

A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens

Ahmad Besaratinia^{1,*}, Haiqing Li², Jae-In Yoon¹, Albert Zheng¹, Hanlin Gao³ and Stella Tommasi¹

¹Department of Cancer Biology, ²Department of Molecular Medicine and ³DNA Sequencing/Solexa Core Facility, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA

Received November 2, 2011; Revised May 30, 2012; Accepted May 31, 2012

ABSTRACT

Many carcinogens leave a unique mutational fingerprint in the human genome. These mutational fingerprints manifest as specific types of mutations often clustering at certain genomic loci in tumor genomes from carcinogen-exposed individuals. To develop a high-throughput method for detecting the mutational fingerprint of carcinogens, we have devised a cost-, time- and labor-effective strategy, in which the widely used transgenic Big Blue[®] mouse mutation detection assay is made compatible with the Roche/454 Genome Sequencer FLX Titanium next-generation sequencing technology. As proof of principle, we have used this novel method to establish the mutational fingerprints of three prominent carcinogens with varying mutagenic potencies, including sunlight ultraviolet radiation, 4-aminobiphenyl and secondhand smoke that are known to be strong, moderate and weak mutagens, respectively. For verification purposes, we have compared the mutational fingerprints of these carcinogens obtained by our newly developed method with those obtained by parallel analyses using the conventional low-throughput approach, that is, standard mutation detection assay followed by direct DNA sequencing using a capillary DNA sequencer. We demonstrate that this high-throughput next-generation sequencing-based method is highly specific and sensitive to detect the mutational fingerprints of the tested carcinogens. The method is reproducible, and its accuracy is comparable with that of the currently available low-throughput method. In conclusion, this novel method has the potential to move the field of carcinogenesis forward by allowing high-throughput

analysis of mutations induced by endogenous and/or exogenous genotoxic agents.

INTRODUCTION

The human cancer genome is shaped by assaults from endogenous and exogenous mutagens (1). Many carcinogens are mutagens or turn into mutagenic derivatives through biotransformation. Of these, some are known to leave a unique mutational fingerprint in the human genome (2,3). These mutational fingerprints manifest as specific types of mutations (e.g. induced base substitution/deletion/insertion), often clustering at certain nucleotide positions in cancer-related loci, in tumor genomes from carcinogen-exposed individuals (4,5). Establishing the mutational fingerprint of carcinogens is important because (i) from a mechanistic point of view, it can help infer human cancer etiology; and (ii) from a standpoint of public health, it can help reinforce hazard removal/reduction strategies for perilous environmental agents. Until recently, the mutational fingerprint of carcinogens has only been investigated in a few cancer-related genes or housekeeping genes (4). With the advent of next-generation sequencing technologies, however, a comprehensive mutational fingerprint of carcinogens can now be determined on a genome-wide scale (6,7). These breakthrough technologies are poised to survey the landscape of human cancer genome and reveal mutational fingerprints, which may be ascribed to environmental carcinogens (8). However, to verify causality, the identified mutational fingerprints need to be experimentally recapitulated in validated model systems and under strictly controlled exposure conditions (4,9).

Transgenic rodents are extensively validated model systems for establishing the mutational fingerprint of carcinogens (10). However, the mutation detection assays incorporated into these transgenic systems are only

*To whom correspondence should be addressed. Tel: +1 626 359 8111 (Ext 65918); Fax: +1 626 358 7703; Email: ania@coh.org

amenable to conventional DNA sequencing analysis (4). This feature is prohibitive because it only allows 'low-throughput' detection of mutational fingerprint using direct DNA sequencing of phenotypically expressed individual mutants. Such an approach is relatively costly, extensively time consuming and extremely laborious (4). To develop a 'high-throughput' method for detecting the mutational fingerprint of carcinogens, we have devised a cost-, time- and labor-effective strategy, in which the widely used transgenic Big Blue[®] mouse mutation detection assay (Stratagene, La Jolla, CA) is made compatible with the Roche/454 Genome Sequencer FLX Titanium next-generation sequencing technology (454 Life Sciences, Branford, CT). As proof of principle, we have used this novel method to establish the mutational fingerprints of three carcinogens with varying mutagenic potencies, including sunlight ultraviolet radiation, 4-aminobiphenyl (4-ABP) and secondhand smoke (SHS) that are known to be strong, moderate and weak mutagens, respectively (11–13). For verification purposes, we have compared the mutational fingerprints of these carcinogens obtained by our newly developed method with those obtained by parallel analyses using the conventional low-throughput approach, that is, standard mutation detection assay followed by direct DNA sequencing using the capillary ABI-3730 DNA Analyzer (ABI Prism, PE Applied BioSystems, Foster City, CA). We have also performed similar analyses to establish the spontaneous mutation spectra in control (sham-treated) samples using both the new next-generation sequencing-based method and the conventional DNA sequencing.

MATERIALS AND METHODS

Selection of carcinogens and experimental treatments

To demonstrate the sensitivity and specificity of our method for detecting the mutational fingerprint of carcinogens, we chose three distinct agents with high, moderate and low mutagenic potencies, respectively. As an extensively studied environmental physical carcinogen, the ultraviolet B (UVB) fraction of sunlight (λ : 280–320 nm) is implicated in the etiology of human skin cancer and proven to be a highly potent mutagen (14–16). The aromatic amine 4-ABP is a widespread environmental contaminant, which is present in various occupational settings and tobacco smoke and considered an etiologic agent in human bladder cancer (17,18). 4-ABP is known to be moderately mutagenic (13). SHS is an environmental pollutant, which is etiologically implicated in human lung cancer, and possesses relatively weak mutagenic potency (12,19).

Detailed information on the experimental treatment of Big Blue[®] mouse embryonic fibroblasts with UVB and the chronic exposure of Big Blue[®] mice to 4-ABP or SHS are provided in our previously published reports (11,13,20). Briefly, the UVB irradiation of Big Blue[®] mouse embryonic fibroblasts was performed *in vitro* at a single biologically relevant dose of 75.6 mJ/cm², and under physiologic conditions (11). *In vivo*, 4-ABP (Sigma-Aldrich Inc., Saint Louis, MO) was administered intraperitoneally to male

adult Big Blue[®] mice on a weekly basis for a duration of 6 weeks at increasing doses of 25–100 mg/kg bw (13). The SHS treatment of male adult Big Blue[®] mice was performed *in vivo* in exposure chambers of a TE-10 smoking machine (Teague Enterprises, Davis, CA) for 5 hr/day, 5 days/week for a duration of 4 months (20). Subsequent to all experimental treatments, genomic DNA was isolated using a standard phenol extraction-based protocol (21). The DNA was dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and kept at –80°C until further analysis.

Modification of the Big Blue[®] mouse mutation detection assay for compatibility with a next-generation sequencing platform

Transgenic Big Blue[®] rodent system is an extensively validated model for studying spontaneous or experimentally induced mutagenesis (4). The genome of these transgenic animals contains multiple copies of a chromosomally integrated λ LIZ shuttle vector, which carries two bacterial reporter genes, including the *cII* and *lacI* (10). To investigate the experimental induction of mutagenesis, transgenic rodents or cell cultures derived from their organs of interest are treated with a test agent *in vivo* or *in vitro*, respectively. Following a latency period needed for the expression of mutations, genomic DNA is isolated, and the λ LIZ shuttle vectors are recovered. The recovered vectors are then used in a bacterial phenotypic expression assay to identify mutants, that is, cells harboring mutations in the reporter gene(s) (10). To find the type and distribution of induced mutations in the *cII* or *lacI* genes, which reflect the mutational fingerprint of the tested agent, each phenotypically expressed mutant needs to be isolated individually and subjected to direct DNA sequencing. This is a time-, cost- and labor-intensive process, and as such, precludes 'high-throughput' generation of mutational fingerprints (4). To address this issue, we have devised a novel strategy, in which a pool of phenotypically expressed mutants, in lieu of single mutants, can be sequenced using a next-generation sequencing platform.

