White Paper

Leveraging High-Performance Compute Infrastructure to Address the Genomic Data Challenge in Life Sciences

Sponsored by: Lenovo and Intel

Nimita Limaye Gerald Wang

May 2022

THE EVOLVING PARADOX — FINDING RAPIDLY SCALABLE SOLUTIONS FOR INCREASINGLY PERSONALIZED THERAPIES

The lens of the life sciences industry has shifted from the development of blockbusters, which addressed the needs of the masses, toward developing niche, high-value, personalized solutions for patients. Increasingly Asia/Pacific public sector investments in precision medicine, genomics, and imaging advances, as well as the broad use of electronic health records and the proliferation of medical Internet of Things (IoT) and mobile devices, have resulted in a massive explosion of structured and unstructured data. According to IDC's Global DataSphere forecast, the global healthcare and life science datasphere accounts for over 7% of the total enterprise datasphere worldwide, and is set to grow exponentially. Asia/Pacific, the world's most populous region, is expected to contribute significantly to the expanding data volume, variety, and velocity challenges of the global health and life science datasphere.

Against the backdrop of the COVID-19 pandemic, which has accelerated the need for innovation, to discover new drugs and vaccines, and to repurpose existing ones, two challenges are evident. First is the need to store the exponentially growing volume of genomic data; second is the need for ultra-high-performance compute infrastructure, with zero downtime, and high concurrency. The ability to handle diverse 'multi-omic' structured and unstructured data, the ability to handle large and small files, and the ability of solution providers to provide high-performance compute, with flexible and scalable solutions and pricing models, are going to be key for the life sciences industry to keep pace with the demand for accelerated innovation fueled by genomics data.

To better understand the industry's needs, IDC conducted a genomics survey, sponsored by Lenovo, with 150 pharmaceutical and biotech companies across five markets – Japan, Korea, Singapore, Thailand, and India – from December 2021 through January 2022.

Drawing on the findings of the survey, this paper highlights the challenges and drivers transforming the landscape, and what life sciences companies require as they build their genomic data infrastructure strategy to keep pace with their expanding needs.

Respondent demographics

- C-Suite (5%), lines of business (78%), IT (17%)
- Primary decision makers (14%) and key decision-making influencers involved in procuring genomics solutions (86%)
- Company size: 100-500 (39%) employees, less than 500 employees (15%), and more than 500 employees (45%)

May 2022, IDC #AP15021X

THE GREAT GENOMIC DATA CHALLENGE FOR ASIA/PACIFIC'S NASCENT INDUSTRY

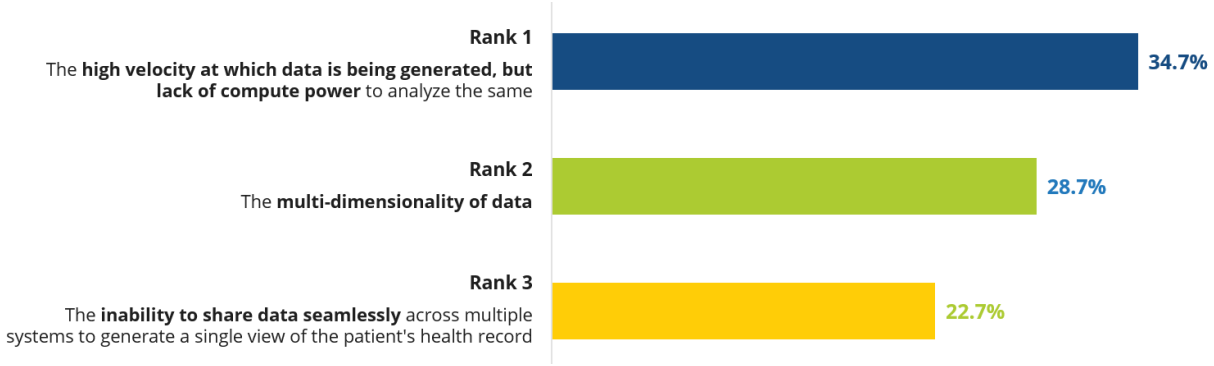
The exponential growth in the volume of genomic data is set to continue. The Global Alliance for Genomics and Health (GA4GH), the standards-setting body in genomics for healthcare, has reported that over 60 million patients will have their genomes sequenced by 2025, fueled by growth in therapeutic areas such as rare diseases and oncology. With cell and gene therapies representing the fastest-growing therapeutic area, posting a double-digit compound annual growth rate (CAGR) and with a heightened focus on precision medicine, the real challenge will be managing the infrastructural challenges associated with the growth of genomic data, especially for Asia/Pacific's nascent life sciences industry.

