



PERSPECTIVE

Current and future trends in whole genome sequencing in cancer

Yuki Katsuya

Department of Experimental Therapeutics, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

Cancer remains a formidable global health challenge affecting millions of lives annually. For decades, conventional cancer treatments used a one-size-fits-all approach, overlooking the intricate genetic variations that drive tumors. The emergence of cancer genomics has ushered in a new era of personalized and targeted cancer therapies. The Human Genome Project, which was launched in 1990 and completed in 2003, and the advent of next-generation sequencing technology, which has dramatically increased the efficiency of genome decoding, have dramatically accelerated cancer genomics. Cancer genome sequencing efforts began in 2005, including The Cancer Genome Atlas (TCGA). In 2007, the first whole-human genome sequence, which was generated through next generation sequencing, was published¹. The first clinical application of whole genome sequencing (WGS) for the rearranged during transfection (*RET*) fusion gene in tongue cancer was reported in 2010². By 2030, hundreds of millions of patients with cancer will have had their genomes sequenced.

As Dr. Yoshida and Dr. Ando have shown in this issue, cancer genome sequencing in clinical settings currently uses targeted exome next-generation sequencing panels (such as the FoundationOne CDx and OncoGuide NCC OncoPanel). However, protein-coding exons make up only 1.2% of the human genome. Limited information is available on somatic

mutations in non-coding regions such as untranslated regions, introns, promoters, regulatory elements, non-coding functional RNAs, repetitive regions, and mitochondrial genomes, which make up 98% of the human genome. Structural variants and viral incorporation into cancer genomes have not yet been widely investigated^{3,4}.

From this perspective, we discuss the current status of, and future directions in, cancer genomics, with particular emphasis on WGS. Additionally, we examine research efforts in various countries to provide a comprehensive view of the landscape of WGS. Furthermore, we introduce the WGS project initiated by the Japanese government.

WGS: current state

Technological advancements

Cancer genomics has witnessed remarkable technological progress enabling researchers to decode the genetic basis of cancer with unprecedented precision. WGS has emerged as a cornerstone technology allowing for the comprehensive analysis of an individual's entire genome.

The evolution of sequencing technologies can be classified into 3 distinct generations. The first generation was Sanger sequencing in the 1980s, which served as the cornerstone for DNA sequencing. The second generation was massively parallel sequencing through platforms such as Illumina and Ion Torrent, which revolutionized high-throughput sequencing capabilities. Currently, the third generation technologies have long-read and single-molecule sequencing abilities⁵. The expense of sequencing a genome remarkably decreased, from an estimated \$1 million in 2007 to \$1000 in 2014, and is currently approximately \$600. This precipitous decline mirrors the trend in decreasing computing power costs described by Moore's Law⁶.

Correspondence to: Yuki Katsuya

E-mail: ykatsuya@ncc.go.jp

ORCID ID: <https://orcid.org/0000-0002-2127-3601>

Received November 1, 2023; accepted November 30, 2023.

Available at www.cancerbiomed.org

©2024 The Author. Creative Commons Attribution-NonCommercial 4.0 International License

Advancements in long-read sequencing technology have enabled intact decoding of lengthy DNA (tens of thousands of bases). When coupled with WGS, current technologies enable analysis of a broader spectrum of regions and the intricacies of acquired structural abnormalities. Long-read sequencing analysis has revealed structural anomalies such as extensive deletions, gene fusions, and various chromosomal rearrangements. Additionally, it has highlighted medium-scale structural abnormalities comprising complex combinations of local duplications, inversions, and microdeletions⁷.

The tremendous volume of data generated by cancer genomics, particularly by WGS, poses substantial challenges. For example, the sequencing of each patient's genome can produce terabytes of data, thus necessitating a robust data storage and management infrastructure. Cloud-based platforms are increasingly used for storing and analyzing data and facilitating global collaboration among researchers. The program used for mutation calling from the sequencing data is also extremely important, because proper algorithm configuration can address problems such as false negatives due to shallow read depth⁸.

