Part of the conference series
Transforming our future

Machine learning and AI in biological science, drug discovery and medicine

Held on 1 March 2023

Conference report

THE ROYAL SOCIETY

Supported by
AstraZeneca

Cover Image:
Digital Biology 02 by Khyati Trehan is part of a series of artworks commissioned by Visualising AI, a project that invites cutting-edge artists to spend time with scientists and engineers to discuss key themes in AI, before using those conversations as a catalyst to create new artworks.

Registered Charity No 207043
Issued: August 2023 DES8186_6

Introduction

On 1 March 2023 the Royal Society hosted a hybrid conference on Machine learning and AI in biological research, drug discovery and medicine. This meeting, supported by AstraZeneca, forms part of the Royal Society's Transforming our future series.

The Transforming our future conferences are unique, high-level events that address scientific and technical challenges of the next decade and which bring together leading experts from the wider scientific community including industry, academia, government, and charities. They are organised with the support of the Royal Society Science, Industry and Translation Committee.

The conference series forms part of the Royal Society's Science and Industry programme which demonstrates the Society's commitment to integrate science and industry across its activities, promote science and its value, build relationships, and foster translation.

This meeting brought together senior stakeholders from industry, academia, charities and funding bodies. The programme was organised by Dr Harren Jhoti OBE FMedSci FRS, Astex Pharmaceuticals, Dr Claus Bendtsen, AstraZeneca, and Professor Mihaela van der Schaar, University of Cambridge. Three sessions – focussing upon biology, chemistry and medicine – offered talks on machine learning for target discovery and 'omic technologies, AI-based chemoinformatics, and computational clinical trial design, as well as the classical prediction of protein folding.

The conference concluded with a panel discussion that addressed how machine learning and AI might be used to achieve future advances in medicine and healthcare for societal benefit, followed by a keynote address from Dr Demis Hassabis CBE FREng FRS, DeepMind.

This report is not a verbatim record, but a summary of the discussions that took place during the two days and the key points raised. Comments and recommendations reflect the views and opinions of the speakers and not necessarily those of the Royal Society.

> "We're now at the really exciting stage where we are starting to see the fruits of using AI and machine learning to accelerate scientific discovery and understand the potential of where that can take us"
>
> Dr Demis Hassabis CBE FREng FRS, DeepMind

Image: Delegates during the conference.

Executive summary

Machine learning and artificial intelligence (AI) technologies are proving transformative for biological, medical and pharmaceutical research, contributing to significant advances in basic science, diagnostics, therapeutic target identification and clinical practice. The UK has pioneered a number of developments and breakthroughs in these areas. However, significant challenges remain, in particular around skills, research design, ways of working and accessibility.

Recent successes

- Increased availability of public data has offered additional opportunities to train, test and benchmark the performance of machine learning models
- Self-supervised deep learning models are increasingly capable of transforming complex imaging data for hard-to-treat diseases into formats that can be interpreted by computer algorithms
- Large language models such as OpenAI are increasingly capable of eliciting key information, enhancing descriptions of protein sequences and mining unstructured data in medical records
- Machine learning technologies have already helped accelerate certain diagnostic and treatment pathways and identify patients most likely to benefit from a given therapeutic intervention
- Digital Twins are now able to stratify patients according to risk profile and expected treatment response. They can also elucidate variability in tumour cells and improve understanding of associated mechanistic processes
- The AlphaFold model represents the most significant machine learning breakthrough in recent years, capable of detecting patterns in highly complex systems and identifying protein sequences with the potential to drive major advances in diverse fields including enzyme design and antimicrobial resistance.

> "The potential benefits of AI-driven real-world evidence to improve drug development and patient care across the board are vast. It is up to us to harness this power and to persuade regulatory bodies and other healthcare stakeholders of the immense potential impact"
>
> Dr Aditya Nori, Microsoft Health Futures

Image: Dr Harren Jhoti OBE FREng FRS networking during the conference.

Future priorities

- Encouraging data-sharing between public- and private-sector stakeholders will be critical to bring together multiple sources of data and to accelerate breakthroughs in model performance
- Machine learning specialists should be involved in early-stage research design to establish study protocols. This will maximise the potential to achieve useful answers to appropriate questions – helping to accelerate therapeutic development and licensing
- Collaboration with domain experts, community stakeholders and under-represented patient groups will be increasingly critical in reducing bias and ensuring that machine learning technologies can benefit diverse groups of individuals
- Generating sufficient volumes of high-quality synthetic data can help to minimise model bias and improve the ability of models to handle incomplete and complex real-world information. Drawing upon real-world evidence can further enrich comparative analysis and refine eligibility criteria for clinical studies
- Developing advanced organoids and organs-on-a-chip may help to improve the quality and accuracy of bridging models to reflect in vivo activity
- Undertaking causal machine learning will help to identify intervention points in biological pathways at which therapeutics are likely to have maximum efficacy, and will help us understand variability across patient groups and dosing regimens
- Applying machine learning technologies to the study of zoonotic pathogens (transmitted from animals to humans) and mental health disorders may help to drive breakthroughs in diseases with high burdens.

"I think that many of the audience were keen to engage and challenge the machine learning community represented by our speakers"

Professor Mihaela van der Schaar,
University of Cambridge

Image: Networking during the conference.

KEYNOTE ADDRESS

Using AI to accelerate scientific discovery

Dr Demis Hassabis CBE FREng FRS highlighted how DeepMind's innovations in safe, ethical machine learning and AI over the past decade have paved the way for breakthroughs in scientific discovery for societal benefit.

DeepMind Technologies Ltd. was established in 2010 at a time when few individuals or companies were actively pursuing research and development into AI. DeepMind sought to plumb the depths of human intelligence and recreate superior artificial constructs to accelerate scientific breakthroughs for the benefit of humanity.

DeepMind focussed initially upon testing and developing their algorithms in gaming environments, focusing on "Grand Challenges" in the field of AI. In 2016, DeepMind's AlphaGo programme deployed novel tactics to defeat incumbent Go champion Lee Sedol. Subsequently some of the core strategies were generalised for other two-player games and released as the AlphaZero programme.

Two years later, AlphaZero had, from scratch, mastered Go, chess and shogi (Japanese chess). AlphaZero models drew upon neural networks and large volumes of information to generate datasets of incrementally higher quality and to elicit optimal outputs within a few hours. Two years after that, DeepMind introduced MuZero, which mastered Go, chess, shogi and the video game Atari without needing to be instructed in the rules.

The AlphaStar programme, released in 2019, prevailed at the highly complex Starcraft II online game and further advanced the limits of artificial learning.

Ultimately, DeepMind sought to deploy models and methodologies from this increasingly sophisticated succession of computer programmes to address real-world scientific problems. The most notable innovation to date is unquestionably the AlphaFold system, whose methods and code were published in 2021 and which was awarded Science's 2021 Breakthrough of the Year and Nature Methods' Method of the Year 2021.

Image: Dr Demis Hassabis CBE, FREng, FRS, DeepMind.

AlphaFold is an AI model that can predict the 3D folded structure of a given protein in seconds using the protein's unique amino acid sequence. Despite the dynamic nature of spontaneous protein folding, AlphaFold's predictions are typically accurate to within the width of a single atom. The AlphaFold Protein Structure database[2], hosted by the European Bioinformatics Institute (EMBL-EBI), contains structure predictions for all 200 million proteins currently known to the global scientific community and is already being deployed extensively by over one million researchers worldwide.

AlphaFold illustrates the potential for AI to detect patterns and correlations in dynamic and highly complex systems and to identify optimal solutions in vast combinatorial spaces. Within the past two years the database has facilitated breakthroughs in a range of fields. Research teams developing malaria vaccines have been able to identify the structure of specific proteins that enable disease vectors to thrive. Structure predictions for several critical proteins in the SARS-CoV-2 virus were identified in 2020, and researchers continue to use AlphaFold to study antibodies that can neutralise SARS-CoV-2 and associated variants. The AlphaFold database offers the opportunity to rapidly drive and disseminate major advances within sustainability, enzyme design, antimicrobial resistance, drug development and cell biology.

In the coming years AI may serve as a 'description language' to help fathom the intricacies of biological systems in a similar vein to how mathematics has served to advance the frontiers of physics. DeepMind is already applying AI models across the full range of scientific disciplines including nuclear fusion, quantum chemistry, and weather prediction.

