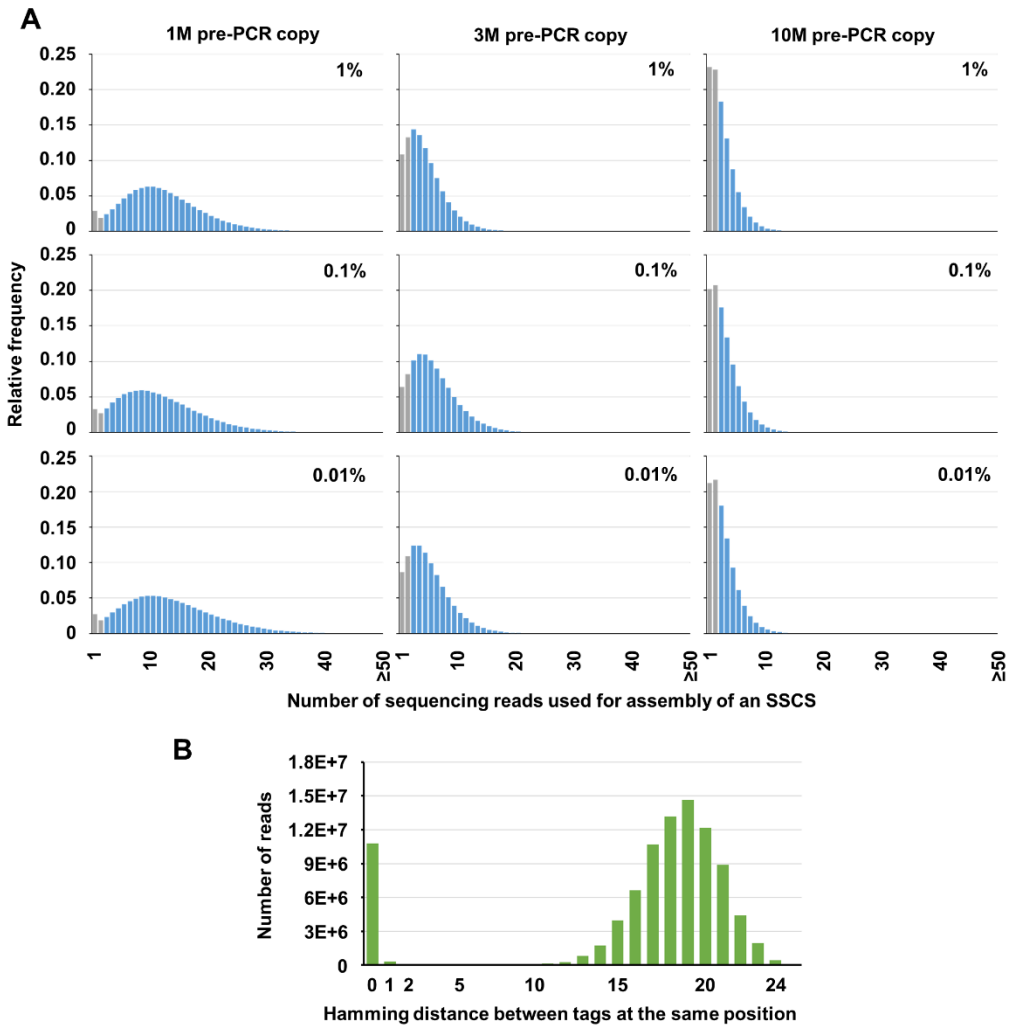

Fig. S2. Efficiency to assemble sequencing reads. (A) Sequencing reads used for SSCS assembly. Using nine EcoSeq libraries prepared from model DNA samples, distribution of the number of sequencing reads used to create an SSCS was visualized as a histogram. Three or more reads (blue bar) were needed to create an SSCS. The upper right number shows a mixing ratio of HPDE-4 to TK6 to prepare artificial 'mutation'. 1M pre-PCR copy number showed the highest efficiency to create SSCSs with a peak at 9-11 reads. SSCS, single-strand consensus sequence. (B) Hamming distance of a EcoSeq library. The Hamming distance of one of the model DNA samples between any pairs of UMIs showed a distribution at zero or ≥10. 4% of reads showing in-between values indicates the ratio of sequencing errors of UMIs, but is rather small and not used for DCS assembly.

