第 17 回全ゲノム解析等の推進に関する専門委員会	参考資料 4
令和 5 年 10 月 3 日	

Action Plan for Whole Genome Analysis 2022

30 September 2022

For people, for life, for the future
Ministry of Health, Labour and Welfare

Table of Contents

0. Foreword

The Action Plan for Whole Genome Analysis (Version 1) (hereafter, Action Plan [Version 1]) for cancer and rare/intractable diseases was formulated in December 2019 to promote whole genome analysis in Japan.

While the present project was affected by the global spread of COVID-19, in FY2020, the government's Basic Policy on Economic and Fiscal Management and Reform 2020 (Cabinet decision of 17 July 2020) stipulated that the Action Plan for Whole Genome Analysis was to be steadily promoted, and that systems were to be put in place to enable extensive analysis and use of data by personnel from industry, the government, and academe to provide new, personalized medical care to patients for whom no treatment is available. On this basis, the relevant study groups compiled policies for further promotion of the Action Plan (Version 1).

In FY2021, the previous study groups were organized into the Expert Committee on Promotion of Whole Genome Analysis (tentative name) (hereafter, Expert Committee), which was established within the Health Sciences Council Science and Technology Committee as the top decision-making body for the Action Plan (Version 1). In addition, to make steady progress with promotion of initiatives based on the Action Plan (Version 1), at its meeting of June 2021, the Expert Committee formulated the Action Plan for Whole Genome Analysis Roadmap 2021 (hereafter, Roadmap 2021), which summarizes the items for implementation during FY2021 and FY2022.

Research using whole genome data has advanced globally over the last few years, and from the perspective of facilitating research and drug discovery, as well as protecting the genome information of citizens, the present project has taken on greater significance in Japan. The Basic Policy on Economic and Fiscal Management and Reform 2021 (Cabinet decision of 18 June 2021) stated that, while taking into consideration the initiatives based on the Japan-U.S. Joint Leaders' Statement, the Government would steadily promote the Action Plan for Whole Genome Analysis and the Roadmap 2021 under the principle of putting patients first and return of results to patients to provide new, personalized medical care to patients for whom no treatment was previously available, and would promote the creation of systems to enable extensive analysis and use of data by personnel from industry, the government, and academe. The Expert Committee held discussions on this basis, and from the perspective of steady promotion of whole genome analysis, it decided to formulate the Action Plan for Whole Genome Analysis 2022 (hereafter, the present Action Plan).

The Basic Policy on Economic and Fiscal Management and Reform 2022 (Cabinet decision of 7 June 2022) states that information infrastructure linking clinical information to information such as the results of whole genome analysis is to be constructed, and an environment for its utilization is to be put in place promptly to promote drug discovery relating to cancer and rare/intractable diseases. On this basis, the Action Plan specifies the directionality, in terms of the patients targeted by the project and its implementation systems, over a period of around five years starting from FY2022. In addition, it includes the management strategy with regard

to return of results to patients and utilization, as well as items relating to ethical, legal, and social issues (ELSI) and patient and public involvement (PPI).

The Expert Committee will continue its discussions in the future, taking into account any changes in the surrounding environment and including the viewpoints of patients' families and citizens under the principle of putting patients first and return of results to patients, with the aim of creating information infrastructure that links clinical information to the results of whole genome analysis to promote drug discovery relating to cancer and rare/intractable diseases. The Expert Committee will create an environment for the use of this infrastructure at an early stage to promote the utilization of the information in research and drug discovery and the introduction of new, personalized medical care, while also steadily promoting the prompt return of results to patients.

* In the present Action Plan, the term "cancer" refers to intractable cancers, rare cancers, pediatric cancers, hereditary cancers, and other types of cancer for which whole genome analysis is expected to have a certain effect, but for which the private sector alone cannot conduct research and drug discovery.

Reference: Meetings of the Relevant Study Groups

Subcommittee on Promotion of Whole Genome Analysis in Relation to Cancer (tentative name)
Subcommittee chair: Ken Yamaguchi, President, Shizuoka Cancer Center
1st meeting: 16 October 2019
2nd meeting: 20 November 2019
3rd meeting: 3 December 2019
4th meeting: 10 December 2020
5th meeting: 16 February 2021

Action Plan for Whole Genome Analysis (Version 1) (compiled December 2019)

Liaison and Coordination Council for Whole Genome Analysis of Cancer (tentative name)
General Manager: Hitoshi Nakagama, President, National Cancer Center
1st meeting: 25 September 2020
2nd meeting: 27 October 2020
3rd meeting: 7 December 2020
4th meeting: 5 February 2021

Study for Promotion of the Action Plan for Whole Genome Analysis (compiled in February 2021) (tentative name)

Study Group on Promotion of Genomic Medicine Analysis in Relation to Rare/Intractable Diseases (tentative name)
Chair: Hidehiro Mizusawa, President, National Center of Neurology and Psychiatry
1st meeting: 8 October 2019
2nd meeting: 19 November 2019
3rd meeting: 3 December 2019
4th meeting: 8 December 2020
5th meeting: 10 December 2020
6th meeting: 16 February 2021

Study Council for Promotion of the Action Plan for Whole Genome Analysis (tentative name)
Chair: Ken Yamaguchi, President, Shizuoka Cancer Center
1st meeting: 10 December 2020
2nd meeting: 16 February 2021

System Development for Further Promotion of Whole Genome Analysis (compiled in March 2021) (tentative name)

Expert Committee on Promotion of Whole Genome Analysis (tentative name)
Chair: Hitoshi Nakagama, President, National Cancer Center
1st meeting: 14 May 2021
2nd meeting: 31 May 2021
3rd meeting: 21 July 2021
4th meeting: 30 July 2021 (held on rotating basis)

5th meeting: 17 September 2021
6th meeting: 18 November 2021
7th meeting: 18 January 2022
8th meeting: 2 March 2022
9th meeting: 23 May 2022 (held on rotating basis)
10th meeting: 7 July 2022
11th meeting: 19 August 2022

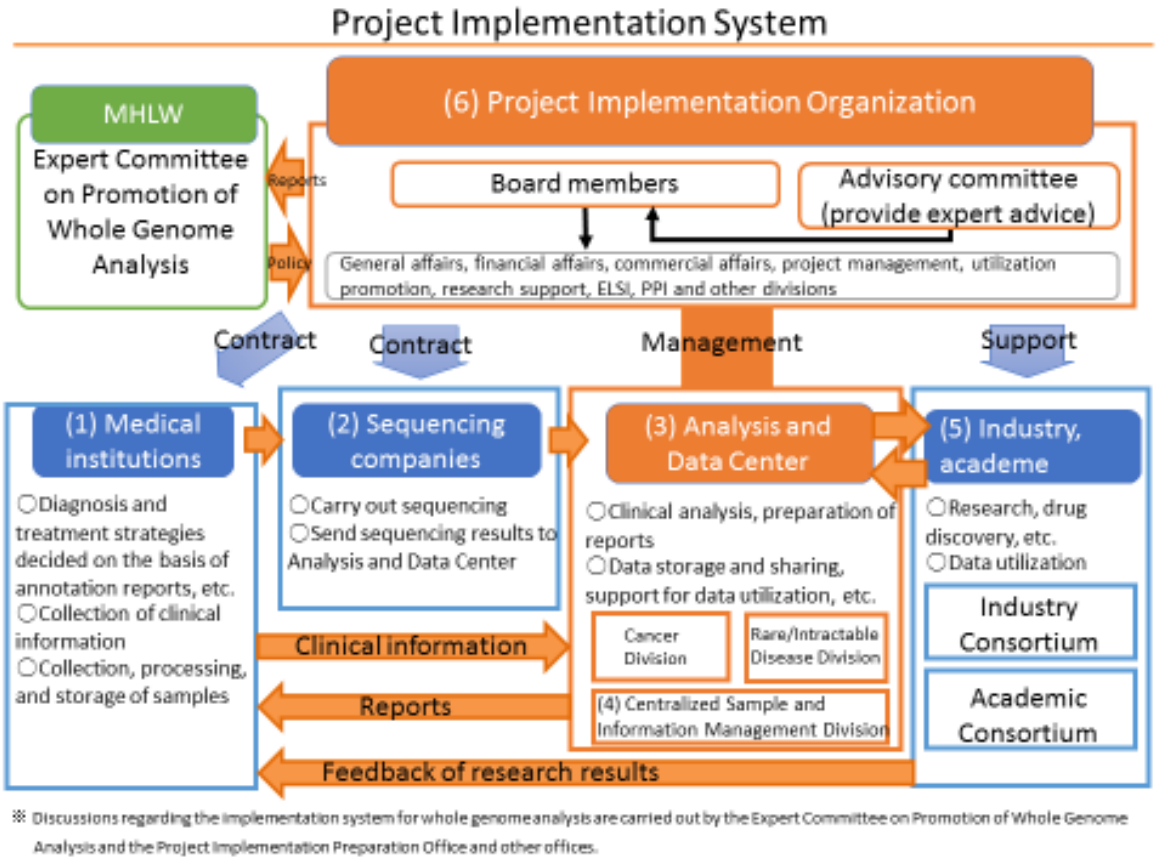
Action Plan for Whole Genome Analysis (compiled in September 2022)

1. Outline of the Project

(1) Project implementation system

The basic direction for implementation of the present project was decided by the Expert Committee established by the Ministry of Health, Labour and Welfare (MHLW), and deliberations are currently under way on the assumption that the Project Implementation Organization will be responsible for the concrete operations. In the implementation of the project, the implementation organization will promote the return of results to patients and the utilization of data in collaboration with medical institutions, sequencing companies, and the Analysis and Data Center, as well as with industry and academe.

Figure 1. Project implementation system



(2) Return of results to patients

The present Action Plan will be implemented steadily to promote the use of the accumulated data (this refers to genome data, clinical data, etc.; the same applies hereafter) in research and drug discovery and the introduction of new, personalized medical care, and also to ensure that the results are returned to patients at an early stage.

The information that can be returned to patients, and the specific ways in which it can be returned, are envisaged as follows.

(a) Information relating to the results of research and drug discovery

New results obtained through research or drug discovery that uses the data accumulated by the present project will be made available as appropriate to industry and academe to promote the use of the Analysis and Data Center. Generalized information on these results will be made available to the general public over the Internet, etc.

(b) Information that can be introduced into clinical practice

Information relating to the condition that prompted a medical examination

- Information from the analysis results that is useful for medical care of the patient, such as pathogenic variants determined at the time of whole genome analysis to be of medical significance, will be returned to the patient promptly to the greatest extent possible.
- With information that is not determined at the time of whole genome analysis to be of medical significance, but is subsequently determined to be medically significant as a result of advanced cross-sectional analysis, etc., efforts will be made to return this information to the patient in line with the patient's wishes if it is judged to be useful for the treatment of the patient following discussions by an expert panel.

Information not relating to the condition that prompted a medical examination

- Among the information on germline pathogenic variants, etc. that is determined at the time of whole genome analysis to be of medical significance, the information that can affect the health of the patient, despite being unconnected to the patient's current medical condition, is to be returned to the patient in line with the patient's wishes following discussions by an expert panel, and with proper considerations such as provision of genetic counseling.
- Of the information on germline pathogenic variants, etc. that is not determined at the time of whole genome analysis to be of medical significance but is subsequently determined to be medically significant as a result of advanced cross-sectional analysis, etc., the information that can affect the health of the patient, despite being unconnected to the patient's current medical condition, is to be returned to the patient in line with the patient's wishes following discussions by an expert panel, and with proper considerations such as provision of genetic counseling.

(c) Information relating to new personalized medical care

Systems will be put in place to enable as many patients as possible to have opportunities to take part in clinical research or clinical trials that use the results of whole genome analysis.

Figure 2. Information that can be returned to patients

Information that Can Be Returned to Patients and How It Is Returned

1. Information relating to the results of research and drug discovery
• Provision of novel therapeutic methods through medical product development*1

2. Information that can be introduced into clinical practice
• Provision of information with clear medical significance, provision of diagnosis and treatment if indicated*2 • Provision of information that was of unclear medical significance at the time of analysis but was subsequently shown to be medically significant, and provision of diagnosis and treatment if indicated*3
3. Information relating to new personalized medical care
• Provision of opportunities for participation in clinical research or clinical trials*4

*1 Information relating to new results obtained through research or drug discovery that uses the data accumulated by the present project will be made available as appropriate to industry and academe to promote the use of the Analysis and Data Center. Generalized information on these results will be made available to the public over the Internet, etc.

*2 *3 In either case, information on germline pathogenic variants, etc. that may affect the health of the patient will be returned to the patient in line with the patient's wishes following discussions by an expert panel, and with proper considerations such as provision of genetic counseling.

*4 Systems will be put in place to enable as many patients as possible to have opportunities to take part in clinical research or clinical trials using the results of whole genome analysis.