These rapid advances must be underpinned by robust safety and ethical principles. Prior to the publication of the AlphaFold database DeepMind consulted with 30 experts in bioethics, human rights and pharmaceutical development to ensure that the release of data would indeed be of net benefit to the global scientific community. Ongoing reinforcement learning with human feedback seeks to train systems to obey specific rules to reduce the risk of harms and inappropriate outputs. DeepMind will continue to offer thought leadership in this space as innovation in AI and machine learning systems develop apace over the coming decade.

> "We can potentially use these AI and machine learning systems to go beyond human capabilities and beyond what we are able to do as human experts to push the boundaries of scientific knowledge"
>
> Dr Demis Hassabis CBE FREng FRS

> "AI has enormous potential to solve some of our biggest challenges, but it must be built responsibly, safely and for the benefit of everyone"
>
> Dr Demis Hassabis CBE FREng FRS

AI-augmented target identification

Dr Anne Phelan, BenevolentAI, examined how AI and machine learning models can enhance the process of target identification and offer opportunities to advance precision medicine.

Identifying and validating appropriate targets at an early stage of the drug discovery process significantly increases the likelihood of clinical success. BenevolentAI's drug discovery pipeline contains targets for multiple complex diseases.

BenevolentAI's platform is built on a knowledge graph that combines over 85 multimodal data sources including 'omics data, information on biological systems and scientific papers. It seeks to integrate disease traits and mechanisms and remove therapeutic silos to connect shared mechanisms across diseases and generate endotype-specific target predictions.

Knowledge graph integration

To overcome problems associated with handling varied data of inconsistent quality, BenevolentAI uses processing pipelines to unify and integrate data into their knowledge graph. Natural language processing consolidates unstructured data from literature, patents and clinical trials. Genetics data (eg from genome-wide association studies) and 'omics data have variant-to-gene annotations layered over them, and this is built back into the data infrastructure. Data provenance is monitored and quality assessment performed on an ongoing basis.

Target identification follows a standard process:

1. BenevolentAI's drug discovery scientists pose a series of specific biological questions to the suite of AI models to explore relevant disease biology. Questions are carefully formulated to identify potential hypotheses.
2. Multiple machine learning models – including large language, genomic, and transcriptomic models – are run to generate complementary target predictions. Smart aggregation models draw upon previous experience to help refine predictions and shortlist targets with a high probability of success. These typically range from relatively well-known to entirely novel targets.

Image: Dr Anne Phelan, BenevolentAI.

3. Targets are evaluated and triaged systematically based upon biological rationale, safety, druggability, and precision medicine potential, using supporting evidence and underlying logic displayed by the tools. Novel targets require a particularly robust biological rationale in order to progress to the validation stage. Decisions are captured in a structured and unstructured form to provide an audit trail and to help models to learn.
4. Triaged targets are validated in laboratory experiments using whatever modality is the most relevant to best test the hypothesis, such as induced pluripotent stem cells, patient-derived samples, CRISPR, or small molecules. These data are used to decide whether the target will pass into BenevolentAI's drug discovery pipeline, and are then integrated back into the knowledge graph to help the system learn.

The future of AI-augmented target identification

AI-augmented drug discovery is driven by three key elements: the ever-growing availability of biomedical data, continual improvements to machine learning algorithms, and greater computational power. A key challenge for the next 5-10 years will be to exploit increasing swathes of data, formulating progressively complex and iterative queries to identify commercially viable targets with acceptable risk profiles.

> "The use of AI and machine learning tools that enable us to extract maximum benefit and information from the totality of data available to us is really on the brink of a revolution"
>
> Dr Anne Phelan, BenevolentAI

Advances are already being shaped by:

- Large language models such as OpenAI which offer great potential to rapidly identify salient information
- Structural tools such as AlphaFold to improve binding pocket elucidation, and enhanced understanding of epitope definition and binding sites. These allow consideration of every target, irrespective of the drug modality required to prosecute that target
- Developments in high-content imaging, single-cell RNA sequencing and pharmacokinetic/pharmacodynamic modelling to inform tissue-specific drug delivery
- Genetic and epigenetic analysis of patient-derived samples, longitudinal studies and patient-derived biomarkers to create a more patient-centric view of drug discovery
- Clinical trial data, electronic health records and transcriptional profiling to help understand the risk-benefit profile of targeting specific genes
- Trial design optimisation, diagnostics and precision medicine to harness commercial opportunities

Maximising the potential of these tools in conjunction with patient-specific contextual information is likely to prove instrumental in enhancing possibilities for genuinely personalised medicine over the coming decade.

Mapping and navigating biology and chemistry with genome-scale imaging

Dr Imran Haque, Recursion Pharmaceuticals, provided a detailed overview of how Maps of Biology can contribute to the discovery of novel therapeutics for disease.

Image: Dr Imran Haque, Recursion Pharmaceuticals

Maps of Biology are an exciting tool for the discovery of new therapeutics to treat disease. Maps are constructed using deep learning models which process images from experiments conducted in Recursion's laboratories. These experiments involve undertaking cell perturbations, such as knocking out individual genes or dosing hundreds of thousands of molecules at multiple concentrations.

Images are created using standardised assays (phenomics) which stain six common cellular structures. The images are then analysed by deep learning models which compare cellular phenotypes to expose commonalities and differences between perturbations (Figure 1). Images are typically data-rich and cheap to produce, which means that the technology can be scaled more quickly than molecular 'omics techniques such as RNA sequencing.

Although phenomics is capable of sensitive detection and able to quantify thousands of mechanisms, unstructured images must be translated into data to enable models to learn from them. Custom-trained deep learning algorithms can extract biologically meaningful representations of images and automatically correct for inherent experimental "noise" such as batch effect.

'One of the huge advantages of these algorithms is that they learn on data and can thus very rapidly accelerate the development of new therapeutics"

Dr Imran Haque, Recursion

FIGURE 1

Confocal microscopy images, such as these phenomics-stained cells, are used by Recursion's deep learning models to create Maps of Biology. The models compare disease phenotypes with healthy controls, as well as phenotypes of cells when dosed with varying concentrations of compounds.

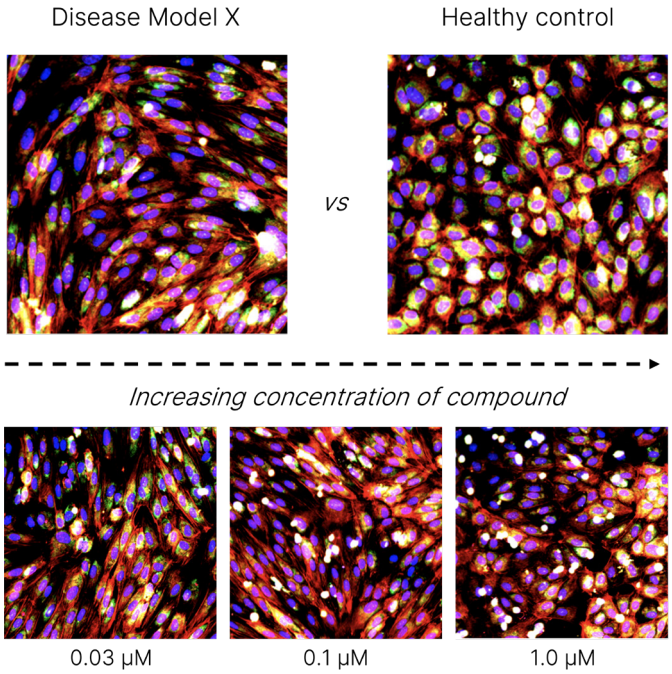


Image: Courtesy of Recursion Pharmaceuticals.

"One of our goals is to find differences between cells that humans themselves may not be able to identify"

Dr Imran Haque, Recursion

Recursion has amassed a vast library of images. Three case studies illustrate the potential to uncover novel therapeutics:

De novo pathway reconstruction

Maps of Biology can help develop models for biochemical pathways. Recursion has used multiple CRISPR guides to knock out over 17,000 genes in different cells. Deep learning models then identify which knockouts generate statistically significant similar cellular phenotypes. This knowledge is used to infer cellular pathways de novo, as models have no prior exposure to scientific literature.

Primary and alternative target selection

Maps of Biology can also prove instrumental in selecting primary and alternative targets. They have helped to identify a new CDK12-adjacent target, RBM39 (or 'Target Gamma') that may offer a novel means to circumvent the biochemical roadblock presented by the structural similarity between CDK12 and CDK13 proteins, which can inhibit the efficacy of treatments for homologous recombination-proficient cancers.