As the preparatory step, we performed the *cII* mutagenesis assay on the genomic DNA of carcinogen-treated cells/mice and control (sham-treated) to phenotypically express the induced and spontaneously derived *cII* mutants, respectively. The assay was performed using the commercially available Transpack Packaging Extract kit (Stratagene) according to the instructions of the manufacturer. Following the expression assay, 150 *cII* mutant plaques obtained from the analysis of genomic DNA from each of the experimental or control group were cored individually and placed in a microtube containing 500 μ l double-distilled water. We note that the pool of 150 mutants per sample is comparable with the number of mutants sequenced individually by the conventional low-throughput method for establishing the mutational fingerprint of carcinogens (4). For verification of reproducibility, two or more independent pools of 150 mutants were prepared from each experimental or control group simultaneously. The microtubes containing pools of 150 mutant *cII* plaques were boiled for 5 min, and

subsequently, centrifuged at 18 000g for 5 min. Ten microliters of the supernatant were immediately transferred to a new microtube containing 40 μ l of a polymerase chain reaction (PCR) mastermix in which the final concentrations of the reagents were 1 \times PCR buffer, 1 \times Q solution, 200 nM each of the forward and reverse primers, 50 μ M dNTP and 2.5^U Taq DNA Polymerase (Qiagen, Valencia, CA). The oligonucleotide primers were custom designed to contain the forward and reverse *cII* sequences (needed for amplification of the entire *cII* gene and its flanking regions) together with tagged linker sequences (required for downstream application of the Genome Sequencer FLX Titanium next-generation sequencing). The forward and reverse primers were 5'-cgtatcgccctcctcgccatcagcccttaccattccagc (t_m : 72.6°C) and 5'-ctatgctgcttccagccgctcagccttgcgaagtgtgagtat (t_m : 72.8°C), respectively. The thermocycling conditions were as follows: denaturation at 95°C for 3 min, 10 cycles of amplification consisted of 45 s at 95°C, 1 min at 60°C, and 1 min at 72°C, 25 cycles of re-amplification consisted of 45 s at 95°C and 2 min at 73°C and finally 7 min of extension at 73°C. The PCR amplified product (526 bp) was purified using the QIAquick PCR purification kit (Qiagen) and kept at -80°C until further analysis.

Genome sequencer FLX titanium next-generation sequencing and bioinformatics data processing and analysis

Ultradeep pyrosequencing was performed using a 454 GS FLX (454 Life Sciences). The amplified PCR products encompassing the entire *cII* gene and its flanking regions plus the 454 (A) and (B) linkers (454 Life Sciences) were further purified by the MinElute PCR purification kit (Qiagen). The resultant was clonally amplified on capture beads in water-in-oil emulsion microreactors (454 Life Sciences). The enriched-DNA beads were deposited onto the wells of a full Roche 454 FLX Titanium PicoTiter Plate device and pyrosequenced in both forward and reverse directions. The 200-nucleotide cycles were carried out in a 10-hr sequencing run, according to the manufacturer's instructions (454 Life Sciences). Sequence reads were generated in FASTQ format using the Data Processing Pipeline (v2.3) of the GS FLX System software (454 Life Sciences). The sequence reads were filtered by read length (within 350–550 bp limit). The filtered sequence reads were aligned to the reference sequence using CLCBIO Genomic Workbench's (v4.5) long read alignment tool. The variations of each sample were detected using CLCBIO Genomics Workbench's (v4.5) SNP/DIP analysis tools, which is based on the Neighborhood Quality Standard algorithm (22). Based on the total number of mutants per sample, minimum variation frequency and read number were set as thresholds to detect base substitutions and insertions and deletions (Indel). Because each sample contained a pool of 150 mutants, we used 0.66% (1 of 150) as minimum threshold for detecting base substitutions or Indels. The minimum variation frequency was also used as benchmark to calculate the total number of each specific type of mutation in

each sample. For example, a 6.6% of C→T base substitution in an individual sample was counted as 10 mutants that have this type of mutation in that sample. The variation distribution was calculated based on the total number of mutated amplicons in each sample. The variation spectrum of each sample was plotted on the reference sequence with a heatmap to improve the visualization of variations. To minimize homopolymer sequencing errors, which may cause assembly errors, false variations (23) or reduced quality score of the reads (24), we implemented a filtration step that uses the 'high-quality' read coverage threshold and the variation status. The filter initially scans the reference sequence and identifies homopolymer region (length: greater than or equal to 3). Subsequently, it counts the coverage of these 'high-quality' reads (Phred quality: greater than or equal to 30) relative to the variation in homopolymer region and filters out the variations with low 'high-quality' read coverage. The filter then compares the variation within the homopolymer and at its neighboring nucleotide. If the variation within the homopolymer is the same as that at its adjacent nucleotide, the filter drops this variation and considers it to have arisen from a homopolymer sequencing error and/or assembly error. Supplementary Figure S1 shows an example of this type of false variation caused by a homopolymer error. In this example, there are two variations, including A→C and C→A. The A→C variation within the homopolymer region is followed by the C→A variation at its neighboring base.

Statistical and bioinformatics analyses

The results are expressed as mean \pm 95% confidence interval. Comparison of mutant frequency data between an experimental group and its corresponding control group was made using the Wilcoxon rank-sum test. To determine the reproducibility of mutation spectra established in duplicate/multiplicate samples analyzed by the next-generation sequencing-based method, we performed both the hierarchical clustering analysis and the principle component analysis (PCA) using the Partek Genomics Suite v6.11.1116 (<http://www.partek.com>). To further analyze the comparability of mutation spectra established in duplicate/multiplicate samples analyzed by the next-generation sequencing-based method, we performed correlation analysis to calculate the similarities between the frequency and position of each mutation detected in the respective samples. In addition, we used correlation analysis to compare the mutation spectra of each set of two matching samples established by the next-generation sequencing-based method and the conventional DNA sequencing, respectively. We note that this correlation analysis takes into account the similarities between the frequency and position of each mutation detected by the respective methods in two counterpart samples. The applied correlation analysis uses stringent criteria to compare two mutation spectra with respect to both the frequency and type of each specific mutation occurred in the entire length of the *cII* gene. This comparative analysis takes into consideration the frequency and type of mutations in the *cII* gene (as a whole) but not at certain

nucleotide positions only. So, the overall mutation frequency and pattern across the full length of the *cII* sequence are compared when the above correlation analysis is performed on two mutation spectra. All statistical tests were two sided. Values of $P < 0.05$ were considered statistically significant. The S-Plus 7.0 for Windows software (Insightful Corp., Seattle, WA) was used for statistical analysis.

RESULTS

Mutant frequency and mutation spectrum

As a potent mutagen implicated in skin carcinogenesis (11), sunlight UVB caused significant mutagenicity in Big Blue[®] mouse embryonic fibroblasts irradiated *in vitro* with this environmental carcinogen. The strong mutagenicity of UVB was demonstrated by a 92.6-fold increase in background *cII* mutant frequency from $3.01 \pm 0.68 \times 10^{-5}$ in control (non-irradiated) cells to $278.92 \pm 20.63 \times 10^{-5}$ in UVB-irradiated cells ($P = 0.0002$). *In vivo* treatment of Big Blue[®] mice with 4-ABP, a known bladder carcinogen with moderate mutagenic potency (13), resulted in a 9.9-fold increase in background *cII* mutant frequency from $2.09 \pm 0.20 \times 10^{-5}$ in bladder DNA of control (solvent-treated) mice to $20.62 \pm 4.77 \times 10^{-5}$ in bladder DNA of 4-ABP-treated mice ($P = 0.0079$). *In vivo* exposure of Big Blue[®] mice to SHS, a known pulmonary carcinogen with comparatively weak mutagenic potency (12), resulted in a 2.1-fold increase in background *cII* mutant frequency from $2.00 \pm 0.29 \times 10^{-5}$ in lung DNA of control (clean-air-treated) mice to $4.09 \pm 0.79 \times 10^{-5}$ in lung DNA of SHS-treated mice ($P = 0.0011$).

To determine what specific type(s) of mutation have caused the significant increase in *cII* mutant frequency in carcinogen-treated cells/mice relative to control, we computed the absolute mutant frequency of each type of mutation in the *cII* gene (i.e. transitions, transversions, deletions and insertions) in the genome of carcinogen-treated cells/mice and control. As shown in Figure 1A, the absolute mutant frequencies of G:C→C:G transversions, G:C→T:A transversions, G:C→A:T transversions, A:T→T:A transversions, A:T→G:C transitions, A:T→C:G transversions and insertions/deletions were all increased, although to different extents, in the *cII* gene in genomic DNA of UVB-irradiated cells relative to control. The percentage contributions of the respective types of mutation to the overall increase in *cII* mutant frequency in UVB-irradiated cells were 1.0, 0.2, 87.0, 3.1, 0.6, 1.4 and 6.7 (Figure 1B). More specifically, mutations occurring at dipyrimidine sites account for nearly all the induced *cII* mutations in UVB-irradiated cells. Of these, G:C→A:T transition mutations, which comprise the majority of all the induced *cII* mutations (87.0%), are the main contributor to the overall increase in *cII* mutant frequency in UVB-irradiated cells (Figure 1A and B and Supplementary Table S1).

