

Towards Transparent and Reproducible Bioinformatics Analyses: the EPIGEN-Brazil Scientific Workflow

Gilderlanio S Araújo¹, Wagner CS Magalhaes¹, Paula JS Viriato¹, Mauricio L Barreto^{2,3}, Bernardo L Horta⁴, M Fernanda Lima-Costa⁵, Alexandre C Pereira⁶, Eduardo Tarazona-Santos¹, Maíra R Rodrigues^{1*} and the Brazilian EPIGEN Consortium.

¹Department of General Biology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil; ²Instituto Gonçalo Muniz, Fundação Oswaldo Cruz, Salvador, BA, Brazil; ³Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal da Bahia, Salvador, BA, Brazil; ⁴Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, RS, Brazil; ⁵Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG, Brazil; ⁶Instituto do Coração, Universidade de São Paulo, SP, Brazil

Two barriers in the dissemination and validation of knowledge today are lack of transparency and reproducibility of all the scientific process. It is argued that one way of tackling these issues is to have the complete scientific process as a research product in itself, which should be fully accessible and interactive. Following this direction, here we show how the EPIGEN-Brazil initiative website is being conceived as a Scientific Workflow and, at the same time, as a repository of bioinformatics resources for investigators in the areas of population genetics and genetic epidemiology. Both concepts are implemented as web tools and are fully and freely accessible online. Our Scientific Workflow is divided in self-contained components that cover different stages of the scientific process. In addition to Scientific Publications that comprise research hypotheses and results, a broad visualization of the research tasks is given by Flowcharts, the practical execution of such tasks is given by Masterscripts, and the intermediate steps, such as workshop discussions and technical reports, by Documents. A content manager connects conceptual and practical tasks and gathers comments from visitors and users. The latter feature allows user feedback to improve existing scientific analyses. We present examples in population genetics and genetic epidemiology that show how our Scientific Workflow can be intuitively navigated by researchers looking for guidance on similar projects or replication of existing results. Finally, we believe that our simple and intuitive approach to disclosing our scientific processes can encourage similar practices that increase the value of research.