High-dimensional genome-wide screening

The maps allow comparisons between the effects of a compound and the effects of any gene knockout on cell phenotype. Comparisons can be graphically visualised using AI, as demonstrated in Recursion's publicly available compound intelligence tool MolRec, which offers a 'keyhole view' into relationships between small molecules and gene knockouts within a subset of Recursion's total dataset. Additional analyses can elucidate cellular toxicity, on- and off-target similarity, compound similarity and dose response.

Recursion has also released one of the largest publicly available datasets of perturbative cellular imaging, RxRx3. This consists of 2.2 million samples including images, metadata and deep learning embeddings of knockouts. It features around 17,000 genes and multiple concentrations of approximately 1,700 compounds. This dataset is likely to play a critical role in facilitating collaborative advances in machine learning technology and drug discovery during the next decade.

Digital twins for personalised oncology

Professor Walter Kolch, University College Dublin, discussed digital twin technologies that are designed to improve cancer diagnosis and treatment.

Cancer incidence is rising steeply, but survival rates are not keeping pace. This is due in part to the challenges associated with isolating effective therapeutics for a given patient. Digital twins – which offer a virtual representation of a physical entity – could help address these difficulties by creating a virtual patient blueprint to help achieve a diagnosis and trial potential therapies in-silico.

Digital twins are already able to identify high-risk cancer patients and synergistic drug combinations by modelling molecules and intracellular interactions. The next step will be to move beyond the cell level to develop 3D models of tumours. The ultimate goal is to model multi-organ interactions and construct a digital blueprint to evolve alongside a patient's treatment journey.

Digital twins to improve diagnosis: identifying high risk neuroblastoma patients

Neuroblastoma is a tumour of the peripheral nervous system, constituting 6% of paediatric cancers but accounting for 15% of childhood cancer deaths. Half of neuroblastomas are high-risk and require aggressive treatment. Currently, only 50% of these high-risk patients are reliably identified from known molecular biomarkers (such as amplification of the gene MCYN). Digital twins can help identify the remaining 50% from an alternative biomarker. A key difference between conventional biomarkers and digital twin biomarkers is that they are not a molecule or set of molecules, but a computer model.

To create digital twin models, molecular data (such as protein interactions, protein phosphorylation, gene expression, and drug responses) generated from cultured neuroblastoma cells are developed into a computer model of the molecular disease in a two-step process. The first involves integration of various molecular data using graph attention networks. In the second step, control modules of the molecular cancer network are identified and extracted as mechanistic ODE (ordinary differential equation) models using a new method developed by the Kolch lab, cSTAR (cell State Transition Assessment and Regulation). By then populating these ODE models with patient-specific data, digital twin models of individuals are created.

Image: Professor Walter Kolch, University College Dublin.

These personalised models are fully tractable and can predict dynamic changes such as disease progression or response to therapy. They accurately find high risk patients and can even be used to identify whether a patient will respond well to chemotherapy, or whether they need another form of treatment.

Predicting tumour response to chemotherapy

Tumour cells can vary widely in their responses to the same drug, and this heterogeneity can make cancer treatment difficult. Mapping thousands of tumour cells can help.

Experiments from the Kolch laboratory found that while some tumour cells positively respond to chemotherapy by triggering the JNK pathway (leading to cell death), other cells fail to activate this stress response and are 'resistant' to drugs. Ensemble modelling of thousands of cells in the same tumour can accurately predict whether the population of cells will react positively or negatively to a treatment.

Using the predictions for how a tumour will react to a given treatment, patients can be sorted into groups based on their best expected response to a particular drug. This enables personalisation of the therapy given to each individual patient.

Digital twins to design therapies

Managing the fates of cancer cells rather than killing a cancer is a new principle that could improve patient prognosis, and cSTAR can help with this. cSTAR is a method that uses machine learning to classify cell states based on 'omics data (or integrated 'omics data). It then identifies key network components that control the cell states and provide instructions how to convert one cell state into another. Through predictive simulations it can develop drug targets - for example, converting proliferating cancer cells into non-proliferating, differentiating `normal' cells. In this way, a malignant proliferating neuroblastoma cell could be targeted to become a benign differentiating cell. Manipulating cancer cell fates can potentially avoid the development of cancer relapse arising from the selection of resistant cells by cytotoxic treatment regimens.

A vision for the future: digital twins as life-long health companions

Generating a holistic digital twin of a person at birth, that is consistently updated and evolves with the individual through their lifetime, would have a major impact across healthcare. Amongst other things, it could assist in early diagnosis, guide personalised treatment, influence patient lifestyle, and reduce healthcare costs.

> "Digital twin models can break the deadlock of hard-to-treat cancers by personalising diagnosis and therapy"
>
> Professor Walter Kolch, University College Dublin

FIGURE 2

The future: a digital twin modelling platform.

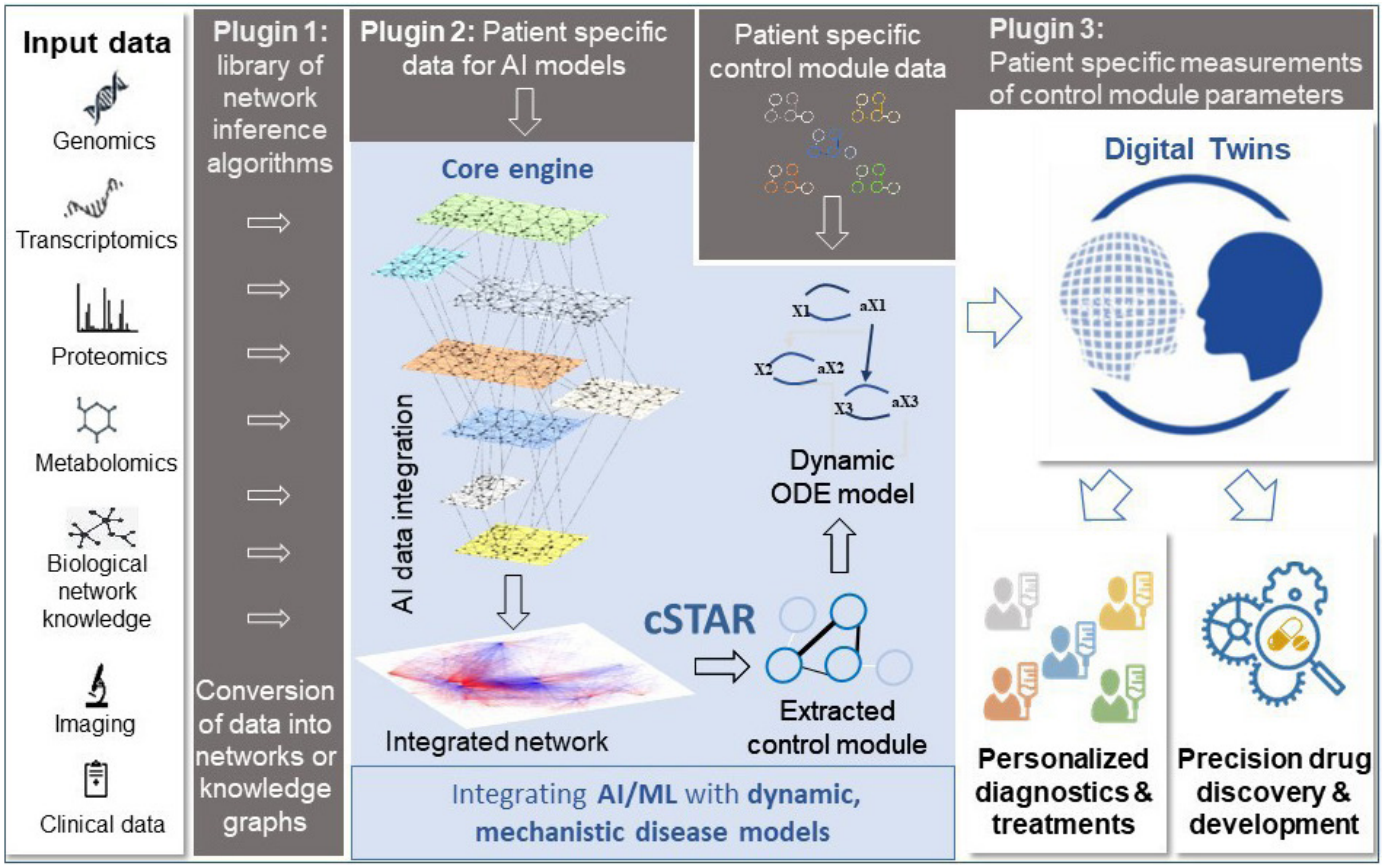


Image: Courtesy of Professor Walter Kolch, University College Dublin.

