CLAIMS

I/We claim:

1. A method for detecting and quantifying genomic mutations developed *in vivo* in a subject following the subject's exposure to a mutagen, comprising:

providing a sample from the subject, wherein the sample comprises double-stranded DNA molecules, and wherein the subject was exposed to the mutagen 30 days or less prior to extracting the sample comprising the double-stranded DNA molecules from the subject;

generating an error-corrected sequence read for each of a plurality of the double-stranded DNA molecules in the sample, comprising:

generating a set of copies of an original first strand of the double-stranded DNA molecule and a set of copies of an original second strand of the double-stranded DNA molecule;

sequencing the set of copies of the original first and second strands to provide a first strand sequence and a second strand sequence; and

comparing the first strand sequence and the second strand sequence to identify one or more correspondences between the first and second strand sequences; and

analyzing the one or more correspondences to determine a mutation spectrum for the double-stranded DNA molecules in the sample.

2. The method of claim 1, further comprising calculating a mutant frequency for the double-stranded DNA molecules in the sample by calculating the number of unique mutations per duplex base-pair sequenced.

3. The method of claim 1, wherein the double-stranded DNA molecules were extracted from liver, spleen, blood, lung or bone marrow of the subject.

4. The method of claim 1, wherein the mutation spectrum is generated by unsupervised hierarchical mutation spectrum clustering.

5. The method of claim 1, wherein the mutation spectrum is a triplet mutation spectrum.

6. The method of claim 1, wherein generating an error-corrected sequence read for each of the plurality of the double-stranded DNA molecules includes generating error-corrected sequence reads of one or more targeted genomic regions.

7. The method of claim 6, wherein the one or more targeted genomic regions is a mutation-prone site in the genome.

8. The method of claim 6, wherein the one or more targeted genomic regions is a known cancer driver gene.

9. The method of claim 1, wherein the subject is a transgenic animal, and wherein at least some of the double-stranded DNA molecules include one or more portions of a transgene.

10. The method of claim 1, wherein the subject is a non-transgenic animal, and wherein the double-stranded DNA molecules comprise endogenous genomic regions.

11. The method of claim 1, wherein the subject is a human, and wherein the double-stranded DNA molecules are extracted from a blood draw taken from the human.

12. A non-transitory computer-readable storage medium comprising instructions that, when executed by one or more processors, performs the method of any one of claims 1-11 for determining if the subject is exposed to at least one genotoxin and/or determining an identity of at least one genotoxin.

13. The non-transitory computer-readable storage medium of claim 12, further comprising computing the mutation spectrum, mutant frequency, and/or triplet mutation spectrum of a detected agent, from which the identity of the at least one genotoxin is determined.

14. A computer system for performing the method of any one of claims 1-11 for determining if the subject is exposed to and/or an identity of at least one genotoxin, the system comprising: at least one computer with a processor, memory, database, and a non-transitory computer readable storage medium comprising instructions for the processor(s), wherein said processor(s) are configured to execute said instructions to perform operations comprising the methods of any one of claims 1-11.

15. The system of claim 14, further comprising a networked computer system comprising:

a. a wired or wireless network;

b. a plurality of user electronic computing devices able to receive data derived from use of a kit comprising reagents to extract, amplify, and produce a polynucleotide sequence of the subject's sample, and to transmit the polynucleotide sequence via a network to a remote server; and

c. a remote server comprising the processor, memory, database, and the non-transitory computer readable storage medium comprising instructions for the processor(s), wherein said processor(s) are configured to execute said instructions to perform operations comprising the methods of any one of claims 1-11; and

d. wherein said remote server is able to detect and identify mutagenic events and/or nucleic acid damage events resulting from genotoxic exposure of a sample.

16. The system of claim 15, wherein the database and/or a third-party database accessible via the network, further comprises a plurality of records comprising one or more of a genotoxin profile of known genotoxins, a genotoxin profile of at least one subject's sample, and wherein the genotoxin profile comprises a mutation or a site of DNA damage.