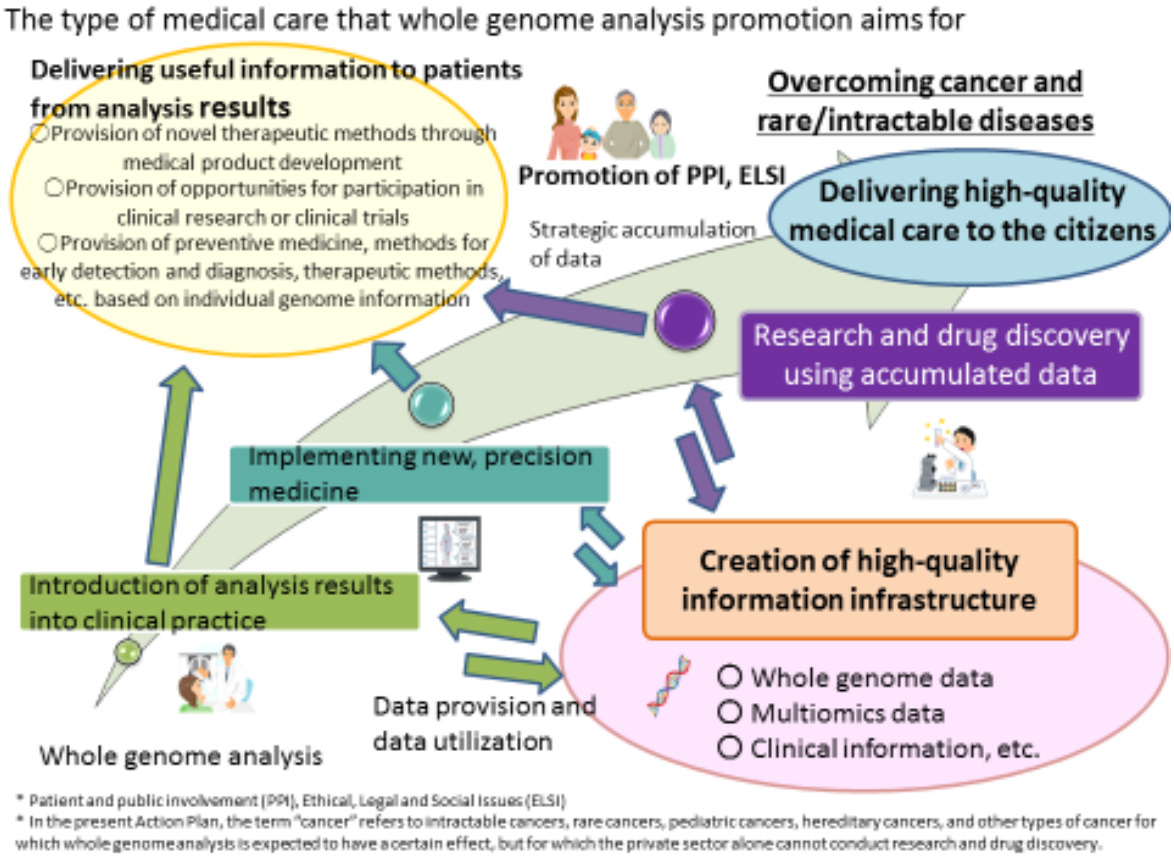
(3) The type of medical care that whole genome analysis promotion aims for

The present project aims for medical care that will overcome cancer and rare/intractable diseases in the future by promoting whole genome analysis that will accumulate data strategically and by facilitating the use of these data in research and drug discovery to deliver high-quality medical care to the citizens.

In addition, the present project will further promote the early introduction of the results of analysis into clinical practice and the implementation of new, personalized medical care.

*In the present Action Plan, the term “cancer” refers to intractable cancers, rare cancers, pediatric cancers, hereditary cancers, and other types of cancer for which whole genome analysis is expected to have a certain effect, but for which the private sector alone cannot conduct research and drug discovery.

Figure 3. The type of medical care that whole genome analysis promotion aims for



2. Objectives of the Project

In analyses that have been carried out to date, the analysis results have been introduced to clinical practice at an early stage and have been used in the promotion of new, personalized medical care.

The full-scale analyses to be carried out from now on will deliver high-quality medical care to the citizens, with the aim of overcoming cancer and rare/intractable diseases in the future. Promoting the strategic accumulation of data and facilitating research and drug discovery that use these data will be essential for achieving this aim, and the present Action Plan therefore promotes the application of the results of whole genome analysis to research and drug discovery.

3. Basic Strategies

The basic strategies for achieving the objectives of the project have been determined, and practical application will start in fields where results are obtained, with the aim of overcoming cancer and rare/intractable diseases in the future. Promoting the strategic accumulation of data and facilitating research and drug discovery that use these data will be essential for achieving this aim, and the basic strategies for using the results of whole genome analysis in research and drug discovery have therefore been drawn up as follows.

(1) Basic strategies for using the results of whole genome analysis in research and drug discovery

To ensure comprehensive return of the results of whole genome analysis to patients, it is important to put in place an environment that stimulates research and drug discovery using the accumulated genome data. Initiatives in collaboration with industry and academe will therefore be promoted with the aim of nurturing industries through the creation of innovation originating in Japan, while at the same time new treatments will be made available to patients.

To enable this, a system is needed to allow open and fair use of the accumulated whole genome analysis information by researchers in domestic and overseas research institutions and companies. An Industry Consortium and an Academic Consortium will therefore be established, and the Project Implementation Organization will create a mechanism to support their collaboration.

In addition, the basic strategies for early introduction of analysis results into clinical practice and for implementing new personalized medical care are as follows.

(2) Basic strategies for early introduction of analysis results into clinical practice

Where the results of whole genome analysis relate to diagnosis or treatment with efficacy that has already been confirmed, existing drugs, etc. will be made available promptly to patients through clinical research.

In addition, when the results of whole genome analysis are returned to patients, confirmation tests with a certain level of accuracy will be performed, and systems for this will be put in place within medical institutions. Where the results of genome analysis give a certain degree of new evidence for technologies to select appropriate treatment methods or diagnose diseases, the goal will be for them to be covered by national health insurance in the future.

These measures are expected to make it possible to provide patients with appropriate treatment at an early stage, thus contributing to enhanced treatment opportunities and more efficient provision of medical care.

(3) Basic strategies to achieve new personalized medical care

New clinical studies and trials will be conducted, in addition to which real-world evidence will be compiled to achieve more personalized medical care through advanced, efficient diagnosis and treatment based on whole genome analysis and multiomics analysis, etc. Efficient clinical studies and trials will also be promoted through the use of clinical study support tools.

Cancer

Methods for prevention and early detection of cancer and for early diagnosis of recurrence using multiomics data in addition to whole genome data will be established, and research and development of new treatment methods, including immunotherapy, will be carried out, with the aim of achieving new, personalized medical care.

(a) Prevention methods

Effective cancer prevention methods with few side effects will be established by using whole genome analysis data.

Example: Prevention by pre-emptive medical care

Every day, the body produces many cells with genome mutations that can become the cause of cancer. These cells are removed by the body's own immune system, and they progress to clinical cancer when this immune system fails. It has so far been considered difficult to achieve medical care that prevents this failure of the immune system, but methods to control immunosuppression while avoiding exhaustion of the immune response will be established through the use of whole genome analysis data, thus preventing the onset of clinical cancer.

(b) Methods for early detection and early diagnosis of recurrence

Simple, inexpensive methods for early detection and early diagnosis of recurrence of cancer that combine whole genome analysis with the development of innovative biomarkers will be established.

Example: Large-scale comparative study using a prospective screening cohort with liquid biopsy

Research will be carried out with the aim of dramatic improvement in the cure rate of refractory cancers. Technology for reliable early detection of refractory cancers such as pancreatic cancer will be established by combining disease onset risk prediction using whole genome analysis with simple, inexpensive liquid biopsy, and this technology will then be linked to medical treatment.

(c) Novel therapies

The aim will be to develop therapies capable of radical cure regardless of the stage of the cancer, with a focus mainly on pediatric cancers, rare cancers, and refractory cancers such as pancreatic cancer. Systems will be established for high-quality diagnosis and treatment using joint drug discovery and joint clinical studies through searching for innovative therapeutic targets that are conducted all the way from basic to clinical research by academe and the corporate world, as well as artificial intelligence, etc.

Examples:

- Development of immunotherapies based on whole genome analysis of tumor-infiltrating lymphocytes, etc.
- Development of personalized medical care and breakthrough new drugs using multiomics data, including tumor microenvironment driver

mechanisms and immune responses, as well as AI analysis and digital medicine.

Rare/Intractable diseases

Whole genome analysis of rare/intractable diseases will be carried out to promote better medical care for patients with rare/intractable diseases, such as early diagnosis and development of novel therapies. Whole genome analysis will facilitate early diagnosis of rare/intractable diseases, elucidation of the nature of rare/intractable diseases, and development of effective treatment and diagnosis methods.

(a) Early diagnosis of rare/intractable diseases

For patients with diseases that are difficult to diagnose through genetic tests other than whole genome analysis, but that can probably be narrowed down through whole genome analysis, a system will be put in place for a physician with expertise in whole genome analysis in the field of rare/intractable diseases to fully explain the significance and purpose of the test, as well as the impact of the test results on the patients and their relatives, so that the patients can undergo whole genome analysis after having understood the implications.

For rare diseases with small numbers of patients in particular, a framework for joint international collection of cases and data sharing will be put in place to enable early diagnosis.

(b) Elucidation of the nature of rare/intractable diseases

In the field of rare/intractable diseases, genome information obtained through whole genome analysis, as well as high-quality clinical information (including disease-specific clinical findings, biomarkers, etc.) based on the characteristics of each rare/intractable disease, will be managed in an integrated fashion.

Carrying out whole genome analyses will allow the accumulation of information such as introns, regulatory regions, and even genome structure that cannot be obtained through gene panel tests or whole exome analysis. At the same time, it will also further the elucidation of the true nature of rare/intractable diseases and provide knowledge that will contribute to the establishment and improvement of objective diagnostic criteria, leading to early diagnosis of rare/intractable diseases.

© Facilitating the development of effective treatment and diagnostic methods

In addition to elucidating the nature of rare/intractable diseases, collecting and analyzing high-quality clinical information and registries of rare/intractable diseases, as well as genome information, will promote the development of more effective therapeutic and diagnostic methods for patients with rare/intractable diseases.

Figure 4. Basic Strategies of the Action Plan for Whole Genome Analysis

Basic Strategies of the Action Plan for Whole Genome Analysis

The present project has determined the basic strategies shown below, and with the aim of overcoming cancer and rare/intractable diseases, practical application will start in fields where the results of whole genome analysis are obtained.

1) Basic strategies for use in research and drug discovery

- To ensure comprehensive return to patients of the results of whole genome analysis, it is important to stimulate research and drug discovery using the accumulated genome data. Initiatives in collaboration with the Industry Consortium and the Academic Consortium will therefore be promoted, with the aim of nurturing industries through the creation of innovation originating in Japan, while at the same time new treatments will be made available to patients.
- A system will be put in place to allow open and fair use of the accumulated whole genome analysis information by researchers in domestic and foreign research institutions and companies.

2) Basic strategy for early introduction of results into clinical practice

- Where the results of whole genome analysis give a certain degree of evidence for technologies to select appropriate treatment methods or diagnose diseases, the goal will be for these to be covered by national health insurance in the future.

3) Basic strategy for achieving new, personalized medical care

- New clinical studies and trials will be conducted, in addition to which real-world evidence will be compiled, to achieve more personalized medical care through advanced, efficient diagnosis and treatment.

4. Initiatives to Date

Cancer

In the field of cancer, the Action Plan (Version 1) stipulates that, “First, the characteristics of mutations in the genome of Japanese people will be clarified by preliminary analysis, in order to advance policy decisions and development of systems for full-scale analysis.” With regard to the preliminary analysis, the Action Plan (Version 1) stipulates that, “For the present, samples will be selected for use that meet the conditions, such as whether the patient’s consent has been obtained for the use of analysis results, whether the stored samples are of sufficient quality for analysis, and whether clinical information is available. Taking into account the opinion of an advisory council, whole genome analysis of refractory cancers with relatively low 5-year survival rates, rare cancers (including pediatric cancers) that are often caused by rare genetic alterations, and hereditary cancers (including pediatric cancers) will be conducted, to the extent possible with the current human resources and facilities.”

On this basis, whole genome analysis was conducted on previously stored samples from 550 cases of refractory cancer and 3,247 cases of hereditary cancer between FY2019 and FY2020. The preliminary analyses[1] included verification of technical issues when performing whole genome analysis,[2] creation of a shared platform (ensuring data of sufficient quality in sequences obtained using a single sequencer), and creation of a unified analysis pipeline (from FASTQ file to mutation calls). Furthermore, results such as the identification of pathological mutations that could not be detected by conventional genetic testing methods[3] demonstrated the significance of whole genome analysis.

At the same time, issues were identified. These included the establishment of a pipeline for structural aberrations, which are results that whole genome analysis is particularly expected to deliver, and splicing mutations that require integrated analysis with transcriptome analysis; speeding up the process of determining the clinical significance of whole genome analysis results; establishment of an expert panel to handle whole genome analysis results; and creation of a system for verification and analysis of whole genome analysis results.