Q&A DISCUSSION

Biology

Dr Anne Phelan, Dr Imran Haque and Professor Walter Kolch discussed technical questions relating to bias and error, public data, and causality, and explored broader issues around patient-specific models, rare diseases, and funding.

Technical questions

- Engines such as the Benevolent Platform™ are informed by training data, which inevitably generate some level of bias towards better understood diseases. The Benevolent team have dedicated considerable time to refining data representation and distilling signals from noise. Their disease-agnostic knowledge graph covers all cell types, multiple biological signals, and does not prioritise specific types of dysregulation.
- Even the most sophisticated language models can offer incorrect predictions. Mitigating strategies include building substantial data dictionaries, examining the accuracy of data behind openAI summaries, and profiling compounds and genetic knockouts on an individual basis. Hybrid machine learning models, which complement traditional models with novel workflows, are more easily tractable. They facilitate defensible decision-making, essential for therapeutics to progress to clinical trial and secure regulatory approval.
- While large public datasets offer a valuable opportunity for model training, they can present considerable variation in structure and in availability of metadata. This is one of the most significant challenges which, if adequately addressed, could transform the potential of machine learning methodologies. Academic research groups who have generated detailed data on specific diseases can add significant value. Seeking instances of consensus in multiple datasets can generate confidence. EU funding for FAIR (findable, accessible, interoperable and reusable) data remains available for restructuring activities.
- Establishing causality is challenging, especially for novel and clinically relevant target-disease pairs. Individualised models and, at smaller scales, mechanistic models can mitigate the effect of harmful feedback loops. Identifying pivotal points in signalling cascades and undertaking experimental ablation (remodelling biological pathways to try to establish therapeutic benefit) can help determine causal relationships. Ascertaining which genes produce similar phenotypes and distinguishing mechanisms of gene interaction can assist in identifying therapeutic interventional pathways.

> "Using public data to benchmark and tune model performance could be the next great frontier for what these machine learning methods could do, were data made more available and more consistent"
>
> Dr Imran Haque, Recursion

> "We need to bring stakeholders together: the data scientists, the biologists, the clinicians, and the patients in order to design solutions that are fit for purpose"
>
> Professor Walter Kolch, University College Dublin

Broader considerations

- Many machine learning models are based primarily upon biological data at population level. Designing effective therapeutics for specific patient groups will require incorporation of additional layers of data into patient-specific models. Integration of socioeconomic data is not yet widespread but is an important future consideration.
- Recursion initially focussed its discovery portfolio upon rare genetic diseases. Although it can be difficult to justify initial investment into clinical trials for these diseases the Recursion platform profiles relevant compounds and genetic knockouts relatively rapidly, accelerating the discovery process in conjunction with patient advocacy groups. Where insufficient data is available on a specific rare disease it can be possible in certain – though not many – cases to extrapolate mechanistic data from related diseases.
- It would be valuable if future research grants awarded to academic groups included funding to procure, structure and preserve data, to develop skills to unify datasets, and to refine and publish software to enable algorithms to be used more widely. Small pools of disease-specific data generated by academic institutions and some biotechnology companies remain challenging to unify. Relevant lessons can be drawn from the RECOVERY trial[1] and recent initiatives within the UK Biobank.

"Academic groups – who produce very detailed, very precise data on a particular disease – can add massive value to our capacity to process and integrate data into unified structures. Bespoke datasets that can become more publicly available would enrich the whole process"

Dr Anne Phelan, BenevolentAI

"Building partnerships between the public and the private sector to unify disparate data sources and to encourage data sharing is really valuable"

Dr Imran Haque, Recursion

Image: Left to right: Professor Walter Kolch, Dr Claus Bendtsen, Dr Anne Phelan and Dr Imran Haque.

1. https://www.recoverytrial.net/

Integrating chemical and biological data: A focus on relevance and translation will boost *in vivo*-relevant drug discovery

Dr Andreas Bender, University of Cambridge and Pangea Botanica, highlighted the importance of a focus on clinical translation when using machine learning on life science data.

Significant sums of money have been invested in machine learning and AI for drug discovery over the past two decades. Research teams worldwide have access to large volumes of relatively clean data, defined ontologies and powerful algorithms. The number of projects advancing to successful clinical trials (and hence drug approval) has nonetheless been smaller than anticipated, and with fewer novel drug targets identified. In many cases, safety and efficacy data generated by models and algorithms are not yet of sufficient quality to facilitate translation to the clinic, and often do not adequately account for variability between patient groups and different drug administration routes.

Image: Professor Andreas Bender, University of Cambridge and Pangea Botanica.

Advances in drug discovery will be driven not only by refinements to machine learning models but by improvements to the quality of in vivo hypotheses and model utility for practical decision-making. An enhanced understanding of the limitations of available data will be equally important. The most impactful work is likely to be achieved by developing high-quality therapeutic compounds targeted to specific patient groups within Phase 2 trials (i.e., matching the right compound to the right patient). Key challenges are likely to include:

1. Identifying and predicting relevant in vivo endpoints

It will be important to reduce dependence upon biochemical proxy assay data to train models, and work to understand relationships between assay endpoints and specific in vivo effects.

Considerations of pharmacokinetics and dose exposure should be incorporated into models, and better disease characterisation should be undertaken to reduce dependence on the symptom level.

Improving understanding of relevant aspects of systems biology and pharmacology will help researchers characterise and quantify links between drugs, targets and disease.

It will also be important to recognise the limitations of conditional data in machine learning models that can over-simplify therapeutic processes and effects, failing to account for differences in efficacy or toxicity between patients. This is especially relevant for adverse reactions which can vary considerably over time and according to dose, gender, or the presence of certain disease phenotypes.

2. Acceptance that model validation is not equivalent to process validation

Chemical space – which encompasses all possible compounds and drug molecules – is large and multi-dimensional. Validating machine learning models within narrow segments of chemical space is challenging; statistical analysis and conclusions from such models must be treated with caution. In particular, validating a machine learning model is not the same as improving the outcome of a drug discovery process, which must take disease context and experimental validation into account to achieve practical impact.

3. Human psychology

The Western scientific approach tends to be highly analytical, dissecting complex systems and problems to aid understanding. Too often, research operates within specialised silos, whereas human biology is heterogenous, dynamic, interconnected, and partially stochastic. To drive advances in drug discovery experts from multiple scientific disciplines must collaborate intensively, and scientific education at all stages must foster interdisciplinary and translational approaches.

> "The key questions are always: how can we make better decisions based on the data? What can we do with predictions from models to enable translation to the clinic? The questions must come first."
>
> Professor Andreas Bender, University of Cambridge and Pangea Botanica

Future priorities

Advances in mRNA technology, oligonucleotide therapeutics and antibody drug conjugates are likely to increase the number of druggable targets in the next decade. Research must, however, remain guided by scientific needs, not simply by technological capability ('science pull', instead of 'technology push'). Tools such as cell painting and digital twins can help research teams to extrapolate results and improve their knowledge of biological systems. Multi-organisational consortia will be needed to generate data in a pre-competitive context using experimental design. Ultimately, patient-centred questions must guide model development and data acquisition at all stages of the drug discovery process.

Machine learning to predict protein function from sequence: therapeutic applications

Dr Lucy Colwell, University of Cambridge, explored how deep learning models uncover the functionality of unknown proteins from amino acid sequences, and how this can contribute to the development of novel therapeutics.

Image: Dr Lucy Colwell, University of Cambridge.

Predicting the functional properties of a protein from its amino acid sequence is important for two key challenges in biochemistry: protein optimisation and protein discovery. Structure prediction can accelerate the identification of new proteins with specific functionality and enhance understanding of the functional effect of genomic mutations. Deep learning (DL) models can now accurately predict functional domains within protein sequences, and large language models can generate textual descriptions of protein sequences, adding millions of annotations to public databases.

Protein optimisation

Protein optimisation underpins the development of biotechnological tools such as antibodies and nanobodies that are used to treat disease.

A familiar method is model-based optimisation (MBO). This involves synthesising and measuring designed sequence variants, using measurement data to build and fine-tune a model capable of learning sequence representations or embeddings from protein sequences where proximity encodes functional activity, and optimising models to select subsequent sequences.