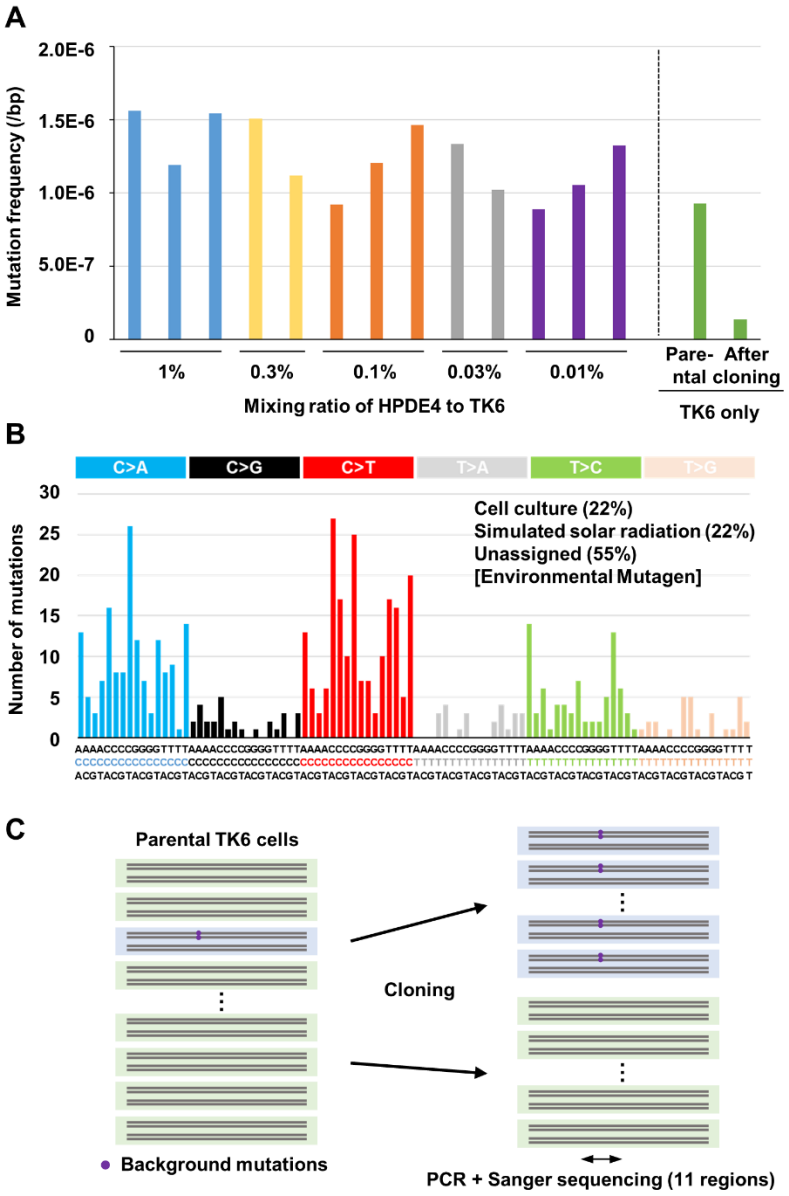

Fig. S3. Unexpectedly detected but real mutations present in background cells. (A) Frequencies of mutations located outside of SNP positions in 13 model DNA samples used in Fig. 2*E* (background mutations), and those in TK6 cells after cloning and in parental TK6 cells. The mutation frequency without cloning was $12.4\pm2.3\times10^{-7}$ per bp, in accordance with parental TK6 cells, but was higher than that in TK6 cells after cloning. (B) The mutational signatures of the background mutations. A mutational signature associated with cell culture was observed. (C) Schema of detection strategy of the background mutations present in parental TK6 cells. Using 30 independent TK6 clones, PCR and Sanger sequencing were performed for 11 background mutations. If the mutant cell population was 1%, 30 × 11 × 0.01 = 3.3 background mutations were expected to be detected.