In addition, the Liaison and Coordination Council for Whole Genome Analysis of Cancer (tentative name) and the Study Council for Promotion of the Action Plan for Whole Genome Analysis (tentative name) held discussions on policy for the full-scale analysis and system development, and the results of these discussions are

[1] Practical Research for Innovative Cancer Control Research Group (Research representative: Teruhiko Yoshida, National Cancer Center Hospital), MHLW funded Comprehensive Research Project for the Promotion of Cancer Countermeasures Research Group (Research representative: Noboru Yamamoto, National Cancer Center Hospital)

[2] Evaluation of sequencing depth of tumor and normal areas in whole genome analysis, integrated analysis with transcriptome analysis, cross-tissue analysis of somatic mutations, global genome mutation analysis (including driver gene frequency analysis compared to Europe and the U.S.), chronological sample analysis, etc.

[3] For example, in cases in which the possibility of hereditary tumor was clinically suspected, but there was no definitive diagnosis, whole genome analysis detected a deep intronic mutation in the ATM gene (which encodes a protein involved in cell cycle control and DNA repair and is thought to be involved in the development of breast and ovarian cancer), and a splicing abnormality was confirmed by combined RNA sequencing. In addition, whole genome analysis was able to identify mutations characteristic of breast cancer, bone and soft tissue tumors, pancreatic cancer, and leukemia.

compiled in “Study for Promotion of the Action Plan for Whole Genome Analysis (tentative name)” (February 2021) and “System Development for Further Promotion of Whole Genome Analysis (tentative name)” (March 2021), respectively.

Since FY2021, the Expert Committee has been set up under the Science and Technology Division of the MHLW’s Health Sciences Council to formulate the Roadmap 2021 and to discuss the establishment of systems for return of results to patients, the operation of the Analysis and Data Center, measures for utilization of the data, measures for storage and utilization of samples, the operation of the Project Implementation Organization, and the review system of the MHLW. At the same time, whole genome analysis of approximately 10,000 samples, comprising both stored and new samples, has been carried out.

Rare/Intractable diseases

In the field of rare/intractable diseases, samples were selected for use in research that met the conditions, such as whether the patient’s consent had been obtained for the use of analysis results, whether the stored samples were of sufficient quality for analysis, and whether clinical information was available. Taking into account the opinion of an advisory council, the samples were classified into monogenic diseases, multifactorial diseases, and difficult-to-diagnose diseases, and to the extent possible with the current human resources and facilities, whole genome analysis was conducted for diseases for which results could be expected (16approx. 2,500 cases in FY2020 and 3,000 cases in FY2021).

It was found that, whereas some patients were diagnosed with rare/intractable diseases on the basis of clinical findings or existing genetic tests, among the patients who were unable to receive a diagnosis through these methods, there were some whose disease was identified as a result of the whole exome analysis in the research.

In addition, it was shown that, in 9.4% of the patients whose disease was not identified by whole exome analysis, the disease was identified as a result of whole genome analysis.[4]

With whole exome analysis and whole genome analysis for the diagnosis of patients who were not diagnosed by conventional methods, it is important to determine the scope of applicable subjects and the methods of analysis, and further research is needed for patients whose diseases are not identified by these analyses.

Based on these initiatives, the basic policy for whole genome analysis from FY2022 onward is summarized in the following chapter.

[4] Practical Research Project for Rare/Intractable Diseases Research Group (Research representative: Norihiro Kokudo, National Center for Global Health and Medicine)

5. Basic Policy Based on Initiatives to Date

(1) Target patients for whole genome analysis

Based on the findings obtained from whole genome analysis to date, the target patients for whole genome analysis are patients who are difficult to diagnose or have a low probability of radical cure through existing medical care, but who may be expected to obtain more accurate diagnosis and effective treatment through the use of whole genome analysis or multiomics analysis. Specifically, the target patients are as shown below.

Cancer

A. Target patients

As a general rule, the targets for analysis will be patients who fulfill all of the following three conditions and who give new consent following a full explanation.

(a) Sufficient samples can be obtained through surgery, biopsy, blood sampling (hematological tumors), etc.
(b) The patient has refractory cancer with a low probability of radical cure through surgery (unresectable advanced cancer, cancer with high probability of recurrence, etc.).
(c) The patient is alive at the start of the analysis and may be expected to receive some form of treatment in the future.

Other patients may be targets for analysis only in the event that the Expert Committee gives its approval in view of the importance of the disease.*

* In the search for markers for future drug development based on the genome database of cancers in Japanese people, important diseases are expected to include rare cancers, adolescent and young adult (AYA) cancers, pediatric cancers, hereditary cancers, treatment-resistant and refractory cancers, cancers with a small number of cases but that are characteristically Japanese (such as adult T-cell leukemia), and cancers with a large number of cases but for which genome information on Japanese cases has not been sufficiently accumulated.

B. Fields in which results may be expected

From among the target patients, the following two fields are expected to yield results.

(a) Cancer types with many structural abnormalities that are difficult to detect by conventional gene panel analysis or whole exome analysis
In addition to structural abnormalities and abnormalities in non-coding regions such as transcriptional regulatory regions, which are difficult to detect by conventional whole exome analysis and gene panel analysis, additional analyses such as multiomics analysis including DNA methylation and other epigenetic irregularities will lead to identification of candidate therapeutic targets.

(b) Cancer types for which stratification by genomic profiling, including germline genomic variation, can lead to treatment
Integration of cancer genome profiling based on whole genome analysis and wholescale refinement of existing subtype classifications will lead to

further application to personalized medical care, such as selection of treatment methods and prediction of recurrence.

Table 1. Cancer: Areas and specific examples of types of cancer (domains) where results may be expected

Areas where results may be expected

Area		Expected results	Specific examples of types of cancer (domains)
i	Cancer types with many structural abnormalities that are difficult to detect by conventional gene panel analysis or whole exome analysis	In addition to structural abnormalities and abnormalities in non-coding regions such as transcriptional regulatory regions, which are difficult to detect by conventional whole exome analysis and gene panel analysis, additional analyses such as multiomics analysis including DNA methylation and other epigenetic irregularities will lead to identification of candidate therapeutic targets.	○ Hematological tumors ○ Bone and soft tissue tumors ○ Brain tumors ○ Some respiratory organ tumors*1 ○ Some digestive organ tumors*2
ii	Cancer types for which stratification by genomic profiling, including germline genomic variation, can lead to treatment	Integration of cancer genome profiling based on whole genome analysis and wholescale refinement of existing subtype classifications will lead to further application to personalized medicine, such as selection of treatment methods and prediction of recurrence.	○ Pediatric, AYA cancers ○ Hereditary cancers ○ Some gynecological and breast cancers*3

*1 Includes driver gene-negative lung cancer, etc.
*2 Includes esophageal cancer, stomach cancer (scirrhous stomach cancer), colorectal cancer (unresectable metastasis), pancreatic cancer, etc.
*3 Includes triple-negative breast cancer, etc.

Rare/Intractable diseases

Rare/Intractable diseases are classified into monogenic diseases, multifactorial diseases, and diseases that are difficult to diagnose. For each of these, cases will be targeted for which results can readily be expected in accordance with the characteristics of the disease.

- Monogenic diseases: Diseases that have been diagnosed as genetic diseases, but for which no known causative gene can be found by whole exome analysis.
- Multifactorial diseases: Diseases, including those that do not require genetic analysis for diagnosis, for which the development of therapies that use whole genome information can be expected and for which a certain number of cases can be secured.
- Diseases that are difficult to diagnose: Cases that are difficult to diagnose even with existing genetic analysis, etc.

(2) Numbers of target cases

Whole genome analysis of approximately 19,200 cases (cancer: approx. 13,700, rare/intractable diseases: approx. 5,500) was carried out from 2019 to FY2021. In FY2022, the plan is to analyze 4,500 cases (cancer: approx. 2,000, rare/intractable diseases: approx. 2,500) of new patients giving consent and to return the results to patients.*

* As well as the results from the analysis targeting 100,000 genomes, this includes the results of multiomics analyses (comprehensive information on biomolecules), etc.

Cancer

In FY2021, whole genome analysis of around 600 prospective new patients was carried out at three medical institutions. From FY2022 onward, analyses will be carried out at medical institutions that are recognized by the Expert Committee as having the necessary systems in place for appropriate return of the results of whole genome analysis to patients, from among the 12 designated core hospitals for cancer genomic medicine and 33 designated hospitals for cancer genomic medicine （tentative name）(as of January 2022). In addition, from FY2023 onward, systems will be put in place in a stepwise fashion that will include examining the participation of medical institutions that can ensure a cancer genomic medicine treatment system, through collaboration with medical institutions approved by the Expert Committee.

In the field of cancer, analysis of approximately 2,000 cases will be carried out in FY2022 on the basis of the situation up to FY2021. Specific figures for the number of analyses from FY2023 onward will be considered on the basis of the situation to date, including any changes in the number of patients being examined.

Rare/Intractable diseases

In the field of rare/intractable diseases, analysis of approximately 2,500 cases will be carried out in FY2022 on the basis of the situation up to FY2021. As with the field of cancer, specific figures for the number of analyses from FY2023 onward will be considered on the basis of the situation to date.

(3) MHLW implementation system for whole genome analysis

(a) Expert Committee on Promotion of Whole Genome Analysis

The Expert Committee, which was established under the Science and Technology Division of the MHLW's Health Sciences Council, is the top decision-making body with regard to the promotion of whole genome analysis. The Expert Committee will hold discussions aimed at steady promotion of the present Action Plan and will check the progress of initiatives based on the Action Plan and make necessary decisions. In addition, the Expert Committee will hold consultations as necessary on review of the present Action Plan. These measures will ensure a highly workable review framework for the present project, in which the responsibility of the government is made evident.

The Expert Committee will continue to be the top decision-making body for the government's basic policy with regard to promotion of whole genome analysis even after the launch of the Project Implementation Organization.

(b) MHLW science research group

The MHLW science research group comprises experts well versed in the practical aspects of whole genome analysis, who will study the technical

aspects of specific operational methods relating to return of results to patients, the Analysis and Data Center, ELSI, etc. and will draft basic policy that will be submitted to the Expert Committee for discussion. In addition, if the Expert Committee requires technical advice on whole genome analysis, the MHLW science research group will submit its opinions to the Expert Committee.

In April 2022, new Project Implementation Preparation Office working groups were set up within the Science Research Group to examine organizational design and recruitment for the launch of the Project Implementation Preparation Office during FY2022, and to ensure that systems are in place for the launch of the office. The Project Implementation Preparation Office will develop a specific system for the launch of the Project Implementation Organization.

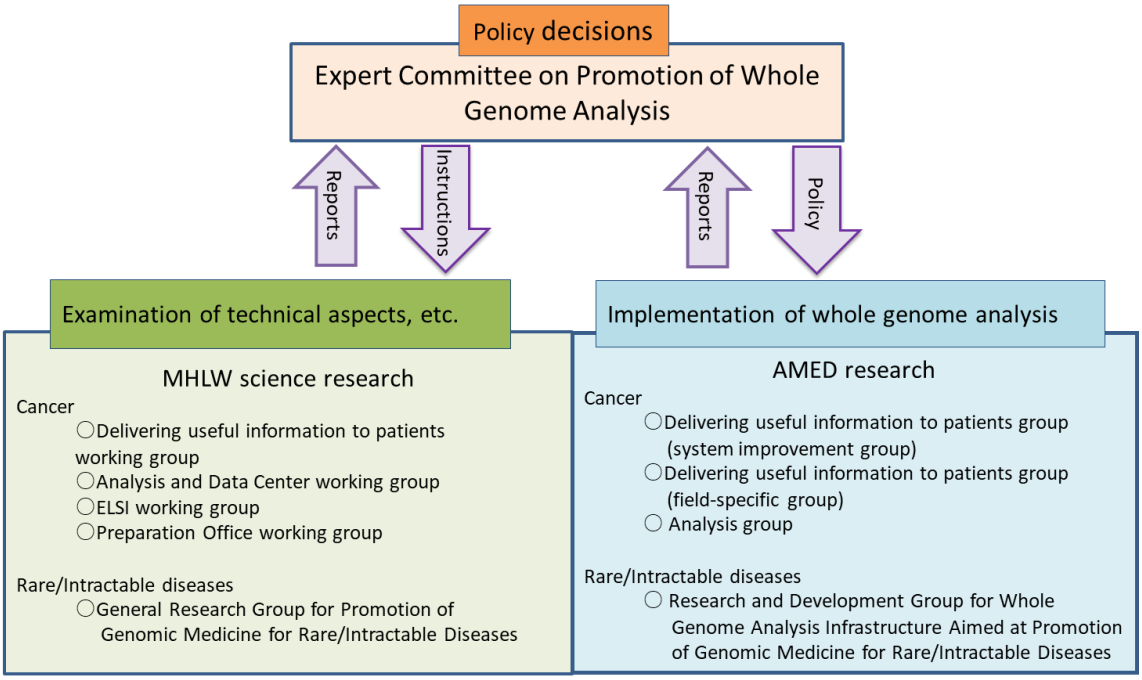
Following the launch of the Project Implementation Organization, the working groups within the MHLW science research group will cooperate as necessary for smooth transition to the advisory committee within the Project Implementation Organization.

