KATCHY

Clinical Knowledge for Humanity

Position Paper

PUBLICATIONS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017453.

KATY Consortium

KATY Publications

Published by Università degli Studi di Roma "Tor Vergata", December 2021. This publication is part of WP8 of the KATY project, led by Università degli Studi di Roma "Tor Vergata" (UNITOV).

Author: Fabio Massimo Zanzotto

Design: EURICE GmbH

The author is solely responsible for its content, it does not represent the opinion of the European Commission and the Commission is not responsible for any use that might be made of data appearing therein.

To cite: Fabio Massimo Zanzotto 2021. Position Paper. KATY Project Publications.

Table of Contents

Abstract

AI-empowered Personalized Medicine promises to find tailored cures for patients. Such therapies are starting to be adopted but AI-empowered Personalized Medicine promises a transformation for the cancer community. However, no matter how precise it is and no matter how many lives it can save, if clinicians do not understand its suggestions and decisions, AI-empowered Personalized Medicine will not be a game changer impacting everyday clinical decisions. Hence, the real challenge is building AI-empowered Personalized Medicine systems that can be accepted by clinicians and clinical researchers. Knowledge at the tips of your figures (KATY) grasps the above challenge and proposes an AI-empowered Personalized Medicine system that can bring medical "AI-empowered knowledge" to the tips of the fingers of clinicians and clinical researchers. AI-empowered knowledge is human interpretable knowledge that clinicians and clinical researchers can: understand, trust and effectively use in their day-to-day activities. KATY is an AI-empowered Personalized Medicine system built around two main components: A Distributed Knowledge Graph and A pool of eXplainable Artificial Intelligence predictors. As a stress test and due to the lack of personalized clinical responses, KATY will be experimented in a low prevalence and complex cancer: Clear cell renal cell carcinoma (ccRCC).

Introduction

Over a million people died from cancer in the EU in 2016, a number expected to increase as a result of undiscovered cancers during the pandemic. With so many patients in the EU undergoing treatment for cancer on a yearly basis, the best cures for patients may eventually become possible by linking the patient's data to huge genomics repositories, banks of -omics, health related data and biobanks. Understanding the best next clinical decision for individual patients will rely on new artificial intelligence systems that place individual patients, or some characteristics of individual patients, in the context of the wealth of clinical and biological information emerging for a disease. Such artificial intelligence systems, which also place this Knowledge within reach of clinicians and the medical research community, stand to transform clinical treatments. However, a large number of technological and ethical breakthroughs are needed to reach this vision. As part of the 1 Million Genomes European initiative, Knowledge at the Tips of your Fingers is a pan-EU consortium dedicated to developing the infrastructure that will one day enable these key insights for patients. The first project undertaken by this consortium project is a pilot study in Renal cancer to develop the infrastructure and partnerships that will usher in this new European initiative.

Renal (kidney) cancer represents more than 3% of all cancers, with around 115,000 new diagnoses in Europe each year and a rising incidence. Renal cancer presents with metastatic disease in up to 1/3 patients and of those who have treatment with curative intent, 1/3 of these subsequently develop metastatic disease. Renal cancer is therefore the most lethal urological malignancy with metastatic disease affecting almost 50% of patients (Bhatt and Finelli 2014). However, unlike the most common cancers (breast, prostate, lung and colorectal), it was only relatively recently that useful drug therapies have been found for renal cancer, perhaps because of intrinsic resistance to chemotherapies. Clear cell renal cell carcinoma (ccRCC) is the most frequent subtype of renal cancer, accounting for 80% of cases, and treatments developed in ccRCC have then been tested and used in the rare non-ccRCC subtypes.

Current clinical practice. The way that Renal cancers are currently treated does not leverage a wealth of genomic data. Nephrectomy is potentially curative in localized renal tumours and may also have a role in metastatic patients with good prognosis factors (according to IMDC classification), where it is not curative but may have symptomatic benefit and delay the need for systemic therapy. There may also be a role for surgery for isolated metastases. However, effective drug treatments are the key to improving survival. During the last fifteen years, three types of drug treatment have been developed that have improved the standards of care of metastatic ccRCC. In the following, the three types of drug treatment:

1. **anti-angiogenic drugs** (bevacizumab, sorafenib, sunitinib, pazopanib, axitinib, tivozanib, cabozantinib and lenvatinib) that generally target VEGFR signalling but may also have effects on other potentially important tyrosine kinases such as PDGFR, MET, FGFR, CSF1R and others. These were the first real revolution in the treatment of renal cancer.

2. **Drugs that target the mTOR metabolic signalling pathway** (everolimus and temsirolimus), although use of these has now largely been displaced as it is hard to identify the minority of patients that have most benefit from them.

3. **Immunotherapies** (nivolumab, ipilimumab, pembrolizumab, avelumab, as well as the largely displaced cytokines, Il-2 and interferon-α), given as monotherapy or in combination, have recently been having a similar transformative impact to the antiangiogenics.

