Facilitate Data Curation in Translational Medicine: NCER-PD Data and Computing Platform for the National Centre of Excellence in Research on Parkinson's Disease Project

Venkata P. Satagopam[*], Wei Gu[*], Kirsten Roomp, Peter Banda, Adriano Barbosa-Silva, Kavita Rege, Sascha Herzinger, Reinhard Schneider[*]

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg

** Presenting authors*

Knowledge-based translational medicine is a rapidly growing discipline in biomedical research and aims to expedite the discovery of new diagnostic tools and treatments by using a multi-disciplinary, highly collaborative, "bench-to-bedside" approach. It involves the integration of multiple high dimensional datasets that capture the molecular profiles of patients, as well as detailed clinical information. Curation, harmonization of retrospective and prospective clinical data from several studies and application of controlled terminologies and standards in order to facilitate cross study comparisons is a challenge. A variety of computational approaches are currently being used to harmonize and relate molecular data to clinical outcomes in order to better understand disease conditions. These methods also have the potential to discover biomarkers for early detection of disease, and targets for drug discovery, and to be used predictively to help to suggest personalised therapeutic strategies for patients.

We have developed the Data and Computing Platform for the research project NCER-PD that focuses on improving the diagnosis and stratification of Parkinson's disease (PD) by combining detailed clinical and molecular data of patients to develop novel disease biomarker signatures. The Data and Computing Platform provides key infrastructure for the integration, curation and interrogation of anonymized clinical and experimental data. The platform manages multidimensional data associated with clinical research, including patient data, sample-associated information, and high-throughput molecular readouts from these samples. These different data flows are integrated at their source with the help of advanced data capture and transfer approaches. Clinical data can be entered remotely, via electronic forms at the time of collection, assuring their integrity and standardization. All clinical data collected within the study is translated into the terminology and standards provided by Clinical Data Interchange Standards Consortium (CDISC)[1]. High-throughput experimental data are uploaded directly to the database service: the tranSMART[2] system. This system enables sharing, integration, standardization and analysis of heterogeneous data from collaborative translational research. It has been used in pharmaceutical industry and in large PPP translational research projects (e.g. IMI- eTRIKS[3], IMI-AETIONOMY[4]) to store and share curated phenotypic data such as clinical observations, omics data and imaging data.

[1]CDSC(2017) https://www.cdisc.org
[2]tranSMART(2017) http://transmartfoundation.org
[3]eTRIKS(2017) http://www.etriks.org
[4]AETIONOMY(2017) http://www.aetionomy.eu