

IBM Research

research.ibm.com/teams/healthcare-and-life-sciences



IBM Research uses advanced computing to accelerate therapeutic and biomarker discovery

Focusing on accelerated discovery, IBM Research is leveraging next-generation computing technologies—artificial intelligence, the hybrid cloud, and quantum computing—to streamline and optimize research in the healthcare and life sciences.

Over the past decade, artificial intelligence (AI) has emerged as an engine of discovery by helping to unlock information from large repositories of previously inaccessible data. The cloud has expanded computer capacity exponentially by creating a global network of remote and distributed computing resources. And quantum computing has arrived on the scene as a game changer in processing power by harnessing quantum simulation to overcome the scaling and complexity limits of classical computing.

In parallel to these advances in computing, in which IBM is a world leader, the healthcare and life sciences have undergone their own information revolution. There has been an explosion in genomic, proteomic, metabolomic and a plethora of other foundational scientific data, as well as in diagnostic, treatment, outcome and other related clinical data. Paradoxically, however, this unprecedented increase in information volume has resulted in reduced accessibility and a diminished ability to use the knowledge embedded in that information. This reduction is caused by siloing of the data, limitations in existing computing capacity, and processing challenges associated with trying to model the inherent complexity of living systems.

IBM Research is now working on designing and implementing computational architectures that can convert the ever-increasing volume of healthcare and life-sciences data into information that can be used by scientists and industry experts the world over. Through an AI approach powered by high-performance computing (HPC)—a synergy of quantum and classical computing—and implemented in a hybrid cloud that takes advantage of both private and public environments, IBM is poised to lead the way in knowledge integration, AI-enriched simulation, and generative modeling in the healthcare and

life sciences. Quantum computing, a rapidly developing technology, offers opportunities to explore and potentially address life-science challenges in entirely new ways.

“The convergence of advances in computation taking place to meet the growing challenges of an ever-shifting world can also be harnessed to help accelerate the rate of discovery in the healthcare and life sciences in unprecedented ways,” said Ajay Royyuru, IBM fellow and CSO for healthcare and life sciences at IBM Research. “At IBM, we are at the forefront of applying these new capabilities for advancing knowledge and solving complex problems to address the most pressing global health challenges.”

Improving the drug discovery value chain

Innovation in the healthcare and life sciences, while overall a linear process leading from identifying drug targets to therapies and outcomes, relies on a complex network of parallel layers of information and feedback loops, each bringing its own challenges (Fig. 1). Success with target identification and validation is highly dependent on factors such as optimized genotype-phenotype linking to enhance target identification, improved predictions of protein structure and function to sharpen target characterization, and refined drug design algorithms for identifying new molecular entities (NMEs). New insights into the nature of disease are further recalibrating the notions of disease staging and of therapeutic endpoints, and this creates new opportunities for improved clinical-trial design, patient selection and monitoring of disease progress that will result in more targeted and effective therapies.

Powering these advances are several core computing technologies that include AI, quantum computing, classical computing, HPC, and the hybrid cloud.

Different combinations of these core technologies provide the foundation for deep knowledge integration, multimodal data fusion, AI-enriched simulations and generative modeling. These efforts are already resulting in rapid advances in the understanding of disease that are beginning to translate into the development of better biomarkers and new therapeutics (Fig. 2).

“Our goal is to maximize what can be achieved with advanced AI, simulation and modeling, powered by a combination of classical and quantum computing on the hybrid cloud,” said Royyuru. “We anticipate that by combining these technologies we will be able to accelerate the pace of discovery in the healthcare and life sciences by up to ten times and yield more successful therapeutics and biomarkers.”

Optimized modeling of NMEs

Developing new drugs hinges on both the identification of new disease targets and the development of NMEs to modulate those targets. Developing NMEs has typically been a one-sided process in which the *in silico* or *in vitro* activities of large arrays of ligands would be tested against one target at a time, limiting the number of novel targets explored and resulting in ‘crowding’ of clinical programs around a fraction of validated targets. Recent developments in proteochemometric modeling—machine learning-driven methods to evaluate *de novo* protein interactions *in silico*—promise to turn the tide by enabling the simultaneous evaluation of arrays of both ligands and targets, and exponentially reducing the time required to identify potential NMEs.