(c) AMED research groups

The AMED research groups of Whole Genome Analysis will report on the status of analysis to the Expert Committee and will carry out research in line with the policy of the Expert Committee under appropriate progress management by AMED. Following the launch of the Project Implementation Organization, the return-to-patients group (system improvement group) will become a medical institution that returns results of the whole genome analysis to patients, and the return-to-patients group (field-specific group) and the analysis group will cooperate as necessary for smooth transition into the Academic Consortium and the Analysis and Data Center, respectively.

Figure 5. MHLW implementation system for whole genome analysis (before launch of the Project Implementation Organization)

MHLW Implementation System for Whole Genome Analysis



(4) Organizations that make up the project

(a) Medical institutions that return results to patients

To return the results of whole genome analysis to patients in an appropriate fashion, medical institutions that return the results to patients are required to have expert whole genome analysis personnel, a system for storage and management of samples, ELSI compliance, advanced medical functions, and systems for conducting clinical studies and trials.

Cancer

The main requirements expected of medical institutions that return results to patients are as follows.

The Expert Committee will examine and approve new medical institutions meeting the requirements once a year, and institutions that are approved will be added to the list of medical institutions that return results to patients the following fiscal year. In addition, the Expert Committee will evaluate the systems and performance of medical institutions that return results to patients once a year.

The Expert Committee will review the main requirements for medical institutions that return results to patients as necessary.

- Main requirements for medical institutions that return results to patients
  - The institution is a designated core hospital for cancer genomic medicine or a designated hospital for cancer genomic medicine.
  - The institution has medical staff engaged in cancer genomic medicine and has treatment systems that allow the return of results to patients.
  - Medical staff engaged in cancer genomic medicine receive mandatory training that includes whole genome analysis to improve their genomic literacy.
  - There are systems in place that can evaluate the analytical validity and clinical usefulness of the results of whole genome analysis. Specifically, there are multiple genome researchers who are proficient in the following items engaged in the project.
    - Confirmation of analysis data (including original data) and detection of different types of call error
    - Interpretation of genomic alterations and attaching clinical significance
    - Verification of genomic alterations by confirmation tests with a certain level of accuracy
  - There are systems in place for the appropriate storage and management of samples with the consent of patients.
  - The institution has an appropriate system in place for conducting clinical studies and trials, either on its own or in collaboration with other institutions, and has a certain level of results.

Rare/Intractable diseases

In FY2021, the Verification Project for Whole Genome Analysis of rare/Intractable Diseases (tentative name) carried out verification of methods relating to return of results to patients, with the cooperation of five medical institutions that have experts in the genome of rare/intractable diseases and are able to play a central role in

handling genome information and in the diagnosis and treatment of rare/intractable diseases. From FY2022 onward, the number of medical institutions cooperating in this verification will be increased in stages, and specific requirements for medical institutions conducting whole genome analysis of rare/intractable diseases will be discussed on the basis of the results of whole genome analysis conducted to date.

(b) Sequencing companies

Under the present Action Plan, uniform, high-quality analysis data will be collected using a standardized methodology. Technologies for which analysis methods have been established will be outsourced to companies that satisfy all of the following five conditions:

1) The company has an analysis base in Japan, and security is ensured by measures that include restricting the range of persons with access privileges, monitoring access, strengthening identity authentication (introduction of multi-factor authentication), data sanitization, and real-time detection of unauthorized access. Risk and security assessments are conducted on a regular basis by third parties, and the person in charge takes appropriate action if issues are identified.
2) The company has a proven record of sequencing more than a certain number of samples and is capable of sequencing large numbers of samples.
3) The company is a clinical laboratory, etc. that carries out accuracy management for genetic testing.[5]
4) The company is capable of sequencing that meets the specifications for international joint research with countries that are advanced in genome analysis.
5) From the perspective of obtaining uniform data, the company has multiple next-generation sequencers with a standardized system.

In addition, to ensure and to improve sequencing accuracy, sequencing companies are required to conduct their own regular quality evaluation and verification, and to also have external accuracy management.[6]

[5] The company must be ISO 15189 certified (Japan Accreditation Board), CAP-LAP certified (College of American Pathologists Laboratory Accreditation Program), or CLIA certified (Clinical Laboratory Improvement Amendments).

[6] During FY2022, external accuracy management will be conducted by the MHLW Science Research Group for promotion of whole genome analysis of cancer and establishment of a system for technology assessment related to whole genome analysis of each individual patient and its clinical application, a center for data analysis and storage, information security and patient confidentiality, and Ethical Legal and Social Implications (ELSI) (tentative name), which will use the sequencing accuracy, the relevant summary values, and the pre-mapping quality control values of the sequencing companies. After the Project Implementation Organization has been established, this organization will carry out external accuracy management of companies.

(c) Analysis and Data Center

The Analysis and Data Center will store, share, and support the utilization of sequencing results collected from sequencing companies and clinical information collected from medical institutions. The Analysis and Data Center is required to carry out the following four roles:

1) Analysis of genome data
2) Collection of clinical data
3) Utilization of data
4) Development of human resources

1) Analysis of genome data

- Genome database

  A genome database will be constructed in the Analysis and Data Center that can appropriately collect and store sequencing results from sequencing companies. This will need to have systems compatible with international research, while ensuring data security by enabling efficient data transfer via cloud computing.

- Unified pipeline

  A unified pipeline will be created and operated to carry out primary analysis, from FASTQ files to the creation of VCF files, by means of a unified method.

- Report preparation system

  A system will be established to prepare reports that are easy for both physicians and patients to understand, which will include the clinical significance after primary analysis and clinical trial information that reflects the clinical information of individual patients, etc.

- Advanced cross-sectional analysis

  In cooperation with industry and academe, the Analysis and Data Center will use the collected genomic data and clinical information to conduct advanced cross-sectional analysis by domain or across domains to establish systems able to promptly return new findings to patients.

2) Collection of clinical data

- Clinical data collection system

  A system will be created within the Analysis and Data Center for management of clinical information collected from multiple medical institutions in a form that will allow comparison.

  During FY2022, the AMED research group will take the lead in examining the standardization of clinical information at multiple medical institutions and the establishment of a clinical information collection system.

  Following the launch of the Project Implementation Organization, the aim will be to construct a system that extracts necessary data from electronic medical records using an application programming interface (API), with no need to manually transcribe and input data into the system, so that the Analysis and Data Center can access data directly.

3) Utilization of data

Systems will be put in place to enable rapid, fair, and safe utilization of the data collected by the present project (genome data, clinical information, etc.) to promote drug discovery and research and development of diagnostic technologies. During FY2022, the Project Implementation Preparation Office will take the lead in formulating data utilization policy and the rules on data sharing (data sharing policy), as well as examining the establishment of a Data Utilization Review Committee, and will construct a data sharing system (research support system) and commence operations on a trial basis, with the aim of starting full data-sharing during FY2023.

The Analysis and Data Center will use the data sharing system based on the data sharing rules, etc. drawn up by the Project Implementation Preparation Office, and will conduct specific operations related to data utilization.

- Data utilization policy and rules on data sharing (data sharing policy)

The Project Implementation Preparation Office will formulate the data utilization policy and the rules on data sharing (data sharing policy) during FY2022.

The data utilization policy will set out the basic concepts for data utilization and items relating to the Data Utilization Review Committee, to ensure fair, smooth use of data. The rules on data sharing (data sharing policy) will set out necessary items regarding data sharing with the Industry Consortium and the Academic Consortium to ensure prompt utilization of the data that are collected.

The Analysis and Data Center will share data with the Industry Consortium and the Academic Consortium in accordance with the data sharing rules to promote utilization of the data in industry and academe. In addition, the Analysis and Data Center will manage the starting point for the data sets in both consortiums (the point at which about 100 cases of data are registered, excluding rare cancers), and after the starting point is reached, the center will manage the overview of data, applications for use, granting access rights, the status of use, etc. by consortium members.

When members of either consortium need to apply to use the data, their applications will be reviewed and approved by the Data Utilization Review Committee, which will be set up within the Project Implementation Organization.

In addition, the Analysis and Data Center will manage the restricted period of the collected data (a period greater than 24 months from the starting point but not exceeding 30 months), and will register data that have passed the restricted period in a public database.

- Data sharing system (research support system)

On-premise and cloud data sharing systems will be created to ensure seamless utilization of the accumulated data, and the Analysis and Data Center will share data with users belonging to either of the consortiums via these systems. Mechanisms will be examined to promote active utilization of the data, such as displaying the clinicopathological significance of gene mutations and introducing functions to support clinical studies at medical institutions.

4) Development of human resources

Operating, maintaining, and improving the Analysis and Data Center will require the involvement of personnel with a wide diversity of specialized skills, and systems will therefore be examined to develop and secure human resources for the Analysis and Data Center through collaboration with graduate schools and personnel exchanges with academe and industry.

In particular, for the development of experts in information analysis and artificial intelligence for genome analysis, educational seminars, etc. will be held to widely disseminate knowledge related to genome analysis in collaboration with the Project for Development of Human Resources Relating to Whole Genome Analysis of Cancer (tentative name) being implemented by the MHLW, and will also develop human resources through on-the-job training (OJT) at divisions that conduct genome analysis.

(d) Centralized Sample and Information Management Division

A Centralized Sample and Information Management Division will be set up within the Analysis and Data Center for centralized management and utilization of whole genome data, clinical information, samples, and information on samples. A prototype system for the development of the division's centralized management system will be completed in FY2022.

In addition, there is a need to establish a sample transfer system that allows third parties not only to use genome data, but also, when necessary, to carry out omics analyses that combine samples of tissue, etc. (surplus samples and residual samples) with genome information and clinical information. The sample management system for this will be created and operated as follows.

- Sample management system (Centralized Sample Management Center) and rules for storage and management

  For samples from new patients, a system will be established to enable collective management using existing facilities. At the same time, samples can be stored at individual medical institutions as long as the quality of the storage and management is the same as that of the collective management, and if necessary, a system is in place allowing samples to be transferred through the same procedures.

  In addition, a system will be established to enable the Centralized Sample Management Center to grasp the type of samples, the amount remaining, and the contents of consent (if sample transfer only to industry is possible, etc.), which will also include samples stored at individual medical institutions.

  These systems will be established on a trial basis during FY2022, with the aim of full-scale operation from FY2023 onward.

  In addition, standard operating procedures for sample storage and management rules (detailed SOPs for each organ) will be developed during FY2022 with the cooperation of experts from the Japan Registered Clinical Laboratories Association, to comply with international standards.

(e) Industry Consortium, Academic Consortium

Consortiums in which industry and academe can participate will be formed and the industry-academic collaborative utilization of data will be promoted to

facilitate research and development of drug discovery and diagnostic technologies and deliver the results to patients promptly.

1) The role of industry

The aim is to develop new diagnostic technologies and therapeutic drugs based on data obtained from whole genome analysis. For this, an Industry Consortium hosted by industrial circles will be established to enable industry to participate proactively in the project, including the data collection process, from the very beginning.

The main objective of the Industry Consortium will be to promote research and development of drug discovery and diagnostic technologies using whole genome analysis data. The consortium will be organized by the pharmaceutical and other industries as an organization in which many different medical and non-medical industries, as well as venture capital companies, can participate, and the aim is to launch it during FY2022. In addition, specific operating rules that will include setting incentives to utilization the data according to personnel, technological, and financial cooperation from each company will be decided.

2) The role of academe

The aim is to promote research into genomic medicine on the basis of the data obtained from whole genome analysis. For this, an Academic Consortium hosted by the academic world will be established as a nationwide Japanese scholarly institution to hold proactive academic discussions regarding whole genome analysis.

The Academic Consortium will be expected, in return for sharing data relating to whole genome analysis and being granted the authority to make widespread utilization of the data, to play the roles of establishing groups of experts in different fields, holding discussions on the clinical and pathological significance of findings on new mutations identified as a result of advanced cross-sectional analyses, and after accumulation of a range of necessary data, giving judgments by expert panels on whether the data merit being returned to patients.

Specific operational rules for the Academic Consortium, such as reviewing applications for membership on an organization-by-organization basis, registration of affiliated researchers, and coordination of joint research, will be established with the aim of launching the consortium in FY2022.

3) Support for industry and academe by the Project Implementation Organization

The Project Implementation Organization will provide operational support to the Industry Consortium and the Academic Consortium to facilitate mechanisms for the prompt return of new knowledge to the general public.

Specifically, the Project Implementation Organization will set up a division to provide the Industry Consortium with operational support for further promotion of development projects using the data. This division will give support for data utilization, intellectual property management, proposals for new research, collaboration with the Academic Consortium, and matching companies for collaboration.