This is exemplified by the recent development of two nanobodies against SARS-CoV-2. The Colwell laboratory optimised 46 of 125 nanobody residues to improve binding affinity to the virus. Sequence libraries were created by combining classification and regression models to identify nanobody sequences able to bind to specific targets. Experimental validation using AlphaSeq, a multi-objective optimisation tool, permitted many sequences to be analysed in a single experiment. Several rounds of optimisation produced multiple nanobody sequences with high binding affinity to SARS-CoV-2.

To validate these sequences, those with the highest affinity were evaluated for neutralisation capacity in several COVID variants. Certain sequences designed by the machine learning model demonstrated a greater than 50-fold improvement in neutralisation and binding to SARS-CoV-2.

Protein discovery

Protein families can be predicted from amino acid sequences. The PFam seed database – a human-verified set of protein domain sequences – was used by the Colwell lab to construct a series of ResNets, deep neural networks that use 'residual connections' to recognise patterns and determine the most important elements of a given sequence. Clustered splits, which divide data into groups based on similarity, enabled models to identify sequences that did not resemble training data and thus offered additional information to complement existing protein discovery methods and augment the PFam database.

Using a simple masking scheme, which conceals certain words in a text, the machine learning framework BERT (Bidirectional Encoder Representations from Transformers) was used to increase a transformer classifier's understanding of language. Transformer classifiers are computer programmes that can classify text into different categories. The masking removed the need for sequence alignment to achieve protein annotation in DL models.

In a collaboration between the Colwell laboratory and the European Bioinformatics Institute (EMBL-EBI), this technique was used to create annotations for 1.5 billion proteins. These annotations were released into databases in May 2022.

Beyond classification

UniProt is a freely accessible protein database that provides information about millions of protein sequences, functions and interactions. At present, one third of proteins within the database are uncharacterised. Although standard models cannot determine protein structure and role, language models have demonstrated the potential to generate free text descriptions of proteins based upon amino acid sequences.

> "Many mutations developed by machine learning models may not be deemed by a human to be intuitively beneficial, but they turn out to create the best sequences. This attests to the fact that models can really see through sequence space"
>
> Dr Lucy Colwell, University of Cambridge

It remains challenging to generate a protein name from a sequence. The Colwell lab trained models to perform multiple protein tasks such as 3D structure prediction, function prediction or mutation effect. They were then instructed to translate between amino acid sequences and text. UniProt curators ascertained that in almost 90% of cases the model matched or exceeded the semantic content of existing annotations. Consequently, nearly 50 million 'uncharacterised proteins' were assigned a DL-generated name.

These findings indicate that deep learning models have significant potential to unravel protein structure and function and to develop nanobody sequences with high binding affinities to pathogens, helping to accelerate the development of novel therapeutics. This will be critical for future global health crises. Community curation and feedback on model accuracy will help drive similar advances in the machine learning field in the next decade.

Exploring the ability of machine learning-based virtual screening models to identify the functional groups responsible for binding

Professor Charlotte Deane MBE, University of Oxford, discussed how a novel neural network can predict protein-ligand binding affinity and how biases in data can influence predictions.

Image: Professor Charlotte Deane, University of Oxford.

PointVS is a group-invariant graph neural network developed by the Deane laboratory at the University of Oxford. It is trained to predict binding affinity, to reveal the key nodes which influence predictions, and to identify important binding sites in a protein pocket, regardless of differences between the binding target and the training data.

Common challenges

A significant proportion of data in the public domain was not generated specifically for machine learning. Datasets often lack sufficient metadata and may not exist at the preferred scale. Researchers must consider appropriate methodologies and remain mindful of limitations in model generalisability. This is particularly relevant for models to predict binding affinity, which are often trained on relatively selective datasets, and which therefore exhibit bias towards particular proteins or molecules.

Close similarities between training sets and test sets mean that many models – especially those with 'black box' algorithms – perform well only on targets which strongly resemble those in relevant training sets.

Difficulties associated with integrating physics into machine learning models make it challenging to predict binding affinity between proteins and small molecules. PointVS seeks improvements to model generalisability, especially when handling novel data.

Testing PointVS

Tests were undertaken to ascertain the potential for PointVS to achieve pose selection and affinity prediction. A docking power test evaluated the ability to identify a close-to-native binding pose of a ligand amongst decoys. A scoring power test assessed the extent to which PointVS could produce binding scores with a linear correlation to binding data. These tests were underpinned by advanced bias correction exercises.

FIGURE 3

Top five edges by attention (from PointVS).

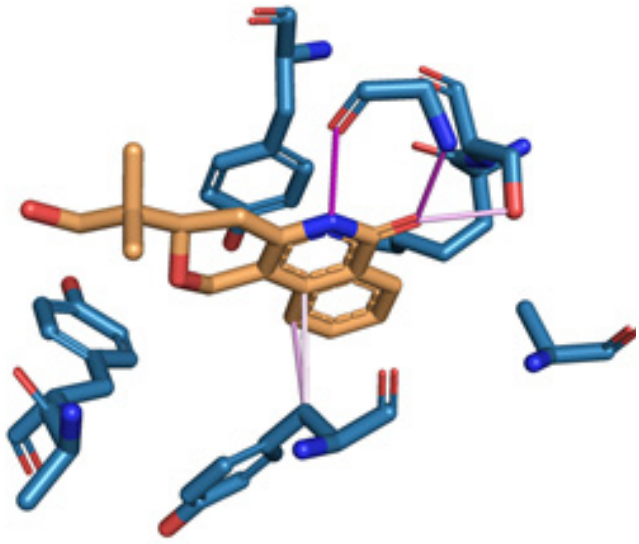


Image: Courtesy of Prof Charlotte Deane.

Identifying important protein atoms using input attribution

A sophisticated online tool called PLIP (protein-ligand interaction profiler) was used to benchmark the capacity of PointVS to identify covalent, hydrophobic, and other bonds. This showed that PointVS demonstrates a strong correlation with binding interactions and suggests that reductions in bias help PointVS to identify the physics of protein-ligand interactions.

PointVS was tested on highly simplified synthetic data to help eliminate noise and bias and to maximise model performance under controlled conditions. The synthetic dataset was formulated to ensure that the model would be most likely to deliver successful predictions when presented with data of adequate breadth and quality.

Attribution-inferred hotspots

Using fragment screen data from crystallography, PointVS was able to use attribution techniques to identify and map important ligand binding atom pharmacophores, or 'hotspots'. Hotspots identified in this way can be used to elaborate fragments to generate molecules that satisfy imposed pharmacophoric constraints, which in turn can assist in developing drugs with strong affinity for a particular protein.

However, hotspots identified by the machine learning model were similar, but not identical, to hotspots generated by physics models. This implies, for instance, that drug development outcomes will depend on the method used to identify hotspots. PointVS was shown to generate molecules with superior standardised ligand efficiency and ranking capabilities when compared to molecules built by STRIFE, a tool for fragment elaboration.

Such deep learning-based models and methodologies may be able to further unpick the physics of binding. Deploying synthetic data and working to remove bias will be critical to advance the field and to develop trustworthy models capable of predicting functional groups responsible for binding.

> "Using clean synthetic data to test how well models perform can really help when later trying to understand how models react to incomplete real data"
>
> Professor Charlotte Deane MBE, University of Oxford

Q&A DISCUSSION

Chemistry

Dr Andreas Bender, Professor Charlotte Deane and Dr Lucy Colwell explored a number of technical elements of machine learning models within chemoinformatics including experimental design, algorithms, and explainability.

Image: From left to right: Dr Andreas Bender, Professor Charlotte Deane, Dr Lucy Colwell.

- Simulated data is relatively easy to control and test. It can be used to explore algorithms' full range of capabilities, but tends to offer simpler insights compared to more complex real-world data, which must be used to refine and test models. A future priority will be to generate sufficient quantities of structured in vivo data from clinical trials to inform improvements to machine learning models.
- Experiments must be carefully designed to collect the right types of data. Involving specialists in machine learning and computation trial design from the start is crucial. Exposing domain experts to examples of successful protocols and trials informed by machine learning is a valuable strategy to foster interdisciplinary collaboration.
- Establishing at the outset which questions a model seeks to address will enable researchers to narrow the list of data types (such as nucleotide sequences, codon sequences, structural data) that are required. Ablations can indicate the value of incorporating additional data, but in principle models should remain as simple as possible to address the questions and obtain relevant clinical endpoints. Comparing baseline results with operationalised algorithms can help identify the minimum quantity and quality of data required to produce an instructive model.
- In some cases, machine learning tools (such as orthogonal assays) and data from historical models can help eliminate false positives generated during high-throughput screening. Compounds that appear positive across a range of target classes are unlikely to be true positives. Further improvements might be achieved by informing models of the estimated error rate in a given dataset.