Advantages and drawbacks of WGS

WGS is extremely valuable, because it facilitates a variety of analyses, such as the detection of changes in non-coding regions (promoters, regulatory regions, etc.); detection of large-scale structural variants (large deletions/insertions, inversions, duplications, translocations, etc.); identification of the presence and location of viral genomes; detection of mitochondrial gene changes; genomic immune analysis such as HLA typing; and diagnosis and treatment of hereditary tumors according to germline gene information, pharmacogenomics, and carcinogenesis determined by mutation signature analysis. However, the disadvantages of WGS include a shallow read depth, which can lead to false negative findings, particularly in cases with high intra-tumor heterogeneity; the need for frozen tumor tissue in the analysis, thus hindering clinical implementation at many institutions; the large volume of data generated, and the high cost of data analysis and storage; and the difficulty in evaluating unknown genetic alterations⁸. To overcome these drawbacks, long-read platforms decrease the unambiguity of read mapping and aid in the detection of structural variants. In addition, advanced deep-learning techniques such as convolutional neural networks enable the identification of gene alterations⁹.

Current achievements of WGS

Cancer genomics has been led by national and international collaborative projects.

The largest WGS study, led by the International Cancer Genome Consortium (ICGC)/TCGA Pan-Cancer Analysis of the Whole Genome (PCAWG) consortium, has yielded several key findings. The first is that the cancer genome contains an average of 4 to 5 driver mutations in protein-coding and non-coding regions. However, in approximately 5% of cases, such as in rare cancers, no driver mutations are identified. Second, chromothripsis, the localized concentration of many chromosomal structural abnormalities caused by a single disruptive event, is often recognized as an early event in tumor evolution. Third, cancers with abnormal telomere maintenance often arise from tissues with low replication activity and present several molecular mechanisms that prevent telomere shortening to critical levels. Germline genetic polymorphisms affect the patterns of somatic mutations, including point mutations, chromosomal structural abnormalities, and retrotransposon rearrangements. Fourth, a small number of non-coding regional mutations have been found to cause cancer, similarly to the non-coding regional mutations in the *TERT* promoter. Overall, this project has provided a framework for standardized WGS analysis methods, cloud computing, and integrated analysis with epigenomics. Although Japanese patients were included in the study, the number of samples from Japanese patients was insufficient (286 of 2,658 cases); therefore, more data on refractory cancers common in Asian and Japanese populations must be acquired in future studies¹⁰.

Another study on non-coding somatic driver point mutations and structural variants in cancer conducted by the PCAWG consortium has revealed that WGS yields additional discoveries in cancer genomes, although point mutations and structural variants driving cancer occur less frequently in non-coding genes and regulatory sequences than in protein-coding genes. Regarding point mutations, the researchers analyzed 2,583 tumors from 27 individual tumor types, and developed a rigorous strategy integrating 13 discovery algorithms and a filtering strategy. The 5'-end mutations in *TP53*; 3' UTR mutations in *NFKB1* and *TOBI*; and rearrangements involving the *AKR1C1* and *BRD4* genes were found to be candidate non-coding driver genes. Considering structural variants, the researchers used 2 driver discovery methods to identify regions significantly affected by significantly recurrent

breakpoints and significantly recurrent juxtapositions, thus accounting for the genomic heterogeneity in the frequency of DNA breakage and repair, and the three-dimensional structure of the genome¹¹.

Clinical implementation of WGS in the UK and the Netherlands

A well-known national project on WGS is the UK's 100,000 Genomes Project, which was launched in 2012. This project included 26,488 patients with cancer. As part of the project, WGS was offered as a routine medical service for all seriously ill children, including those with suspected genetic cancers, and to adults with certain rare diseases from 2019. Samples were acquired at 13 centers in the UK, and WGS and central automated analysis (processing, calling, quality checking, storing, presenting, annotating, and prioritizing the variants identified by sequencing) was performed by Genomics England, a national company that analyzes genomic information. Local interpretations and clinical reports were also described. Other epoch-making systems have also been established, including collaborations with national and international organizations, development of the structure of the database (Genomics England Clinical Interpretation Partnership), and collaborations with industry¹².