As illustrated in Figure 2A and B, the percentage contributions of G:C→C:G transversions, G:C→T:A transversions, G:C→A:T transitions, A:T→T:A

transversions, A:T→G:C transitions, A:T→C:G transversions and insertions/deletions to the overall increase in *cII* mutant frequency in bladder DNA of 4-ABP-treated mice were 15.3, 40.0, 20.4, 5.7, 8.7, 1.8 and 8.1, respectively. Specifically, mutations occurring at G:C basepairs account for 81.2% of all the induced *cII* mutations in bladder DNA of 4-ABP-treated mice. Of these, G:C→T:A transversion mutations, which constitute 40% of all the induced *cII* mutations, dominate the overall increase in *cII* mutant frequency in bladder DNA of 4-ABP-treated mice (Figure 2A and B and Supplementary Table S1). As shown in Figure 3A and B, the percentage contributions of G:C→C:G transversions, G:C→T:A transversions, G:C→A:T transitions, A:T→T:A transversions, A:T→G:C transitions and A:T→C:G transversions to the overall increase in *cII* mutant frequency in lung DNA of SHS-exposed mice were 6.3, 15.7, 49.2, 7.4, 15.9 and 10.1, respectively. Thus, G:C→A:T transition mutations account for nearly half of all the induced *cII* mutations in the lung DNA of SHS-exposed mice (Figure 3A and B and Supplementary Table S1).

We then mapped the locations of induced mutations in the *cII* gene in the genome of carcinogen-treated cells/mice by plotting the induced mutations versus control (spontaneously derived) mutations along the reference *cII* sequence. As shown in Figure 1C and D, the UVB-induced mutations occurred at specific nucleotide positions in *cII* gene, which were distinct from those loci at which spontaneous mutations occurred in control. These UVB-specific mutations clustered at several nucleotide positions, predominantly within dipyrimidine-sequence contexts, and were almost exclusively G:C→A:T transitions (Figure 1C and Supplementary Figure S2). The overall spectrum of induced mutations in the *cII* gene of UVB-irradiated cells is comparable with that previously found in the same model system using the conventional low-throughput method ($P < 0.0001$, Figure 1A–C, Supplementary Figure S2 and Supplementary Table S1).

Mapping of the induced *cII* mutations in the genome of 4-ABP-treated mice showed that the majority of mutations were located at G:C basepairs (Supplementary Figure S3). Of these, G:C→T:A transversions clustering at several codon positions in the *cII* gene were specific for 4-ABP treatment (Supplementary Figure S3). This spectrum of induced mutations in the *cII* gene of 4-ABP-treated mice is also comparable with that previously found in the same model system using the conventional low-throughput method ($P = 0.004$, Figure 2A–C, Supplementary Figure S3 and Supplementary Table S1).

Furthermore, mapping of the induced *cII* mutations in the genome of SHS-treated mice revealed that most mutations were localized to G:C basepairs (74.9%, Supplementary Figure S4). There were subtle differences between the locations of SHS-induced mutations and the spontaneously derived mutations in the *cII* gene in lung DNA from SHS-treated mice and control, respectively. In addition, the frequencies of mutation at certain loci along the *cII* gene in SHS-treated mice were slightly different from those in control. These subtle differences in the type and location of SHS-induced and control *cII* mutations concur with the weak mutagenicity of SHS and

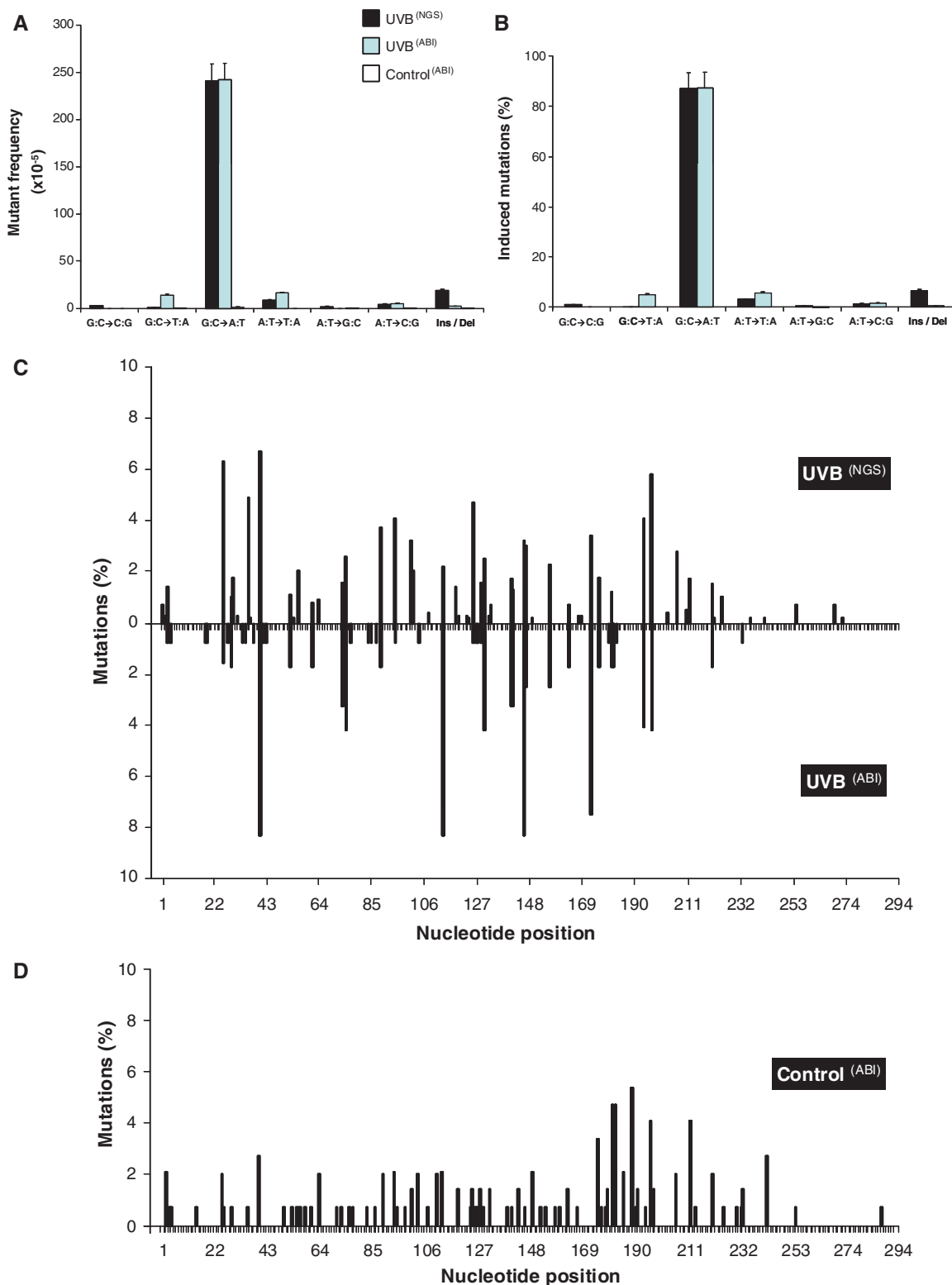


Figure 1. *cII* mutant frequency and mutation spectrum in UVB-irradiated cells versus control. Mutation analysis of the *cII* gene in mouse embryonic fibroblasts irradiated with UVB or control was performed using the *cII* mutagenesis assay, as described in 'Materials and Methods'. **(A)** Absolute mutant frequency of each specific type of mutation in the *cII* gene of UVB-irradiated cells or control as determined by both the new method (NGS) and the conventional method (ABI). Average results (bars) from multiple analyses plus 95% confidence interval (error bars) are shown. **(B)** Percentage increase in frequency of each specific type of mutation in the *cII* gene of UVB-irradiated cells or control as determined by both the new method (NGS) and the conventional method (ABI). Results are expressed as 'induced mutation (%)', which is calculated as [(mutant frequency of each type of mutation in UVB group – mutant frequency of the respective type of mutation in control group × 100)/(overall induced mutant frequency in UVB group – overall spontaneous mutant frequency in control group)]. Distribution of mutations in the *cII* gene of UVB-irradiated cells **(C)** and control **(D)** as determined by the new method (NGS) and/or the conventional method (ABI).

