Mining Electronic Health Records to Validate Knowledge in Pharmacogenomics

Adrien Coulet, Malika Smaïl-Tabbone

Mining Electronic Health Records to Validate Knowledge in Pharmacogenomics.

by Adrien Coulet and Malika Smaïl-Tabbone

The state of the art in pharmacogenomics (PGx) is based on a bank of knowledge resulting from sporadic observations, and so is not considered to be statistically valid. The PractiKPharma project is mining data from electronic health record repositories, and composing novel cohorts of patients for confirming (or moderating) pharmacogenomics knowledge on the basis of observations made in clinical practice.

Pharmacogenomics (PGx) studies how individual gene variations cause variability in drug responses. A state of the art of PGx is available and constitutes a basis for implementing personalized medicine, i.e., medical treatment tailored to each patient taking into account his/her genomic context. However, most of the state of the art in this domain is not yet validated, and consequently not yet applicable to medicine. Most information results from studies that do not fulfil statistics validation standards and are difficult to reproduce because of the rarity of gene variations studied (making it hard to recruit sufficiently large cohorts) and of the multifactorial aspects of drug responses [1]. The increasing use of electronic health records (EHRs) generates large repositories that offer new opportunities, such as composing patient cohorts for the study of clinical hypotheses hard to test experimentally. Typically, EHR repositories make it possible to assemble cohorts of patients to study the impact of gene variations on drug responses on the basis of practice-based data [2].

In October, 2015, The French National Research Agency (ANR) committed funding to a project called PractiKPharma (**Pr**actice-based evidence for **acti**oning **K**nowledge in **Pharma**cogenomics). The project aims to validate or moderate PGx state-of-the-art (SOTA) knowledge on the basis of practice-based evidence, i.e., knowledge extracted from EHRs. Units of knowledge in PGx typically have the form of ternary relationships *gene variant–drug–adverse event*, and can be formalized to varying extents using biomedical ontologies. To achieve this goal, the PractiKPharma consortium will focus on four objectives, illustrated in Figure 1: (1) to extract SOTA knowledge from PGx databases and literature; (2) to extract observational knowledge (i.e., knowledge extracted from observational data) from EHRs; (3) to compare knowledge units extracted from these two origins, to confirm or moderate SOTA knowledge, with the goal of enabling personalized medicine; (4) Finally, to emphasize newly confirmed knowledge, omics databases will be investigated for molecular mechanisms that underlie and explain drug adverse events. This investigation will use and contribute to the biomedical Linked Open Data [3].

The PractiKPharma consortium comprises four academic partners: two computer science laboratories, the LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications, Nancy, France) in Nancy, the LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier, France) in Montpellier; and two University Hospitals, the HEGP (The Georges Pompidou European Hospital, Assistance Publique – Hôpitaux de Paris (AP-HP), France) in Paris, specialized in EHRs management and pharmacogenomics, and the CHU Saint-Etienne (University Hospital (CHU) of Saint Etienne, France), specialized in pharmacovigilance.

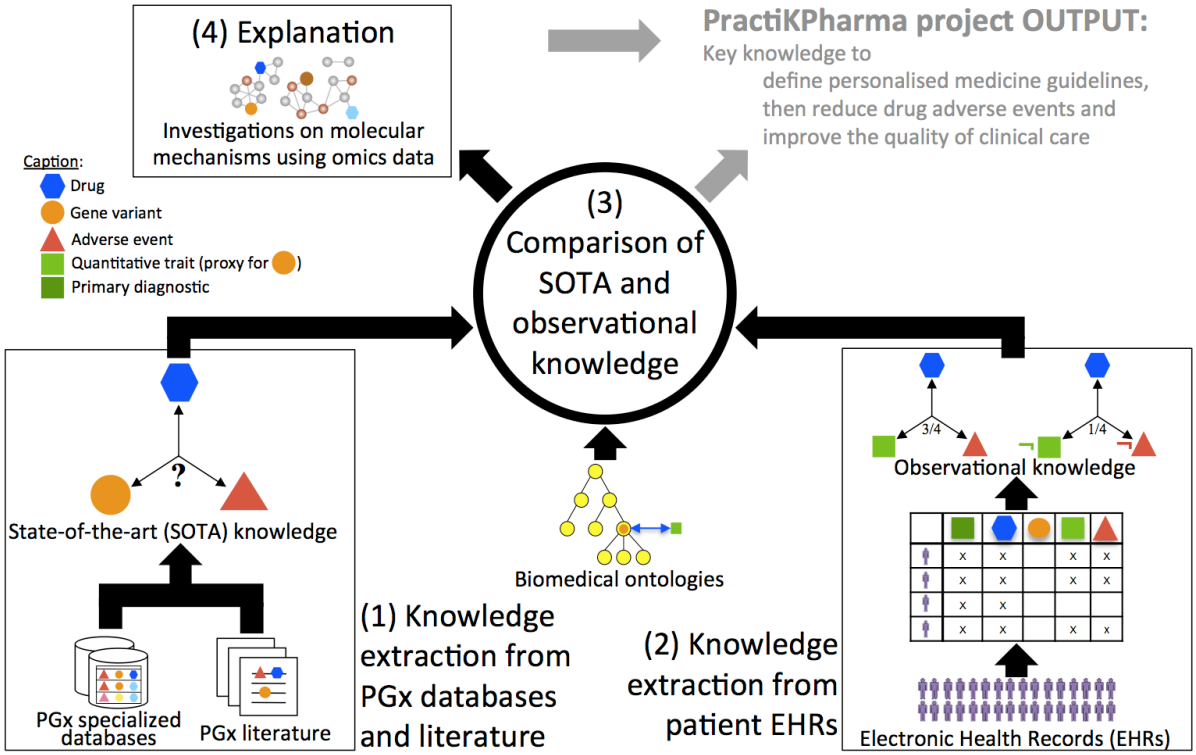

Figure 1: Outline of the four objectives of the ANR PractiKPharma project.

PractiKPharma will have impacts both in computer science and biomedicine, including:

- the development of novel methods for knowledge extraction from text and EHRs;
- enabling multilingual semantic annotation of EHRs;
- methods for representing and comparing SOTA and observational knowledge;
- a database that maps genotypes to quantitative traits to facilitate the study of PGx with EHRs;
- the completion and connection of Linked Open Data related to PGx;
- methods for hypothesizing on mechanisms of adverse events and validated PGx knowledge units.

Overall, the final goal of PractiKPharma is to provide clinicians with actionable PGx knowledge to establish guidelines that when implemented in personalized medicine will reduce drug adverse events, and improve the quality of clinical care.

The PractiKPharma consortium will hire two PhD students, two post-docs, one pharmacist and one research engineer to contribute to the project. In addition, PractiKPharma will foster collaboration and dissemination throughout the EU to take into consideration EHRs of various populations and to establish international projects.

Link:
PractiKPharma project: http://practikpharma.loria.fr

References:
[1] J. P.A. Ioannidis: "To replicate or not to replicate: The case of pharmacogenetic studies", *Circulation: Cardiovascular Genetics*, 6:413–8, 2013.
[2] A. Neuraz et al.: "Phenome-wide association studies on a quantitative trait: Application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics, *PLoS Computational Biology*, 9(12), 2013.
[3] K. Dalleau, N. Coumba Ndiaye, A. Coulet: "Suggesting valid pharmacogenes by mining linked data", in *Proc. of the Semantic Web Applications and Tools for the Life Sciences (SWAT4LS) Conference, 2015.*

Please contact:
Adrien Coulet, LORIA, Nancy, France, +33 3 54 95 86 38, adrien.coulet@loria.fr