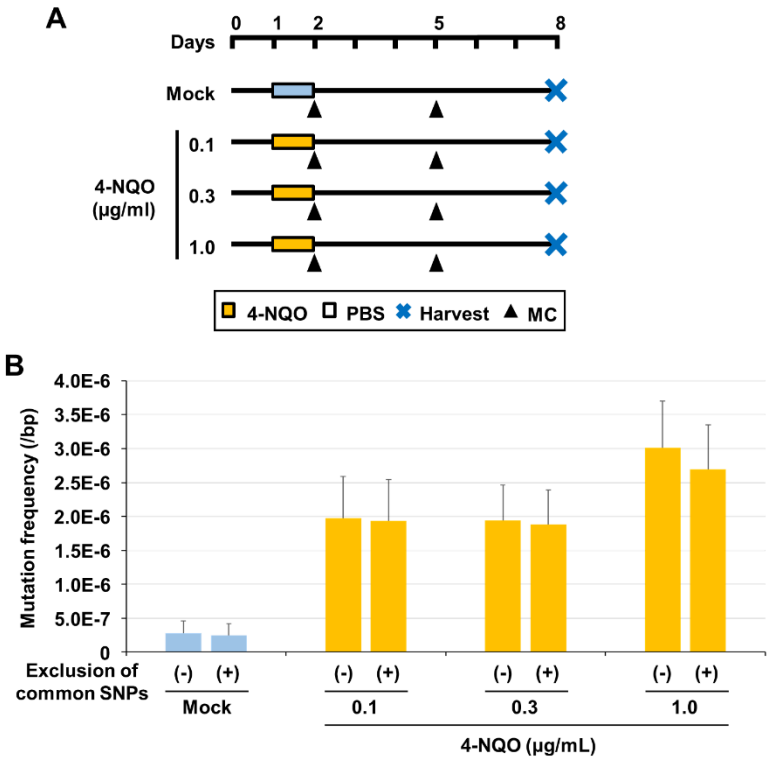

Fig. S4. Mutagen treatment and the effect of common SNPs exclusion in cloned cells. (A) Treatment protocol with 4-nitroquinoline 1-oxide (4-NQO). A 293FT clone was treated with 4-NQO for three doses (0.1, 0.3, and 1 μg/mL) for 24 hours on Day 1, and DNA was harvested on Day 8, securing sufficient time for DNA lesions to be fixed as mutations. MC, medium change. (B) Mutation frequencies before (re-displayed from Fig. 3*A*) and after exclusion of common SNPs based upon a SNP database. The exclusion of common SNPs had little effect on the mutation frequency. Error bar represents 95% confidence interval (CI).

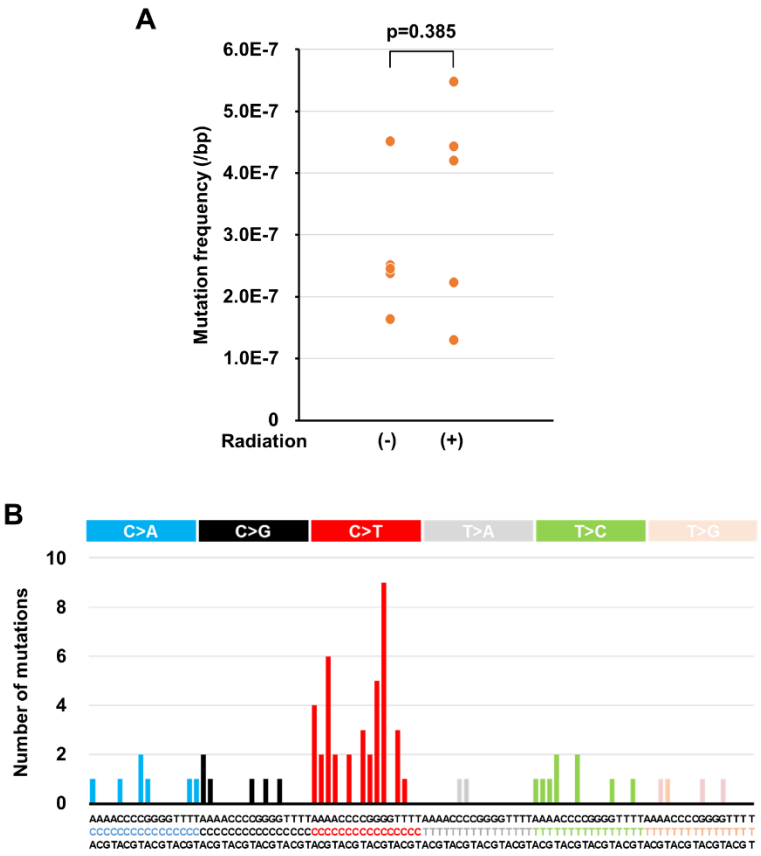

Fig. S5. Mutation accumulation of pediatric blood cells. (A) The effect of radiotherapy. Mutation frequencies of normal peripheral blood cells in pediatric sarcoma patients who received both chemotherapy and radiotherapy (n=5) and those who received only chemotherapy (n=5). Somatic mutations were not accumulated by radiation therapy (p=0.385). Error bar represents the standard deviation (SD). (B) A mutational spectrum of non-treated pediatric blood cells. A mutational profile of 16 peripheral blood cells before treatment are shown. The spectrum was similar to that of granulocytes from healthy donors reported in the previous study (8).

