Hitachi Research Institute Report

Globally Accelerating Use of Biodata

Hiroto Matsumoto, Senior Manager, 3rd Research Department
Masayuki Miyazaki, Senior Researcher, Technology Strategy Group, 3rd Research Department

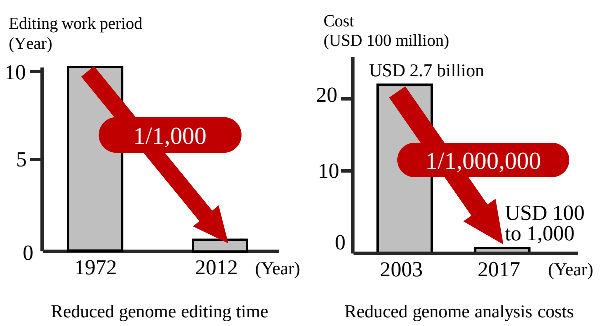
The development of Information and Communication Technology (ICT) including big data and Artificial Intelligence (AI) has promoted the digitization of various industries and changed industrial structures. In the bioindustry, data-driven research and business development activities are also actively conducted thanks to the genome editing technologies and the progress of technologies, including sensors and measurement devices.

This article verifies the U.K. and U.S. bioindustry movements driven by innovation that has been brought about through the expanded use of huge volumes of data, and discusses measures to be taken for strengthening Japan's competitiveness.

1. Increasing amount of biodata

The clustered regularly interspaced short palindromic repeats/ CRISPR associated protein 9 (CRISPR/Cas9) genome editing technology announced in 2012 by the University of California, Berkeley, and Umeå University in Sweden has brought about drastic changes in bio-research fields. Genome editing has conventionally required trial-and-error experiments and has taken an enormous amount of time (more than 10 years) to obtain the desired results. However, the use of CRISPR/Cas9 technology has enabled the genome editing of organisms as planned within a few days (Figure 1).

The development of big data processing technology and genome measuring and sequencing technologies has also accelerated since the late 2000s. When the Human Genome Project was completed in 2003, gene analysis costs amounted to USD 2.7 billion. However, these costs have currently reduced drastically to USD 100 to USD 1,000 (Figure 1).

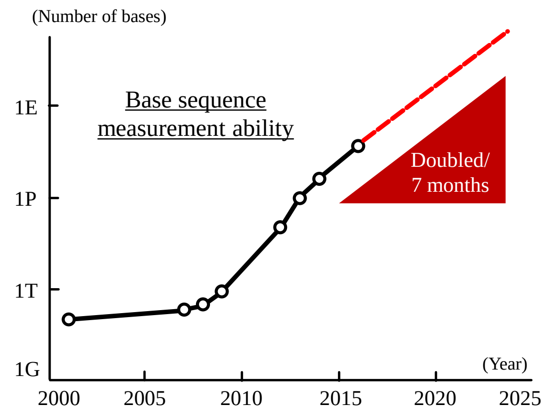


Source: Prepared by Hitachi Research Institute using various materials

Figure 1: Changes in genome editing time and genome analysis costs

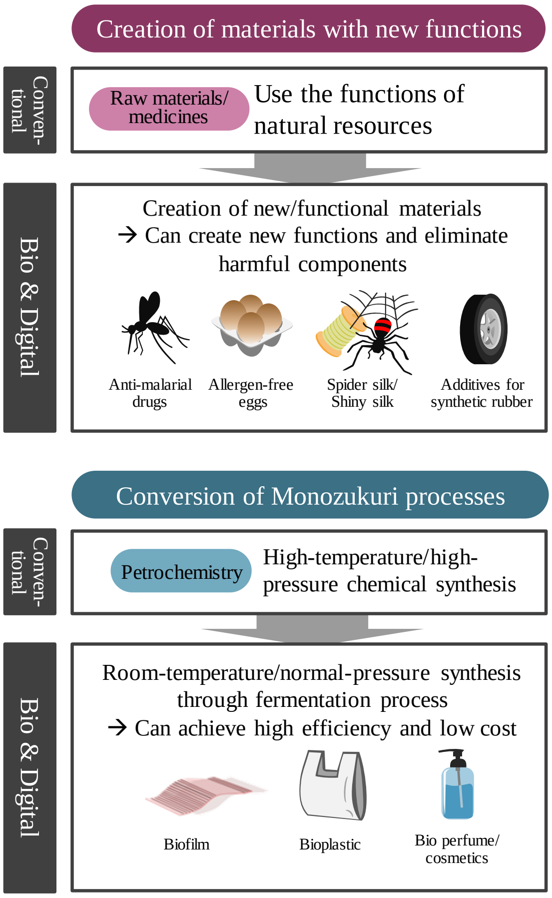
Such drastic reduction of genome editing and analysis time, as well as the relevant costs, will accelerate the development of the bioindustry. Especially, the accumulated volume of biodata (e.g. genetic information and intracellular metabolic reaction information) that are created through editing work or using sequencers is continuously increasing at a good pace (doubled in 7 months) (University of Illinois) (Figure 2). The accumulation of such a huge volume of data has created the possibility of creating new businesses in the raw materials, medical treatment, and petrochemistry fields.