The Project Implementation Organization will also set up a division to provide the Academic Consortium with operational support for further promotion of research and development using the data. This division will give support for new research proposals, collaboration with the Industry Consortium, and matching researchers for collaboration.

4) Arranging data use fees and intellectual property

The basic rules for data use fees and their collection mechanism will be decided during FY2022 through discussions of the Expert Committee. In principle, it would be preferable for the government to manage the data quality at a guaranteed level within the framework of its budget, with other operational costs borne by users.

The ownership of intellectual property obtained through the use of data and rules for handling data will be defined in the data utilization policy and rules on data sharing (data sharing policy).

(f) Project Implementation Organization

The Project Implementation Organization must be a highly autonomous organization with strong governance and full transparency and accountability. To ensure this, the Project Implementation Organization will have public status, and systems will be put in place for board members comprising a broad range of personnel recruited from industry and academe to be able to make flexible and prompt operational decisions based on the latest knowledge.

The Project Implementation Preparation Office will be set up within the Japan Health Research Promotion Bureau (JH) during FY2022 and will examine the organization and its structure. With the MHLW taking a leading role, the most suitable organizational form for the Project Implementation Organization will be decided by around FY2023, for launch of the organization in FY2025.

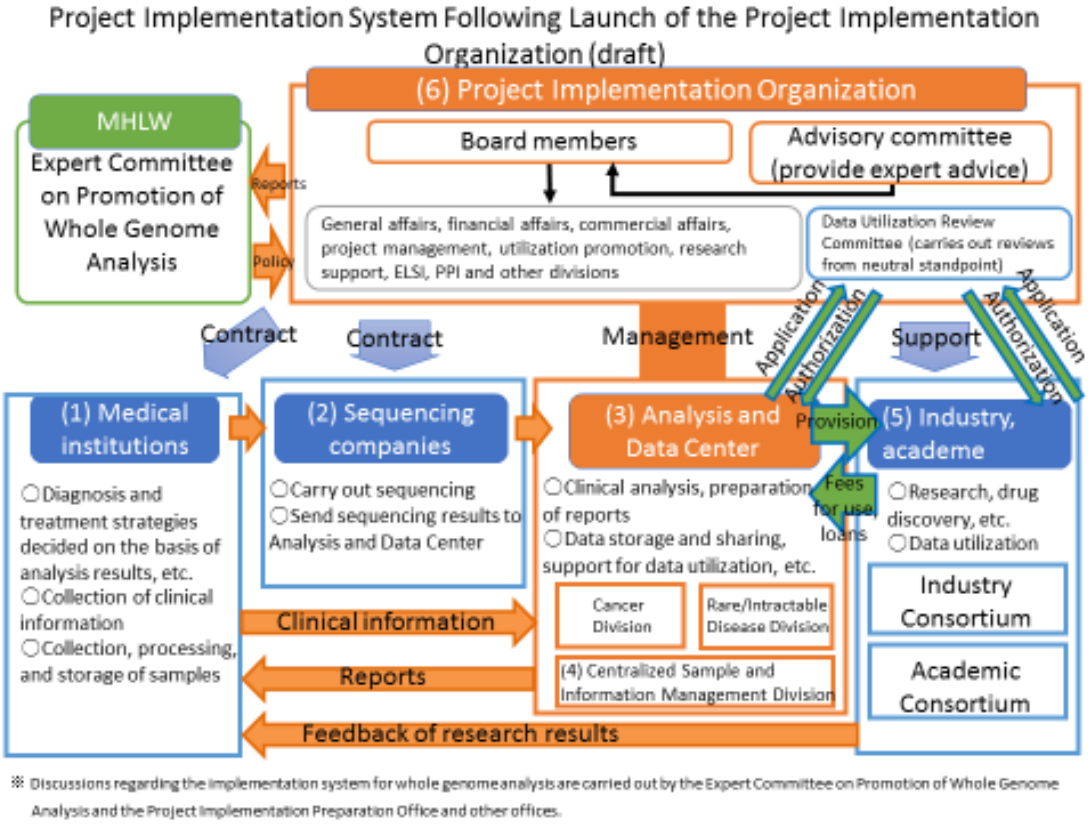
The draft arrangement of divisions and functions of the Project Implementation Organization are as follows.

1) Establishment of board members, including the director, mainly comprising outside experts
2) Establishment of various advisory committees that provide expert advice to the board members:
   - Advisory committee on return of results to patients
   - Advisory committee on genome analysis
   - Advisory committee on ELSI
   - Advisory committee on patient and public involvement, etc.
3) Establishment of divisions necessary for project implementation (organizational operation divisions, project divisions) after examination by the Project Implementation Preparation Office

   Examples of organizational operation divisions:
   - General Affairs Division
   - Finance Division
   - ELSI Division
   - Patient and Public Involvement Division, etc.

   Examples of project divisions:
   - Project Administration Division (management of sequencing, samples, Analysis and Data Center, etc.)
   - Scientific Research Support Division (support for research, academe, etc.)
   - Commercial Affairs Division (support for industry, etc.), Utilization Promotion Division, etc.

Figure 6. Project implementation system following launch of the Project Implementation Organization (draft)



※ Discussions regarding the implementation system for whole genome analysis are carried out by the Expert Committee on Promotion of Whole Genome Analysis and the Project Implementation Preparation Office and other offices.

6. Policy and Details of Project Operations

(1) Return of results to patients

(a) Outline of the system for return of results to patients

To ensure that patients can receive high-quality medical care based on the results of whole genome analysis regardless of where they live, return of results to patients will have to be carried out through standardized methods and also on the basis of analysis and report formats that take into account the particular characteristics of the fields of cancer and rare/intractable diseases.

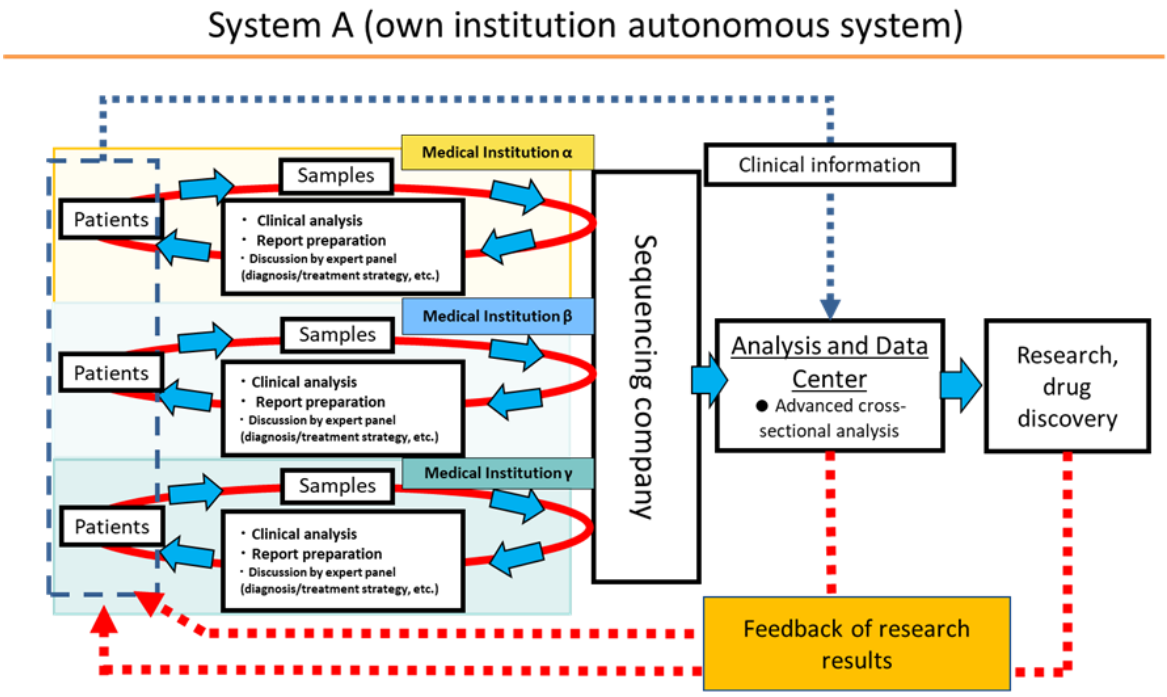
1) System A (own institution autonomous system)

In System A, clinical genome analysis[7] and preparing reports are carried out at the medical institution (or an affiliated medical institution). Specifically, after sequencing is carried out by the sequencing company, the medical institution (or an affiliated medical institution) conducts clinical genome analysis of the FASTQ data that are generated and prepares a report, and its expert panel carries out discussions. The medical institution can then provide patients with personalized medical care.

Medical institutions under System A cooperate with each other to improve the accuracy of their clinical genome analysis and report preparation. In addition, the accuracy of reports prepared by the medical institution (or an affiliated medical institution) is evaluated by the Analysis and Data Center to guarantee the quality of reports.

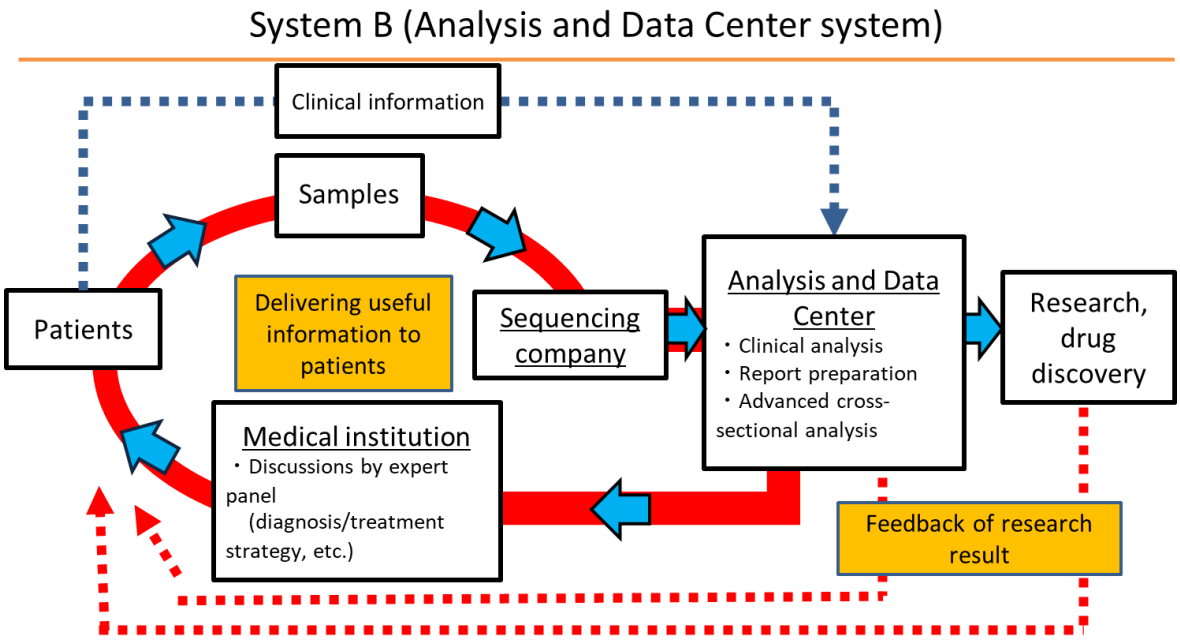
[7] This refers to the analysis (mapping, variant calling) and determination of clinicopathological significance of FASTQ data generated after sequencing by a sequencing company.

Figure 7. System A (own institution autonomous system)



2) System B (Analysis and Data Center system)
In System B, clinical genome analysis and report preparation are carried out at the Analysis and Data Center. Specifically, after sequencing is carried out by the sequencing company, the FASTQ data that are generated are sent to the Analysis and Data Center, where clinical genome analysis is carried out and reports are prepared. The medical institution uses these reports as the basis for discussion by the expert panel, and then provides patients with personalized medical care.

Figure 8. System B (Analysis and Data Center system)



3) Points to note when implementing the system for return of results to patients
With both systems, the following points should be noted when returning results to patients.

- There may be times when clinical genome analysis is not fully accurate; therefore, when performing actual therapeutic interventions on patients, confirmation tests should be carried out with a different analysis method that guarantees a certain level of accuracy (different types of genetic test, companion diagnostics, cancer gene panel tests, etc.).
- FASTQ data that are created following sequencing by a sequencing company are to be sent to the Analysis and Data Center within 2 weeks.

4) Approach to return of results to patients in the different fields

Cancer
Return of results to patients in the field of cancer will be carried out with both systems. The aim is to create the optimum framework by appropriately aligning the systems so that, while specialist analysis is carried out under System A, standardized analysis methods are expanded at nationwide level under System B.

- FY2021: For System A, research on return of results to patients commenced at three medical institutions. For System B, research commenced with 6 fields (groups).
- FY2022: The research system put in place in FY2021 will be maintained, while medical institutions with systems in place will be added to those implementing System B, taking their specialties and regional characteristics into

consideration. In addition, appropriate creation of System B will be examined.

- FY2023 onward: System A will be maintained for medical institutions carrying out specialist analysis, and System B will be further expanded.

