A development framework for data analytics in genomics

A. Ardeshirdavani *, Y. Moreau
STADIUS – ESAT, KU LEUVEN, Leuvne, 3000, Belgium
amin.ardeshirdavani@esat.kuleuven.be

Genomic medicine is currently the main component of personalized medicine. This rapidly developing science-driven approach to healthcare holds great benefits for patients, clinicians, healthcare providers and society as a whole. It promises to change healthcare from being reactive to disease to being predictive to disease onset, from being general for all to being tailored to the individual. Health will become personalized and will change from cure to prevention, partly by avoiding severe, genetic disorders. This new approach towards the use of genomic information is highly disruptive to current medical procedures, to the IT infrastructure used in medicine, and towards the role of genetic specialists in medical organizations. Medical professionals, patients, and policymakers have to develop mechanisms to approach apparently healthy individuals and the society has to develop strategies to obtain the right genomic information at the right time during life. We developed a framework to store, mine, organize and interact with genomic information, from a technical, a clinical, a patient and a societal perspective. To maximise the potential of genomic medicine we have to: a. Improve genome-wide genetic analysis and diagnostic strategies. b. Investigate the infrastructure and strategies to detect, store and report findings from genomic information to patients. c. Explore the feasibility of integrating medically actionable genomic variants, including carrier status, into the health care system. d. Develop strategies for genomic data sharing to foster a learning health environment. Therefore, the goal of this project has been to develop a bioinformatics workflow (pipeline) for multi-step processing of human genome sequencing data, from the data generated by the sequencing instrument located in different locations and making the data ready for interpretation by a genomics specialist. We propose a methodology named *NGS-Logistics*[1] for federated analysis of sequence variants from personal genomes. We have also considered prediction applications, such as gene prioritization by *Endeavour*[2], Variant prioritization by *eXtasy*[3] and disease-gene interaction prediction *Begeel*[4]. The *NGS-Logistics* platform, which fulfills all requirements for a successful application has been installed in KU Leuven, UZ Leuven, CMG Antwerp, CHU Liege, UZ Brussels, UCL, and UZ Ghent and allow the researcher's process data inclusively and comprehensively from multiple sources while guaranteeing privacy and security.

Keywords: NGS; Genome; Analysis; Annotaiton; Interpretaion.

References

[1] A. Ardeshirdavani, E. Souche, L. Dehaspe, J. Van Houdt, JR. Vermeesch, Y. Moreau: "NGS-Logistics: federated analysis of NGS sequence variants across multiple locations". Genome medicine. 2014 Dec;6(9):71.
[2] L. Tranchevent, A. Ardeshirdavani, S. ElShal, D. Alcaide, J. Aerts, D. Auboeuf, Y. Moreau: "Candidate gene prioritization with Endeavour". Nucleic acids research. 2016 Apr 30;44(W1):W117-21.
[3] A. Sifrim, P. Dusan, L. Tranchevent, A. Ardershirdavani, R. Sakai, P. Konings, J. Vermeesch, J. Aerts, B. De Moor, Y. Moreau. "eXtasy: Variant Prioritization by Genomic Data Fusion." Nature methods. 2013 Nov 1;10(11):1083-4.
[4] S. ElShal, L. Tranchevent, A. Sifrim, A. Adreshirdavani, J. Davis, Y. Moreau. "Beegle: from literature mining to disease-gene discovery". Nucleic acids research. 2016 Jan 29;44(2):e18.