GENOMIC DATA ANALYSIS:
WORKFLOW MANAGEMENT SYSTEMS IN BIOINFORMATICS

Workflow management systems for analysing genomic data.

Analysing Genomics data involves the handling of large amounts of data that are often processed with various essential tools that are chained into complex pipelines. As these pipelines are dependent on the software versions, compute infrastructure, and other parameters, generating identical results from the same data can be challenging. In recent years, there has been a development of advanced workflow management systems for handling such pipelines that ensures high levels of reproducibility and scalability.

Based on their own experience in handling genomic data at the Genome Institute of Singapore (GIS), Ms Laura Wratten, Dr Andreas Wilm, and Dr Jonathan Göke summarised the advantages of such systems for genomic data in a review published in [Nature Methods](). The review illustrates aspects of data provenance, portability, scalability, re-entrancy, and summarises resources for high quality workflows that users can build on without having to develop their own pipeline from scratch.

In addition to the review, the team also provided a [code repository]() with example implementations for different workflow managers that makes it easy to compare the different systems for new users.

"Data is at the core of genomics. How to handle data to ensure that results are reproducible while managing resources optimally is one of the key aspects in our work. These workflow management systems have been available for some time, but many teams are still building their own custom scripts. We wrote this review to provide an easy introduction that should convince experts and non-experts that workflow management systems are a much needed solution that improves the standard for any genomic projects."

Dr Jonathan Göke, Group Leader of the
Laboratory of Computational Transcriptomics at GIS