Rare/Intractable diseases

Uniform, high-quality information will be provided to patients through centralized analysis and report preparation (System B). In addition, a system will be established to provide appropriate explanations to patients based on the characteristics of rare/intractable diseases and the results of the Verification Project for Whole Genome Analysis of Rare/Intractable Diseases (tentative name). For rare diseases with small numbers of patients in particular, a joint international framework for collecting cases and sharing data will be established to enable early diagnosis.

(b) Details of the system for return of results to patients

The basic approach to the various processes in the system for return of results to patients (obtaining informed consent from patients; collection, processing, transfer, and management of samples; sequencing; clinical analysis and preparation of reports at the medical institution or the Analysis and Data Center; and discussion of diagnosis, treatment strategy, etc. by the expert panel at the medical institution) are set out below.

1) Obtaining informed consent from patients

Cancer

The Draft Model for the Action Plan for Whole Genome Analysis Explanatory Document (hereafter, Draft Model),[8] which has been approved by the Expert Committee, will be used for the process of obtaining informed consent from patients who are to participate in the present project. Participation in the present project will be permitted only if a patient has consented to all items included in the Draft Model.

The draft Model was drawn up for use in the AMED project Practical Research for Innovative Cancer Control, and it provides a model of the items that need to be included in the explanatory document as common items relating to the Action Plan for Whole Genome Analysis. It should therefore be noted that the Draft Model alone does not cover all the explanatory items stipulated by the ethical guidelines, and individual medical institutions will need to check carefully whether the explanatory document prepared using the Draft Model contains the necessary items and also whether there are any inconsistencies between the parts of the explanatory document that were inserted from the Draft Model and the other parts.

The various different research groups will give their feedback on the Draft Model, and a standardized Informed Consent Form (ICF) will be drawn up during FY2022. The standardized ICF will be used from FY2023 onward.

Rare/Intractable diseases

Consent will be obtained using an Informed Consent Form based on the Draft Model, with the addition of any extra explanation needed for individual research studies.

- Introduction of information and communication technology (ICT) into the informed consent process

  The introduction of a system for obtaining informed consent (IC) by electronic methods (e-consent) will be examined when the appropriate tools and communication environment are in place. The advantages and points to note regarding use of e-consent must be fully understood.

  In addition, for the introduction of e-consent into medical settings, it will be necessary to determine how hospitals transfer and manage IC information and their means of cooperation with the Analysis and Data Center by identifying specific problems in operability in medical settings and

[8] Draft Model for the Action Plan for Whole Genome Analysis Explanatory Document https://www.mhlw.go.jp/content/10901000/000904765.pdf

studying concrete proposals for commencement. Furthermore, rather than simply computerizing the process of explanation, obtaining consent, and storing the records, the introduction of e-consent should be examined together with the creation of an online platform in which the patient can take a lead role in managing information.

Advantages of e-consent

- Increased efficiency through easier modification, tracking, and centralized management of the content of the ICF.
- The use of explanatory contents matched to the pace of the individual, review and confirmation of consent details, and ease of sharing with family members improve understanding by patients.
- Increase in the rate of participation in the project.
- Recontacting patients is more convenient.

Points to note regarding use

- Difficult for patients who are unfamiliar with ICT to understand the content and express their intentions.
- Fewer opportunities for questions than when patients are dealt with in person, which can often impede attentive patient care.
- Difficulty in handling cases and confirming requirements when a proxy is needed.

Points for examination when introducing e-consent

- Methods of IC information transfer and management at the hospital
- Means of cooperation with the Analysis and Data Center
- Creation of an online platform in which the patient can take a lead role in managing information

2) Collection, processing, transfer, storage, and management of samples

Collection, processing, transfer, storage, and management of samples will be conducted using standardized methods that comply with international norms, so that specimen quality can be assured. In addition, all processes will be completed in Japan.

- Collection, processing, and transfer of samples (transfer within facilities, transport between facilities)

  Collection and processing of samples will be carried out in line with the Rules for Handling Pathological Tissue Samples for Genome Research stipulated by the Japanese Society of Pathology. Standard operating procedures for each research group and each field will be drawn up during FY2022 to ensure that collection and processing are carried out according to standardized methods that comply with international norms.

- Storage and management of samples

  As a general rule, sample storage and management will be carried out as centralized management at the Centralized Sample Management Center managed by the Analysis and Data Center. The methods for

storage and management of samples will be in line with the Rules for Handling Pathological Tissue Samples for Genome Research stipulated by the Japanese Society of Pathology. Standard operating procedures will be drawn up during FY2022 to ensure that storage and management are carried out according to standardized methods that comply with international norms.

Strict compliance with the standard operating procedures is also obligatory when outsourcing to a sequencing company or storage in a cooperating medical institution is desired.

3) Sequencing

Samples submitted by medical institutions will be sequenced by sequencing companies. To ensure standardization of sequencing methods in compliance with international norms, sequencing companies will have to draw up standard operating procedures and submit an outline to the Analysis and Data Center during FY2022.

- Quality control

  Sequencing companies will evaluate the quality and quantity of the data before human genome mapping on the basis of the criteria established by the Expert Committee to obtain data that meet the reference values. Sequencing companies will be under the obligation to cooperate with external accuracy management carried out by the MHLW Science Research Group for promotion of whole genome analysis of cancer and establishment of a system for technology assessment related to whole genome analysis of each individual patient and its clinical application, a center for data analysis and storage, information security and patient confidentiality, and Ethical Legal and Social Implications (ELSI) (tentative name), using the sequencing accuracy values, the relevant summary values, and the pre-mapping quality control values of sequencing companies.

4) Clinical genome analysis and report preparation at medical institutions or the Analysis and Data Center

The Analysis and Data Center will carry out clinical genome analysis of FASTQ data sent by outsourced sequencing companies, and prepare reports. With new mutations, advanced cross-sectional analysis will be performed using the FASTQ data, etc. to determine their medical significance.

The systems needed for this are as follows.

- Information management

  Centralized management system

  The Analysis and Data Center will establish a centralized management system for centralized management and utilization of whole genome data, clinical information, samples, and information on samples. The centralized management system will be capable of managing the status of patients' informed consent, information on samples, information on data transfer, quality control (QC) results, the progress of genome

analysis, the status of clinical information collection, the status of return of reports to medical institutions, etc.

The system infrastructure will be put in place and a prototype will be completed in FY2022. The system will be upgraded in stages starting in FY2023, with completion in about three years.

The following points should be noted in relation to the centralized management system.

- From the perspective of preventing confusion over samples, information on samples will always be managed in such a way that it can be linked to the genome information and clinical information databases.
- A common ID format for use when data are released will be examined.
- With regard to the consent from patients, the implementation of technology to ensure traceability, such as confirmation of identity when consent is obtained, deletion of data when consent is withdrawn, etc. will be examined
- The system will be able to manage the starting point of each data set (the point at which data from around 100 cases are registered [excluding rare cancers]) in the Industry Consortium and the Academic Consortium, and to enable centralized management of overview of data, applications for use, granting of access rights, and status of use, etc. by consortium members after the starting point is reached.
- The restricted period of the genome data and basic clinical information that have been collected (a period greater than 24 months from the starting point but not exceeding 30 months) will be managed, and data that have passed the restricted period will be registered in a public database.
- A mechanism will be established to enable samples from new patients to be managed centrally using existing facilities (Centralized Sample Management Center).
- A mechanism will be examined to enable information on samples to be shared when necessary, such as when a patient participates in a clinical trial.
- An automated system for storing and retrieving samples will be examined.
- Security measures, etc. will be constantly upgraded.
- Linkage of the centralized management system to the system for utilization of samples is essential, so that samples can be shared with academe or industry if an application from a user is approved by the utilization review committee and there is a request for utilization of samples from the centralized management system. Even when samples are outsourced to a sequencing company or stored at a cooperating medical institution, cooperation in the utilization of samples by the centralized management system is essential. The utilization of samples will be conducted in accordance with the data utilization policy and rules on data sharing (data sharing policy).

Various requirements for data management and system creation

Data management and system creation within the Data and Analysis Center are important infrastructure that relate directly to the treatment of patients. It is therefore important to control access to the data, manage the logs, store the data safely using secret sharing technology, and implement a security detection and decision-making process. It will thus be necessary to satisfy the following requirements relating to data management, data storage, system development and environment creation, and security.

- Data management
  - Access to clinical information and genome data is strictly managed not just by logs, but also by management of data users (who is given access to which data, and when, etc.).
- Data storage
  - Data are stored in a form that is sufficiently secure. This requires investigation of multiple technologies such as secret sharing technology to ensure use of the optimal technology, as well as a constant grasp of issues such as interoperability and scalability.
  - Both on-premise and partial cloud storage will be verified in FY2022 for storage of genome data. From FY2023 onward, storage will be implemented as a hybrid operation that takes into account the strengths and weaknesses of both of them.
- System development and environment creation
  - An analysis workflow from primary analysis to secondary analysis and diverse analysis resources will be provided using a hybrid on-premise and cloud framework.
  - Secure systems and a secure network environment will be put in place between medical institutions, sequencing companies, and the Analysis and Data Center. The Analysis and Data Center will be based on the use of servers physically located in Japan.
  - The information system will be created in a highly scalable form so that it is not dependent on any specific technology and can constantly incorporate new technologies.
- Security requirements
  - The cloud is to be registered or in the process of registration with ISMAP (the Japanese government's Information System Security Management and Assessment Program). Firewalls will be deployed in the cloud for security monitoring, detection, and blocking.
  - The system must be capable of monitoring and reporting suspicious behavior through storage of data access logs and correlation analysis rules using these logs. Methods will be examined to automatically perform periodic optimization of correlation analysis to prevent false positives.
  - For security detection, the payload (contents of packet transmission passing through the network) of cloud services has log-based monitoring that cannot be checked or detected, and it is therefore not fully real-time monitoring. Methods to avoid this will be examined.
  - Automation of the security decision-making process to quickly isolate the system, take countermeasures, and make a report after an incident occurs, as well as methods to avoid false positives and over-detection, will be examined.

- Use of cloud computing (external storage, external access) is to be considered in areas that handle workflows directly related to clinical practice. In addition, the requirements of the guidelines for safety management of medical information systems (security design for web applications such as multi-factor authentication for data access, virtual private network [VPN], and cookie acquisition, correction of vulnerabilities, sanitization, etc.) will be met to ensure thorough project security and data quality.

○ Clinical genome analysis and advanced cross-sectional analysis
The Analysis and Data Center will collect sequencing data, etc. from sequencing companies to build a genome database and will also automatically collect clinical information from medical institutions using an API to build a clinical information database. In addition, as well as analyzing sequencing data by means of a unified pipeline, the Analysis and Data Center will collaborate with industry and academe to conduct advanced cross-sectional analyses to clarify the medical significance of new mutations.

Construction of the genome database
The sequencing data collected from sequencing companies (FASTQ data), the bam (or cram) files generated from the unified pipeline, and the mutation information (VCF files) will be compiled into a database at the Analysis and Data Center.

Systems and infrastructure will be created that are capable of storing the results of primary analysis (mapping, variant calls) and quality checks of whole genome sequencing data (tumor 120×, normal 30×) from around 10,000 cases per year.

Collection of clinical information
Information relating to outcomes such as prognosis or mortality and information on drugs or therapeutic methods that correspond to mutations will be important clinical information for collection. In FY2022, the data format of clinical information will be standardized as much as possible at the various medical institutions, and the input format (template) will be standardized. At the same time, there will be development aimed at the future implementation of an automatic clinical information collection method using APIs. A cloud computing system will also be constructed to create a database of the clinical information that is collected.

- API automatic collection method
Clinical information will be automatically collected using APIs. The aim is ultimately to lessen the burden on physicians of clinical data collection by using AI for natural language processing of the free-written content in templates and medical records. An open data source API for transferring clinical information from electronic medical records and an API for data access will be developed in FY2022. In this regard, the following points should be noted.

- Rather than being limited to a particular electronic record vendor, prototyping of programs to convert diverse electronic record data held by multiple medical institutions to a standard format will be carried out.
- Push type and pull type systems for sending and receiving clinical information will be compared, and issues or areas for improvement will be verified through creation of a prototype.
- The security conditions for data management and system construction will be met.

- Clinical information database
  The clinical information collected by the Analysis and Data Center will use a cloud service. A prototype system will be constructed in FY2022, and any issues or points for improvement will be verified, with the aim of implementation from FY2023 onward. In this regard, the following points should be noted.
  - For the method of description within the clinical database structure, multiple methods will be compared, and one that meets the performance requirements will be selected.
  - For the unstructured database, advanced methods will be compared and verified, taking into consideration the patterns of utilization.
  - The database will be of a type that can ensure user-friendliness for the expert panel, as well as search and other functional capabilities.
  - The database will satisfy the security requirements for data management and system construction.
  - The clinical information items that are collected will need to encompass the information required for various applications, including care of the corresponding patient, clinical research, and drug discovery. Items common to cancer and rare/intractable diseases will be collected, as well as additional items of information that depend on the characteristics of the disease.[9]
  - The system will be capable of adding additional items of clinical information as necessary.