Challenges with the velocity and volume of data

The lack of compute power to deal with the speed at which data is coming in, the complexity resulting from the multidimensionality of the data, and the need for seamless interoperability, were the top 3 challenges in the context of genomics (see Figure 1). This is not surprising considering the increasing shift toward a multi-omics strategy, deep phenotyping, and the need to integrate multiple sources of data to derive a truly holistic view of a patient's health profile.

FIGURE 1

Top 3 IT Challenges in the Context of Genomics



N=150

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

The growing reliance on the processing power of high-performance systems to speed up precision medication and genomics is much needed, but it is still early days for the industry. Slightly over 50% of Asia/Pacific respondents stated that they were still at a startup stage with respect to their genomics high-performance computing (HPC) infrastructure, having less than a year of experience. A third indicated having 1–3 years of experience, and only 15% have more than 3 years of experience.

To put context to this, the industry's maturity in the use of HPC varies across the Asia/Pacific region. In general, socioeconomically stronger first-tier jurisdictions generally having better access to good healthcare and life science services which support their local communities. These are countries like Singapore, South Korea, Taiwan, Australia, New Zealand, South Korea, Thailand, Malaysia, and India, which have strong public and private sector investments for cutting-edge digital health initiatives that rival that of the world's best. While many of the region's digital health ecosystems are motivated by public health mandates, widening pockets of opportunities are arising from the business of life

sciences in less-matured jurisdictions, such as Thailand and India, that have keen eyes on becoming regional medical tourism hubs.

Increasing genomic workloads and storage capacities

Asia/Pacific is reported to show the highest CAGR in the genomics market, creating an accelerated need for the Asia/Pacific life sciences industry to rapidly build out storage capacity.

A third of Asia/Pacific respondents consider drug discovery and diagnostic genomics workloads as critical to their business, while 15–20% are prioritizing vaccine research and personalized medicine. This is in alignment with the renewed focus on accelerating drug discovery, developing diagnostic biomarkers to support precision medicine strategies, and driving vaccine development, fueled by the COVID-19 pandemic. IDC believes that even if the demand for COVID-19 testing slows down, the ongoing government funding for genome projects and start-up businesses will continue to drive the increase in workloads.

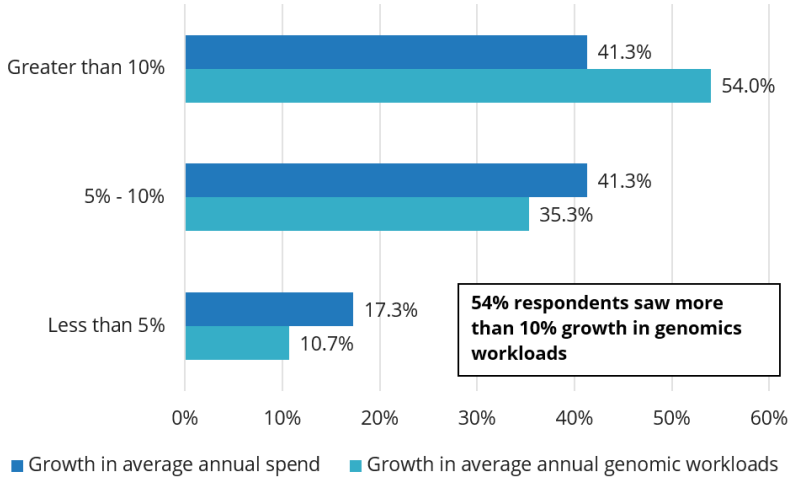
About half of Asia/Pacific respondents have a current genomic data storage capacity of 1–5 PB, and a fifth have a capacity of more than 5 PB. If left unaddressed, the inability to scale storage capacity could pose other challenges, such as industry and regulatory compliance in a highly regulated industry such as life sciences.

Geographies expecting accelerated growth need to scale up their infrastructure to support their genomic workloads and storage capacities, or risk stifling innovation and limiting growth opportunities. Despite the obvious and urgency, there is still the question of whether firms are investing enough to ensure they have the right technology infrastructure to meet the growth of their data.

