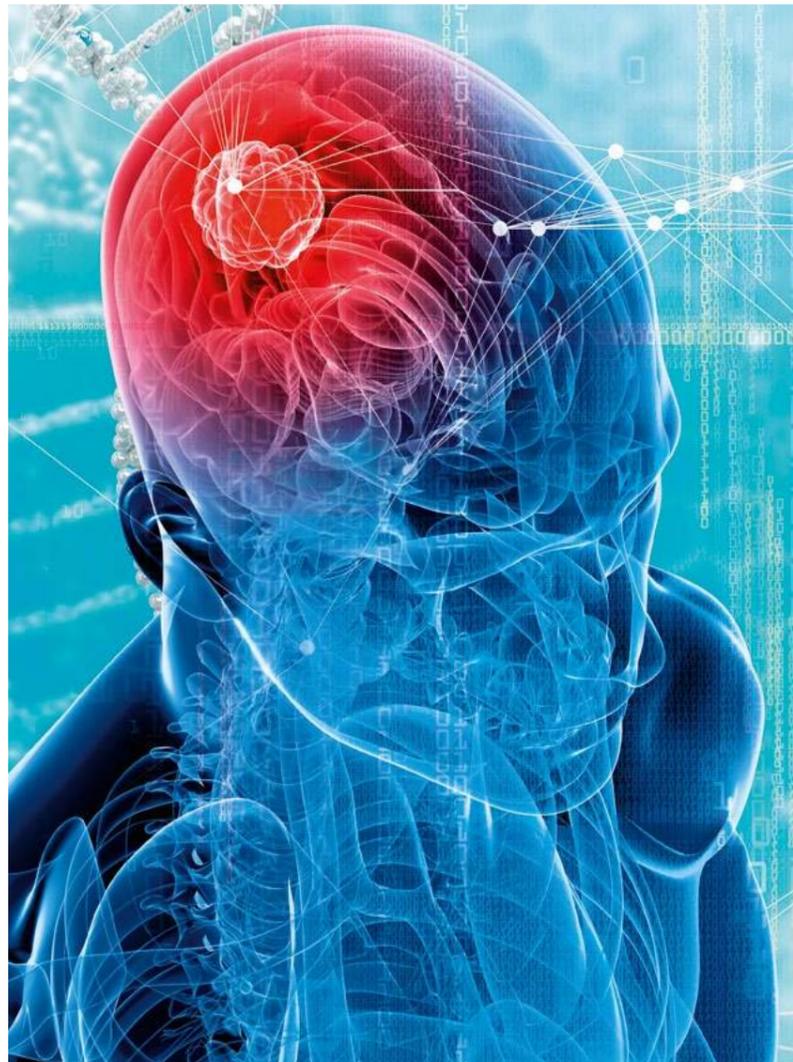


Computational methods of researching cancer treatments

Artificial Intelligence (AI) and machine learning algorithms have the potential to bring substantial advances in the fields of research exploring complex diseases and trying to identify effective treatments. Dr Benjamin Haibe-Kains, working at The Princess Margaret Cancer Centre in Toronto, has spent over a decade developing machine learning tools and databases that could help scientists gain a better understanding of different sub-types of cancer and their response to pharmacological treatments, in order to identify more effective drugs for individual patients.



The computational analysis of big data is set to bring huge advances in terms of the methods used to carry out medical and basic research. A number of researchers worldwide have already started developing algorithms that can analyse large amounts of biological and medical data, identifying new hypotheses for future investigation. A particular field of medical research that could highly benefit from the use of machine learning algorithms is that of exploring cancer genomics and trying to determine the best treatments for different subtypes of the disease.