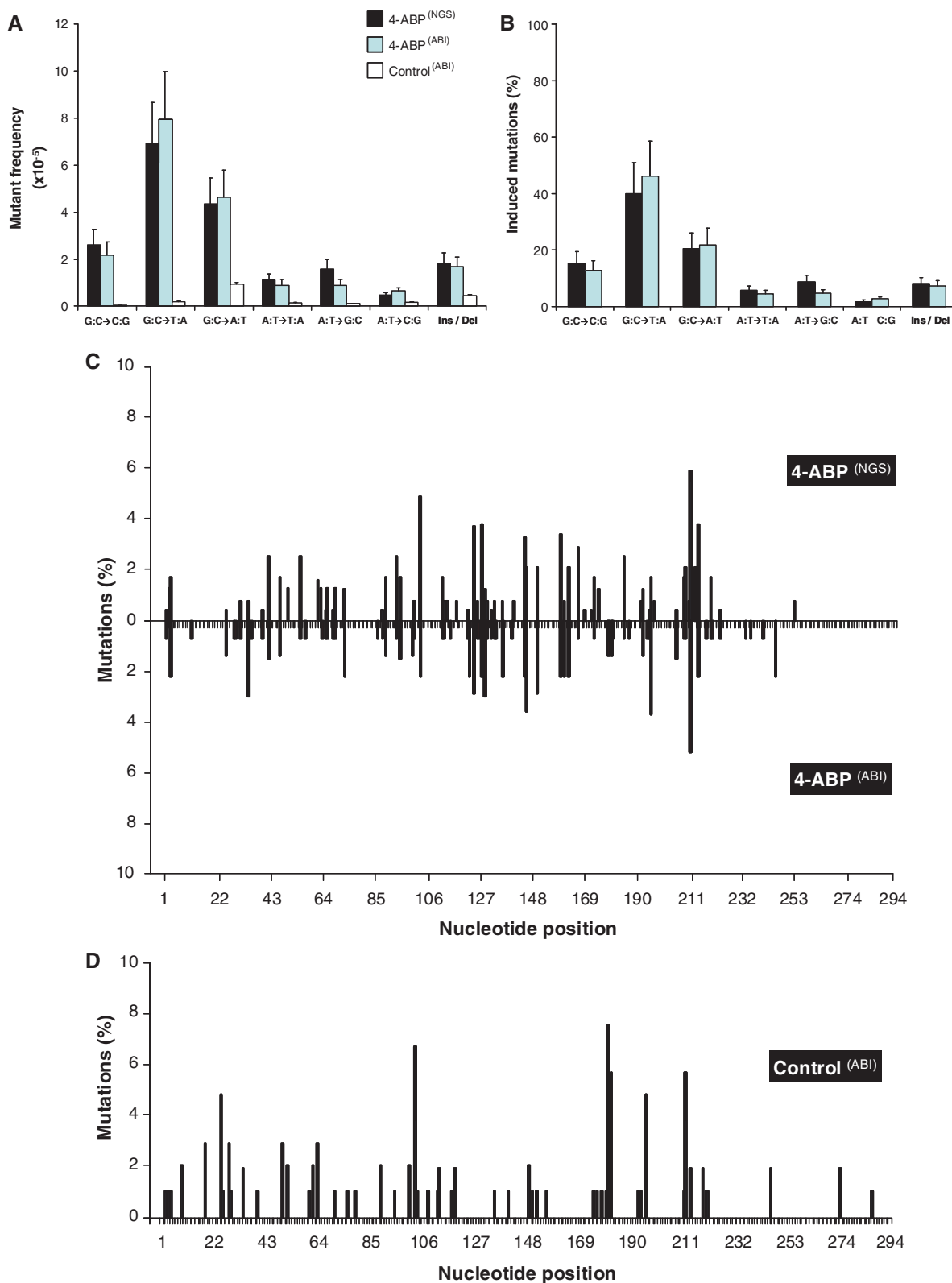


Figure 2. *cII* mutant frequency and mutation spectrum in 4-ABP-treated mice versus control. Mutation analysis of the *cII* gene in bladder DNA of mice treated with 4-ABP or control was performed using the *cII* mutagenesis assay, as described in ‘Materials and Methods’. **(A)** Absolute mutant frequency of each specific type of mutation in the *cII* gene of 4-ABP-treated mice or control as determined by both the new method (NGS) and the conventional method (ABI). **(B)** Percentage increase in frequency of each specific type of mutation in the *cII* gene of 4-ABP-treated mice or control as determined by both the new method (NGS) and the conventional method (ABI). Results are expressed as ‘induced mutation (%)’, which is calculated as [(mutant frequency of each type of mutation in 4-ABP group – mutant frequency of the respective type of mutation in control group $\times 100$)/(overall induced mutant frequency in 4-ABP group – overall spontaneous mutant frequency in control group)]. Distribution of mutations in the *cII* gene of 4-ABP-treated mice **(C)** and control **(D)** as determined by the new method (NGS) and/or the conventional method (ABI).

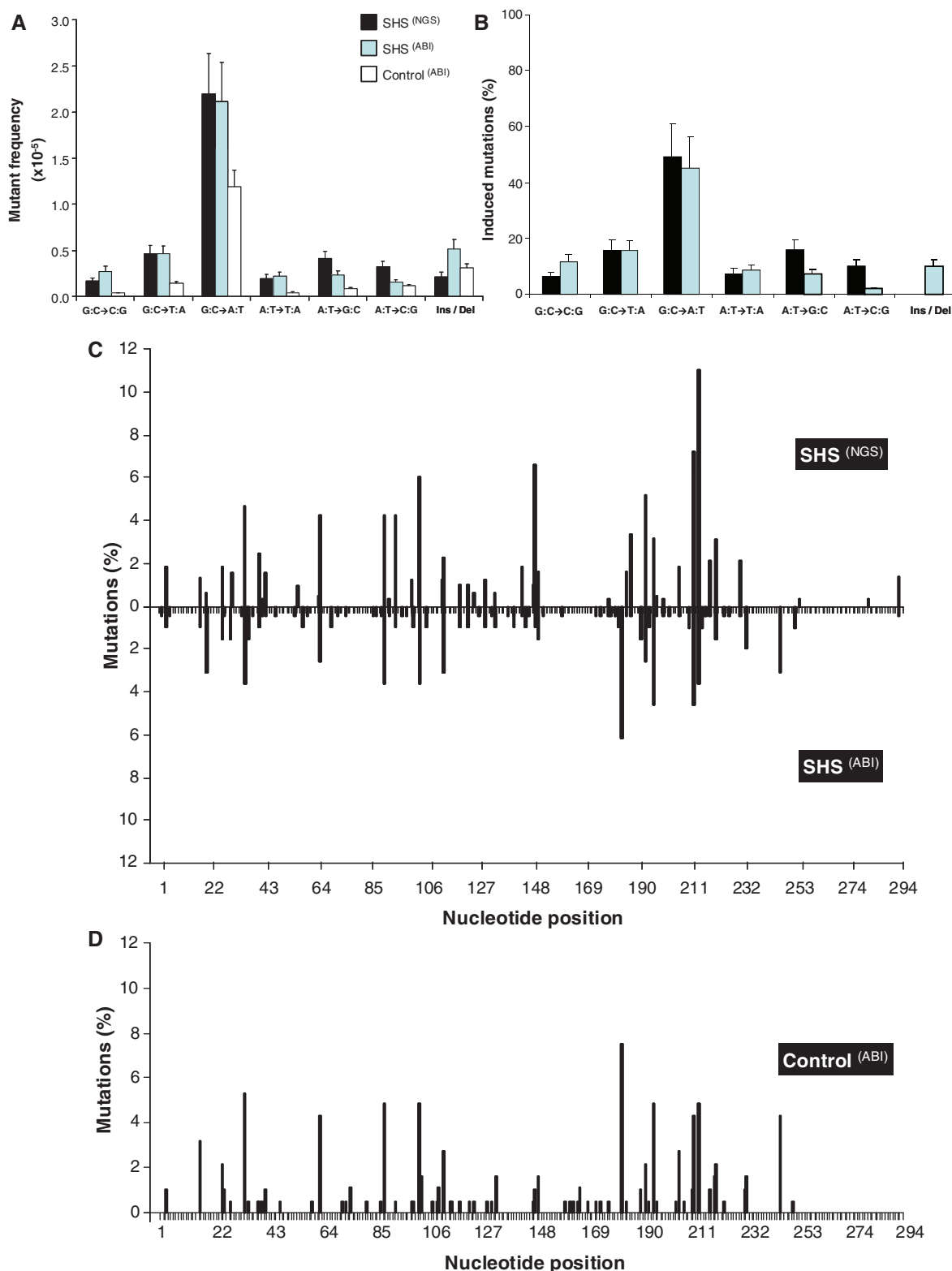


Figure 3. *cII* mutant frequency and mutation spectrum in SHS-treated mice versus control. Mutation analysis of the *cII* gene in lung DNA of mice exposed to SHS or control was performed using the *cII* mutagenesis assay, as described in 'Materials and Methods'. **(A)** Absolute mutant frequency of each specific type of mutation in the *cII* gene of SHS-exposed mice or control as determined by both the new method (NGS) and the conventional method (ABI). **(B)** Percentage increase in frequency of each specific type of mutation in the *cII* gene of SHS-exposed mice or control as determined by both the new method (NGS) and the conventional method (ABI). Results are expressed as 'induced mutation (%)', which is calculated as [(mutant frequency of each type of mutation in SHS group – mutant frequency of the respective type of mutation in control group × 100)/(overall induced mutant frequency in SHS group – overall spontaneous mutant frequency in control group)]. Distribution of mutations in the *cII* gene of SHS-exposed mice **(C)** and control **(D)** as determined by the new method (NGS) and/or the conventional method (ABI).