Proteochemometric modeling relies on the application of deep machine learning tools to determine the combined effect of target and ligand

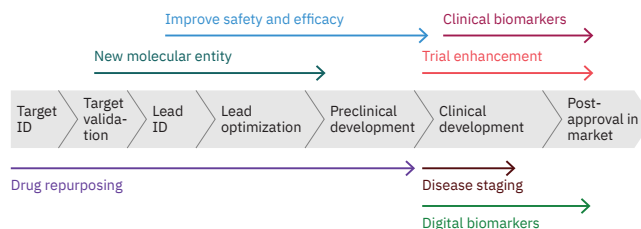
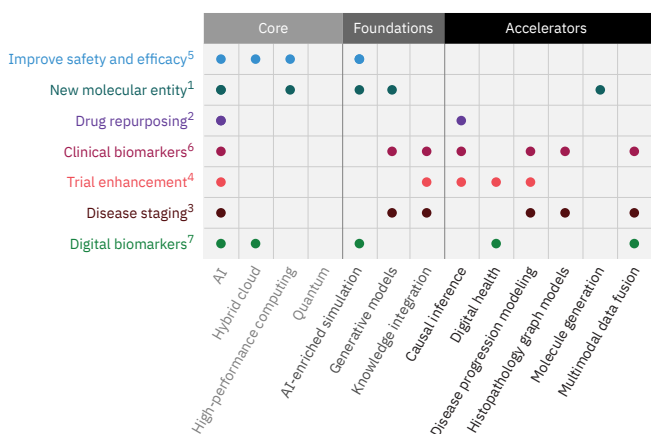


Fig. 1 | Accelerated discovery at a glance. IBM is developing a computing environment for the healthcare and life sciences that integrates the possibilities of next-generation technologies—artificial intelligence, the hybrid cloud, and quantum computing—to accelerate the rate of discovery along the drug discovery and development pipeline.

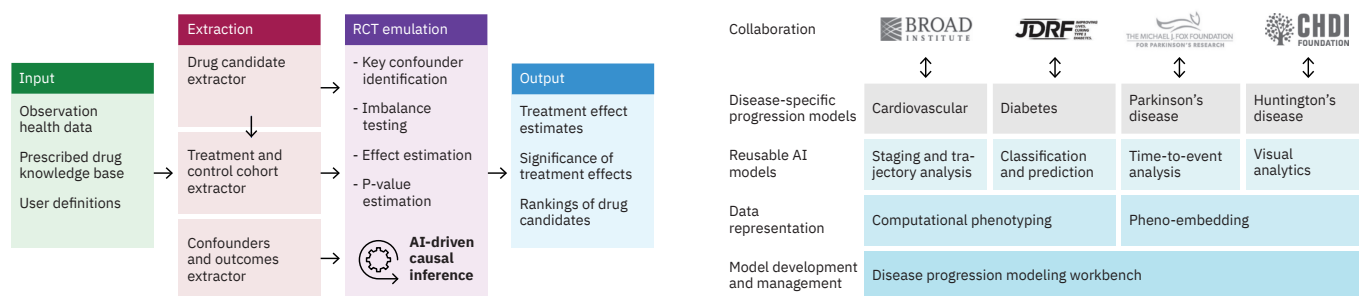


Fig. 2 | Focusing on therapeutics and biomarkers. The identification of new molecular entities or the repurposing potential of existing drugs², together with improved clinical and digital biomarker discovery, as well as disease staging approaches³, will substantially accelerate the pace of drug discovery over the next decade. AI, artificial intelligence.

parameter changes on the target-ligand interaction. This bimodal approach is especially powerful for large classes of targets in which active-site similarities and lack of activity data for some of the proteins make the conventional discovery process extremely challenging.

Protein kinases are ubiquitous components of many cellular processes, and their modulation using inhibitors has greatly expanded the toolbox of treatment options for cancer, as well as neurodegenerative and viral diseases. Historically, however, only a small fraction of the kinome has been investigated for its therapeutic potential owing to biological and structural challenges.

Using deep machine learning algorithms, IBM researchers have developed a generative modeling approach to access large target-ligand interaction datasets and leverage the information to simultaneously predict activities for novel kinase-ligand combinations¹. Importantly, their approach allowed the researchers to determine that reducing the kinase representation from the full protein sequence to just the active-site residues was sufficient to reliably drive their algorithm, introducing an additional time-saving, data-use optimization step.

Machine learning methods capable of handling multimodal datasets and of optimizing information use provide the tools for substantially accelerating NME discovery and harnessing the therapeutic potential of large and sometimes only minimally explored molecular target spaces.

Drug repurposing from real-world data

Electronic health records (EHRs) and insurance claims contain a treasure trove of real-world data about the healthcare history, including medications, of millions of individuals. Such longitudinal datasets hold potential for identifying drugs that could be safely repurposed to treat certain progressive diseases not easily explored with conventional clinical-trial designs because of their long time horizons.

Turning observational medical databases into drug-repurposing engines requires the use of several enabling technologies, including machine learning-driven data extraction from unstructured sources and sophisticated causal inference modeling frameworks.