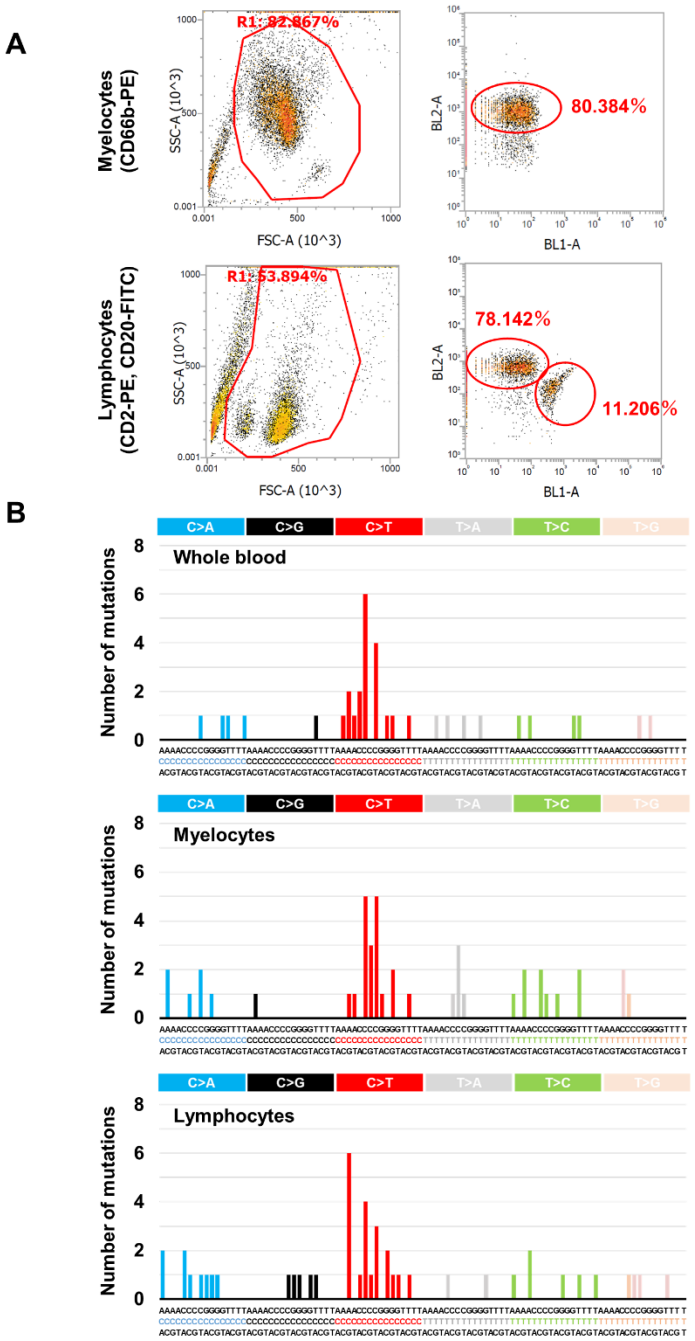

Fig. S6. Mutation accumulation in different cell types. (A) Cell isolation of myelocytes and lymphocytes of patient 3 (*SI Appendix*, Table S6). Flow cytometry confirmed enrichment of approximately 80.4% of CD66b-positive myelocytes from peripheral blood. Flow cytometry also showed that the contents of CD2-positive and CD20-positive lymphocytes of the isolated fraction were 78.1% and 11.2%, respectively. (B) Mutational profiles of peripheral mononuclear cells, myelocytes and lymphocytes of two pediatric patients after chemotherapy. Both mutational profiles of myelocytes and lymphocytes were similar to that of whole blood cells.

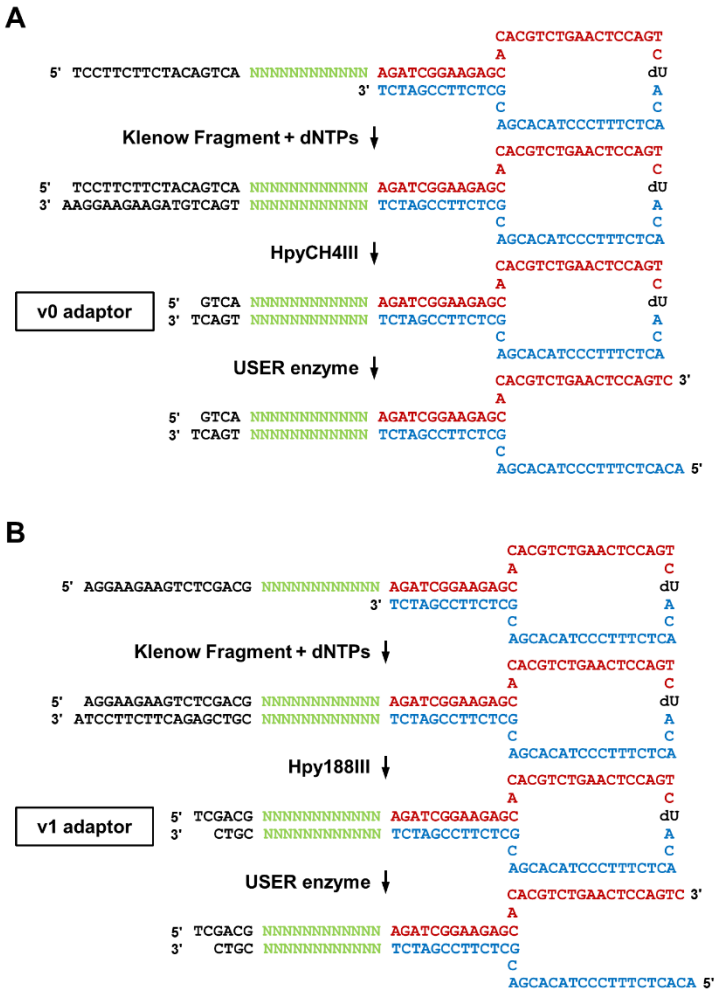

Fig. S7. Flow of Duplex-loop adaptor synthesis. (A) Flow of v0 adaptor synthesis. A 94-mer oligonucleotide was synthesized with one deoxyuridine (dU) base, 12-nt randomized bases as a unique molecular identifier (UMI), and 13-nt complementary 3'-end. Self-annealed oligonucleotide was converted to double-strand using Klenow fragment (3'->5' exo-), and the product was digested by HpyCH4III restriction enzyme. Synthesized Duplex-loop adaptor could be cut at a dU-base using uracil-specific excision reagent (USER) enzyme. (B) Flow of v1 adaptor synthesis. A 95-mer oligonucleotide was synthesized. After self-annealing and conversion to double-strand, the product was digested by HpyCH188III restriction enzyme. Synthesized adaptor could be cut at a dU-base using USER enzyme.