As shown in Figure 2, while 54% of Asia/Pacific respondents expect more than 10% growth in their genomic workloads, a third noted that their current annual spend on genomic data compute ranged between $500,000 and $1 million, while 20% noted a spend of over $1 million. About 41% cited an increase in annual spend by 5–10%, and 41% of respondents indicated an increase of more than 10%.

FIGURE 2

Growth in Annual Spend on Genomic Data Storage & Compute Versus Growth in Genomic Workloads in the Coming Two Years



N=150

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

ACCELERATING GENOMICS TRANSFORMATION WITH HPC

Close to 90% of Asia/Pacific respondents are using high-performance workstations, while over 50% also use laptops for data visualization. Interestingly, 36% are using 3D augmented reality/virtual reality (AR/VR) solutions, indicative of the growing shift toward immersive visualization techniques, complemented by deep learning to enable molecular modeling and simulations. This has been fueling the creative instincts of scientists, allowing scientists to play around with data and explore molecular interactions in a 3D mode and identify drug targets.

Seeking high-speed scalability to overcome current limitations

About 28% of Asia/Pacific respondents cited the lack of scalability of their existing solution as the primary driver for evaluating alternative solutions. The drivers ranked second and third are that the existing solution required too much customization and that it was very complex. The lowest ranked drivers included the fact that the pricing models were not desirable and that there were challenges with support or that the support costs were too high.

About 50% of Asia/Pacific respondents are planning on transforming their HPC landscape by acquiring a new solution, with 25% doing so within 3–6 months, and 10% in less than 3 months, demonstrating an extremely rapid pace of transformation. As organizations in the more developed and better-funded markets invest in next-generation sequencing (NGS) machines to improve their productivity, they will need to transform their HPC environment to accommodate the faster and newer work processes. An NGS machine with the ability to create 1 TB of data a day will also increase the storage requirements for genomic data dramatically.

Support for concurrency and research collaboration

Over 50% of Asia/Pacific respondents have 10–50 users of their genomics solutions and 20% have 50–200 users. About 6% have more than 200 users of their genomics solutions. This is indicative of the growing size of genomics teams within R&D organizations and calls attention to the need for solutions to support concurrency and drive collaboration to accelerate genomic research.

According to survey data, almost three-quarters of respondents are participating in a national, regional or global genome initiative, a reflection of progress in global industry efforts to leverage genomic data for game-changing innovation in healthcare and life sciences.

About 30% of the respondents are participating in the US All of US initiative, which is a historic initiative launched in 2018 to collect the genomic and health data of over a million people in the United States to accelerate biomedical research. Over 13% participated in Europe's 1+ Million Genomes, an initiative which unites 22 European Union countries, the United Kingdom and Norway, and plans on sequencing at least 1 million genomes by 2022.

About 12% of the respondents are part of the GenomeAsia 100K project, hosted by Nanyang Technological University, Singapore. Three other members of this initiative include Macrogen, South Korea, Genentech, US, and MedGenome from India/US. Launched in 2016, its goal is to understand the genome diversity of Asian ethnicities by sequencing the genomes of 100,000 of people living in Asia. This is particularly important since Asians represent 40% of the world's population but account for barely 6% of the world's recorded genome sequences. About 8% of the respondents have joined Genomics Thailand, around 5% are in the UK Biobank initiative, and less than 1% participated in the China Kadoorie Biobank (CKB) initiative. The increasing focus on sequencing the Asian genome is only going to increase the genomic data produced, stored and used by Asian organizations.

COUNTRY SNAPSHOT

India

Genomics has moved to the center stage, not only because of the Genome India project, which aims to sequence 10,000 Indians. Contract research and manufacturing organizations in India are growing their drug discovery partnerships and eyeing pharmaceutical firms looking to move away from China. The nation’s genomic sequencing capacity will increase, with private laboratories joining the Indian SARS CoV-2 Genomic Consortia (Insacog), a network of 28 government laboratories formed to boost COVID-19 genome sequencing efforts. This lines up with the survey showing diagnostic genomics and drug discovery as the major workloads, and 83% respondents expecting workloads to grow by more than 10% over the next two years. About 50% of India organizations have more than 5 PB of storage, and 80% respondents expect their storage spend to grow by more than 10% over the next two years.