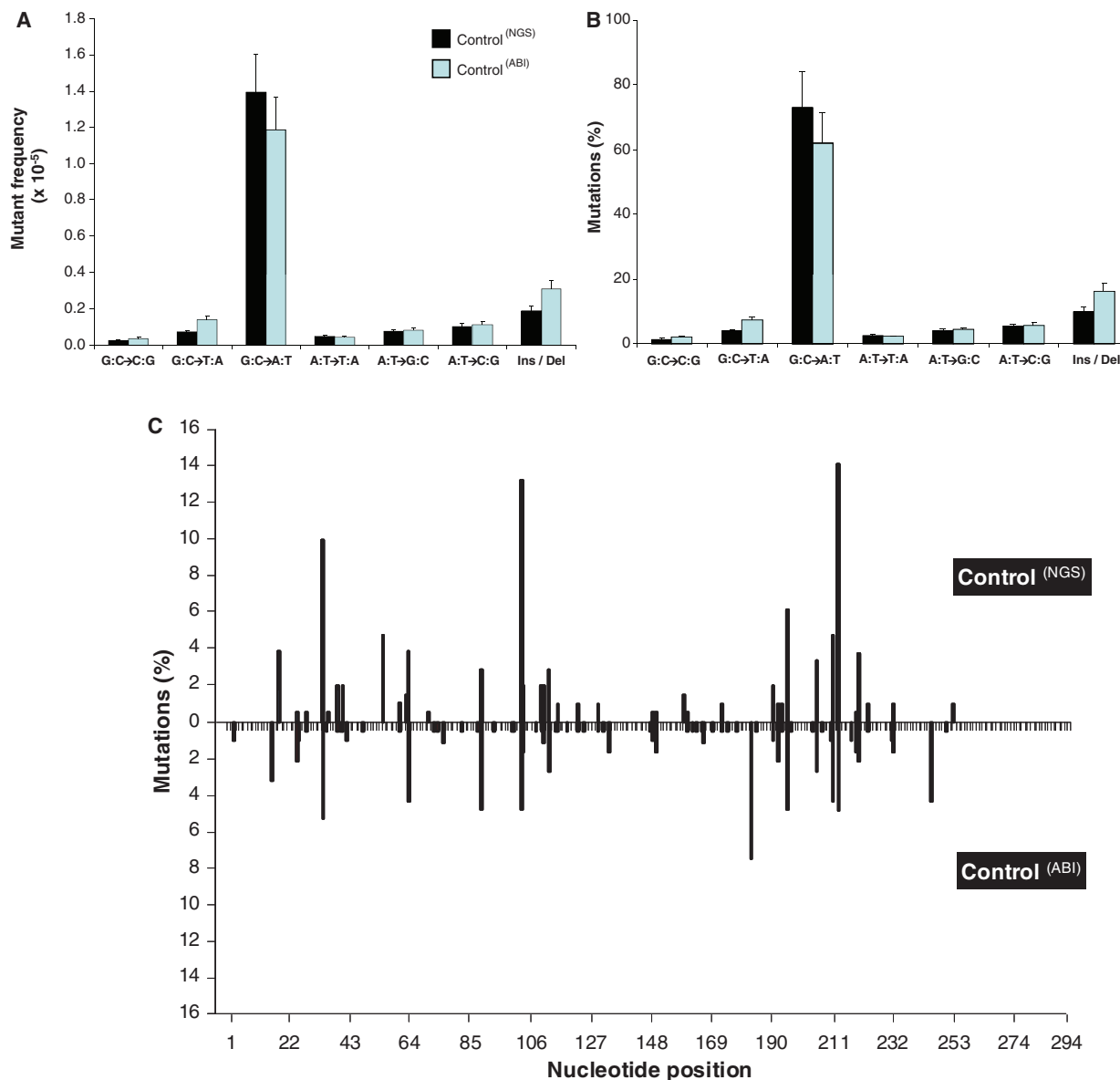


Figure 4. Spontaneous *cII* mutation spectrum in control (sham-treated) mice. Mutation analysis of the *cII* gene in lung DNA of control mice (clean air exposed) was performed using the *cII* mutagenesis assay, as described in 'Materials and Methods'. **(A)** Absolute mutant frequency of each specific type of mutation in the *cII* gene of control mice as determined by both the new method (NGS) and the conventional method (ABI). **(B)** Percentage of each specific type of mutation in the *cII* gene of control mice as determined by both the new method (NGS) and the conventional method (ABI). **(C)** Distribution of mutations in the *cII* gene of control mice as determined by both the new method (NGS) and the conventional method (ABI).

are consistent with the results found previously in the same model system using the conventional low-throughput method ($P < 0.0001$, Figure 3A–C, Supplementary Figure S4 and Supplementary Table S1).

Finally, we established the spontaneous mutation spectrum in control (sham-treated) samples using our next-generation sequencing-based method. As shown in Figure 4A and B and Supplementary Table S1, the percentages of G:C→C:G transversions, G:C→T:A transversions, G:C→A:T transitions, A:T→T:A transversions, A:T→G:C transitions, A:T→C:G transversions and insertions/deletions in the *cII* gene of control genomic DNA were 1.4, 3.8, 72.9, 2.5, 4.0, 5.5 and 9.9, respectively. Of these, the majority occurred at 5'-CpG

dinucleotides, with G:C→A:T transitions being the predominant type of mutations (i.e. over 90% of all mutations occurring at CpG-containing sequences were G:C→A:T transitions) (Supplementary Figure S5). This spectrum of spontaneous *cII* mutations in the genomic DNA of control (sham-treated) cells is comparable with that previously found in the same model system using the conventional low-throughput method ($P < 0.0001$, Figure 4A and B, Supplementary Figure S5 and Supplementary Table S1). Altogether, the data indicate that our new method can sensitively and specifically detect the mutational fingerprint of three prominent carcinogens with varying mutagenic potencies, including UVB, 4-ABP and SHS, as well as establish the spectrum of spontaneous

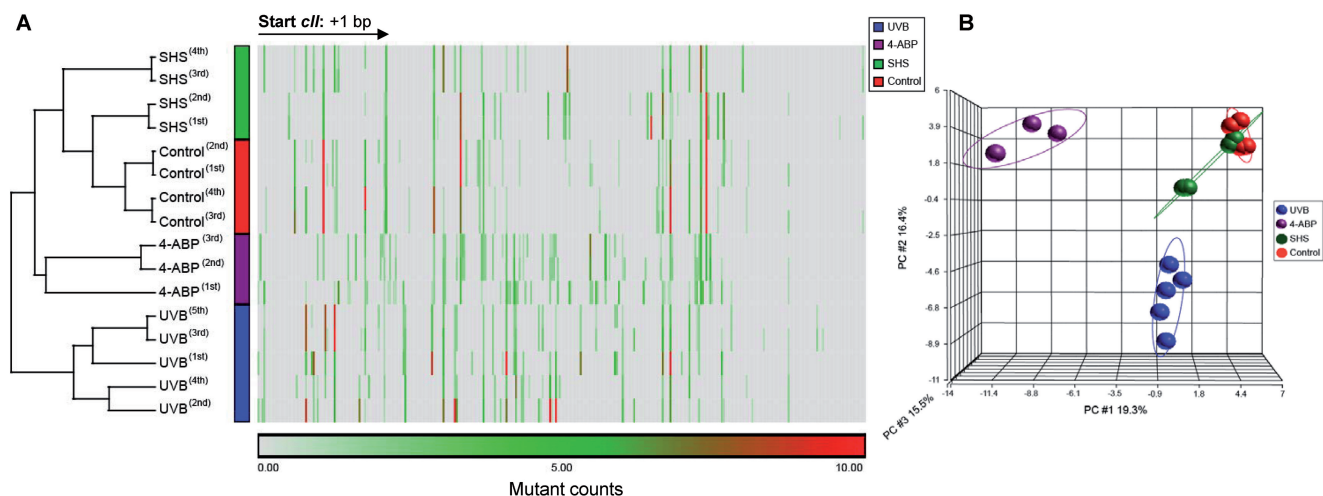


Figure 5. Reproducibility of the *cII* mutation spectra established in duplicate/multiple samples of carcinogen-treated mice/cells versus control. The hierarchical clustering analysis (A) and the PCA (B) were performed using the Partek Genomics Suite v6.11.1116 (<http://www.partek.com>).

mutations in control. The levels of sensitivity and specificity of our new method are comparable with those of the currently available low-throughput method.

