Cell Research

www.nature.com/cr
www.cell-research.com

RESEARCH HIGHLIGHT

Somatic mutation accumulation seen through a single-molecule lens

Lovelace J. Luquette[1] and Peter J. Park[1]✉

Cell Research (2021) 31:949–950; https://doi.org/10.1038/s41422-021-00537-2

Somatic mutations (SMs) accumulate over the lifetime of cells and can lead to cancer and other diseases; however, SMs can be difficult to detect, especially when they are present in very few cells. In a recent *Nature* paper, Abascal et al. develop a protocol capable of detecting SMs present on only a single molecule of DNA and apply it to both mitotic and post-mitotic human tissues.

In multicellular organisms, somatic mutations (SMs) may occur after the zygote stage, to be inherited by only a fraction of the cells comprising the mature organism. SMs have long been profiled in tumor tissues, but recent studies have also uncovered their roles in noncancerous tissues, e.g., in neurodevelopmental disease[1] and in the aging of the hematopoietic system.[2]

Unlike germline mutations that are inherited by essentially all cells in an individual, the fraction of cells harboring any particular SM is highly variable. The earliest possible SM would occur after the first division of the zygote and could be inherited by around half of all cells in the organism; a tumorigenic SM initially present in only a single cell could expand into a tumor and become shared by a large fraction of the tumor cells. At the opposite extreme, an SM that occurs in a post-mitotic cell will not be inherited by other cells, leaving the SM at an infinitesimally low frequency.

Detection of low-frequency SMs presents substantial challenges. For example, at standard whole-genome sequencing depths of 30×, most heterozygous SMs present in 1% of cells would not be present on even a single sequencing read. A heterozygous SM at 20% frequency would be present on only three reads on average and must be differentiated from artifacts that can be supported by similar numbers of reads such as endogenous single-stranded DNA lesions or technical artifacts (e.g., sequencer errors and DNA damage that may occur during tissue collection, storage, and library preparation[3]). Increasing sequencing depth can help in detecting lower-frequency mutations and in differentiating SMs from artifacts, but this strategy is cost-prohibitive at the whole genome scale. As a result, several specialized methods to detect low-frequency SMs have been developed. One approach is to sequence DNA from a single cell, either directly, via enzymatic amplification of a single genome,[4,5] or indirectly, by in vitro clonal expansion of a single cell.[6] Another approach is redundant sequencing of multiple copies of a single DNA molecule from a bulk population, sometimes referred to as single-molecule consensus sequencing (SMCS).[7]

One form of SMCS termed "duplex sequencing" differentiates technical artifacts from low-frequency SMs by ensuring that reads are obtained from both the Watson and Crick strands of the original DNA molecule. True SMs should be supported by all reads derived from a single molecule and on both DNA strands, whereas technical artifacts and DNA lesions should appear on a subset of reads from the same molecule or be isolated to a single DNA strand. One such method is BotSeqS,[8] a duplex sequencing protocol with whole-genome coverage. This protocol calls for random shearing of the input DNA by sonication which can lead to stretches of single-stranded DNA originating from nicks in the DNA backbone or in the form of 5′ and 3′ overhangs. However, these single-stranded DNA stretches could be filled in by DNA polymerases later in the protocol,[7] potentially copying single-stranded artifacts into both DNA strands and making them indistinguishable from true mutations.

With this in mind, Abascal et al.[9] first assessed the accuracy of BotSeqS by comparing it to standard whole-genome sequencing of single cell-derived colonies from a similar cell population. Surprisingly, BotSeqS libraries yielded an average of 1240 mutations per cell compared to only 66 mutations per cell from single-cell-derived controls. This, along with an abundance of C>A and C>G single nucleotide substitutions specific to BotSeqS, suggested that the majority of BotSeqS SMs were artifactual. This motivated the development of NanoSeq, which improved BotSeqS in two ways. The first innovation in NanoSeq is that DNA is fragmented by restriction enzyme digestion rather than sonication. By choosing a restriction enzyme that leaves blunt DNA ends, the creation of single-stranded 5′ and 3′ overhangs is avoided. The second innovation is the addition of non-A chain-terminating dideoxynucleotides during A-tailing, which helps to prevent DNA extension originating at nicks in the DNA backbone. After correcting for detection sensitivity, NanoSeq libraries contained only 109 SMs per cell on average (compared to 66 from single cell-derived colony controls), reflecting a dramatic reduction in false positives compared to BotSeqS.

The first application of NanoSeq was to compare the SM burdens of stem cells and their terminally differentiated progeny. In both the hematopoietic system and colonic crypts, the SM burden of stem cell populations was comparable to that of the differentiated cells, suggesting that the cell divisions required to produce terminally differentiated progeny from stem cell ancestors incur relatively few SMs. NanoSeq was then used to measure the rate at which SMs accumulate with age by comparing SM burdens from individuals spanning 0–100 years of age. Included in this analysis were cortical neurons, which are long-lived and post-mitotic, and visceral smooth

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ✉email: peter_park@hms.harvard.edu

Published online: 27 July 2021

muscle cells, which can divide albeit relatively infrequently. This analysis largely confirmed previous reports of age-related accumulation of SMs in colon crypt[10] and neurons.[11] An unexpected finding, however, was that the rates of SM accumulation in colon, blood, neurons, and smooth muscle were similar despite large differences in the rate of mitosis. This striking observation calls into question the extent to which cell division causes SMs and is consistent with Abascal et al.'s first finding that few SMs were generated during cell division.

An important caveat of the NanoSeq method is that only ~29% of the human genome flanking the restriction enzyme recognition sequences is accessible. An alternative NanoSeq protocol for greater genome coverage was also developed, but little data was presented. The SMCS paradigm exemplified by NanoSeq may also be less well-suited to certain analyses where single-cell DNA sequencing has proven effective, such as lineage tracing[4] and the detection of larger mutations such as structural rearrangements and copy number mutations.[12] Thus, SMCS and single-cell approaches are likely to provide important complementary information as well as orthogonal confirmation of results.

In summary, NanoSeq is a powerful improvement over previous SMCS techniques, enabling detection of low-frequency SMs with high specificity. The results reported by Abascal et al. have raised provocative questions concerning the relative contribution of cell division in producing SMs, and future work on this subject will lead to a deeper understanding of DNA damage and repair processes in vivo and the role played by SMs in the aging of normal human tissues.

REFERENCES

1. Poduri, A. et al. *Science* **341**, 1237758 (2013).
2. Jaiswal, S. et al. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
3. Costello, M. et al. *Nucleic Acids Res.* **41**, e67 (2013).
4. Lodato, M. A. et al. *Science* **350**, 94–98 (2015).
5. Chen, C. et al. *Science* **356**, 189–194 (2017).
6. Blokzijl, F. et al. *Nature* **538**, 260–264 (2016).
7. Salk, J. J. et al. *Nat. Rev. Genet.* **19**, 269–285 (2018).
8. Hoang, M. L. et al. *Proc. Natl. Acad. Sci. USA* **113**, 9846–9851 (2016).
9. Abascal, F. et al. *Nature* **593**, 405–410 (2021).
10. Lee-Six, H. et al. *Nature* **574**, 532–537 (2019).
11. Lodato, M. A. et al. *Science* **359**, 555–559 (2018).
12. Umbreit, N. et al. *Science* **368**, eaba0712 (2020).

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to P.J.P.

Reprints and permission information is available at http://www.nature.com/reprints