  Unified pipeline
  A unified pipeline will be constructed to centrally and collectively perform the common parts of the analysis process carried out by many researchers. Commonly used analysis tools and parameters will be selected, based on domestic and international trends, to ensure future data sharing and compatibility with various databases. In addition, periodic reviews will be conducted.

  For the cloud-based genome analysis system infrastructure, multiple services will be selected from the cloud services that can meet the performance, availability, security, and scalability requirements of the unified pipeline. These will be compared in terms of performance and cost, any other issues will be identified, and a prototype will be built during FY2022. The configuration will be upgraded in stages from

[9] Table of Items of Clinical Information Collected in Relation to Cancer Whole Genome https://www.mhlw.go.jp/content/10901000/000833423.pdf

FY2023, and completed within approximately three years. Modifications to security measures, etc. will be made on an ongoing basis.

Advanced cross-sectional analysis
An analysis infrastructure (including the use of AI) that can handle secondary analysis, including advanced cross-sectional analysis, will be constructed. A prototype system for researchers to perform secondary analysis will be built and verified in FY2022. This will use advanced cloud computing technologies, and research studies will be carried out to improve analysis performance. In addition, the data format will be standardized to enable use of AI, and AI model development leading to improvements in the accuracy of analysis based on the characteristics of diseases and return of results to patients will be promoted in collaboration with the different research groups. The quality and performance of these AI models will be evaluated.

- Report preparation
  Report preparation system
  The Analysis and Data Center will create a prototype report preparation system during FY2022 in collaboration with medical institutions that return results to patients, with the aim of implementation from FY2023 onward. In this regard, the following points should be noted.
  - APIs for extracting genome mutation and clinical information will be put in place, and a framework that allows the development of a report preparation system without being limited to one particular vendor will be examined.
  - This examination will identify issues such as the division of roles between the Analysis and Data Center and medical institutions, responsibility demarcation points, operational procedures such as visual confirmation of reports and response to inquiries, and the personnel system required for operation.
  - Issues will be identified, such as the establishment of a knowledge database to effectively use mutation information and return it to patients, as well as the personnel system and the expertise of staff in charge of development required for updating this database.
  - For variants of unknown significance (VUSs), evaluation methods in various forms, including machine learning and AI, will be examined.
  - For clinical trials, a system will be examined that can search for information not only in Japanese databases (University Hospital Medical Information Network Clinical Trials Registry [UMIN-CTR], Japic Clinical Trials Information [JapicCTI], Japan Medical Association Center for Clinical Trials Clinical Trials Registry [JMACCT-CTR], Japan Registry of Clinical Trials [jRCT]), but also in the US ClinicalTrial.gov database.

  The requirements for the reports themselves are as follows.
  - The report will set out the clinical significance of pathogenic variants and will include clinical trial information, etc. that reflects the clinical information of individual patients.

- With regard to setting out the clinical significance of the pathogenic variants, reference will be made to the four-tier classification of somatic mutations put forward by the American Society of Clinical Oncology (ASCO), the College of American Pathologists (CAP), and the Association for Molecular Pathology (AMP), and to the germline mutation pathogenicity classification guidelines of the American College of Medical Genetics and Genomics (ACMG) and the AMP. Reflection of these classifications in reports will be examined.
- Database information (ClinVar database provided by the National Center for Biotechnology Information [NCBI], Catalogue Of Somatic Mutations In Cancer [COSMIC], etc.) should also be provided with ease of use by the expert panel in mind.

Furthermore, after the report is prepared, it should be divided into different levels for stepwise return to patients, such as: (1) the level of existing gene panel tests; (2) the level of whole exome analysis; and (3) the level of structural aberrations and research analysis beyond the known range.

5) Expert Panel discussions and return of results to patients at medical institutions

Cancer

Starting in FY2022, the analysis results of whole genome analysis will be returned to patients, mainly by the designated core hospitals for cancer genomic medicine and designated hospitals for cancer genomic medicine that have in place the main requirements for medical institutions that return results to patients (see (4) (a)).

- Expert panel

  Since reports on the results of whole genome analyses are expected to contain more content relating to secondary findings than reports on cancer genome panel tests, medical institutions that return whole genome analysis results to patients need to have systems in place that can judge such content appropriately.

  Therefore, for the functions of expert panels in cancer genomic medicine, in addition to the requirements indicated in the “Requirements for Implementation of Expert Panels” (Notice by the Director of Cancer and Disease Control Division, Health Service Bureau, MHLW), expert panels at medical institutions that return the results of whole genome analysis to patients are also required to have the participation of experts who can judge the scientific validity, interpretation, and significance of the results of whole genome analysis.[10]

- Confirmation tests

[10] For example, this assumes full-time staff, etc. who have taken introductory- and advanced-level courses of the MHLW Project for Development of Human Resources Relating to Whole Genome Analysis of Cancer and have received certification as experts in bioinformatics, or full-time physicians who have authored peer-reviewed papers in English related to whole genome analysis.

When therapeutic interventions are actually carried out on patients, confirmation tests will be carried out using other analysis methods with guaranteed accuracy (different types of genetic tests, companion diagnostics, cancer gene panel tests, etc.).

- Evaluation of the results of whole genome analysis
  As a technical issue, the analytical validity of analysis results obtained by whole genome analysis is unknown at this stage. The analytical validity and clinical usefulness of the results will therefore be evaluated by comparing them with already confirmed cancer gene panel tests, etc. to consider how the confirmation tests and whole genome analysis should be conducted to better benefit patients.

(2) Utilization

It would be desirable to promptly put in place a system allowing fair, safe utilization of data and samples collected through the present project to promote research and development for drug discovery and diagnostic technologies.

The Analysis and Data Center will share data with the Industry Consortium and the Academic Consortium on the basis of the data utilization policy and the rules on data sharing and will also promote utilization of the data using the data sharing system (research support system). In addition, a Data Utilization Review Committee established within the Project Implementation Organization will review applications from users for uses that require an application to be made, such as detailed analysis, and will decide whether to grant a license for use.

The Project Implementation Preparation Office will formulate the data utilization policy and the rules for data sharing, hold discussions on establishment of the Data Utilization Review Committee within the Project Implementation Organization, and conduct pilot operations during FY2022, with the aim of commencing full-scale data sharing during FY2023.

(a) Data utilization policy
The data utilization policy stipulates the following for fair, seamless utilization of data.

- Basic concepts for data utilization
  - Utilization of the data shall be limited to academic research, development of pharmaceutical products, etc., and utilization for preventive purposes based on scientific evidence.
  - The users shall be domestic and foreign companies and academic research institutions belonging to the Industry Consortium or Academic Consortium. However, for use outside Japan, users must be from countries or regions with systems to protect personal information that are recognized as being of a similar level to those in Japan.
  - In the event that a user violates the data utilization policy, the Project Implementation Organization may take measures such as publicizing the user’s name, suspending their license to use data, refusing new applications to use data, seeking an injunction, or claiming compensation for damages.

- Data users shall not provide or resell, etc. the relevant data to any third party other than themselves unless the user imposes on the third party the same confidentiality obligations that are imposed on the user by the data utilization policy, the rules on data sharing, and the licensing agreement, and provides the data within the scope of the specified purposes of use.
- In principle, it would be preferable for the government to manage the data quality at a guaranteed level within the framework of its budget, while other operational costs are borne by users.

- Data Utilization Review Committee

  The Data Utilization Review Committee will be established within the Project Implementation Organization and will review applications for the utilization of data to ensure fairness in granting licenses for data use to applicants. The Data Utilization Review Committee will include both men and women, and the members shall be subject to appropriate conflict of interest (COI) management. They will also be bound by the obligation of confidentiality with respect to information they obtain in their capacity as members.

  Applications for data use will be reviewed in accordance with the following review items.

  - The purpose of use is academic research, development of pharmaceutical products, etc., and use for preventive purposes based on scientific evidence. The user may not use the data for any purposes other than those for which authorization was received from the Data Utilization Review Committee.
  - The data do not permit identification of individuals or blood relatives, or confirmation of the existence or non-existence of blood relatives, except in cases where the purpose is to return results to patients, such as participation in clinical studies. The research does not have the potential to disadvantage individuals, small groups of people, or specific geographical regions.
  - The research plan is scientifically valid, and the scope of the data to be used is appropriate.
  - The applicant has sufficient experience or ability to carry out the research plan.
  - The applicant has sufficient research facilities and a management system for the storage and disposal of information.
  - Other items that the Data Utilization Review Committee members consider necessary.

- Other items
  - Intellectual property rights
    - Intellectual property and intellectual property rights arising from the use of the data will belong to the user who created them.
  - Publication
    - Users may publish the results of research using the data.
    - Clinical information of patients within these data may be included in articles, etc. to the extent necessary for the publication of the results.
    - Users must give due consideration to the possibility that the published material may lead to personal identification. If there is a possibility that

the published material may lead to personal identification, this must be reported to the Data Utilization Review Committee for reviewing again before publishing.

- Confidentiality
  - Data users shall manage the information they receive from the relevant data as confidential information, and they may not disclose or provide this information to third parties unless it is disclosure or provision to an outsourcing company, and the user imposes on the third party the same confidentiality obligations that are imposed on the user by the data utilization policy, the rules on data sharing, and the licensing agreement, and provides the data within the scope of the specified purposes of use.
  - When handling the data, the user shall take all reasonable safety measures with regard to risks such as unauthorized access, loss, destruction, leakage, etc.

- Disclosure of information
  - The Data Utilization Review Committee will disclose the names of parties who have been licensed to use the data.

(b) Rules on data sharing (data sharing policy)

FASTQ data, etc. generated after sequencing by the sequencing company are to be sent to the Analysis and Data Center within two weeks. To ensure prompt utilization of the data, including clinical information, that are collected, the Analysis and Data Center will share the data with the Industry Consortium and the Academic Consortium simultaneously. This aims to promote academic discussions relating to whole genome analysis, facilitate clinical understanding of cancer based on whole genome analysis, and improve cancer prevention and prognosis by academe, as well as to promote research and development for drug discovery, diagnostic technology, and prevention by industry.

The rules on data sharing stipulate the following items, which are necessary for the above purpose.

The rules will be revised as necessary in accordance with international trends and the progress of research.

- Basic items
  - The data to be shared are genome data, which are clinical information and sequencing information collected from medical institutions, analyzed in detail by the Analysis and Data Center, and compiled into a database.
  - An Industry Consortium hosted by industry and an Academic Consortium hosted by academe will be formed for sharing the data.
  - In principle, data users will bear the costs associated with utilization of the data. However, a mechanism will be examined to reduce the burden for data that the Data Utilization Review Committee recognizes can only be used in the Academic Consortium.
  - The Analysis and Data Center will register genome data and basic clinical information in public databases after a limited period (24–30 months) has elapsed. However, detailed clinical information will not be registered in public databases because it may lead to the identification of individuals.

After the data are registered in public databases, users who wish to conduct research using detailed clinical information will still be required to participate in the Industry Consortium or the Academic Consortium. Public database users may possess intellectual property. There is no time limit for publication of research results based only on the use of public databases.

- Industry Consortium
  - Medical institutions, research institutions, and companies (regardless of the size of the company and whether it belongs to the medical industry or not) may participate in the Industry Consortium.
  - The data can be used for overviews or simple analysis for the purpose of planning.
  - A license for use granted by the Data Utilization Review Committee is needed for detailed analysis.
  - Free application for and possession of intellectual property are permitted. No notification to the consortium is required for application.
  - The results can be published, but with a fixed restricted period (24–30 months)[11] before publication will be set.
  - The Consortium is to be notified of publication of results, regardless of the timing of data transfer to public databases.

- Academic Consortium
  - The Consortium will consist of institutions that acquire data (medical institutions), data analysts, and institutions (including companies) that conduct joint research with them.
  - The data can be used for overviews[12] or simple analysis[13] for the purpose of planning.
  - Institutions that acquire data (medical institutions) are free to conduct analyses of the data they have collected.
  - Institutions wishing to conduct detailed analyses of data other than their own data will need to have an agreement with the institution that collected the data and will need to notify the Data Utilization Review Committee in advance.
  - Free application for and possession of intellectual property are permitted.
  - The Consortium is to be notified of publication of the results in a paper or application for intellectual property, regardless of the timing of data transfer to public databases.

(c) Data sharing system

As a base for research and drug discovery for conducting data analysis, provision, and storage, the Analysis and Data Center will create a data sharing system that

[11] This is the period calculated from the starting point of each cancer type (except cases in which the content of the application is published after a fixed period of time has elapsed under the application publication system). The restriction period may be shortened to less than 24 months if the data acquirer or analyzer judges that there is little need to keep the results unpublished due to the publication of a paper, etc. The period of limitation for rare/intractable diseases will be determined separately.

[12] Overview of the status of accumulation of data for each cancer type.

[13] Simple analysis: Conducting data analysis necessary for the planning of detailed analysis, such as the identification of the number of cases of pathogenic variants, etc.

promotes seamless research and development. The data sharing system will be part of the centralized management system, and will include the research support system and the system to facilitate the utilization of genome databases and clinical information databases through open APIs indicated below.