Hierarchical clustering analysis and the principle component analysis

We determined the reproducibility of the results obtained by our new method by assaying duplicate/multiple samples in a single run (to examine intra-assay variation) and/or in two different runs (to examine inter-assay variation). We prepared two independent pools of 150 *cII* mutants from UVB-irradiated cells and analyzed them in a single assay run. In addition, we used another three independent pools of 150 mutants from the UVB-irradiated cells and analyzed them in a subsequent assay run. In all cases, we verified reproducible results for the duplicate/triplicate samples analyzed in a single run as well as in two independent runs (Figure 5). More specifically, the heatmap generated by the Hierarchical Clustering Analysis, which uses the Pearson's Dissimilarity to measure differences between frequency and position of each mutation occurred amongst different samples, showed that all the UVB-irradiated samples clustered very closely together (Figure 5A). This observation was further confirmed by the PCA mapping that showed that all the UVB-irradiated samples grouped together and remained distant from other differently treated or control samples (Figure 5B).

Furthermore, we analyzed duplicate 4-ABP-treated samples in a single assay run as well as in a subsequent run. As shown in Figure 5, comparable results were obtained from the analysis of the above-specified samples. The heatmap generated by the Hierarchical Clustering Analysis (Figure 5A) and the PCA mapping (Figure 5B) revealed that all the 4-ABP-treated samples clustered closely together and stayed separated from other differently treated or control samples. Moreover, we examined duplicate SHS-treated samples in a single assay run and in a subsequent run. As shown in

Figure 5, comparable results were obtained from the analysis of the above SHS-treated samples, as reflected by the clustering of all SHS-treated samples together (Figure 5A), as well as mapping of these samples closely to each other, while being apart from other differently treated samples (Figure 5B). Given the weak mutagenicity of SHS, it is also of note that the SHS-treated samples did not map too far from the control samples (Figure 5B).

We have also analyzed duplicate control samples in a single assay run as well as in a subsequent run. As shown in Figure 5, comparable results were obtained from the analysis of the above control samples. The heatmap generated by the Hierarchical Clustering Analysis (Figure 5A) and the PCA mapping (Figure 5B) showed that all the control samples clustered together and stayed distant from other differently treated samples. These data validate the reproducibility of our new high-throughput next-generation sequencing-based method for detecting both the carcinogen-induced and control mutation spectra.

Read coverage analysis

To demonstrate the sensitivity and specificity of our next-generation sequencing-based method and its efficient read coverage for the detection of experimentally induced/control mutations we performed a read coverage analysis on all differently treated samples and control. For each sample, we used the total number of reads as benchmark and then randomly selected 5×, 10×, 20×, 35×, 50× and 100× coverage (e.g. 5× coverage for each sample containing a pool of 150 mutants is 750 reads). Except for 100× coverage, where the full reads were used once, for all other coverage analyses (5×, 10×, 20×, 35× and 50×), we performed the random selection of reads 2–4 times and used the average results. Note that 'coverage' here refers to the average number of reads sequenced per sample. For each coverage analysis, we calculated (i) the minimum true mutation, which is the percentage of mutations detected in the *x* randomly selected reads that can also be found in the full reads (100×) and (ii) the maximum false mutation,

which is the number of mutations detected in the x randomly selected reads that are not detectable in the full reads. As shown in Figure 6, in all cases, $20\times$ coverage was sufficient to yield $\geq 97\%$ minimum true mutation and ≤ 2 maximum false mutation. From $20\times$ coverage onward, the minimum true mutation began to approach 100%, whereas the maximum false mutation started to reach negligible level.

To specifically address the effect of depth of coverage on the detection of mutation, we then grouped the true mutation ratios and false mutation counts into four categories, including (i) $5\times$ coverage, (ii) $10\times$ coverage, (iii) $20\times$ coverage and (iv) $35\times$ and plus coverage. The analysis of variance and Robust Test of Equality of Means results revealed that the depth of coverage (up until $20\times$) has a significant effect on the true mutation ratio, $F(3, 20.53) = 20.27$ ($P < 0.0001$) (Supplementary Table S2). *Post hoc* comparisons using Games-Howell's test showed that the mean of true mutation ratio at $10\times$ coverage ($M = 0.9429$, $SD = 0.041$) was not significantly different from that at $5\times$ coverage ($M = 0.9269$, $SD = 0.054$) ($P = 0.797$), whereas this value differed significantly from that at $20\times$ coverage ($M = 0.9845$, $SD = 0.017$) ($P = 0.014$). However, the mean of true mutation ratio at $20\times$ coverage was not significantly different from that at $35\times$ and plus coverage ($M = 0.9999$, $SD = 0.0003$). Together, these data indicate that the true mutation ratio is improved with increasing depth of coverage up until $20\times$, after which there is no significant improvement. Likewise, similar analysis confirmed that the depth of coverage (up until $20\times$) has a significant effect on the false mutation count, $F(3, 28.34) = 26.55$ ($P < 0.0001$); the false mutation count continues to reduce significantly with increasing depth of coverage up until $20\times$, after which there is no significant reduction (Supplementary Table S2). Altogether, these data indicate that our next-generation sequencing-based method has more than sufficient read coverage (~ 5 times higher than required) to detect the experimentally induced and control mutations with high sensitivity and specificity.

DISCUSSION

In this study, we have developed a high-throughput method for detecting the mutational fingerprint of carcinogens by devising a cost-, time- and labor-effective strategy in which a widely used transgenic mutagenesis assay is made compatible with a next-generation sequencing platform. Accordingly, we have modified the Big Blue[®] mouse mutation detection assay and incorporated it into the Roche/454 Genome Sequencer FLX Titanium next-generation sequencing technology. In addition, we have set up a detailed bioinformatics approach to process and analyze the high volume sequencing data. We have used this novel method to detect the mutational fingerprints of three prominent environmental carcinogens with varying mutagenic potencies, including sunlight UVB, 4-ABP and SHS that are known to be strong, moderate and weak mutagens, respectively (11–13). Here, we demonstrate that our new method can detect the

mutational fingerprints of these three carcinogens with high sensitivity and specificity. Furthermore, we verify that the accuracy and reproducibility of this method are comparable with those of the currently available low-throughput method.

Using this new method, we have successfully established the mutational fingerprints of sunlight UVB, 4-ABP and SHS by detecting three distinct mutation spectra in the *cII* gene in the genomes of Big Blue[®] mice/cells treated with the respective carcinogens. The mutational fingerprint of sunlight UVB was characterized by the preponderance of dipyrimidine-targeted mutations, which were predominantly G:C→A:T transitions, and clustered at several codon positions in the *cII* gene in the genome of UVB-irradiated cells (Figure 1B and C, Supplementary Figure S2 and Supplementary Table S1). The 4-ABP-induced mutational fingerprint manifested as the prevailing G:C basepair-localized mutations, which were mostly G:C→T:A transversions, and occurred frequently at several codons in the *cII* gene in the genome of 4-ABP-treated mice (Figure 2B and C, Supplementary Figure S3 and Supplementary Table S1). In the case of SHS, a subtle, yet, distinguishable mutational fingerprint was established as the induced *cII* mutations, mostly being G:C→A:T transitions, were localized to G:C basepairs in the genome of SHS-treated mice (Figure 3B and C, Supplementary Figure S4 and Supplementary Table S1). The above-specified mutational fingerprints of these three carcinogens are comparable with those found previously in the same model system using the conventional low-throughput method (11–13).

The tested carcinogens in this study are known to induce predominantly base substitutions, a type of mutation that is effectively detectable by the Big Blue[®] mutation assay (10,25). The successful application of our new method for the detection of mutational fingerprint of tested carcinogens indicates that this method is suitable for establishing the mutational signature of a wide range of carcinogens. In addition, the method is flexible to be coupled with other transgenic or non-transgenic mutation detection assays if the modifications described here are implemented, accordingly. For instance, the *gpt* delta transgenic mutation assay, which is optimized for the detection of small/large deletion mutations and point mutations (26), can easily be incorporated into this method to allow establishing the mutational fingerprint of other classes of carcinogens, for example, clastogens. Likewise, the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) mutation assay (27) can be introduced into this new method to offer high-throughput detection of mutational fingerprint of carcinogens in an endogenous reporter gene of the human genome.