Japan

One of six countries involved in the internationally acclaimed Human Genome Project, Japan is focused on functional genomics, and the Japan government is a strong proponent of genomics medicine. In pursuit of breakthrough innovation, its large pharmaceutical sector’s drug discovery spending makes Japan one of the top 3 largest pharmaceutical markets globally. This explains drug discovery as the top HPC workload, and over half of Japan respondents expect such workloads to grow by 5-10% in the next two years. About 80% have storage capacities up to 5 PB while 63% of India respondents are expecting their storage spend to grow by more than 5% in these next two years.

Korea

The two major genomic initiatives are the Ulsan 10,000 Genome Project and the multi-ministry genome project 2014-2021 for genomic research into health, agriculture, and marine. Korea has also set up a national biobank to store genomic information. Diagnostic genomics and drug discovery are the two major workloads with almost three-quarters of respondents expecting their workloads to grow by 5-10% next two years. About 90% respondents have up to 5 PB of storage, with two-thirds of respondents expecting their storage spend to increase by more than 5% in the next two years.

Singapore

The island-state has more biomedical researchers per capita than the United States, with large-scale funding from the government for COVID-19 testing and diagnostics. The government is the largest spender for R&D, with a vision for Singapore to become a regional biomedical hub for Asia. The Genome Institute of Singapore (GIS) was established in June 2000 by the Agency of Science, Technology and Research, to gain insights on the genetic diversity of multiethnic Asian populations. Completed in December 2021, the genetic databank of whole-genome sequenced data involved over 5,000 Singaporeans, enabling the country's healthcare ecosystem to be equipped with a more accurate diagnosis of genetic diseases. Diagnostic genomics remains by far the top workload, with 97% respondents expecting their workloads to increase by over 10% in the next two years. About 30% of organizations have more than 5 PB of storage with 77% of respondents expecting their storage spend to grow by more than 10% in the next 2 years.

Thailand

In line with Thailand's strategy to strengthen its a healthcare tourism sector, Genomics Thailand is a collaborative research network established to support the National Strategic Initiative on Precision Medicine: 2019-2023. Its focus areas include the Genomics Thailand Initiative to sequence the genes of 50,000 Thais and to support foreign pharmaceutical companies. In line with this, drug discovery, vaccine research, and personalized medicine emerged as the top 3 workloads with 53% respondents expecting their workloads to grow by over 10% in the next two years. About 97% of organizations have storage capacities of up to 5 PB, with 87% expecting their storage spending to increase by more than 5% over the next two years.

CONCLUSION

Genomics could serve as a game changer to accelerate precision medicine for treating chronic illnesses and lifestyle disorders, which ranks among humanity's top challenges. A third of Asia/Pacific respondents put highest priority to the development of a precision medicine strategy to support drug discovery and improve care delivery. Against the backdrop of the COVID-19 pandemic, the increasing funding and demand for genomics has made the Asia/Pacific region an attractive market for genomics research. Life sciences organizations will need to increase their pace of innovation by building out hyper-scalable, ultra-high-performance infrastructure to keep pace with their fast-growing genomic workloads.

MESSAGE FROM THE SPONSOR

With Genomics being at the forefront of medical discovery, it is critical for researchers and scientists to process population scale bioinformation in record time and drive solutions that can solve humanity's greatest challenges. Lenovo's Genomics Optimization And Scalability Tool (GOAST), powered by Intel® Xeon Scalable processor, does just that. Optimized specifically to manage population scale genome analytics, Lenovo GOAST is a proven solution that accelerates genomic workflows and data processing. GOASTv3.0 with software optimizations on the workflow increased the performance by 55-65% over previous GOAST v2.0. This increased the analysis throughput by 188X, enabling researchers and scientists to complete a month's worth of research in under 48 mins.

Unlike the more expensive GPU-based solutions, Lenovo GOAST, is built on a powerful CPU-based architecture that makes high-throughput and high-performance analytics affordable, and accessible. Loaded with pre-tuned hardware and pre-configured software, GOAST is a turnkey appliance solution that has been optimized with extensive in-house testing to ensure you don't have to re-invent the wheel. To know more about Lenovo GOAST, read the detailed solution brief, here.

APPENDIX

FIGURE 3

Organizations' top IT challenges in the context of genomics

Respondents were asked to rank their top IT challenges. E.g., cybersecurity risks was ranked 5th by 44% of Asia/Pacific respondents.