Current treatment guidelines, with updates due to the latest trial data, advise first line combination therapy (either ipilimumab plus nivolumab, or axitinib with pembrolizumab or avelumab) for most ccRCC patients, with a single agent VEGFR-inhibitor for some good prognosis patients or those for whom immunotherapy is contra-indicated. Second- and third-line therapy is currently a relatively diverse set of generally monotherapy options from other VEGFR- or mTOR-inhibitors (or immunotherapy if not previously had).

The only stratification currently used is the relative blunt tool of IMDC prognostic scoring, and the very broadest histology. Disappointingly, in this era of precise molecular techniques, for a tumour type where all the therapies are targeted agents, there are still no molecular tests being used to select the most appropriate treatment for an individual patient.

Study setup

We believe that artificial intelligence is the right tool to translate genetic, transcriptomic and/or morphological data into therapeutic decisions. Indeed, there is data suggesting that Molecular selection may be possible. Transcriptomic profiles in the IMmotion150 and IMmotion151 trials show that benefit from anti-angiogenics or an immunotherapy combination correlate with Angiogenesis and T-effector signatures, respectively. MET expression is associated with sensitivity to cabozantinib (a VEGFR-inhibitor with additional MET inhibition) in its pivotal registration trials. mTOR pathway mutations associated with mTOR-inhibitor sensitivity in case series. Immunotherapy sensitivity correlates with PBMR1 mutation but not with markers, such as unselected mutation burden, that are predictive in other cancer types.

Planned Advances in KATY Confidently identifying low risk of disease recurrence in ccRCC as well as predicting response to various targeted and immunotherapies could be a gatekeeper for patient selection in trials, to prevent unnecessary treatment and reduce the number of cases needed for enrolment.

The planned advances in KATY are summarized as follows (see also Figure 1):

- KATY will develop a series of predictive AI models leveraging publicly available genomics and multi-omic data for decision making in personalized medicine to predict disease recurrence from transcriptomic and histological data, response to targeted therapy and response to immunotherapy leveraging proteogenomics
- KATY will disseminate all this knowledge with a sophisticated Distributed Knowledge Graph that incorporates the results of artificial intelligence models of use to researchers and clinicians and a simple patient-centric interface aimed to support patients alongside clinicians in guiding therapeutic choices along their path to treatment
- KATY will provide evidence-based information on optimal treatment selection that can support health policy decision making, whilst ensuring patient centricity in decision making.

In addition to the wealth of publicly available data collected in the KATY knowledge graph, a patient cohort of over 500 renal cancers subjected to combination targeted and immunotherapy will be generated. The resulting patient cohort will be used to refine the predictive models developed within the project.

Technological backdrop and study workflow

Knowledge at the tips of your fingers (KATY) is the mandate – reaching this vision means emphasizing simplicity and explainability to the end user. So, while a great deal of sophistication lies at the heart of the KATY project, the core vision of the project is to make the fundamental insights accessible in a simple manner to the research and clinical community. Of course, simplicity is in the eye of the beholder, and the level of abstraction and focus must be tailored to each community separately. The overall technological backdrop behind the KATY project is highly cooperative across the 20 institutes involved in the project (Figure 2).

Problem definition, data processing and data engineering. To reach this vision of simplicity in communication in the context of an EU-wide collaboration, aimed at conveying research findings with simplicity requires a great deal of engineering. Clinicians in the consortium work closely with bioinformaticians to identify research problems of critical and unmet need. Datasets most relevant to the field are identified and used to develop harmonized data processing and engineering strategies to support artificial intelligence methods.

Making data FAIR. Data engineering teams within KATY begin to make this data approachable by incorporating FAIR practices. Existing ontologies connect the datasets developed to the broader community into a knowledge graph which makes querying the data simpler. Essentially, KATY data engineers re-work flat tables to define knowledge graphs that: (1) define real-world entities of a domain, (2) provide relationships between them (3) define rules for possible classes of entities and relations via some schema (4) they enable reasoning to infer new knowledge. This fourth element of the knowledge graph also makes them a perfect resource for the artificial intelligence community, which can then use the datasets within the knowledge graph iteratively to facilitate intuition and discovery processes and even add to the knowledge graph.

Data visualization. Data visualization and analytical workflows are at the heart of scientific discovery. The bioinformatics teams in KATY work closely with clinicians and the medical research community to identify new analytical workflows aimed at delivering insightful visualizations in different contexts. A web-development team brings these visualizations to life using a REACT interface aimed at making data abstractions understandable to researchers, clinicians and patients.

Predictive modelling. In the information age, scientific discovery is formalized through computational modelling. Machine learning and artificial intelligence techniques develop predictive models that are capable of explaining and predicting over multi-omic datasets using fewer bits of information. AI techniques formalize what has been learned through the accumulation of datasets brought together within the project. Data engineers take this data as input in the KATY project and work together to develop new ontologies and databases to house the data being developed by engineers.