Our overall findings show that the new method is superior to the conventional method for establishing the mutational fingerprint of carcinogens. Most importantly, the new method offers great advantages over the traditional method as it saves significant amounts of time, labor and cost. For example, the conventional method requires preparation, processing and analysis of a large number of mutants (individually), whereas the new method achieves this same objective by a single analysis of a pool of

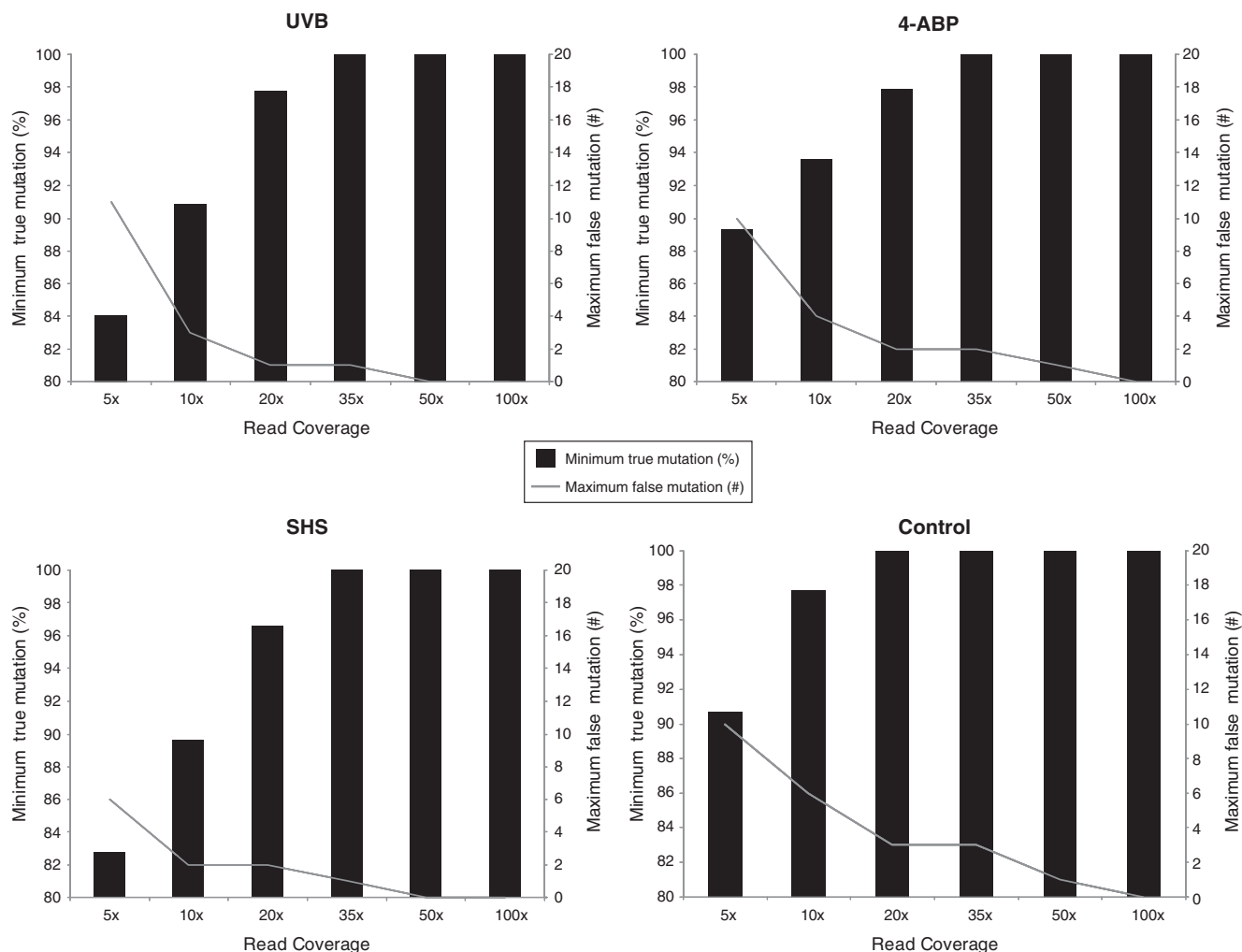


Figure 6. Read coverage analysis for *cII* mutations in carcinogen-treated mice/cells versus control. For each sample, total number of reads was used as benchmark and subsequently, randomly selected 5x, 10x, 20x, 35x, 50x and 100x coverage analyses were performed as described in the text.

mutants (simultaneously). We have calculated the amounts of time and expenses that we spent on the analysis of all our tested samples using both the conventional and the new methods. According to our calculations, the new method is approximately 20 times less time consuming and 3.5 times less costly than the conventional method. If the reduced workload of personnel is factored into these calculations, the savings will become even greater. To reduce the cost of sequencing, one can also use barcoding for multiplexing, that is, an auxiliary technique in which sample-specific barcoding adapters that include unique sequence tags and a restriction site are attached to individual samples. After pooling the tagged DNA samples, library preparation and sequencing, the tag sequences are used to identify the generated sequences that correspond to each original sample (28,29). Currently, work in our laboratory is underway to use 12 different barcoded adapters in a single assay run, which will enable us to pool 12 independent samples together and analyze them simultaneously in each of the 8 lanes of a Roche/454 Genome Sequencer. Prospectively, the incorporation of the barcoding approach into our new next-generation sequencing-based method

will save greater amounts of time, labor and cost in future sequencing projects.

As the next-generation sequencing technologies are constantly evolving and rapidly undergoing refinements, the cost of such analysis is expected to drop significantly (6–8). Due to financial constraints and lack of bioinformatics support, small laboratories may not be able to perform on-site next-generation sequencing work. Currently, however, many universities, research institutes and private companies have core facilities, which provide competitive next-generation sequencing services and bioinformatics data analysis to outside investigators. The accuracy and reproducibility of our new method, which is consistent with the known low error rate of the Roche/454 platform (30), and its comparable sensitivity and specificity with those of the existing method, together with the above-mentioned prospects are all indicative of the potential of the new method for becoming the mainstay of mutational fingerprinting for carcinogens.

Recently, Gilles *et al.* (31) have shown that the total error rates of the 454 GS-FLX Titanium instrument for the first 101 bases and for the full-length sequence

(average: 550 bases) are 0.49% and 1.07%, respectively. The majority of these errors could be ascribed to false insertions and deletions. Of note, insertions and deletions were the minor types of mutation detected in this study, whereas base substitutions comprised the predominant type of the detected mutations. Gilles *et al.* (31) have also demonstrated that high coverage can help to correct false mutations caused by random errors, and $5\times$ coverage was determined as the minimum coverage needed to achieve this goal. In our study, the sequence reads were in the range of 100–400 bases after being trimmed for the primer sequences at the beginning and end of the reads. Given the average length of the reads for our target sequence, high coverage and minor occurrence of insertions/deletions relative to base substitutions, we feel confident that the 0.66% minimum threshold used for the detection of mutations in this study is sufficient to distinguish between true mutations and sequencing errors generated by the GS-FLX Titanium analysis. The reproducibility of the results obtained by our new next-generation sequencing-based method, as well as their comparability with those obtained by the conventional method reassure the sensitivity and specificity of the new method for detecting the mutational fingerprint of carcinogens. Of note, we have also used a minimum threshold of 0.33% and obtained similar results to those found using the 0.66% minimum threshold (data not shown). Altogether, the read-out of interest in this study is the mutagen signature, which is not substantially affected by the minimum threshold criteria. We stress that the comparable mutational signatures of all the tested carcinogens established by our new next-generation sequencing-based method and the conventional DNA sequencing confirm the adequacy of the minimum threshold criteria used for the detection of mutations in this study.

We would like to acknowledge that, for comparability purposes, we have analyzed a pool of 150 mutants per sample by our new method, which is consistent with the conventional DNA sequencing approach, in which similar number of mutants is sequenced individually for establishing the mutational fingerprint of carcinogens. We note that in our preliminary studies, we have used pools of 50 and 150 mutants, respectively, per sample, and analyzed them by our next-generation sequencing-based method, which yielded similar results in both cases. This observation together with the finding that our next-generation sequencing-based method has more than sufficient read coverage (~ 5 times higher than required; Figure 6), and the fact that obtaining larger number of mutants, especially in case of weak mutagens, may not necessarily prove practical, indicates that the pool of 150 mutants analyzed in this study is large enough for establishing the induced and control mutation spectra. We note that given the numerous mutable nucleotide positions in the *cII* gene, sequencing different number of mutants may reveal slightly different mutations detectable at various nucleotide positions in this gene; however, our data indicate that the overall spectrum of mutations in the full-length *cII* gene remains the same as long as an average of 150 pooled mutants is used for DNA sequencing.

Thus far, few studies have used next-generation sequencing technologies for the detection of mutations in foreign DNA (using cell free environment), for example, shuttle vector or RNA template or yeast (32–34). These elegant studies have confirmed the applicability of next-generation sequencing platforms for mutagenicity analysis (32–34). However, the mutation detection systems employed in these studies may not necessarily represent some of the key determinants of mutagenesis in mammalian cells, for example, chromatin structure, DNA sequence contexts, fidelity and efficiency of DNA polymerases and DNA repair (4,35–38). In addition, these systems are not suitable for investigating organ-specific mutagenicity in relation to tumorigenesis, which is a unique property of certain carcinogens (12,13). The latter is reflective of the need for studying target-organ mutagenesis in animal models of tumorigenicity. The current literature lacks a comprehensive study, in which the application of next-generation sequencing technologies for the detection of mutations in chromosomal genes of a mammalian system is explored. Our study is the first demonstration of the applicability of these technologies for the detection of mutational fingerprint of carcinogens in a chromosomal gene in a validated mammalian model system (4,10). In spite of the increasingly popular use of transgenic model system for mutational analysis of carcinogens, the system remains low-throughput and cost, time and labor ineffective (10). This study offers a new strategy to modify the mutation detection assay in this model system by making it compatible with a next-generation sequencing platform, thus, allowing high-throughput analysis of mutational fingerprint of carcinogens in a cost-, time- and labor-effective manner.