THE VALUE AND CHALLENGES OF BIG DATA

Throughout the years, scientists have collected vast amounts of data from experiments aiming to achieve a better understanding of the molecular dynamics behind complex diseases. However, the complexity of such data makes it difficult for human researchers to run in-depth analyses and extract the relevant information. This is where the use of high performance computers, coupled with the right programs could be of great help. In the last decades, computer scientists have worked hard to develop machine learning algorithms – programs allowing machines to quickly analyse large amounts of data and “learn” models useful to identify the relevant pieces of information and make predictions. These algorithms can then be leveraged to develop artificial intelligence (AI) tools that can assist humans to analyse data that are beyond our reach. As they become increasingly advanced and sophisticated, AI tools are opening up a new world of possibilities for big data analysis, by transforming the way in which studies and investigations are conducted, leading to important discoveries in a much shorter time.

ARTIFICIAL INTELLIGENCE AND CANCER RESEARCH

Artificial Intelligence and their underlying machine learning algorithms could be of particular value for research exploring complex diseases, when trying to identify effective pharmacological treatments for them.

Cancer, one of the leading causes of death worldwide, is a perfect example of this. Scientists have not yet been able to identify a systematic treatment for cancer that is successful in curing the disease in many of its most aggressive subtypes. Cancer derives from an uncontrolled division of abnormal cells in a given part of the body, which can invade and destroy surrounding healthy tissue and organs. It is an extremely complex disease, with more than 200 subtypes, each of which is often diagnosed and treated differently. As cancer arises from aberrations in the genomic materials of the cells, scientists have developed sophisticated profiling technologies to measure these aberrations and use them to personalise therapies. Still, the most common treatments for cancer are chemotherapy, using drugs to kill the most proliferative cells, and radiotherapy, using high energy X-rays. These treatments can sometimes be successful in reducing or eradicating cancerous cells, yet they can be highly toxic and are not tailored toward the specific set of genomic aberrations that make each tumour unique.

Artificial Intelligence tools could assist researchers by analysing the complex genomic make-up of each individual tumour to develop accurate predictors of treatment response. This would in turn help to identify more effective treatments for individual patients, a major step towards personalised medicine.

COMPUTATIONAL METHODS FOR CANCER RESEARCH

Throughout his career, Dr Benjamin Haibe-Kains has explored the potential of machine learning algorithms, bioinformatics and computational

genomics to improve prediction of patient survival and response to therapies. During his graduate training at the Université Libre de Bruxelles in Belgium, he worked on developing predictors of survival in breast cancer patients based on high dimensional gene expression (messenger RNA) data. He continued his research as a postdoctoral fellow at the Dana-Farber Cancer Institute/Harvard School of Public Health, leveraging a large collection of data to develop “gene expression signatures” to robustly identify molecular subtypes of breast and ovarian cancers.

Using large collections of cancer molecular data and machine learning algorithms, Dr Haibe-Kains and his collaborators identified several “prognostic biomarkers”, that are

Dr Benjamin Haibe-Kains explores the potential of machine learning algorithms, bioinformatics and computational genomics to improve prediction of cancer patients’ survival and response to therapies

computational models using specific molecular features to predict the probability of survival of cancer patients treated with standard-of-care treatments. He says: “Amongst many discoveries, my research unravelled the landscape of cancer pathway activities associated with patients’ survival in each of the molecular subtypes, allowing me to further improve my molecular prognostic models.”

EFFECTIVE TREATMENTS FOR INDIVIDUAL PATIENTS

In 2012, Dr Haibe-Kains started his own independent laboratory, broadening his field of research to explore ways in which machine learning algorithms could be used to predict therapy response in patients. With big players like IBM announcing its Watson initiative to develop AI to assist oncologists in their treatment decision process, it was an exciting opportunity for Dr Haibe-Kains to extend the biomarker discovery beyond patients’ survival, and to make personalised medicine a reality. However, Dr Haibe-Kains and his team quickly