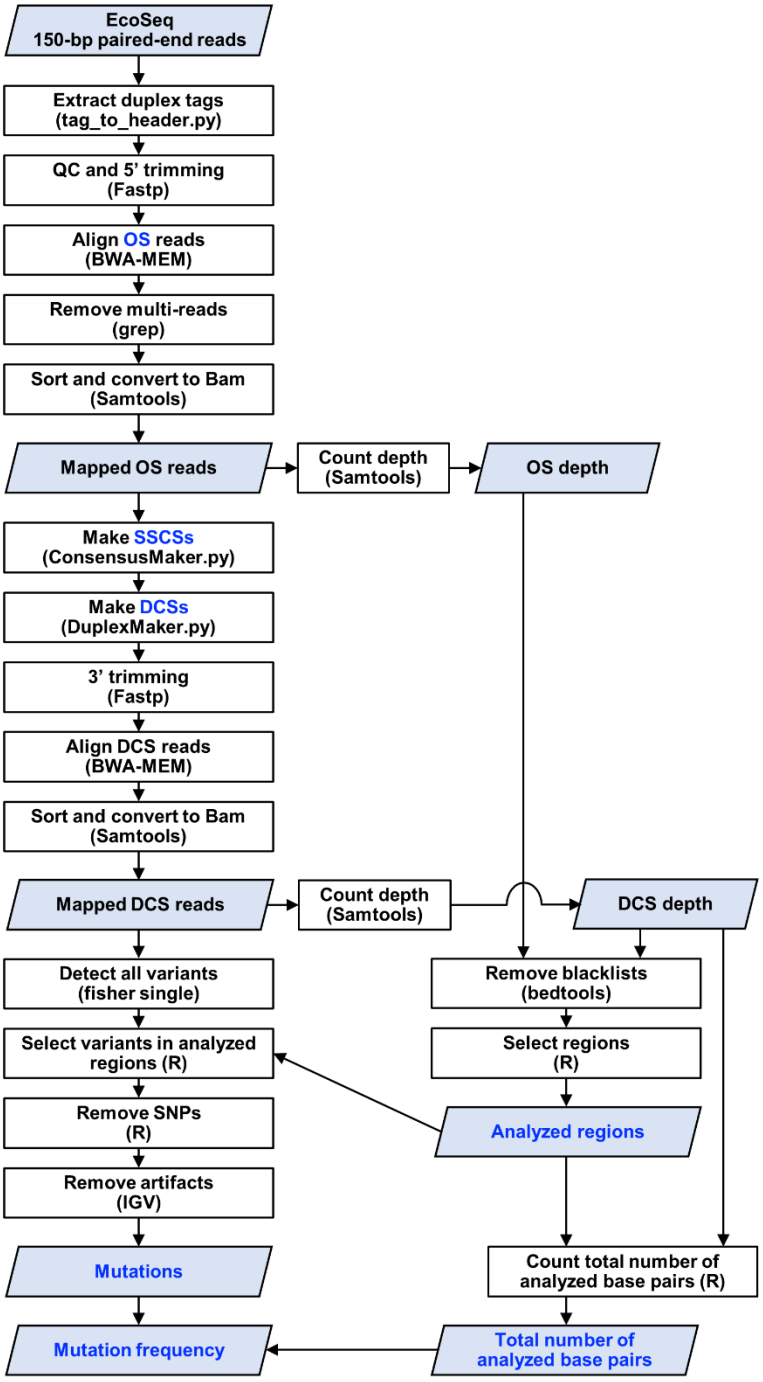

Fig. S8. Flow of data processing to create DCSs and detect mutations. Three or more mapped original sequences (OSs) containing identical UMI sequences were merged into a single-strand consensus sequence (SSCS). Two SSCSs from a single double-strand DNA were merged into a duplex consensus sequence (DCS). Single-strand DNA damage was excluded in the process of DCS assembly, and PCR and sequencing errors were excluded in the process of SSCS assembly. Mutations were detected by mapping DCSs to a reference genome. On the other hand, analyzed regions were obtained based upon the OS depth and DCS depth. Based upon the mutations and the total number of analyzed base pairs, a final mutation frequency was obtained.