New business creation is divided into two types (Figure 3). One is related to the creation of materials with new functions, and the other is related to the conversion of Monozukuri (manufacturing) processes to environmentally less burdensome bioprocesses. The use of biodata for the creation of materials with new functions enables the design of raw materials with new functions that cannot be obtained from natural resources. In addition, it becomes possible to create useful materials by identifying and eliminating harmful components contained in natural resources. Examples include spider silk which is an elasticized silk and chickens that lay allergen-free eggs. Biomedicines are also included in the creation of materials with new functions (e.g. protein), and the market size will be JPY 30 trillion in 2022. Regarding the conversion of Monozukuri processes, materials that have been synthesized through high-temperature/high-pressure petrochemical processes are created through room-temperature/normal pressure bio-fermentation processes. This sector's market size is expected to be JPY 130 trillion in 2030. In this sector, it is expected to



Note: Genetic information accumulated by research institutes and companies is measured using the number of bases
Source: Prepared by Hitachi Research Institute using various materials

Figure 2: Changes in biodata amount



Source: Prepared by Hitachi Research Institute using various materials

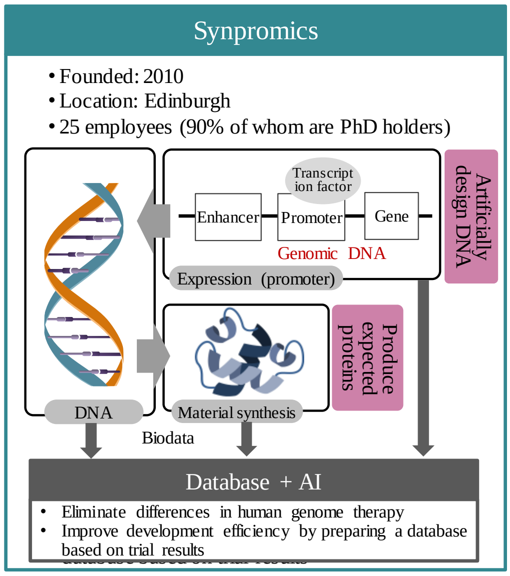
Figure 3: New industries using bio-related data

create materials through the use of biodata, identification of enzymes effective for fermentation and material conversion, and use of metabolic reactions rather than chemical reactions. This will enable highly efficient, low-cost production of plant-derived products such as biofuels and bioplastics.

Along with such growing expectations for new business creation using biodata, investments in startup companies and others in this sector are increasing, mainly in the UK and the US. Well-known IT investor groups (e.g. venture capital firms, Softbank, and a former Google CEO) have formed investment funds and started to invest in startup companies in the bio sector. In the UK, investments in bio venture companies are increasing year by year in terms of both amount and number (GBP 160 million; 15 deals/year), and more companies are entering this sector.

2. The UK accelerates biodata-based research and development

The UK is focusing on research and development of seed technology that creates materials with new functions using biodata. The government has actively made investments, and globally attracted talented researchers in next-generation industries. There are currently about 2.3 million students and 150,000



Source: Prepared by Hitachi Research Institute using various materials

Figure 4: Efforts by UK-based Synpromics Ltd.