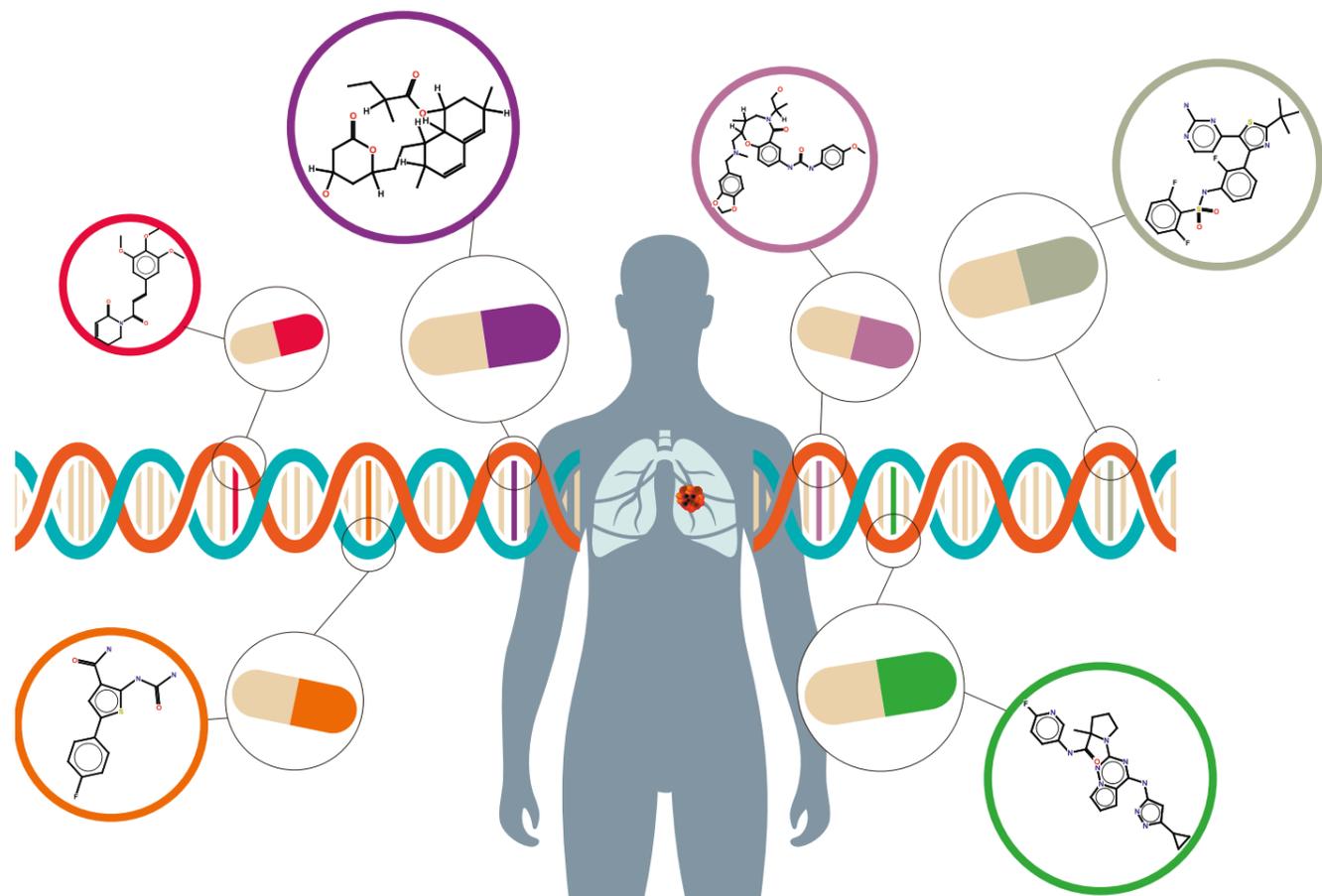
recognised that, due to the high costs of clinical trials, the existing clinical data related to given treatment and cancer subtypes were extremely scarce. Given that machine learning usually requires a large sample size to avoid artefactual discoveries, it was time to investigate “preclinical models”, which are cancer cells derived from patient tumours that one can replicate infinitely. These models therefore provide a fantastic advantage compared to clinical trials as the same cancer cells can be tested with multiple therapies to assess which one is the most efficient, something impossible to do with patients.