The Netherlands is unique in that WGS for the diagnosis of primary unknown cancers has been a reimbursed health care expense since 2021. They have implemented an algorithm that predicts the primary site of unknown primary cancers with high accuracy based on the WGS results¹³. A study on the feasibility of the clinical implementation of WGS in the Netherlands was also conducted. For 1,200 analyses, WGS was compared with the standard of care (SOC) molecular diagnoses for the same patient samples. The analytic pipeline was optimized accordingly, and the concordance rate improved to 98.8%. Of the 848 patients, 602 had actionable genes, of whom 190 (22.4%) had additional treatment options determined as a result of WGS¹⁴. The world largest metastatic tumor WGS database, the Hartwig Medical Foundation database, is also located in the Netherlands. The UK and France are other countries with public programs enabling patients with specific cancers to access WGS¹⁵.

Clinical implementation of WGS in Japan

The Japanese government launched the WGS project “The Action Plan for Whole Genome Analysis for Cancer and Rare/intractable Diseases” in 2019, with the aim of promoting

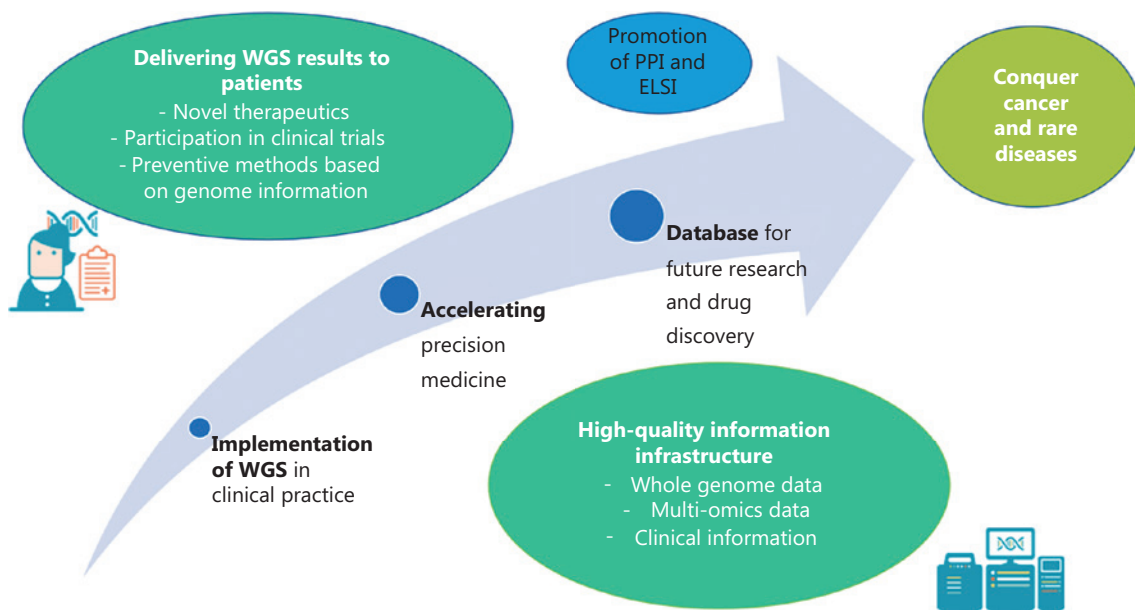


Figure 1 Overview of the “The Action Plan for Whole Genome Analysis for Cancer and Rare/intractable Diseases” WGS project in Japan. WGS, whole genome sequencing; PPI, patient and public involvement; ELSI, ethical, legal, and social issues. Modified from “Action Plan for Whole Genome Analysis 2022”¹⁶.