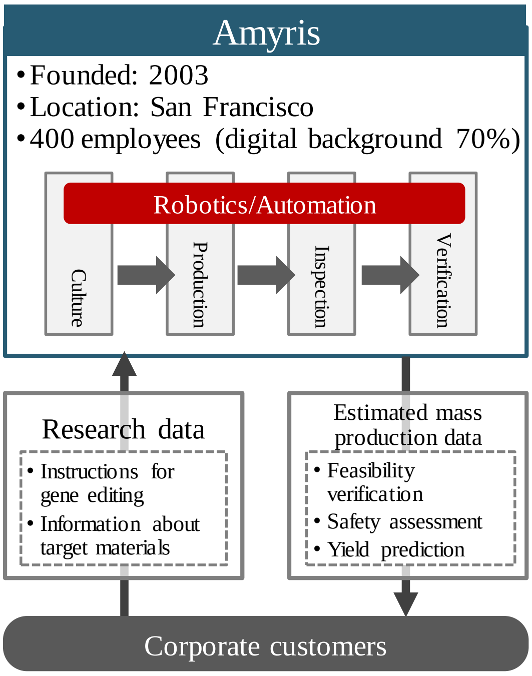
bio-digital researchers in the UK, including 90,000 international students. The UK has become an attractive place for talented people. The Department for Business, Energy and Industrial Strategy (BEIS) is authorized to make grants to fund scientific research focusing on the bio sector. The Biotechnology and Biological Sciences Research Council (BBSRC), a subsidiary organization of the BEIS, gives scientific research grants to universities and other research institutions that accept researchers and other personnel. In addition, Innovate UK, another subsidiary organization of the BEIS, supports the research and development of seed technology by investing in university spin-out bio-venture companies. IT venture capital firms (e.g. MedCity) located in London's financial district are also actively investing in bio-venture companies. The regions where bio-venture companies are actively conducting their businesses are Edinburgh and London. The Roslin Institute, which has succeeded in developing clone sheep technology, and the University of Edinburgh are located in Edinburgh, while the University of Cambridge, which has discovered DNA structures, the University of Oxford, and Imperial College London are located in London.

Venture companies that are actively using genome data and conducting research on new materials by receiving government grants and investments from venture capital firms have emerged in these regions. Synpromics (Figure 4), a venture company from the University of Edinburgh, has artificially created promoters (intra-genome genetic information that commands enzyme-producing function activities of cells) and gotten relevant intellectual property rights. Synpromics also creates promoter libraries using multiple promoter information and provides pharmaceutical and other corporate customers with them as fee-based service. In addition, using machine learning algorithms, Synpromics analyzes correlational relationships between its promoter libraries and genetic functions emerging from such libraries, and extracts and proposes a combination of multiple promoters that are highly possible to obtain desired genetic functions. Synpromics' customers, including pharmaceutical manufacturers which have performed genome editing to create materials with new functions, have been able to shorten their research and development periods by using its promoter libraries.

3. The U.S. promotes biodata-based mass-production technology

In the U.S., industrial agglomerations aiming to develop functional materials using biodata have progressed in San Francisco and Boston. A structure has been built in each region so that venture companies' development risk can be reduced by letting academic and industry sectors share and jointly use necessary raw materials and equipment for the mass production of materials with new functions. Joint experimental facilities (QB3) of the University of California, Berkeley (UCB), a nonprofit biobank in Boston (Addgene) that stores CRISPR/Cas9 data samples, and other organizations assume these functions. Startup companies and others can use QB3 facilities at a low cost. At these facilities, they can conduct experiments on materials that they develop, and verify mass production feasibilities. Addgene provides biological samples to research organizations in the R&D phase at a low cost. Addgene also enables these organizations to verify the quality of target materials through trial production using its biological samples and their design data prior to mass production. In these regions, universities giving technological support and unicorn companies conducting research and development of mass production have gathered around joint facilities providing samples and equipment-related support as hub facilities, having built a structure to enable early commercialization.

Bio-venture companies, such as Amyris, Zymergen, and Ginkgo Bioworks, which specialize in development support for mass production, are playing important roles in this field. Amyris has in-house robot-based automated equipment for microbial culture, production, inspection and verification processes (Figure 5). This enables Amyris to receive biodata related to bio-medicines from pharmaceutical manufacturers (customers) and conduct experiments on mass production in a short period of time. Using its automated equipment, Amyris conducts inspection and verification while gradually increasing production volume, estimates mass production capability before pre-production, and provides corporate customers with such estimated data. In this way, customers are able to determine feasibility of mass production while



Source: Prepared by Hitachi Research Institute using various materials

Figure 5: Efforts by US-based Amyris, Inc.

reducing capital investment. Zymergen has also fully automated its experiment environments, from experiment facilities to sample delivery. It supports customers' biodata-based mass production of materials with new functions. As described above, US-based bio-venture companies specializing in development support for mass production are playing an intermediary role in bridging the gap between research and development of seed materials and development of mass production.

As part of government activities, the U.S. Department of Energy promotes the establishment of Agile Bio Foundry (ABF) to comprehensively accumulate mass production biodata created through research and development in these regions. ABF is a public-private sector consortium. Under its plan, mass production data and equipment and technologies of nine national research centers in the U.S. are made freely available to domestic private-sector companies so that the development of bio-related mass production will be promoted in the U.S. ABF aims to create 100 types of new materials by 2020.

