MOLGENIS Workbench for Systems Medicine

K. Joeri van der Velde[1,2], Mark de Haan[1,2], Cisca Wijmenga[2], Richard J. Sinke[2], Tom J. de Koning[2], Rolf H. Sijmons[2], MOLGENIS team[1,2] and Morris A. Swertz[1,2]

[1] University of Groningen, University Medical Center Groningen, Genomics Coordination Center, Groningen, The Netherlands

[2] University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands

Contact: k.j.van.der.velde@umcg.nl, m.a.swertz@rug.nl

Project website: http://www.molgenis.org, source code: https://github.com/molgenis

License: GNU Lesser General Public License version 3 (GNU LGPLv3)

In 2009 we first presented MOLGENIS at BOSC as software generator for complex life science data. Since then, we developed MOLGENIS to tackle more challenges such as high-throughput analysis pipelines (BOSC 2011). This year we present a fresh and modernized MOLGENIS with a data model that now can be fully changed at runtime and includes easier upload format, data explorer, REST/R APIs, visualization and annotation tools with the focus on systems medicine based research and clinical applications.

High-throughput use cases such as multi-omics integration and NGS variant interpretation can now benefit from MOLGENIS adaptable upload formats and query performance, but also require a pre-filled toolbox to help process and understand these data. Therefore we also added extensible variant 'annotators' that enables easy data enrichment (CADD, FitCon, 1000G, ExAC, ClinVar, CGD, HPO, etc.), application analysis protocols (risk prediction, monogenic diagnostic analysis, etc.) and supporting algorithms (discover de-novo variants, symptom-to-disease matching, genome build liftover, etc.). These annotators are also available as a command-line executable to enable use in routine analysis pipelines before uploading the results. Adding more annotators is currently done by implementing a minimal Java interface class. Visual inspection of genes and variants in a biological context is made possible by a WikiPathway-based viewer and Dalliance-powered genome browser.

MOLGENIS is a collaborative open source platform on a mission since 2002 to generate great software infrastructure for life science research. It has already produced a large variety of applications including patient registries, model organism databases, biobank catalogs and computational script generators. We have refreshed the MOLGENIS platform by moving from generation-time to run-time configuration, allowing the users to upload complete data structures, incorporating popular software tools like Maven, MySQL, SpringMVC, GitHub, Bootstrap, Java 8 and ElasticSearch. The resulting modular software suite generates rich web applications that feature an import wizard for flexible formats, APIs for REST and JSON, user and rights management, cross-dataset ontological harmonization, and of course data exploration tools including plotting, filtering, aggregation, complex queries, and metadata browsing. Many components have runtime extension points, meaning custom R plots and reporting templates in Freemarker can be defined and used to present data. Imported data is indexed using ElasticSearch to eliminate long loading times.

We expect the MOLGENIS community will continue to develop valuable Systems Medicine exploration apps as well as function as a sharing platform for best practice data and pipelines, integration with international sharing platforms such as GA4GH and Cafe Variome (for which pilots are underway), well-curated reference knowledge-bases, and optimal user interfaces, results of which can disseminate into research institutes, clinical software companies and individual labs.