Asia/Pacific (N=150)	
#1 The high velocity at which data is being generated, but lack of compute power to analyze the same	34.7%
#2 The multi-dimensionality of data	28.7%
#3 The inability to share data seamlessly across multiple systems	22.7%
#4 The need for customization of solutions, adding to cost and complexity	32.0%
#5 Cybersecurity risks	44.0%

Japan (N=30)	
#1 The inability to share data seamlessly across multiple systems	40.0%
#2 The need for customization of solutions, adding to cost and complexity	26.7%
#3 The high velocity at which data is being generated, but lack of compute power to analyze the same	40.0%
#4 Cybersecurity risks	26.7%
#5 The multi-dimensionality of data	33.3%

Korea (N=30)	
#1 The high velocity at which data is being generated, but lack of compute power to analyze the same	26.7%
#2 Cybersecurity risks	33.3%
#3 The multi-dimensionality of data	33.3%
#4 The inability to share data seamlessly across multiple systems	33.3%
#5 The need for customization of solutions, adding to cost and complexity	20.0%

India (N=30)	
#1 The high velocity at which data is being generated, but lack of compute power to analyze the same	46.7%
#2 The multi-dimensionality of data	40.0%
#3 The inability to share data seamlessly across multiple systems to generate a single view of the patient's health record	33.3%
#4 The need for customization of solutions, adding to cost and complexity	46.7%
#5 Cybersecurity risks	50.0%

Singapore (N=30)	
#1 The high velocity at which data is being generated, but lack of compute power to analyze the same	50.0%
#2 The multi-dimensionality of data	46.7%
#3 The need for customization of solutions, adding to cost and complexity	26.7%
#4 The inability to share data seamlessly across multiple systems to generate a single view of the patient's health record	36.7%
#5 Cybersecurity risks	53.3%

Thailand (N=30)	
#1 The high velocity at which data is being generated, but lack of compute power to analyze the same	40.0%
#2 The multi-dimensionality of data	30.0%
#3 The inability to share data seamlessly across multiple systems to generate a single view of the patient's health record	30.0%
#4 The need for customization of solutions, adding to cost and complexity	36.7%
#5 Cybersecurity risks	70.0%

N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 4

Most critical workloads in an organization

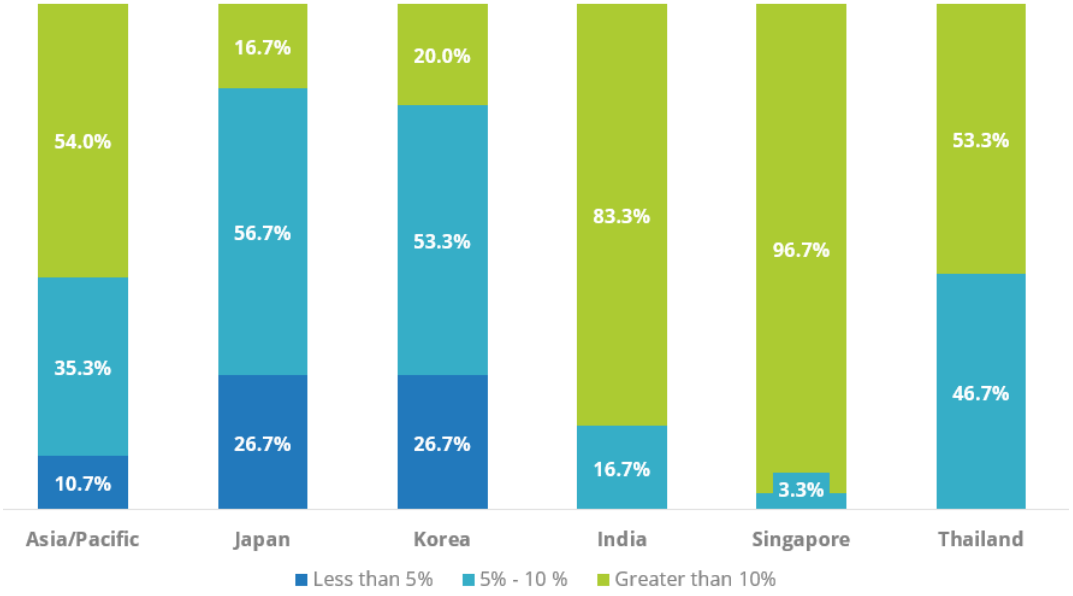
	Asia/ Pacific	Japan	Korea	India	Singapore	Thailand
Drug discovery	**31.3%**	40.0%	26.7%	26.7%	30.0%	33.3%
Diagnostic genomics	**30.0%**	20.0%	33.3%	43.3%	43.3%	10.0%
Personalized medicine	**18.7%**	20.0%	20.0%	13.3%	13.3%	26.7%
Vaccine research	**15.3%**	10.0%	16.7%	13.3%	10.0%	26.7%
Agriculture to drive food security	**4.0%**	10.0%	3.3%	3.3%	0.0%	3.3%
Others	**0.7%**	0.0%	0.0%	0.0%	3.3%	0.0%