Table S1. Mutation frequency in the model DNA samples

Sample	Mixing ratio of HPDE-4 to TK6	Pre-PCR copy number	Expected 'mutation' frequency (/bp)	Number of detected 'mutations'	Total number of analyzed base pairs	Measured 'mutation' frequency (/bp)	Background mutation frequency (/bp)
1%1M	1%	1M	2.82E-06	76	35,187,211	2.16E-06	1.56E-06
1%3M	1%	3M	2.82E-06	56	17,473,459	3.20E-06	1.55E-06
1%10M	1%	10M	2.82E-06	16	3,802,327	4.21E-06	7.89E-07
0.1%1M	0.1%	1M	2.82E-07	11	33,669,834	3.27E-07	9.21E-07
0.1%3M	0.1%	3M	2.82E-07	12	36,501,927	3.29E-07	7.67E-07
0.1%10M	0.1%	10M	2.82E-07	3	8,401,310	3.57E-07	1.07E-06
0.01%1M	0.01%	1M	2.82E-08	1	41,371,006	2.42E-08	9.67E-07
0.01%3M	0.01%	3M	2.82E-08	5	33,302,948	1.50E-07	9.31E-07
0.01%10M	0.01%	10M	2.82E-08	0	7,192,025	0	9.73E-07
1%1M_R1	1%	1M	2.82E-06	66	37,780,537	1.75E-06	1.19E-06
1%1M_R2	1%	1M	2.82E-06	93	23,958,119	3.88E-06	1.54E-06
0.3%1M_R1	0.3%	1M	8.45E-07	28	26,557,487	1.05E-06	1.51E-06
0.3%1M_R2	0.3%	1M	8.45E-07	49	33,046,209	1.48E-06	1.12E-06
0.1%1M_R1	0.1%	1M	2.82E-07	15	27,371,118	5.48E-07	1.21E-06
0.1%1M_R2	0.1%	1M	2.82E-07	10	40,335,514	2.48E-07	1.46E-06
0.03%1M_R1	0.03%	1M	8.45E-08	6	20,980,838	2.86E-07	1.33E-06
0.03%1M_R2	0.03%	1M	8.45E-08	9	32,283,485	2.79E-07	1.02E-06
0.01%1M_R1	0.01%	1M	2.82E-08	1	19,942,569	5.01E-08	1.05E-06
0.01%1M_R2	0.01%	1M	2.82E-08	1	33,261,862	3.01E-08	1.32E-06
Parental_TK6	0%	1M	NA	45	48,476,375	NA	9.28E-07
Cloned_TK6	0%	1M	NA	1	7,299,183	NA	1.37E-07

Table S2. The results of Sanger sequencing of the 11 background mutations in the 30 TK6 clones

	Background mutation				Contained in SNP DB (NCBI dbSNP Build 153)	Number of sequenced clones	Number of clones with a background mutation
	Chromosome	Position (GRCh38)	Reference base	Variant base			
Mutation 1	chr1	35,561,088	C	T	No	30	0
Mutation 2	chr2	65,830,394	C	A	No	30	0
Mutation 3	chr2	159,944,210	C	T	No	30	0
Mutation 4	chr3	140,074,862	C	T	No	30	1
Mutation 5	chr3	186,492,211	A	G	No	30	0
Mutation 6	chr4	54,213,520	T	C	No	30	0
Mutation 7	chr7	24,926,057	C	T	No	30	0
Mutation 8	chr8	131,342,233	C	A	No	30	0
Mutation 9	chr9	86,009,603	C	T	No	30	0
Mutation 10	chr11	61,535,574	G	A	No	30	0
Mutation 11	chr18	35,127,130	G	C	No	30	0

Table S3. Mutation frequency of 293FT cells treated with 4-NQO

Sample	Number of detected mutations	Number of detected mutations (after exclusion of common SNPs)	Total number of analyzed base pairs	Mutation frequency (/bp)	Mutation frequency (/bp) (after exclusion of common SNPs)
Mock	9	8	32,576,943	2.76E-07	2.47E-07
4NQO (0.1mg/mL)	39	38	19,762,968	1.97E-06	1.93E-06
4NQO (0.3 mg/mL)	54	52	27,790,265	1.94E-06	1.88E-06
4NQO (1.0 mg/mL)	73	65	24,226,885	3.01E-06	2.70E-06
4NQO (0.3 mg/mL) (v0AdRep1)	94	93	48,854,288	1.92E-06	1.91E-06
4NQO (0.3 mg/mL) (v0AdRep2)	43	42	32,416,404	1.33E-06	1.30E-06
4NQO (0.3 mg/mL) (v0AdRep3)	64	64	43,098,769	1.48E-06	1.49E-06
4NQO (0.3 mg/mL) (v1AdRep1)	152	150	101,175,509	1.50E-06	1.49E-06
4NQO (0.3 mg/mL) (v1AdRep2)	210	209	117,887,671	1.78E-06	1.78E-06
4NQO (0.3 mg/mL) (v1AdRep3)	226	224	130,237,004	1.74E-06	1.73E-06

