

Component-based framework for bioinformatics analysis

Kristian Ovaska, Marko Laakso, Sampsa Hautaniemi

Computational Systems Biology Laboratory, Institute of Biomedicine and Genome-Scale Biology Program, Biomedicum Helsinki, University of Helsinki



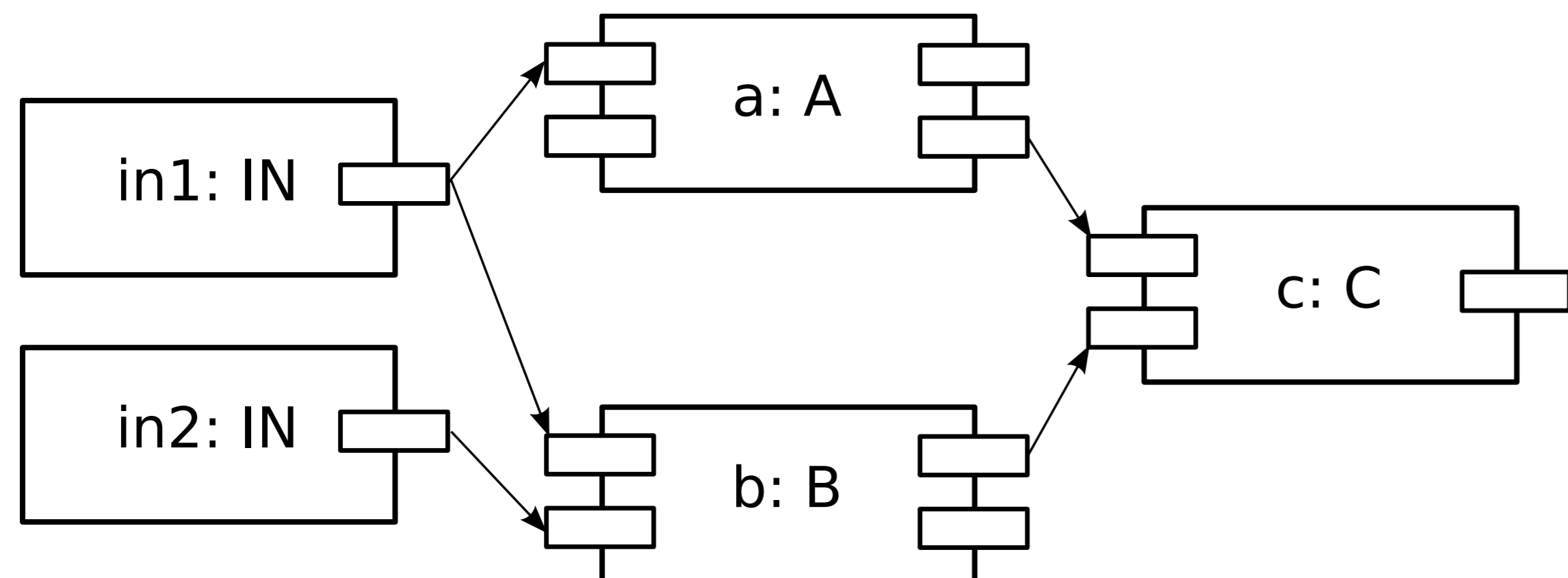
Abstract

Background: High-throughput measurement technologies produce vast amounts of data that must be analyzed computationally. The analysis should be done efficiently while maintaining high quality and repeatability.

Results: We have developed a component-based open source framework for high-throughput data analysis. Each component of the framework implements a well-defined part of the analysis, such as data import or differentially expressed gene selection. Third parties are able to extend the capabilities of the framework by writing new components.