- To date, chemoinformatics research has focussed largely upon understanding protein sequences, and less on non-human pathogens. This is likely to be a future research frontier in machine learning. Another priority may be enhancing understanding of mechanistic disease pathways in mental health disorders, moving beyond disease classification based largely upon symptom clusters.
- Explainability – demonstrating what is happening within a model and working to understand the rationale for predictions and decisions – is especially difficult within chemical space compared to protein sequence space. Continued exploration will be essential.

"We need to place more emphasis upon baselines in our papers to clarify that with a certain dataset, one can achieve a certain amount, but that the potential step change from using machine learning models is significant"

Professor Charlotte Deane MBE, University of Oxford

"We need to show that collecting different types of data can really work well – seeing conclusions and insights from such data can be motivating and empowering"

Dr Lucy Colwell, University of Cambridge

"We need to ask ourselves: do we have enough data, the relevant chemical endpoints and the appropriate chemical space to make the right decision given the practical question being asked?"

Professor Andreas Bender, University of Cambridge and Pangea Botanica

"If you want to solve problems you need to bring people together, and not just from biology and chemistry"

Professor Charlotte Deane MBE, University of Oxford

Transforming the practice of medicine – a human-centred perspective

Aditya Nori, Microsoft Health Futures, spoke about a human-centred approach to AI in medicine. To offer benefit to healthcare systems, the development of AI technologies must prioritise applicability in real-world environments.

Long-term trends such as ageing populations and increasingly complex clinical procedures, exacerbated by backlogs from the COVID-19 pandemic, means that demands upon healthcare services are higher than ever. Productivity concerns, labour shortages and budgetary pressures are placing increasing strain upon healthcare systems worldwide, hampering the development of new treatments and diagnostic tools.

Human-centric AI could help address some of these challenges, improving the capacity of healthcare systems to prioritise the needs of patients, caregivers, and the clinical workforce. It is, however, essential that approaches to integrating AI into healthcare systems remain safe, responsible, and transparent.

"This project demonstrates how AI technology can be used to support clinicians without removing them from the decision-making process. Clinicians remain in control and are accountable for patient care, but benefit from AI which serves to augment their capabilities"

Dr Aditya Nori, Microsoft Health Futures

CASE STUDY 1

Project InnerEye

Project InnerEye exemplifies human-centric AI. It assists healthcare professionals with labour-intensive tasks and accelerates their workflow, enabling them to dedicate more time to patient care.

Radiotherapy treatment plans are created using 3D scans of a patient. These are programmed into a linear accelerator to direct radiation to tumours while sparing healthy tissue. Delineating tumours from healthy organs using existing tools can take several hours, generating a major bottleneck in treatment workflows. Project InnerEye uses 39 million trainable parameters to independently contour tumours in head, neck and prostate CT scans. It achieves clinical-grade accuracy and in many cases does not require any adjustment. It offers the potential to reduce clinician time by up to 90%.

Project InnerEye technology is now being expanded to a wider variety of cancers. An open-source toolkit of the core model architecture – comprising 3D image segmentation and algorithms for computational pathology and radiology – has been made publicly available via GitHub.

CASE STUDY 2

Image: Dr Aditya Nori, Microsoft Health Futures.

Real-world evidence (RWE)

Microsoft Futures' work on RWE uses AI to help to democratise healthcare and drive biomedical innovation by better tracking the effectiveness of treatments over time.

RWE provides a wealth of data from multiple sources, such as wearables, genetic samples and health records providing valuable information for researchers. To take advantage of this data, AI models must be able to process a wide range of unstructured data modalities and account for data biases. Although randomised controlled trials are considered the gold standard for testing new drugs and treatments, RWE can help to increase the timescale of comparative analysis, providing information about longer-term outcomes and adverse effects. In addition, RWE can enhance drug performance following regulatory approval, helping to refine eligibility criteria and move towards a more personalised approach to clinical trials.

To demonstrate the value of AI-processed real-world evidence, any new model must first be validated. This will require replicating steps within existing trials to ensure that the conclusions drawn from AI models are consistent with what is already known. In addition, Microsoft is developing large language models and AI tools to extract information embedded in unstructured medical records. This work aims to increase data accessibility, more easily identify patients who may be eligible for trials, and even create 'virtual patients' on whom interventions could be tested in silico.

Closing the loop with AI

Dr Kim Branson, GSK, explored the potential for machine learning technologies to integrate data and methodologies and to generate new mechanistic insights for the design and development of more effective therapeutics.

Disease in humans is ultimately a system-level problem, shaped by time and context. A considerable proportion of clinical and basic research has traditionally addressed relatively narrow biological questions, often at the microscale. In the last few years, however, researchers have been able to deploy machine learning and AI methods to generate large volumes of high-quality data with the potential to transform drug discovery. This has largely been due to reductions in the cost of sequencing and CRISPR technologies, the increasing prevalence of biobanks and electronic health records and the growth of high-throughput imaging techniques (such as lattice light sheets). In particular, these methods have begun to make it possible to characterise disease in humans at scale by:

- **Integrating disease tensors:** images, gene expression profiles, and rates of disease progression can now be more readily combined. Models can be designed to study large patient populations and narrow cohorts simultaneously, test multiple hypotheses, and offer initial predictions of how disease will progress over time.
- **Closing the learning loop:** findings from personalised clinical studies can be incorporated directly into larger observational and genetic cohort analyses through 'reverse translation', informing subsequent machine learning integration work.

Key challenges for the next decade are likely to include:

- Refining models to enhance their representational capacity, fidelity, and efficacy at different scales. There is also potential to improve a model's ability to reflect disease progression over time.
- Developing increasingly sophisticated organoids and organs-on-a-chip to inform bridging models that more accurately represent in vivo activity. This can allow researchers to study, for example, the pace of clonal evolution in cancerous tissues, cellular responses to chemotherapy, and the impact of tumour growth upon organ function. These observations can then be compared with clinical outcomes during in vivo trials.

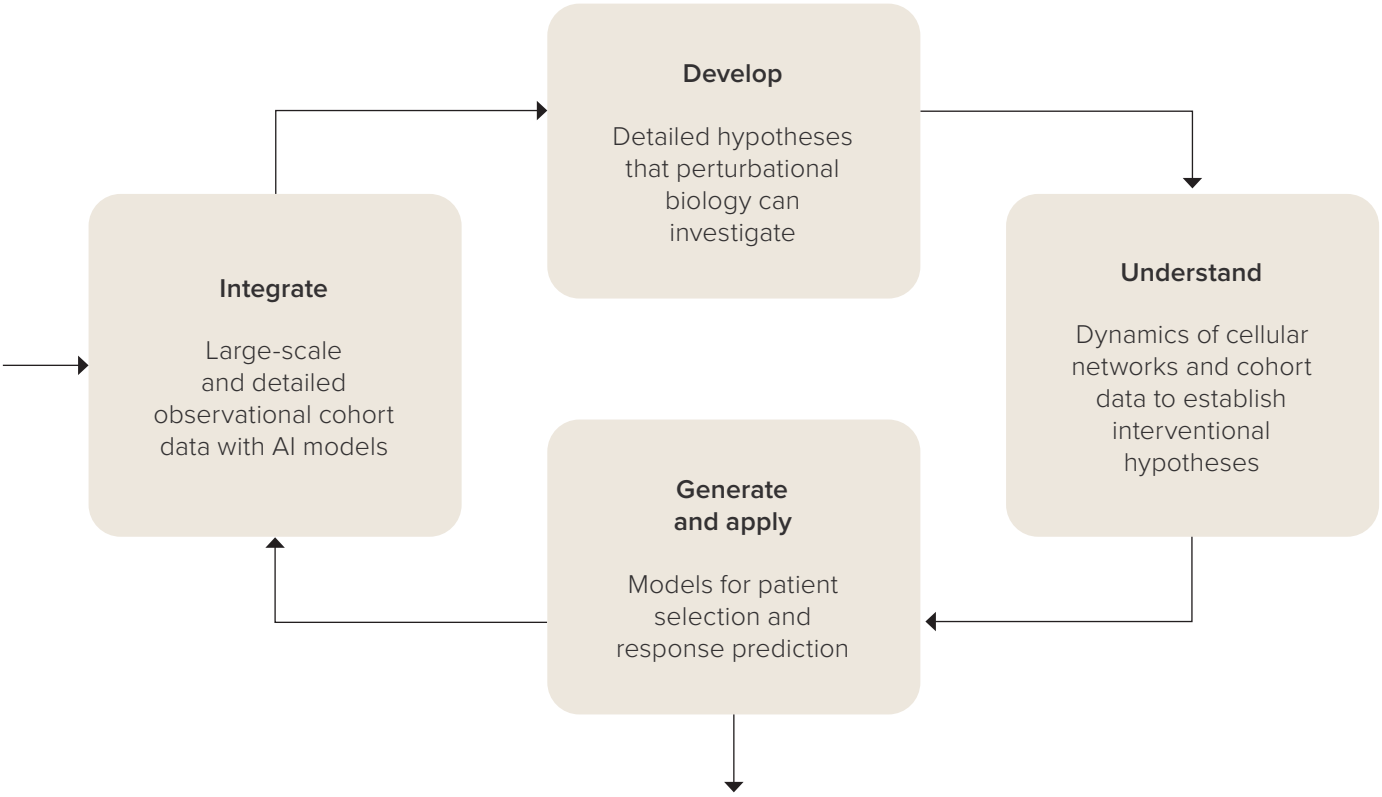
Image: Dr Kim Branson, GSK.