Explainability. A second layer of visualization is needed to interpret the deep learning techniques developed in the project. Simple analytical workflows and explainable artificial intelligence techniques. The technical backdrop to KATY is establishing an EU wide framework for applied bioinformatics problems like those presented for Renal Cancer. The emphasis in the first year has been to establish the partnerships grounded on best-case software engineering practices

Outlook

The KATY project develops a network and collaborative infrastructure to usher in the adoption of artificial intelligence into the EU health infrastructure. The pilot project in Renal cancer will develop a proof of concept of the added value of artificial intelligence and serve as the nucleus to develop the infrastructure to make this happen.

Acknowledgements

KATY Consortium is supported by the EU programme Horizon 2020/H2020-SCI-FA-DTS-2020-1 (contract number 101017453). Authors would also like to thank the European FENIX Research Infrastructure (https://fenix-ri.eu/), the CI-TASK from Gdansk, the PL-Grid Infrastructure from Poland and the CCRT/TGCC from CEA-France for providing their hardware and software resources.

Figure 1 Overview of value added by the KATY project

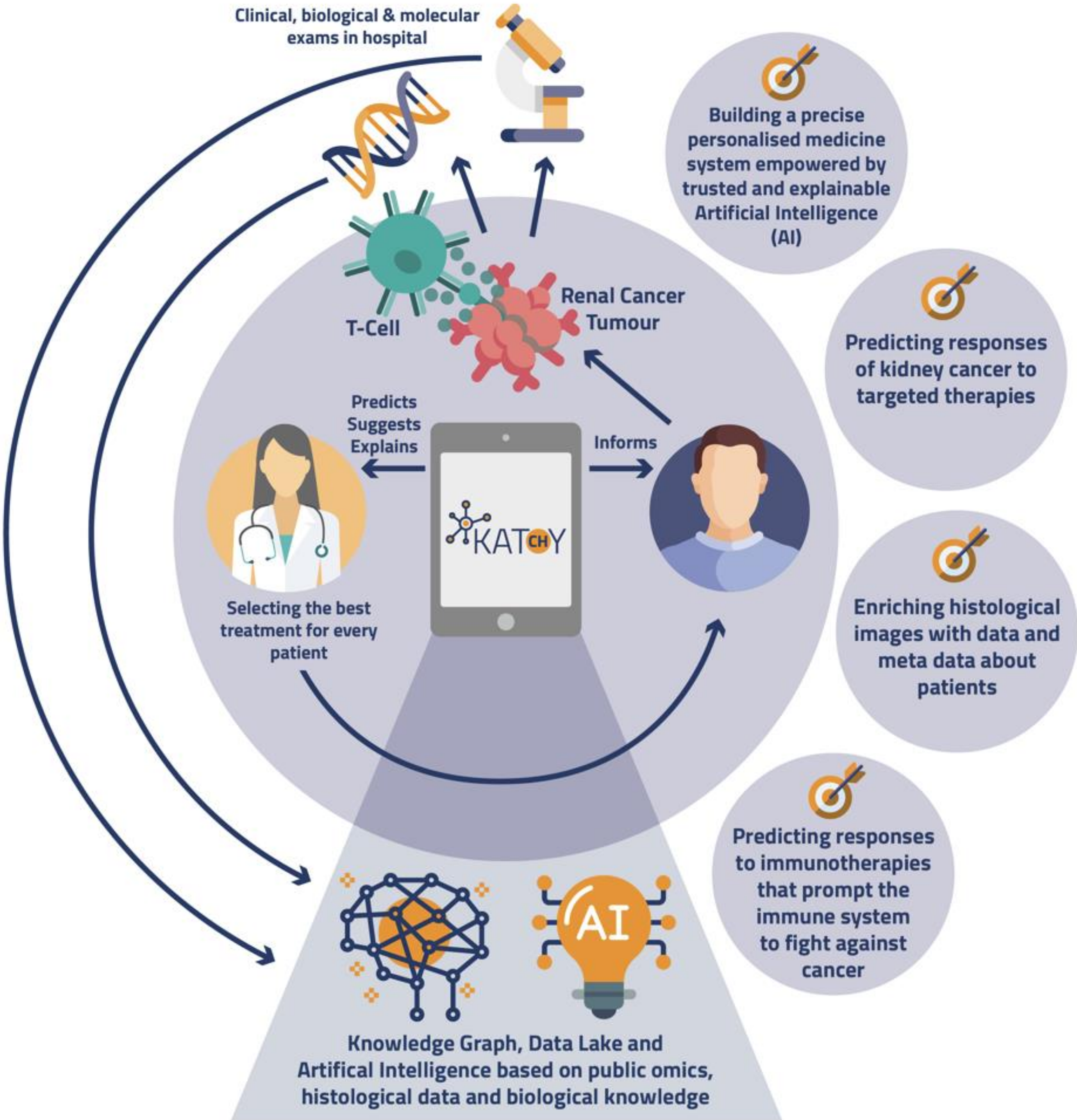

Figure 2 Technological backdrop and study workflow