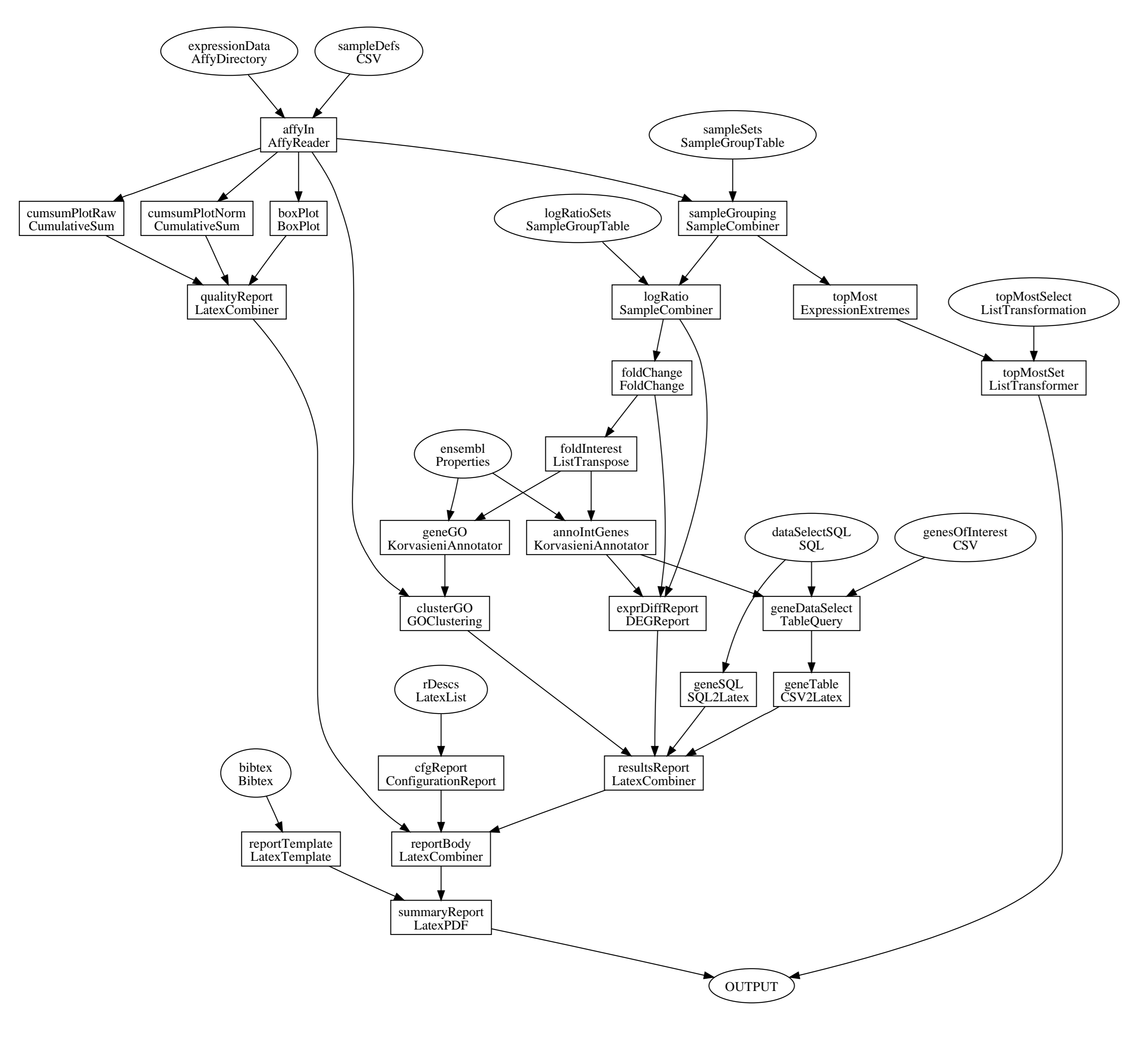
Backbone framework

Each analysis project is configured as a workflow, or a network of components. Components communicate through files. Components can be written in any language, which enables to leverage existing libraries, e.g. Bioconductor.



The workflow engine is invoked with a command-line interface. The engine supports the iterative nature of scientific studies by executing only those components whose configuration has changed when the network is re-executed.