4. Action taken by Japan to strengthen the use of biodata, and future issues

Amid activating data-driven bioresearch and business development, efforts directed toward comprehensively accumulating national and regional biodata are underway in the U.S. and Europe. In the U.S., the National Bioeconomy Blueprint was released as a presidential pledge in 2012, and the Federal Activities Report on the Bioeconomy was released in 2016 as a product of the Department of Energy and an

Table 1: Global biodata accumulation

	Bio-sample types ('18)	Number of genome analysis devices ('12)	Number of genome analysis projects ('18)
U.S.	21,717	785	25,675
China	3,617	200	2,457
U.K.	561	136	388
Germany	535	111	555
Japan	141	34	117

Source: Prepared by Hitachi Research Institute using various materials including JGI databases

Table 2: Composition of COCN "i-Bio"

Working Group	Name	Outline
WG1	Highly functional chemicals	Design highly functional chemical materials through bioprocessing
WG2	Highly functional food materials	Design highly functional food materials through bioprocessing
WG3	New biomaterials	Waste, wastewater disposal and biosensor
WG4	Leading technology	Ascertain biodata-related issues Integration/use/distribution of data

Source: Prepared by Hitachi Research Institute using various materials including COCN reports

interagency collaboration. The importance of expanding the bioindustry through the accelerated development of domestic biodata collection and analysis tools, as well as resolving environmental and other social issues, were proposed. Priority areas include the replacement of petroleum-derived fuels with biomass fuels. ABF described above is part of these activities. In 2012, the EU adopted "Innovation for Sustainable Growth: A Bioeconomy for Europe." Under its research and development program "Horizon2020," the EU has also formulated a 7-year plan to invest EUR 3.7 billion in bio development and establishment of bio databases. In the UK, the Synthetic Biology Leadership Council (SBLC) released the "Biodesign for the Bioeconomy" in 2016. It describes the necessity of standardizing data formats, establishing databases, ensuring data security, and globally circulating data, for example. In 2016, China also launched its national gene bank to accelerate the creation and accumulation of domestic biodata. As described above, the U.S., Europe, China, etc. are actively gathering and accumulating biodata under government initiatives assuming that these data can be applied in extensive fields.

In Japan, on the other hand, genetically modified foods, etc. have not been widely accepted, and as a result, research and development in biotechnology have been delayed, except for some fields (e.g. medicine). According to publicly disclosed information about global biodata accumulation, Japan lags far behind the U.S., Europe and China (Table 1).

These countries' bio-related strategic initiatives and technology innovation in recent years have prompted Japan to start considering data-based bioindustry promotion measures. Bio-related fields are also included in the "Investments for the Future Strategy 2018 (Cabinet Secretariat)" launched by the government. This strategy's objectives are specifically as follows: (1) contribution to resolving social issues through research and development of innovative biotechnology that are made possible by integrated bio-digital technologies, (2) implementation of new labeling for foods with function claims that are expected to contribute to good health, (3) creation of a system to assess and commend biomaterials and biomass resources that can contribute to carbon cycle society, and (4) clarification of the applicability of the "Law Concerning the Conservation and Sustainable Use of Biological Diversity through Regulations on the Use of Living Modified Organisms" relating to genome editing technology.

Universities and research institutions are concurrently carrying out a biodata integration project mainly focusing on academic fields by connecting the National Bioscience Database Center (NBDC) with domestic research centers and universities.

As industrial activities, the "Council on Competitiveness-Nippon (COCN)" has been promoting a theme "strategy of bio-industry through the integration of bio and digital technology (i-Bio)" since 2017 (Table 2). To improve the competitive advantage of Japan's bioindustry, this theme proposes bioindustry strategies using digital data that are being rapidly accumulated in recent years. The Council also holds discussions on promising biotechnology application fields related to chemicals, food, and new industries, as well as system and technology issues and necessary measures for the use of domestic biodata. In the case of biodata in particular, private-sector companies conduct in-house development in many cases. As a result, a problem has been pointed out, namely that biodata created through experiments and development are unevenly held by various companies and are left undisclosed.

To accelerate the use and circulation of biodata domestically in future, the following technology development and system design are necessary. Regarding the use of biodata, the academic sector's ongoing database integration projects need to be expanded to the private sector, and measures for project continuation are necessary. Regarding the circulation of biodata, the concept of bio-information banks needs to be realized. In connection with the establishment of bio-information banks, it is important to implement a data trust system, design a system that can offer incentives to data providers, and form a data distribution market that can determine data value. It is also necessary to develop blockchain, security, and other technologies from the standpoint of data protection.