- Using causal machine learning to identify the optimal intervention point(s) in the biological pathway at which therapeutics are likely to be most effective, and how this may vary across patient groups and dosing regimens. AI methodologies have the potential to enhance understanding of these pathways and biological networks to increase therapeutic efficacy.
- Undertaking advanced perturbation biology at scale, developing models which integrate patient history to explore the dynamics of biological systems and help devise relevant hypotheses for in vivo trials. These models can inform stratification of trial design, and the screening and recruitment of appropriate patients.

The development of future therapeutics must reflect the considerable heterogeneity in the drivers, presentation, and evolution of disease in humans. AI and machine learning tools have significant potential to generate, interpret and integrate data from multiple sources and to help researchers design interventions that can be adapted and targeted to different patient groups for maximum efficacy.

FIGURE 4

Closing the loop with AI. Through effective integration of population-scale datasets, observational cohorts and complex models, we can build a learning loop to transform our approach to drug discovery. As well as optimizing our use of data, insights gained can be used to finetune our models and feed back into our experimentation process – guiding our next set of tests and closing the loop.



Source: GSK plc.

"Engineering can teach us to change the design of bridging models, bringing observational cohorts and patients back into the design. – which is what we really care about"

Dr Kim Branson, GSK

Machine learning for translatable biomarkers and targets

Professor Daphne Koller, co-founder of insitro, highlighted new ways in which machine learning models can identify novel treatments for complex diseases.

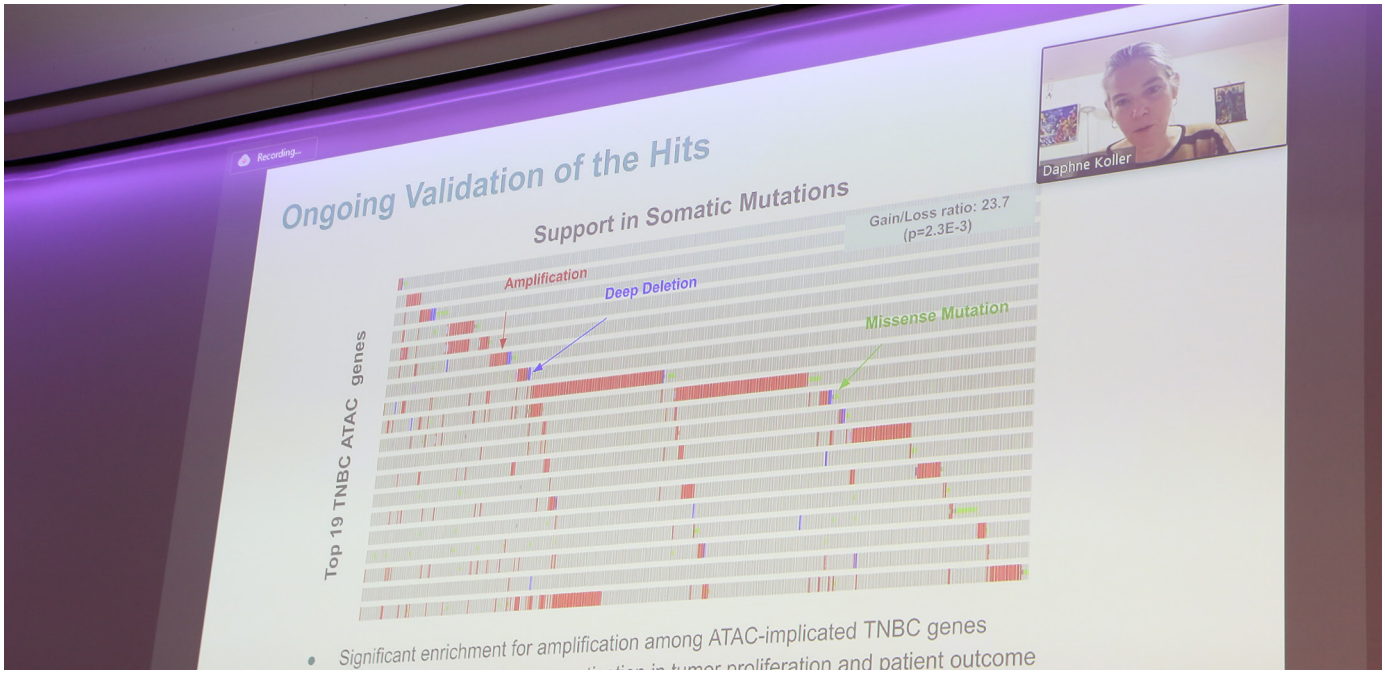


Image: Professor Daphne Koller, insitro.

Many complex diseases are defined subjectively, with insufficient reference to the underlying and highly intricate biology of disease. This presents significant obstacles to the identification of actionable causal nodes and effective treatments.

insitro's machine learning models are designed to address these complexities and challenges, offering an innovative approach to drug discovery. Models integrate large volumes of high-fidelity clinical, genetic, and cellular data at sufficient scale to determine a more representative and unbiased view of disease biology and identify intervention nodes (sites) of potential therapeutic interest through what insitro calls *biological state embedding*.

Using clinical data to determine potential therapeutic targets

Before interrogating any data, self-supervised deep learning (DL) models turn high-content clinical data such as histology images into a simplified numerical format known as an embedding that can be easily processed by computer algorithms. Additional information can be added to, or 'overlaid', onto the numerical data in smaller cohorts.

Histology embeddings can predict elevated mortality risk to significantly improve breast cancer patient risk stratification. Since embedding-based risk and clinical risk for breast cancer are uncorrelated, this implies that models can identify features in histology images that are not currently detectable by clinicians, providing an entirely new perspective on patient outcome.

Machine learning models trained to predict tumour genotype can successfully detect somatic breast cancer mutations with 90-95% accuracy from histological images alone.

insitro used machine learning to predict ATAC-seq signal from histology across a large cohort of patients with 31 different cancer types. ATAC-seq is a non-clinical sequencing method for measuring chromatin accessibility, an indication of the activity of genome segments. It can reveal lowly-expressed driver genes, non-coding driver mutations and epigenetic mechanisms of therapy resistance. insitro's multi-instance deep neural network, trained on 400 patient samples with 23 cancer types and deployed on 11,000 patients, generated statistically significant predictions of ATAC-seq peaks. Machine learning models can subsequently identify peaks which display the highest correlations with outcome-associated genes. This can facilitate the discovery of novel therapeutic targets for cancers.

Using experimental data to determine potential therapeutic targets

Cellular models developed in-house at insitro are used in multimodal assays, creating rich and longitudinal data which is then analysed by machine learning models. Data can be derived from transcriptomics, live-cell imaging, or from morphological and electrophysiological information. Machine learning models use these embeddings to define a disease axis, which may help identify novel therapeutic targets. The combination of machine learning and quantitative phase imaging enables longitudinal phenotyping and exposure response tracking in live cells.