Table S4. Patient characteristics of peripheral blood DNA samples

	Chemotherapy (n=10)	No chemotherapy (n=10)
Sex		
Male	5	5
Female	5	5
Age		
Median	12.0	12.3
Range	6-19	6-18
Diagnosis		
Rhabdomyosarcoma	3	4
Ewing sarcoma	2	2
Osteosarcoma	1	3
Chondrosarcoma	1	0
Round cell sarcoma	1	0
Paraganglioma	1	0
Glioblastoma	1	0
Liposarcoma	0	1
Treatment		
Platinum-based drug	6	0
Alkylating agent	10	0
Topoisomerase II inhibitor	10	0
Radiotherapy	5	0

Table S5. Mutation frequency of peripheral blood cells of pediatric patients treated with and without chemotherapy

Sample	Age	Chemotherapy			Radiotherapy	Number of detected mutations	Total number of analyzed base pairs	Mutation frequency (/bp)
		Platinum-based drug	Alkylating agent	Topoisomerase II inhibitor				
BloodAfterCtx01	6	+	+	+	+	11	26,135,296	4.21E-07
BloodAfterCtx02	8	+	+	+	+	4	30,706,510	1.30E-07
BloodAfterCtx03	8	+	+	+	+	14	31,560,615	4.44E-07
BloodAfterCtx04	16	-	+	+	+	9	40,232,622	2.24E-07
BloodAfterCtx05	18	+	+	+	+	16	29,201,321	5.48E-07
BloodAfterCtx06	6	+	+	+	-	6	23,871,935	2.51E-07
BloodAfterCtx07	8	-	+	+	-	7	29,378,895	2.38E-07
BloodAfterCtx08	15	-	+	+	-	9	36,651,354	2.46E-07
BloodAfterCtx09	16	-	+	+	-	6	36,537,189	1.64E-07
BloodAfterCtx10	19	+	+	+	-	13	28,774,883	4.52E-07
BloodBeforeCtx01	6	-	-	-	-	4	68,865,158	5.81E-08
BloodBeforeCtx02	7	-	-	-	-	2	43,630,027	4.58E-08
BloodBeforeCtx03	11	-	-	-	-	1	28,962,914	3.45E-08
BloodBeforeCtx04	11	-	-	-	-	4	39,839,366	1.00E-07
BloodBeforeCtx05	13	-	-	-	-	2	46,098,452	4.34E-08
BloodBeforeCtx06	13	-	-	-	-	4	29,975,106	1.33E-07
BloodBeforeCtx07	14	-	-	-	-	3	41,443,720	7.24E-08
BloodBeforeCtx08	15	-	-	-	-	5	34,695,859	1.44E-07
BloodBeforeCtx09	15	-	-	-	-	3	30,338,596	9.89E-08
BloodBeforeCtx10	18	-	-	-	-	8	46,165,464	1.73E-07

Table S6. Patient characteristics of peripheral blood DNA samples in multiple time points

Patient	Sex	Diagnosis	Age of 1st blood collection (before chemotherapy)	Chemotherapy (Pla/Alk/TopII)	Chemotherapy period (month)	Period of 1st-2nd blood collection (month)	Period of 1st-3rd blood collection (month)
Patient 1	M	Osteosarcoma	7	+/+/-	8.1	36.0	NA
Patient 2	F	Osteosarcoma	15	+/+/-	9.8	36.6	NA
Patient 3	M	Osteosarcoma	5	+/-/-	7.1	19.6	53.7
Patient 4	M	Osteosarcoma	12	+/+/-	10.3	42.1	75.2
Patient 5	F	Osteosarcoma	13	+/+/-	7.7	20.6	NA
Patient 6	M	Osteosarcoma	17	+/-/-	7.2	19.3	NA

Pla, platinum-based drugs; Alk, alkylating agents; TopII, topoisomerase II inhibitors.

Table S7. Mutation frequency of peripheral blood cells of pediatric patients in multiple time points

	Mutation frequency (/bp)				
	1st blood sampling	2nd blood sampling	3rd blood sampling		
			Whole blood	Myelocytes	Lymphocytes
Patient 1	1.33E-08	1.67E-07	NA	NA	NA
Patient 2	8.00E-08	3.06E-07	NA	NA	NA
Patient 3	5.17E-08	1.83E-07	1.25E-07	1.93E-07	1.62E-07
Patient 4	5.21E-08	2.87E-07	2.14E-07	2.36E-07	3.40E-07
Patient 5	1.22E-07	2.49E-07	NA	NA	NA
Patient 6	5.52E-08	3.03E-07	NA	NA	NA

Table S8. Sequence of digested fragments by restriction enzymes (*in silico* digestion)