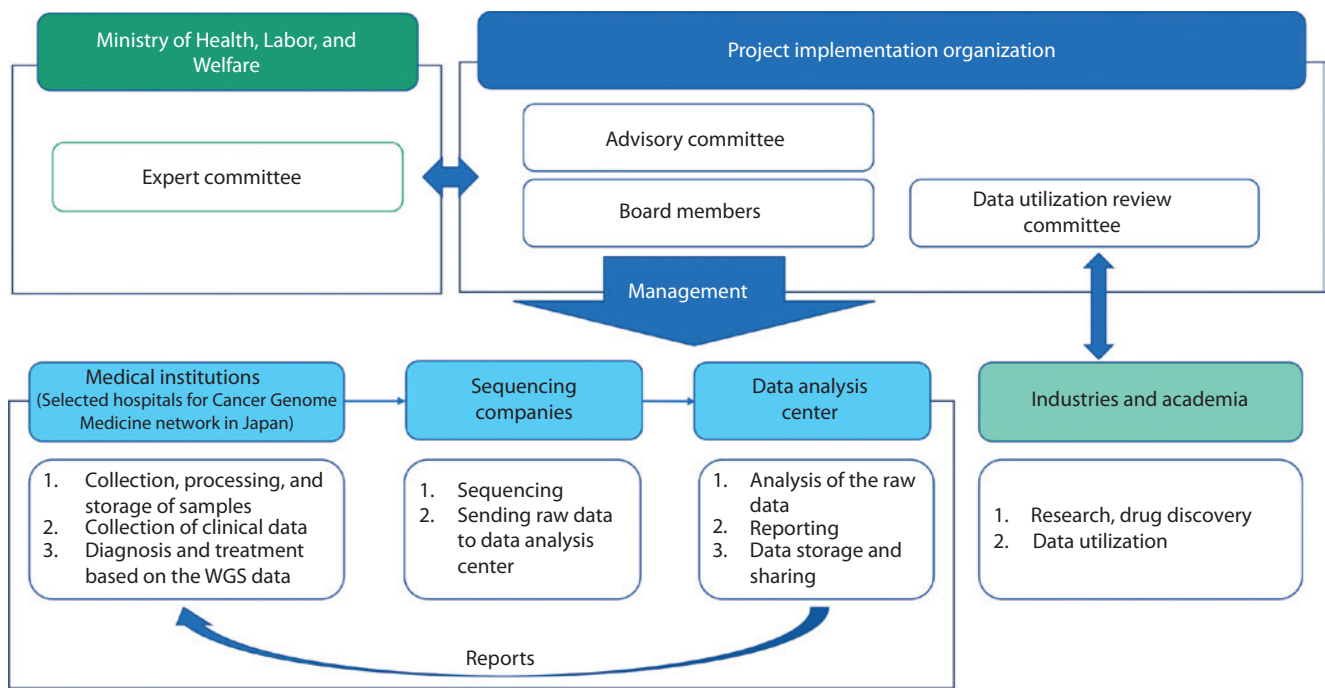


Figure 2 The WGS project's implementation and collaboration system. Modified from "Action Plan for Whole Genome Analysis 2022"¹⁶.

personalized medical care, research, and drug discovery based on high-quality national genomic data. The WGS project focuses on providing the results of the analyses to patients and establishing data for development of pharmaceuticals and academic research (Figure 1).

The project's implementation directions are determined by the Expert Committee under the Ministry of Health, Labor, and Welfare, which collaborates with medical institutions (selected from 12 core base and 33 base hospitals in the Cancer Genome Medicine network in Japan), sequencing companies, and a central data analysis center, to promote personalized medical care for each patient. Additionally, ethical, legal, and social issues, as well as patient and public involvement, are being considered. The data analysis center, established to curate the massive amounts of genomic data produced in this project, is responsible for detecting genomic alterations by using a unified analysis pipeline. The results are returned to physicians after review by an expert panel, thereby providing opportunities for new treatment methods and participation in clinical trials. The data analysis center also provides a mechanism for sharing genomic and clinical information through the cloud. The obtained data will be used in academia and industry with data-sharing policies. Furthermore, accumulated data, such as genomic data, high-quality clinical

information, and multi-omics data, will be used by industries and academia for research and drug discovery (Figure 2)¹⁶.