- Research support system
  - On-premise and cloud data-sharing systems available for use by the research groups and consortiums will be built.
  - A computing environment for analyzing the shared data will be built.
  - Efforts will be made to support the installation of analysis programs and conduct large-scale analysis and to facilitate the utilization of data.

- Facilitation of the utilization of the genome database and clinical information database by APIs
  - Efforts will be made to lower the barriers to participation through the use of APIs, so that companies and academe can prepare reports based on their own technologies, ideas, and data sources.
  - The Analysis and Data Center will put in place an environment in which each medical institution can prepare and return reports on its own responsibility through the development of various analysis resources and environments.

(3) Development of human resources

(a) Development of human resources relating to the Analysis and Data Center

The tasks required for operation, maintenance, and improvement of the Analysis and Data Center are very diverse, and they include the construction of genome analysis infrastructure, the construction of a system for collection of clinical information, the preparation of patient reports, the development of network security, and the development of an information infrastructure for data utilization. There is therefore a need for diverse expertise among the personnel involved in the Analysis and Data Center, including bioinformatics, medical information, information security, clinical genetics, high-performance computing, and cloud computing. The following points should be noted in relation to developing and securing human resources for the Analysis and Data Center.

- While it is essential to develop young researchers who have the ability to take on these tasks to operate the Analysis and Data Center and keep it running, there appears to be a significant shortage of such human resources in Japan. A mechanism therefore needs to be examined whereby the cooperation of graduate schools is sought to enable students to obtain a degree by conducting research on data analysis, construction of information infrastructure, data management, etc. at the Analysis and Data Center.
- The development of human resources using real-world data will be examined. This will require management of security and access logs.
- The analysis and interpretation of whole genome sequencing data is still developing. Moreover, there have been continued innovations in the various related fields, such as measurement technologies including long-

read sequencing and single-cell sequencing, cloud computing, various different web technologies, security, and artificial intelligence. To maintain a sustainable and competitive system for data analysis and infrastructure development, the Analysis and Data Center will be given the capacity to conduct cutting-edge research.

- A mechanism to make it beneficial for companies to work and conduct research at the Analysis and Data Center will be examined, and human resources will be gathered at the Analysis and Data Center rather than being outsourced from companies. This will require attention to maintenance of trade secrets and neutrality.
- Human resources will be secured through mutual personnel exchanges with industry and academe from a long-term perspective, and maintaining motivation and ensuring career paths will be examined to prevent human resources from leaving.

(b) Development of human resources relating to the use of clinical information, etc.

The number of personnel involved in the use of clinical information, such as genetic counselors, will be examined on the basis of actual experience of return of results to patients in FY2021 and FY2022 and in conjunction with the number of medical institutions conducting whole genome analysis. Having genetic counselors present is an essential requirement for medical institutions that return the results of whole genome analysis to patients, and human resources in this field will be developed through OJT, etc.

7. Items Relating to Ethical, Legal, and Social Issues (ELSI)

The present Action Plan stipulates the implementation of whole genome analysis on an unprecedented scale in Japan, and it includes database construction, facilitation of research and development for drug discovery and diagnostic technologies, and return of the results of whole genome analysis to patients. Implementation of these projects may be expected to involve various ethical, legal, and social issues (ELSI). For the present project to be properly implemented based on the understanding and trust of society, it will be essential to appropriately address ELSI and to put in place systems to achieve this.

Specifically, an ELSI Division staffed by personnel with expertise in this field will be set up within the Project Implementation Organization, and this division will examine and take action on initiatives needed to ensure that the plan for the project overall is being implemented with due regard for ELSI.

The following points in particular should be noted with regard to taking action over ELSI.

- A standardized ICF will be used to enable cross-sectional data utilization.
- The use of ICT and AI in clinical implementation of whole genome analysis should be examined from the perspective of reducing the workload in clinical settings.
- The guidance prepared by the MHLW research group will be used to explain the project to patients, and explanations and information will be provided as clearly and carefully as possible to respect the free will of patients and seek their informed consent. The necessary systems, including the use of e-consent, will be considered.
- Opportunities to receive genetic counseling will be ensured and enhanced, which will include the active use of ICT.
- If findings other than the main findings are obtained, the ethical guidelines and the guidance prepared by the MHLW research group will be referred to when taking action.
- The policy on information security and protection of privacy will be clarified, the necessary systems for implementation will be established, and patients will be informed about these systems.
- Institutional issues in the development of a social environment to ensure there are no disadvantages resulting from genome information will be identified, and the policy of the present project for addressing these issues will be examined.
- The consultation system for whole genome analysis will be improved, and the development and expansion of support systems will be promoted through education and awareness activities so that existing consultation organizations can provide primary consultation services.

8. Items Relating to Patient and Public Involvement (PPI)

In the implementation of the present Action Plan, efforts must be made not only to publicize and explain the plan to the target patients, but also to continuously disseminate information to the general public and society at large and to establish a mechanism for patient and public involvement to ensure transparency and to include the viewpoints of patients and citizens.

Specifically, a Patient and Public Involvement Division will be set up within the Project Implementation Organization, and systems to incorporate the viewpoints of patients and the public will be established in the research institutions and medical institutions participating in the present project. Through these measures, broad-based information dissemination and publicity activities aimed at the general public will be carried out, and a system to collect the views of patients and the public and reflect them in the project will be established. In addition, support will be provided to develop the necessary human resources for PPI, and studies and discussions of methods for more appropriate implementation of PPI, including the development of a system for educating the general public, will be carried out.

The following points in particular should be noted with regard to implementing PPI.

- The project will aim to improve literacy with regard to research and medical care using genome information through generalized public awareness and education and awareness activities with clearly defined targets to foster a social environment in which no disadvantage is caused by genome information.
- Measures will be carried out to increase the transparency of implementation of the present project, such as incorporating the viewpoints of patients and the public, formulating policies on data traceability and utilization, and providing the necessary explanations of these policies.

9. Conclusion

The present Action Plan was formulated from the perspective of steady promotion of the Action Plan (Version 1) that was formulated in December 2019, on the basis of the discussions that have been held to date and against the backdrop of recent global advances in research using whole genome information and the increasing importance of projects relating to whole genome analysis in Japan. The present Action Plan clarifies the objectives of the project for whole genome analysis, and also sets out the basic policies.

The individual initiatives will be promoted in line with the objectives of the project, with the aim of overcoming cancer and rare/intractable diseases in the future.

The Expert Committee will continue to check the progress of the initiatives and changes in the environment relating to industry and academe, and will respond as necessary.

Furthermore, the way the Project Implementation Organization operates and matters relating to ELSI and PPI require continuous examination that goes beyond the items for examination in the present Action Plan, and the Expert Committee will continue to discuss these issues.

In addition, from a broader perspective it would be desirable to give urgent consideration to specific exit strategies to put in place an environment for utilization of data by industry and academe, promote innovative drug discovery and research, and deliver high-quality medical care to patients.

10. Glossary*

Genome
The genome of an organism means all of its genetic information. Human genetic information consists of DNA, which is an arrangement of about 3 billion bases (a type of chemical substance) of just four types: adenine (A), guanine (G), cytosine (C), and thymine (T).

Gene
A specific region of the genome that encodes a functional molecule such as a protein (i.e., the base sequence of that region is converted to an amino acid sequence). The human genome contains approximately 20,000 genes that encode proteins. Genetic information on DNA is copied onto a substance called messenger RNA, and proteins are made on the basis of that information.

Genomic medicine
Medical care in which a person's genome information is analyzed to provide more effective and efficient medical treatment that matches the patient's physical make-up (particularly characteristics at the gene level) and medical condition.

Whole genome analysis
Analysis of the entire genome at once. In the past, it was common to narrow the target for analysis down to a few hundred genes, but technological advances have made it possible now to analyze the whole genome in detail. Under the government's Action Plan for Whole Genome Analysis, a system will be created in which whole genome analysis is performed on samples provided by patients, and the results are interpreted and discussed by experts so that they can be used appropriately in the medical care of the patient. In addition, a database with genomic and clinical information will be constructed, and this will be used as a foundation for creating new medical treatments, pharmaceuticals, and diagnostic technologies in the future.

Transcriptome analysis
Analysis of RNA created from the genome. RNA has a structure consisting of an arrangement of four substances (bases): adenine (A), guanine (G), cytosine (C), and uridine (U).

Multiomics analysis
A method of comprehensive analysis of the genome, proteins, metabolic products, transcriptional products, etc. present in a living organism.

Sequencing
Reading the arrangement (sequence) of bases in DNA or RNA.

Sequencer
A device for reading the arrangement (sequence) of bases in DNA or RNA. Dramatic advances in next-generation sequencers, which appeared in the mid-2000s, allowed whole genome analysis at low cost using technology to read the sequences of multiple DNA molecules simultaneously.

Digital medicine
Software and hardware products that measure substances and make interventions on a scientific basis to maintain human health and diagnose diseases.

Analysis and Data Center
A center that collects genomic and clinical data on patients from medical and research institutions and compiles the data into a highly secure database for use in medical treatment, research, and drug discovery at medical and research institutions and companies in Japan and other countries. Collecting data on large numbers of patients allows more detailed comparison and analysis of individual differences or characteristics, so that the data can be used in a variety of research and development activities.

Cancer genomic medicine
Medical treatment in which the genomes of cancers (patients) are studied and the nature of an individual patient's cancer is ascertained on the basis of genetic changes, so that the most suitable treatment can be selected. For example, in Japanese lung cancer patients, changes in a gene called EGFR are seen in 30–50% of cases. It is known that EGFR protein inhibitors (molecular target drugs) are highly effective in patients with such cancers, so when this genetic change is found in a patient, treatment with an EGFR inhibitor first, rather than other anticancer drugs, may be appropriate for that patient. This applies to medical treatment based on the results of whole genome analysis of cancer or the results of cancer gene panel tests.

Cancer gene panel test
A test that examines changes in dozens to hundreds of genes occurring in cancer cells, to understand the characteristics of a cancer. It is partially covered by national health insurance. Depending on the genetic changes, it is sometimes possible to determine which drugs are most likely to be effective. The results of the test are reviewed by a panel of specialists, called an "expert panel," and the doctor in charge recommends a treatment to the patient in line with the results of the discussions of the expert panel.

Liquid biopsy
A method of analyzing DNA and RNA contained in bodily fluids such as blood, urine, or saliva. Cancer gene panel tests covered by insurance are tests using cancer tissue and liquid biopsy using blood.

Expert panel
A panel of experts from various professions who provide medical interpretations of the results of genome analyses. Based on the analysis results, the panel examines whether there are drugs that can be expected to be effective on the basis of the pathogenic variants that were detected and whether there is a possibility of hereditary diseases. In the case of cancer, the expert panel includes a full-time physician specializing in cancer pharmacotherapy, a physician specializing in genetic medicine, medical staff with genetic counseling skills, a physician specializing in pathology, and experts in molecular genetics and cancer genomic medicine.

Hereditary tumor
There are cases in which a person is born with a genetic difference that predisposes them to cancer. Such cases are called “hereditary tumor,” as the predisposition to cancer can be passed on to the next generation. These tumors account for several percent of all cancers. Genome analysis of cancers sometimes gives results that are suggestive of hereditary tumors. Patients who wish to learn more about hereditary tumors, including the risk that they or their family members may suffer cancer in the future, are able to receive genetic counseling. Patients may also request not to be informed in the event that a test gives a result suggesting hereditary tumor.

Rare/Intractable disease
Under the Rare/Intractable Disease Law, there are 333 designated rare/intractable diseases (diseases that meet the following five criteria: unknown mechanism of pathogenesis, no established treatment, the disease is rare, long-term treatment is necessary, and objective diagnostic criteria have been established) that are eligible for subsidized medical care. From a genetic perspective, these diseases may be classified into those consisting of monogenic diseases alone, those consisting of mixed monogenic and multifactorial diseases, and those consisting of multifactorial diseases alone. In addition, diseases that are not designated as rare/intractable diseases and for which a disease concept has not yet been fully established are also included in the present plan as rare/intractable diseases in the broad sense of the term.

Monogenic disease
A hereditary disease caused by mutation of a single gene.

Multifactorial disease
A disease caused by multiple genetic factors, as well as environmental factors, lifestyle, and aging.

Pathogenic variantsA change in the base sequence of the genome. Somatic mutations in cancer cells that are acquired (i.e., not present at birth) are generally referred to as pathogenic variants.

Polymorphism
An individual variation in the germline seen in the genome sequence that occurs at a frequency of 1% or more in a population.

Variant
Various individual differences in the germline seen in the genome sequence. Those that are related to disease are called pathogenic variants.

*This glossary was prepared by the MHLW Science Research Group for promotion of whole genome analysis of cancer and establishment of a system for technology assessment related to whole genome analysis of each individual patient and its clinical application, a center for data analysis and storage, information security and patient confidentiality, and Ethical Legal and Social Implications (ELSI) (tentative name).