> "We consistently find that machine learning is able to identify subtle patterns across diverse imaging modalities that provide a more holistic view of the underlying biology of disease at multiple biological scales."
>
> Professor Daphne Koller, insitro

insitro has also developed POSH (Pooled Optical Screening in Human Cells), a platform that offers rapid, scaled genetic screening. This platform introduces a CRISPR library to cells and phenotypes using both fixed and live cell imaging. POSH then sequences cells in-situ to generate a genotype-phenotype matrix that is rich at the single-cell level. Scaling using automation and microscopy provides rapid, genome-wide optical CRPSIR screening with high content readouts.

insitro has applied these techniques to define a novel disease axis for Tuberous Sclerosis Complex (TSC) and has identified 76 potential screening hits. Initial genetic analysis has identified genetic evidence for related phenotypes in 60% of these hits. Moreover, 14 of the hits were interrogated in a functional electrophysiology assay known to induce a seizure phenotype in TSC cells. Of these screening hits, seven (including six that are entirely novel) successfully reverted the seizure phenotype.

Such advances demonstrate the ability of machine learning models to rapidly identify potential therapeutic targets, paving the way for future treatment pathways in diseases where novel therapeutics are critically needed.

Q&A DISCUSSION

Medicine

Dr Aditya Nori, Dr Kim Branson and Professor Daphne Koller explored some of the technical and ethical considerations associated with the use of machine learning models for clinical data.

Technical questions

- It is suggested that utilising machine learning to identify patients likely to respond to clinical trial interventions may overfit models, favouring therapeutics which benefit only a small sub-set of individuals. In many cases, however, it can be preferable to elicit a positive response in a sub-set of patients and achieve regulatory approval for a given intervention rather than risk a failed trial and no viable therapeutic. Trials in which only certain individuals respond to treatment can indicate the need for further work to understand the underlying biology of the relevant disease.
- Integrating large-scale clinical data with detailed information from narrow patient cohorts is a key priority for the future. Reductions to the costs of many clinical assays offer an opportunity to obtain additional data points and expand the numbers of cohorts within trials. Insights from retrospective machine learning models can in some cases enhance the design and richness of new clinical studies.
- The potential for unstructured, qualitative data such as clinicians' notes to add value to machine learning models remains debated. Such data can offer a powerful means to enrich language models, in particular for complex psychological and psychiatric interventions. However, growth in copy-paste templates risks undermining their validity, and perspectives from both clinicians and patients can prove subjective, introducing bias at an early stage.

Image: Professor Daphne Koller, Dr Aditya Nori, Dr Kim Branson and Professor Mihaela van der Schaar.

"There is a real value in un-anchoring oneself from preconceptions and subjectivities and in building a map of biology that is unbiased and as driven by objective measurements as possible. One can then overlay data gleaned from clinicians' notes"

Professor Daphne Koller, insitro

Broader considerations

- The research portfolios embedded within many NHS hospitals offer numerous opportunities for collaboration with industry organisations, who seek to empower these institutions to train their own machine learning models. Ownership of relevant data usually rests with the hospitals and healthcare providers, who must ensure that data is protected and anonymised appropriately. Accelerating the pace of innovation and enquiry will be a key challenge for the coming decade.
- It appears unlikely that AI and machine learning technologies will replace clinicians in the near future. At present, these innovations have the potential to augment clinical expertise and to help accelerate the design and approval of therapeutic interventions. Technology does not yet offer a full understanding of all human disease drivers and mechanisms.

"It is wonderful to partner with domain and clinical experts in hospitals to co-innovate and design solutions together, but I wish we could achieve these things faster"

Dr Aditya Nori, Microsoft Health Futures

"There are different challenges involved in dealing with unstructured notes, but these are hugely powerful – they contain some of the most valuable information"

Dr Kim Branson, GSK

PANEL DISCUSSION

How can we accelerate real-world breakthroughs in medicine and healthcare through machine learning and AI?

Dr Danielle Belgrave, DeepMind, Professor Eoin McKinney, University of Cambridge, Jia-Yi Har, Cathay Capital, and Dr Tom Callender, University College London, explored diverse questions related to clinical advances, finance, and ethics. The discussion was chaired by Professor Mihaela van der Schaar, University of Cambridge.

Clinical perspectives

- Machine learning can offer opportunities to improve NHS care by accelerating existing diagnostic and treatment pathways. Microsoft's Project InnerEye, for example, substantially reduces lead times for radiology and radiotherapy planning. Machine learning tools also demonstrate potential to revolutionise clinical trial methodologies by modelling disease progression and patient response longitudinally and by reducing dependence on narrow cohorts.
- Machine learning can help identify with greater precision which individuals would benefit from which interventions, and how frequently. This would reduce the impact of overdiagnoses upon patients and the NHS and enable treatments to be targeted more effectively. It is not impossible that future patients may receive personalised risk profiles and mitigation plans for a range of diseases.
- Challenges remain in accurately generating, testing and encoding causal hypotheses for clinical studies and interventions, and in handling associated uncertainty estimates. Causal machine learning could transform clinical predictions and decision-making, but further work is needed to build and benchmark models.
- Overfitting of machine learning models remains one of the most significant obstacles to generalisability and clinical application. Moving beyond a black-box approach will be essential to understand how and why models perform poorly in particular patient sub-cohorts.
- Whilst medical students are likely to benefit from some training in the basic principles of machine learning and AI, most clinicians simply need to understand how to interact with relevant models and tools.
- Clinicians are likely to favour tools that are straightforward to use and that offer improvements to operational efficiency and quality of care. Tools that are regarded positively by well-respected leaders in their field are more likely to gain the trust of healthcare professionals.
- Human expertise and compassion will remain an indispensable part of the clinician's toolkit and will be complemented by machine learning.

Business and finance

- Companies and investors driving advances in machine learning and AI technologies must remain mindful that decision makers and end users are relatively risk-averse in adopting new technologies in the real world, that large-scale data has only recently started to be made available, and that regulatory considerations necessitate robust clinical evidence. Return on investment derived from the implementation of new AI technologies is an equally important consideration in the decision-making process.
- Investors' priorities are often influenced by the scale of unmet need, availability of data, and the availability of interventions. There may be significant potential in the coming decade for AI to enable digital therapeutic interventions to individuals living with mental health challenges, enable early detection and intervention of diseases such as cancer and dementia, and enhance the cost-effectiveness of chronic disease management. Tech companies and investors can collaborate in the investment, acceleration and scaling of these technologies, ultimately benefitting patients on a global scale.

Ethical considerations

- Fairness, accountability and transparency are increasingly important considerations. Many models used today were trained on cohorts of data gathered over many years, in which minority groups are commonly underrepresented. Machine learning specialists must work to mitigate bias at the earliest stages of data collection and algorithm creation. Early collaboration with domain experts and social scientists can add significant value.
- Developing machine learning solutions for specific (especially under-represented) patient groups is likely to be a key priority for the coming years. Models will also need to reflect geographical differences in funding for healthcare and clinical trials.
- In resource-limited settings with low doctor/patient ratios, it may be more realistic to deploy machine learning and AI tools for simpler solutions such as screening, triaging, and basic imaging. Such tools may also add value in policy modelling, forecasting and planning to shape some of the social determinants of human health.
- The next major advances in machine learning and AI will rely upon appropriate framing of the right questions and upon sustained, effective communication between domain experts and machine learning specialists.

"By using real-world data and estimating individual effects from treatments that already exist we can start to craft a much stronger evidence basis for what clinicians actually do"

Professor Eoin McKinney, University of Cambridge

"Machine learning is a community science – it's not just providing technical solutions, it's a case of always trying to understand domains, which means working with local communities and domain experts"

Dr Danielle Belgrave, DeepMind

"We need to be able to build tools in a way that those at the coalface, who are trying to solve clinical problems and who know what they need to address, are supported to answer those questions well"

Dr Tom Callender, University College London

"Big tech has a major role to play in mental health, such as in the development of AI-enabled therapists. This could be immensely scalable given the lack of mental health therapists globally"

Jia-Yi Har, Cathay Capital

Image: From left to right: Dr Tom Callender, Professor Eoin McKinney, Jia-Yi Har, Dr Danielle Belgrave and Professor Mihaela van der Schaar.

The Royal Society

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- The Fellowship, Foreign Membership and beyond
- Influencing
- Research system and culture
- Science and society
- Corporate and governance.

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500
W royalsociety.org

Registered Charity No 207043

Cover Image:
Visualising AI by Khyati Trehan.

Registered Charity No 207043
Issued: August 2023 DES8186_6