As of September 2023, a whole-genome analysis of more than 12,000 cancer cases has been completed. We plan to acquire 100,000 genomic profiles for Japanese patients with cancer in 5 years. Cancer types of interest are those with many structural abnormalities that are difficult to detect by SOC, and cancers stratified by genomic profiling. The former types include hematological tumors, bone and soft tissue tumors, brain tumors, and driver negative non-small cell lung cancers. The latter types include pediatric, adolescent, and young adult cancers; hereditary cancers; and some gynecological and breast cancers.

For the clinical implementation of WGS in Japan, which has universal health coverage, proper cancer types are being analyzed that will show superiority over targeted exome next-generation sequencing panels as current SOC for patients with advanced cancer and are reimbursed health care expenses in Japan.

Future directions

WGS has emerged as a transformative tool in cancer genomics that can provide an unprecedented view of the cancer genetic

landscape. WGS is expected to facilitate the development of highly personalized treatment strategies, the identification of rare mutations, and the elucidation of tumor evolution and heterogeneity. Although regulatory frameworks and research efforts vary across countries, international collaboration is essential to overcome the challenges associated with data sharing, ethics, and access to high quality healthcare. By working together, the global community will be able to fully harness the power of WGS in the fight against cancer. We hope that this issue will improve understanding of Japan and the National Cancer Center, and provide a starting point for international cooperation.

Conflicts of interest statement

No potential conflicts of interest are disclosed.

References

1. Mayor S. Genome sequence of one individual is published for first time. *BMJ*. 2007; 335: 530-1.
2. Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol*. 2010; 11: R82.
3. Hamdan A, Ewing A. Unravelling the tumour genome: the evolutionary and clinical impacts of structural variants in tumorigenesis. *J Pathol*. 2022; 257: 479-93.
4. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016; 534: 47-54.
5. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-generation sequencing technology: current trends and advancements. *Biology (Basel)*. 2023; 12: 997.
6. National Human Genome Research Institute. DNA Sequencing Costs: Data. Available from: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>, Accessed on November 30th, 2023.
7. Sakamoto Y, Xu L, Seki M, Yokoyama TT, Kasahara M, Kashima Y, et al. Long-read sequencing for non-small-cell lung cancer genomes. *Genome Res*. 2020; 30: 1243-57.
8. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci*. 2018; 109: 513-22.
9. Zverinova S, Guryev V. Variant calling: considerations, practices, and developments. *Hum Mutat*. 2022; 43: 976-85.
10. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 2020; 578: 82-93.
11. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. 2020; 578: 102-11.
12. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ*. 2018; 361: k1687.
13. Schipper LJ, Samsom KG, Snaebjornsson P, Battaglia T, Bosch LJW, Lalezari F, et al. Complete genomic characterization in patients with cancer of unknown primary origin in routine diagnostics. *ESMO Open*. 2022; 7: 100611.
14. Samsom KG, Schipper LJ, Roepman P, Bosch LJ, Lalezari F, Klompenhouwer EG, et al. Feasibility of whole-genome sequencing-based tumor diagnostics in routine pathology practice. *J Pathol*. 2022; 258: 179-88.
15. Cuppen E, Elemento O, Rosenquist R, Nikic S, IJzerman M, Zaleski ID, et al. Implementation of whole-genome and transcriptome sequencing into clinical cancer care. *JCO Precis Oncol*. 2022; 6: e2200245.
16. Ministry of health Labor and Welfare. Action Plan for Whole Genome Analysis 2022. 2022. Available from: <https://www.mhlw.go.jp/content/10901000/001073173.pdf>.

Cite this article as: Katsuya Y. Current and future trends in whole genome sequencing in cancer. *Cancer Biol Med*. 2024; 21: 16-20. doi: 10.20892/j.issn.2095-3941.2023.0420