In summary, we have developed a new next-generation-based method that can detect the mutational fingerprint of carcinogens with high sensitivity and specificity. In addition, we have shown that the accuracy and reproducibility of this new method are comparable with those of the currently available low-throughput method. Given the accuracy and reproducibility, great expediency and speed, and labor, time and cost effectiveness of this method, the method is poised to be employed in large-scale screening projects for detecting mutagenic carcinogens and become the method of choice for high-throughput DNA-sequencing analysis. Prospectively, the method will have the potential to move the field of carcinogenesis forward by allowing high-throughput analysis of mutations induced by endogenous and/or exogenous genotoxins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

The authors thank Basilio Gonzalez for library preparation and Dr. Jinhui Wang for analyzing the samples by the Roche/454 Genome Sequencer, S. Kim for technical

support, and their department chair for general support. They also thank the staff and management of the City of Hope Animal Resources Center, in particular, Lauren Ratcliffe, Marie Prez, Armando Amaya, Yvonne Harper, Donna Isbell, Kenneth Golding, and Dr. Richard Ermel for help with the conduct of all mouse experiments.

FUNDING

Funding for open access charge: American Cancer Society [RSG-11-083-01-CNE to A.B.]; University of California Tobacco Related Disease Research Program [18KT-0040 to A.B.].

Conflict of interest statement. None declared.

REFERENCES

- Pfeifer, G.P. and Besaratinia, A. (2009) Mutational spectra of human cancer. *Hum. Genet.*, **125**, 493–506.
- DeMarini, D.M. (1998) Mutation spectra of complex mixtures. *Mutat. Res.*, **411**, 11–18.
- Hussain, S.P. and Harris, C.C. (1998) Molecular epidemiology of human cancer: contribution of mutation spectra studies of tumor suppressor genes. *Cancer Res.*, **58**, 4023–4037.
- Besaratinia, A. and Pfeifer, G.P. (2006) Investigating human cancer etiology by DNA lesion footprinting and mutagenicity analysis. *Carcinogenesis*, **27**, 1526–1537.
- Olivier, M., Hollstein, M. and Hainaut, P. (2010) TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.*, **2**, a001008.
- Marguerat, S., Wilhelm, B.T. and Bahler, J. (2008) Next-generation sequencing: applications beyond genomes. *Biochem. Soc. Trans.*, **36**, 1091–1096.
- Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92**, 255–264.
- Meyerson, M., Gabriel, S. and Getz, G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.*, **11**, 685–696.
- Pfeifer, G.P. and Hainaut, P. (2011) Next-generation sequencing: emerging lessons on the origins of human cancer. *Curr. Opin. Oncol.*, **23**, 62–68.
- Lambert, I.B., Singer, T.M., Boucher, S.E. and Douglas, G.R. (2005) Detailed review of transgenic rodent mutation assays. *Mutat. Res.*, **590**, 1–280.
- Besaratinia, A., Kim, S.I. and Pfeifer, G.P. (2008) Rapid repair of UVA-induced oxidized purines and persistence of UVB-induced dipyrimidine lesions determine the mutagenicity of sunlight in mouse cells. *FASEB J.*, **22**, 2379–2392.
- Kim, S.I., Yoon, J.I., Tommasi, S. and Besaratinia, A. (2012) New experimental data linking secondhand smoke exposure to lung cancer in nonsmokers. *FASEB J.*, **26**, 1845–1854.
- Yoon, J.I., Kim, S.I., Tommasi, S. and Besaratinia, A. (2012) Organ specificity of the bladder carcinogen 4-aminobiphenyl in inducing DNA damage and mutation in mice. *Cancer Prev. Res. (Phila)*, **5**, 299–308.
- Cleaver, J.E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer*, **5**, 564–573.
- de Grujil, F.R. (2002) Photocarcinogenesis: UVA vs. UVB radiation. *Skin Pharmacol. Appl. Skin Physiol.*, **15**, 316–320.
- Pfeifer, G.P. and Besaratinia, A. (2012) UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci.*, **11**, 90–97.
- Beland, F.A. and Kadlubar, F.F. (1990) In: Cooper, C.S. and Grover, P.L. (eds), *Chemical Carcinogenesis and Mutagenesis*, Vol. 1. Springer-Verlag, Berlin/Heidelberg, pp. 297–325.
- Talaska, G., al-Juburi, A.Z. and Kadlubar, F.F. (1991) Smoking related carcinogen-DNA adducts in biopsy samples of human urinary bladder: identification of N-(deoxyguanosin-8-yl)-4-aminobiphenyl as a major adduct. *Proc. Natl. Acad. Sci. USA*, **88**, 5350–5354.
- Besaratinia, A. and Pfeifer, G.P. (2008) Second-hand smoke and human lung cancer. *Lancet Oncol.*, **9**, 657–666.
- Kim, S.I., Arlt, V.M., Yoon, J.I., Cole, K.J., Pfeifer, G.P., Phillips, D.H. and Besaratinia, A. (2011) Whole body exposure of mice to secondhand smoke induces dose-dependent and persistent promutagenic DNA adducts in the lung. *Mutat. Res.*, **716**, 92–98.
- Pfeifer, G.P., Chen, H.H., Komura, J. and Riggs, A.D. (1999) Chromatin structure analysis by ligation-mediated and terminal transferase-mediated polymerase chain reaction. *Methods Enzymol.*, **304**, 548–571.
- Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W.L., Russ, C., Lander, E.S., Nusbaum, C. and Jaffe, D.B. (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res.*, **18**, 763–770.
- Dames, S., Durtschi, J., Geiersbach, K., Stephens, J. and Voelkerding, K.V. (2010) Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes. *J. Biomol. Tech.*, **21**, 73–80.
- Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.*, **8**, R143.
- Jakubczak, J.L., Merlino, G., French, J.E., Muller, W.J., Paul, B., Adhya, S. and Garges, S. (1996) Analysis of genetic instability during mammary tumor progression using a novel selection-based assay for in vivo mutations in a bacteriophage lambda transgene target. *Proc. Natl. Acad. Sci. USA*, **93**, 9073–9078.
- Nohmi, T. and Masumura, K. (2005) Molecular nature of intrachromosomal deletions and base substitutions induced by environmental mutagens. *Environ. Mol. Mutagen.*, **45**, 150–161.
- Albertini, R.J. (2001) HPRT mutations in humans: biomarkers for mechanistic studies. *Mutat. Res.*, **489**, 1–16.
- Lennon, N.J., Lintner, R.E., Anderson, S., Alvarez, P., Barry, A., Brockman, W., Daza, R., Erlich, R.L., Giannoukos, G., Green, L. et al. (2010) A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol.*, **11**, R15.
- Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M. and Fire, A.Z. (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.*, **35**, e130.
- Droege, M. and Hill, B. (2008) The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *J. Biotechnol.*, **136**, 3–10.
- Gilles, A., Meglec, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, **12**, 245.
- Petrie, K.L. and Joyce, G.F. (2010) Deep sequencing analysis of mutations resulting from the incorporation of dNTP analogs. *Nucleic Acids Res.*, **38**, 8095–8104.
- Smith, D.R., Quinlan, A.R., Peckham, H.E., Makowsky, K., Tao, W., Woolf, B., Shen, L., Donahue, W.F., Tusneem, N., Stromberg, M.P. et al. (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res*, **18**, 1638–1642.
- Yuan, B., Wang, J., Cao, H., Sun, R. and Wang, Y. (2011) High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing. *Nucleic Acids Res.*, **39**, 5945–5954.
- Garinis, G.A., van der Horst, G.T., Vijg, J. and Hoeijmakers, J.H. (2008) DNA damage and ageing: new-age ideas for an age-old problem. *Nat. Cell Biol.*, **10**, 1241–1247.
- Guo, C., Kosarek-Stancel, J.N., Tang, T.S. and Friedberg, E.C. (2009) Y-family DNA polymerases in mammalian cells. *Cell Mol. Life Sci.*, **66**, 2363–2381.
- Mellon, I., Spivak, G. and Hanawalt, P.C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*, **51**, 241–249.
- Wogan, G.N., Hecht, S.S., Felton, J.S., Conney, A.H. and Loeb, L.A. (2004) Environmental and chemical carcinogenesis. *Semin. Cancer Biol.*, **14**, 473–486.