N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 5

Percentage of average annual genomic workloads growth in the next two years

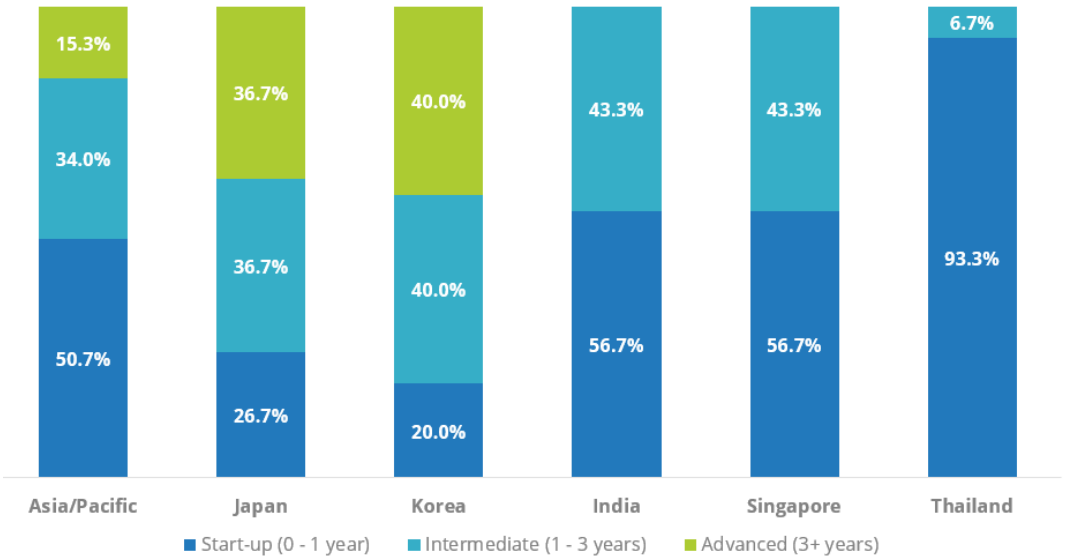


N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 6

Stage of maturity of genomics high-performance computing



N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 7

Reasons for evaluating a new HPC solution

Respondents were asked to rank their reasons for evaluating a new HPC solution. In the figure below, 30.7% of respondents ranked "the pricing models are not desirable" as the #5 reason for evaluating a new HPC solution.

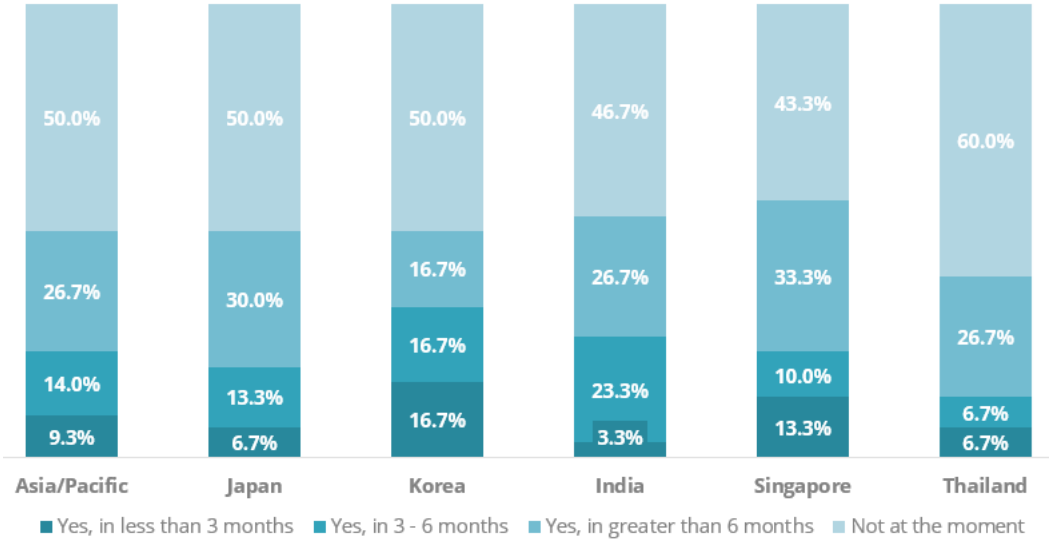
Asia/Pacific (N=75)	
#1 The existing solution is not scalable	28.0%
#2 Too much customization is required	20.0%
#3 The existing solution is very complex	20.0%
#4 Increased throughput required	21.3%
#5 The pricing models are not desirable	30.7%