Dr Haibe-Kains’ laboratory invested most of its resources in compiling and curating the largest anticancer drug screens in preclinical models. Such screens

contain not only the genomic make-up of the cancer cells, but also the way these cells react to chemical treatments; these complex data are referred to as pharmacogenomics. These efforts resulted in the development

of PharmacoDB¹, a web-application allowing researchers to quickly access the pharmacogenomic data to investigate the possible associations between genomic aberrations and drug response.

As part of his current research, Dr Haibe-Kains is testing machine learning algorithms on the database, to try and pinpoint predictors of treatment reaction. Paired with the right machine learning algorithms, PharmacoDB could be used to develop an AI tool assisting in selection of the most effective treatments for each individual cancer patient. Dr Haibe-Kains’ laboratory also leveraged these valuable data to address another important issue in cancer research: how to classify drugs based on their mechanism of action? Although biologists, chemists and pharmacologists teamed up to develop a large portfolio of drugs with high anticancer potential, it is unclear for many of these drugs how they actually inhibit the growth of cancer cells. Dr Haibe-Kains and his team developed the Drug Network Fusion (DNF), a



new technique integrating multiple pharmacogenomic data to design a comprehensive drug similarity map (or taxonomy). DNF allows researchers to assess the similarity between a drug with unknown mechanism of action with well-characterised drugs, therefore providing an efficient tool to identify potential new indications for approved or experimental drugs (a process called “drug repurposing”). As more pharmacogenomic data becomes available and machine learning algorithms are further improved, databases such as PharmacoDB could become extremely valuable resources.

A NEW ERA FOR RESEARCH

If developed and used correctly, the use of machine learning and AI tools in

basic and translational cancer research could mark the beginning of a new era for personalised medicine, characterised by quick and advanced data analysis,

Application of artificial intelligence in basic and translational cancer research could mark the beginning of a new era for personalised medicine

which was previously unattainable. Dr Haibe-Kains’ work is a perfect example of this, as he has introduced computational methods that could speed up cancer research significantly, by analysing large datasets and identifying predictors of treatment response.

Dr Haibe-Kains’ approach to research is highly collaborative and multidisciplinary, merging the expertise of scientists from a number of different fields. In future,

the computational methods developed by him might lead to ground-breaking discoveries, which could inform oncologists on how to select the most effective treatments for individual cancer patients in clinical settings.

1 <http://pharmacodb.pmggenomics.ca/> **2** *PharmacGx: An R package for analysis of large pharmacogenomic datasets.* Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, Freeman M, Selby H, Gendoo DM, Grossman P, Beck AH, Aerts HJ, Lupien M, Goldenberg A, Haibe-Kains B. *Bioinformatics.* 2015 Dec 9. **3** *PharmacDB: an integrative database for mining in vitro anticancer drug screening studies.* Smirnov P, Kofia V, Maru A, Freeman M, Ho C, El-Hachem N, Adam GA, Balalawi W, Safikhani Z, Haibe-Kains B. *Nucleic Acids Res.* 2017 Oct 9, gkx911. **4** *Integrative cancer pharmacogenomics to infer large-scale drug taxonomy.* El-Hachem N, Gendoo DM, Soltan Ghorraie L, Safikhani Z, Smirnov P, Chung C, Deng K, Fang A, Birkwood E, Ho C, Isserlin R, Bader G, Goldenberg A, Haibe-Kains B. *Cancer Res.* 2017 Mar 17. **5** *Gene isoforms as expression-based biomarkers predictive of drug response in vitro.* Safikhani Z, Smirnov P, Thu KL, Silvester J, Lupien M, Mak TW, Cescon D, Haibe-Kains B. *Nat Commun* 2017 oct, in press.



Behind the Bench

Dr Benjamin Haibe-Kains

E: benjamin.haibe.kains@utoronto.ca **T:** +1 416 581 8626 **W:** <http://www.pmggenomics.ca/bhklab/>
<https://www.linkedin.com/in/benhaibekains/> <https://twitter.com/bhaibeka>

Research Objectives

Dr Haibe-Kains’ research focuses on the computational integration of high dimensional molecular data to simultaneously analyse multiple facets of carcinogenesis.