Workflow visualization

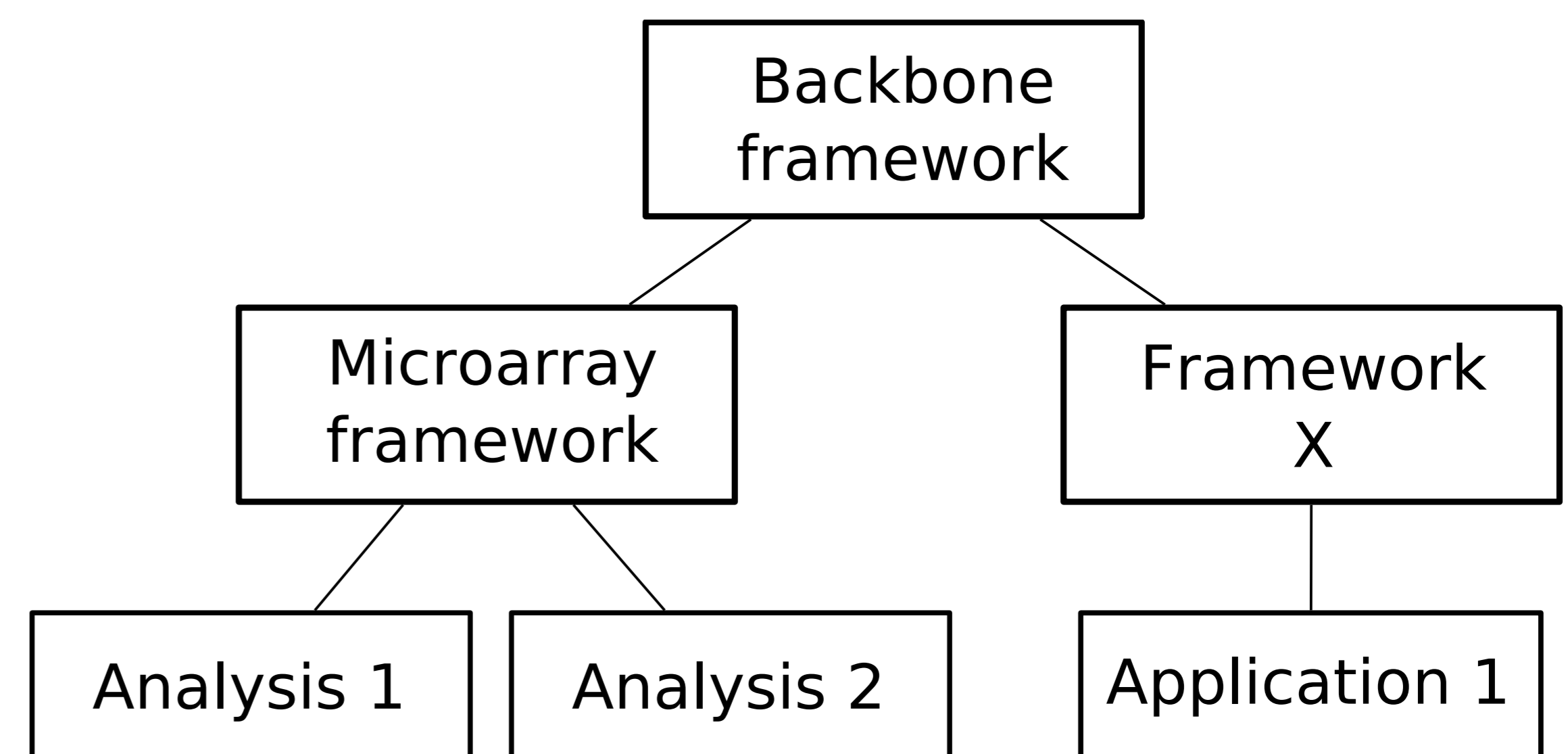


Summary

- ▶ We have developed a systematic and flexible platform for bioinformatics analysis
- ▶ Novel methods can be implemented as components and integrated into analysis workflows
- ▶ The whole analysis process can be automated, including final report generation

Basic architecture

The architecture is based on object oriented frameworks (OOFs), partially implemented applications that are extended to complete applications. OOFs enable the reuse of both code and program design, improving productivity and analysis quality.



The backbone framework provides core facilities, while application area specific frameworks provide components for e.g. microarray analysis.

Microarray framework

We currently provide approx. 70 components for expression, SNP and ChIP-chip microarray analysis. Components include:

- ▶ Data import, preprocessing and quality control
- ▶ DEG selection using fold change and t-test with FDR
- ▶ Annotation using Ensembl, KEGG and PINA (PPI database)
- ▶ Pathway analysis using GSEA and PathwayExpress
- ▶ Gene Ontology analysis: gene clustering and filtering
- ▶ High quality PDF report generation that contains final results as well as workflow configuration

Example outputs of components

