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 makeup 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

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
several “prognostic biomarkers,” that are computational models using specific molecular features to predict the probability of survival of cancer patients treated with standard-of-care treatments. After 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 personalized medicine a reality.

However, Dr. Haibe-Kains and his team quickly recognized that, due to the nature 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 artifact data, it was time to investigate “preclinical models,” which are cancer cells derived from patient tumors, that one can replicate infinitely. These models therefore provide a fantastic advantage compared to clinical trials in many dimensions: cancer cells can be tested with multiple therapies to assess which one is the most efficient, something impossible to do with patients.

As part of his current research, Dr. Haibe-Kains is testing machine learning algorithms on the database, to try and pinpoint predictions of treatment reaction. Paired with the right machine learning algorithms, Pharmacogenomics 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 Pharmacogenomics 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 personalized medicine, characterised by quick and advanced data analysis, which was previously unattainable. Dr. Haibe-Kains is taking advantage of this perfect opportunity to develop new tools and techniques to better classify drugs (DNF) and discover new biomarkers predictive of drug responses in cancer cells. Even though these discoveries will have to undergo further validation before their translation into clinics, they show that machine learning can yield promising results with potential clinical relevance.

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

When and how did you first start being interested in computational models to be used in cancer genomics research? What led you to start your PhD in Computer Science and what did you find about machine learning that inspired you to consider bioinformatics, a booming field in need of researchers with expertise in machine learning? I followed this advice and started a PhD under a co-supervision with Dr. Christos Sozioiu, a breast cancer oncologist at the Institute 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 compendia of gene expression data to develop computational platforms for pharmacogenomic data analysis and pharmacomarker discovery. These open-source tools are further expanding our knowledge of cancer and other diseases.

What are your plans for future research? First, I want to validate our first discoveries – predictors of drug efficacy and new drugs predicted to be efficacious in aggressive cancer types – in animal studies to get them closer to clinical applications. The Princess Margaret Cancer Centre enjoys a strong drug development infrastructure, 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.

**RESEARCH OBJECTIVES**

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

**FUNDAING**

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 Biomedical and Computer Science departments at the University of Toronto (Canada). He earned his PhD in Bioinformatics at the Université Libre de Bruxelles (Belgium), before completing a postdoctoral fellowship at the Dana-Farber Cancer Institute and Harvard School of Public Health (USA).

**CONTACT**

Dr. Benjamin Haibe-Kains, PhD
Princess Margaret Cancer Research Tower, Floor 11, Room 310
101 College Street
Toronto
ON M5G 1L7
Canada
E: benjamin.haibe-kains@utoronto.ca
T: +1 416 581 8626
W: http://www.pmcans.org/bhekikat
F: https://www.linkedin.com/in/benjamin-haibe-kains
W: https://twitter.com/bhaibeka

---

1. http://pharmacogenomics.journals.org/