N=75 Asia/Pacific

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 8

Plans on transforming their HPC landscape by acquiring a new solution



N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 9

Humanity's top 3 challenges in which genomics could be a game changer

Respondents were asked to rank humanity's challenges in which genomics could be a game changer. E.g., forensic genomics was ranked 5th by 40% of India respondents.

Asia/Pacific (N=150)	
#1 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	32.7%
#2 Forensic genomics	24.7%
#3 Environment and climate change	28.0%
#4 Hunger and nutrition food for all, malnutrition	24.7%
#5 Development of drugs and vaccines	17.3%

Japan (N=30)	
#1 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	46.7%
#2 Forensic genomics	40.0%
#3 Environment and climate change	30.0%
#4 Hunger and nutrition food for all, malnutrition	33.3%
#5 Development of drugs and vaccines	10.0%

Korea (N=30)	
#1 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	33.3%
#2 Hunger and nutrition food for all, malnutrition	30.0%
#3 Environment and climate change	36.7%
#4 Development of drugs and vaccines	26.7%
#5 Forensic genomics	30.0%

India (N=30)	
#1 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	40.0%
#2 Hunger and nutrition-food for all, malnutrition	40.0%
#3 Environment and climate change	30.0%
#4 Development of drugs and vaccines	26.7%
#5 Forensic genomics	40.0%

Singapore (N=30)	
#1 Forensic genomics	30.0%
#2 Development of drugs and vaccines	23.3%
#3 Environment and climate change	26.7%
#4 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	40.0%
#5 Hunger and nutrition food for all, malnutrition	23.3%

Thailand (N=30)	
#1 Developing a precision medicine strategy to treat chronic illnesses, rare diseases, lifestyle disorders	26.7%
#2 Forensic genomics	23.3%
#3 Development of drugs and vaccines	26.7%
#4 Hunger and nutrition food for all, malnutrition	33.3%
#5 Environment and climate change	43.3%

N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

FIGURE 10

Most important solution for solving humanity's greatest challenges in which genomics could be a game changer

	Asia/ Pacific	Japan	Korea	India	Singapore	Thailand
Increase productivity of land	26.7%	16.7%	26.7%	36.7%	23.3%	30.0%
Fortification of food grains and fruits and vegetables to address nutritional deficiencies	22.0%	33.3%	13.3%	23.3%	26.7%	13.3%
Lessen the use of pesticides and chemical fertilizers for safer food	20.7%	13.3%	16.7%	23.3%	23.3%	26.7%
Increasing agricultural output	17.3%	20.0%	26.7%	10.0%	10.0%	20.0%
Reduce crop damage due to natural calamities and pest attacks, etc.	13.3%	16.7%	16.7%	6.7%	16.7%	10.0%

N=150 Asia/Pacific; 30 per country

Source: IDC Asia/Pacific Genomics Survey 2022, sponsored by Lenovo

LEARN MORE

- *IDC Perspective: How Data Interoperability and Federated Learning are Fueling a Precision Medicine Strategy and Transforming Drug Repurposing (forthcoming)*
- *IDC Perspective: GPU-Powered Transformer Models Poised to Accelerate Drug Discovery and Disrupt Drug Development* (IDC #US47660321, May 2021)
- *IDC Perspective: Future of Intelligence Defined* (IDC #US45720619, January 2020)

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com