Funding

Princess Margaret Cancer Foundation, Canadian Institutes of Health Research, Terry Fox Research Institute, Stand Up To Cancer Canada, Cancer Research Society, Natural Sciences and Engineering Research Council of Canada, Canadian Cancer

Research Society, Ministry of Economic Development/Ministry of Research and Innovation of Ontario, Ontario Institute for Cancer Research

Bio

Dr Benjamin Haibe-Kains is Scientist at the Princess Margaret Cancer Centre and Assistant Professor in the Medical Biophysics and Computer Science departments of the University of Toronto (Canada). He earned his PhD in Bioinformatics at the Université Libre de Bruxelles (Belgium), before conducting a

postdoctoral fellowship at the Dana-Farber Cancer Institute and Harvard School of Public Health (USA).

Contact

Benjamin Haibe-Kains, PhD
 Princess Margaret Cancer Research Tower,
 Floor 11, Room 310
 101 College Street
 Toronto
 ON M5G 1L7
 Canada

Q&A

When and how did you first start being interested in computational models to be used in cancer genomics research?

When I did my bachelor in Computer Science, I was interested in the development of Artificial Intelligence in robotics. In the early 2000s, AI robotics was still in its infancy and the applications were still limited. This is when my former supervisor at the Université Libre de Bruxelles, Prof Gianluca Bontempi advised me to consider bioinformatics, a booming field in need of researchers with expertise in machine learning. I followed his advice and started a PhD under a co-supervision with Dr Christos Sotiriou, a breast cancer oncologist at the Institut Jules Bordet. This is how I started using computational models in cancer genomics research.

So far, how effective have the machine learning tools developed by you been in improving biomarker discovery and drug selection from pharmacogenomic data?

My lab was definitely not the first to tackle these important challenges. However, we built on our previous experience in meta-analysis of large compendium of gene expression data to develop computational platforms for pharmacogenomic data analysis. PharmacGx2 and PharmacDB3 are open-source and freely available

for the scientific community. With these large amounts of pharmacogenomic data in hand, we applied machine learning techniques to better classify drugs (DNF4) and discover new biomarkers predictive of drug response in cancer cell lines. Even though these discoveries will have to undergo further validation before their translation into clinic, they show that machine learning can yield promising results with potential clinical relevance.

How long do you believe it might take to start witnessing a major introduction of AI technology within medical settings?

Not long, probably a few years from now. As more hospitals have the patients’ electronic health records connected to the experimental data derived from their tumour materials, we will finally have access to the large, high quality data required for AI to show its full potential for biomedical applications. Recognising that large cohorts of patients and derived materials are necessary to make major discoveries and build the new generation of AI-based tools in the clinic, hospitals are joining forces, and even pharmaceutical companies have started to share more and more data related to clinical trials. These are the necessary steps we must take to unleash the power of AI for personalised medicine.

In the years to come, what role do you feel AI will have in terms of research and innovation within the medical field?

It is hard to predict the roles of AI in the future of medicine; the applications are close to limitless. It will help patients

better schedule the series of medical appointments involved in the treatment of complex diseases such as cancer. It will help hospitals better monitor their performance, and ensure the highest standards of safety, diagnosis accuracy and treatment efficacy. The mining of data from wearable devices will allow continuous monitoring and patients to be more proactive when symptoms arise. And of course, AI will enable new discoveries, further expanding our knowledge of cancer and other diseases. Limitless.

What are your plans for future research?

First, I want to validate our first discoveries – predictors of drug response and new drugs predicted to be efficacious in aggressive cancer types – in animal studies to get them close to clinical applications. The Princess Margaret Cancer Centre enjoys a strong drug development group, which will be key in this endeavour. Second, I want to make our pharmacogenomic platforms tools of choice for hospitals and pharmaceutical companies, to further facilitate data sharing and large-scale computational analysis. Finally, I want to integrate molecular data with imaging data, both pathological and radiological images, to predict the best course of treatments and better follow the patients over time. Our first application of deep learning on radiological images is promising, supporting this line of research for future clinical applications.