Parkinson's disease (PD) is one of the most common neurodegenerative disorders in the world, affecting 1% of the population above 60 years of age. Within ten years of disease onset, an estimated 30–80% of PD patients develop dementia, a debilitating comorbidity that has made developing disease-modifying treatments to slow or stop its progression a high priority.

IBM researchers have now developed an AI-driven, causal inference framework designed to emulate phase 2 clinical trials to identify candidate drugs for repurposing, using real-world data from two PD patient cohorts totaling more than 195,000 individuals². Extracting relevant data from EHRs and claims data, and using dementia onset as a proxy for evaluating PD progression, the team identified two drugs that significantly delayed progression: rasagiline, a drug already in use to treat motor symptoms in PD, and zolpidem, a known psycholeptic used to treat insomnia. Applying advanced causal inference algorithms, the IBM team was able to show that the drugs exert their effects through distinct mechanisms.

Using observational healthcare data to emulate otherwise costly, large and lengthy clinical trials to identify repurposing candidates highlights the potential for applying AI-based approaches to accelerate potential drug leads into prospective registration trials, especially in the context of late-onset progressive diseases for which disease-modifying therapeutic solutions are scarce.

Enhanced clinical-trial design

One of the main bottlenecks in drug discovery is the high failure rate of clinical trials. Among the leading causes for this are shortcomings in identifying relevant patient populations and therapeutic endpoints owing to a fragmented understanding of disease progression.

Using unbiased machine-learning approaches to model large clinical datasets can advance the understanding of disease onset and progression, and help identify biomarkers for enhanced disease monitoring, prognosis, and trial enrichment that could lead to higher rates of trial success.

Huntington's disease (HD) is an inherited neurodegenerative disease that results in severe motor, cognitive and psychiatric disorders and occurs in about 3 per 100,000 inhabitants worldwide. HD is a fatal condition, and no disease-modifying treatments have been developed to date.

An IBM team has now used a machine-learning approach to build a continuous dynamic probabilistic disease-progression model of HD from data aggregated from multiple disease registries³. Based on longitudinal motor, cognitive and functional measures, the researchers were able to identify nine disease states of clinical relevance, including some in the early stages of HD. Retrospective validation of the results with data from past and ongoing clinical studies showed the ability of the new disease-progression model of HD to provide clinically meaningful insights that are likely to markedly improve patient stratification and endpoint definition.

Model-based determination of disease stages and relevant clinical and digital biomarkers that lead to better monitoring of disease progression in individual participants is key to optimizing trial design and boosting trial efficiency and success rates.

A collaborative effort

IBM has established its mission to advance the pace of discovery in healthcare and life sciences through the application of a versatile and configurable collection of accelerator and foundation technologies supported by a backbone of core technologies (Fig. 1). It recognizes that a successful campaign to accelerate discovery for therapeutics and biomarkers to address well-known pain points in the development pipeline requires external, domain-specific partners to co-develop, practice, and scale the concept of technology-based acceleration. The company has already established long-term commitments with strategic collaborators worldwide, including the recently launched joint Cleveland Clinic-IBM Discovery Accelerator, which will house the first private-sector, on-premises IBM Quantum System One in the United States. The program is designed to actively engage with universities, government, industry, startups and other relevant organizations, cultivating, supporting and empowering this community with open-source tools, datasets, technologies and educational resources to help break through long-standing bottlenecks in scientific discovery. IBM is engaging with biopharmaceutical enterprises that share this vision of accelerated discovery.

"Through partnerships with leaders in healthcare and life sciences worldwide, IBM intends to boost the potential of its next-generation technologies to make scientific discovery faster, and the scope of the discoveries larger than ever," said Royyuru. "We ultimately see accelerated discovery as the core of our contribution to supercharging the scientific method."

1. Born, J. et al. *J. Chem. Inf. Model.* **62**, 240–257 (2022).
2. Laifenfeld, D. et al. *Front. Pharmacol.* **12**, 631584 (2021).
3. Mohan, A. et al. *Mov. Disord.* **37**, 553–562 (2022).
4. Harrer, S. et al. *Trends Pharmacol. Sci.* **40**, 577–591 (2019).
5. Parikh, J. et al. *J. Pharmacokinet. Pharmacodyn.* **49**, 51–64 (2022).
6. Kashyap, A. et al. *Trends Biotechnol.* (2021). <https://doi.org/10.1016/j.tibtech.2021.11.006>
7. Norel, R. et al. *npj Parkinson's Dis.* **6**, 12 (2020).

CONTACT

IBM Research
Email: HCLS@ibm.com
Web: research.ibm.com/teams/healthcare-and-life-sciences