Restriction enzyme	Sequence	Number of digested fragments per genome	Number of digested fragments per genome (100-700 bp)	Sequence of digested fragments (100-700 bp)			
				Expected reduction rate (%)	GC (%)	Repetitive sequence (%)	Coding region (%)
BmtI	GCTAGC	285,890	21,202	0.22	41.9	62.2	1.1
KpnI	GGTACC	288,002	15,716	0.20	45.4	53.7	2.5
BamHI	GGATCC	366,025	30,905	0.38	48.0	43.2	3.1
BauI	CACGAG	392,615	36,829	0.47	49.1	54.9	2.9
AhlI	ACTAGT	400,321	33,123	0.42	37.4	46.2	0.8
Eco32I	GATATC	446,502	47,322	0.61	36.9	64.8	0.7
MfeI	CAATTG	564,334	68,346	0.88	37.2	56.7	0.6
BglII	AGATCT	774,757	115,648	1.48	39.9	55.2	1.1
HindIII	AAGCTT	860,393	150,480	1.89	39.0	49.5	1.1
TaqI	TCGA	1,603,400	454,754	5.56	44.5	57.6	2.4
DraI	TTTAAA	3,423,962	1,529,283	17.21	35.8	45.2	0.8
Csp6I	GTAC	5,115,133	2,673,383	29.95	39.6	51.0	1.3
MboI	GATC	7,492,833	4,673,551	47.69	41.8	54.6	1.2
BshFI	GGCC	8,562,304	4,109,375	40.87	43.7	53.6	1.4
HpyCH4V	TGCA	14,824,420	8,224,716	71.80	40.8	52.1	1.2

Table S9. Sequences of oligonucleotides for adaptor and primers

	Sequence
Oligonucleotide for Duplex-loop adaptor	
v0 adaptor	5'-TCCTTCTTCTACAGTCANNNNNNNNNNNNAGATCGGAAGAGCACACGTCT GAACTCCAGTC/deoxyU/ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
v1 adaptor	5'-AGGAAGAAGTCTCGACGNNNNNNNNNNNNAGATCGGAAGAGCACACGTCT GAACTCCAGTC/deoxyU/ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Pre-PCR-specific primers	
Forward	5'-ACACTCTTTCCCTACACGACGCT-3'
Reverse	5'-GACTGGAGTTCAGACGTGTGCTC-3'
Post-PCR-specific primers	
Forward	5'-ATACGGCGACCACCGAGATC-3'
Reverse	5'-CAAGCAGAAGACGGCATACGAG-3'
Primers to detect 11 background mutations	
Forward (mutation 1)	5'-CTCTTTTTGCCCCCGACAAC-3'
Reverse (mutation 1)	5'-GCTTCGCTGTCTCCCCATTAT-3'
Forward (mutation 2)	5'-TCGTTCCAAATCGGTGTCCA-3'
Reverse (mutation 2)	5'-TTCAAACATCAGGGCGTGGT-3'
Forward (mutation 3)	5'-GCCGGTGATTCTCTCTTCCT-3'
Reverse (mutation 3)	5'-CCCTGTTGCAAGCATAATGTGT-3'
Forward (mutation 4)	5'-ATGCCTTATTGTCTGGGCCT-3'
Reverse (mutation 4)	5'-AAGCCTTGGCCTCTACTTCAG-3'
Forward (mutation 5)	5'-TGGGATCATCTCAGGGTGGT-3'
Reverse (mutation 5)	5'-CCTGGCTTACTGGTCTGCAA-3'
Forward (mutation 6)	5'-CCTGCCACCAGAAAAGTGGT-3'
Reverse (mutation 6)	5'-TCCCTTCAGCTTGGACGTTG-3'
Forward (mutation 7)	5'-TAGGGAGAGCAATGGGGGTG-3'
Reverse (mutation 7)	5'-GAGTGGCCATTCCAGGTTCC-3'
Forward (mutation 8)	5'-GGAGATGCAGCAGTTTTGTCA-3'
Reverse (mutation 8)	5'-ATGCCCAATTTTGCATGTGTGA-3'
Forward (mutation 9)	5'-TCCCATTTCACCAACTGCCT-3'
Reverse (mutation 9)	5'-ACCCGAACACAGTGAGTAACA-3'
Forward (mutation 10)	5'-TAGGGAGAGCAATGGGGGTG-3'
Reverse (mutation 10)	5'-GAGTGGCCATTCCAGGTTCC-3'
Forward (mutation 11)	5'-AGGGAGAAGGGACAATGGCA-3'
Reverse (mutation 11)	5'-TGGGCTACTGATTTTGTCTTCG-3'

Legends for Datasets

Dataset S1 (separate file). Samples and coverage of EcoSeq libraries. All samples and their coverage analyzed by EcoSeq are listed.

SI References

1. S. R. Kennedy et al., Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* **9**, 2586-2606 (2014).
2. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
3. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
4. J. T. Robinson et al., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26 (2011).
5. J. G. Tate et al., COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucl. Acids Res.* **47**, D941-D947 (2019).
6. J. E. Kucab et al., A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e816 (2019).
7. A. Degasperi et al., A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nature cancer* **1**, 249-263 (2020).
8. F. Abascal et al., Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405-410 (2021).