Customization and Integration of GenePattern Microarray Workflows into a Collaborative Environment

Felix Eichinger[1] and Beth Kirschner[2], Internal Medicine[1] and School of Information[2], University of Michigan, Ann Arbor, MI
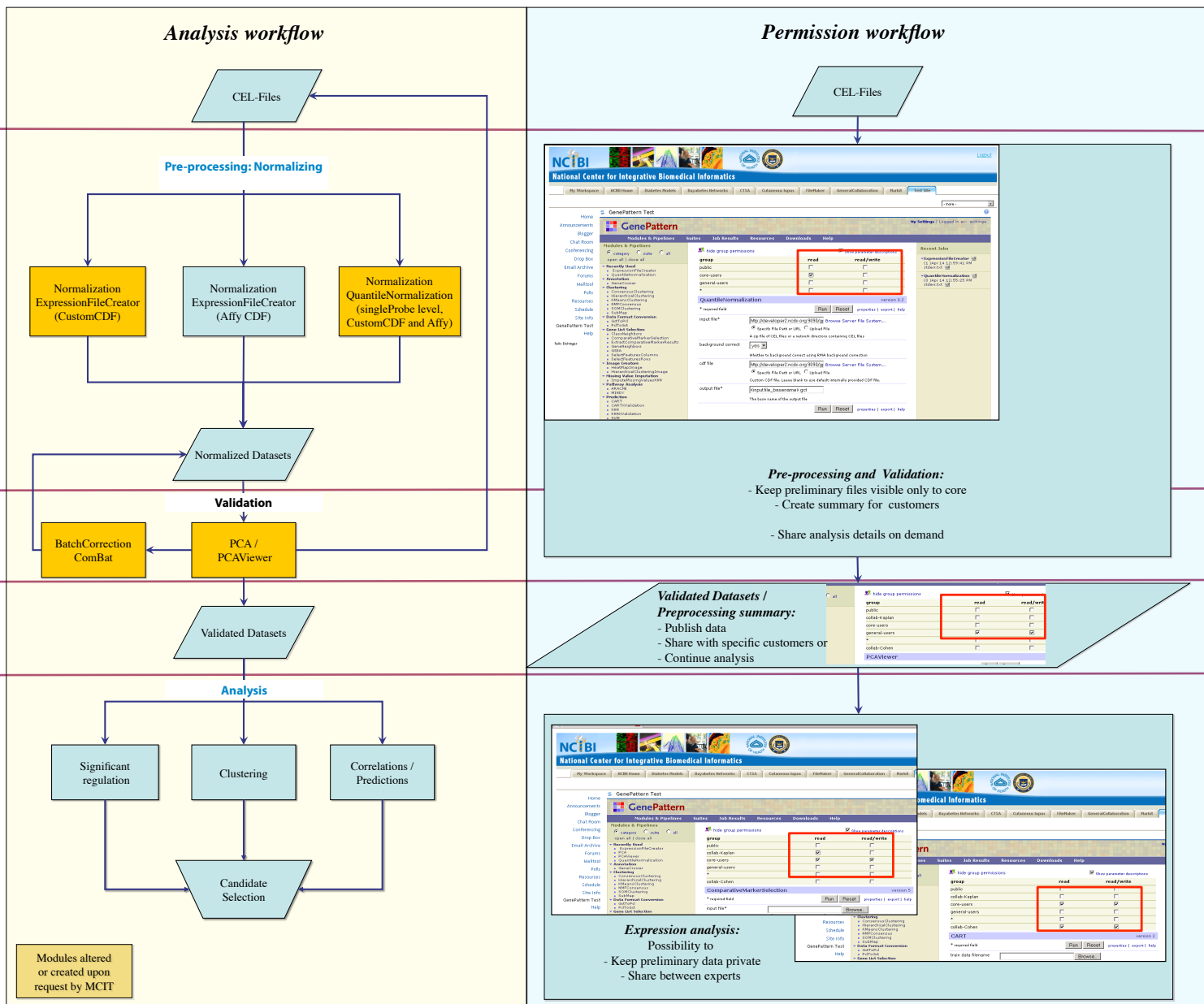
Abstract

In a collaborative effort, the Kretzler Lab, the NCIBI developer team and the MCIT are establishing a framework for effective and collaborative analysis of microarrays. To attain this goal, the workflow for microarray analyses, developed and refined over the years by the Kretzler lab, has been implemented in GenePattern. This ensures that the two main goals for a scientific workflow can be met:

1) Persistence, resulting in improved efficiency, reproducibility and the possibility to effectively communicate the current stage of the work between different members of the lab or to collaborators.

2) Flexibility, in adapting for the need to respond to the current data set while staying in a global analysis strategy and the capability to improve and extend the original workflow as new tools and techniques emerge.

With GenePattern it is possible to fulfill both of these requirements, the support for pipelines enhances effectiveness and reproducibility, while the modular architecture enabled the MCIT team to adapt and de novo implement analysis modules to meet our needs.

In parallel, the MCIT and the NCIBI developer team are joining forces in integrating GenePattern into the NCIBI portal system, creating the basis for effective collaboration through integrated and group-specific sharing of files and results.

Analysis workflow — CEL-Files; Pre-processing: Normalizing; Normalization ExpressionFileCreator (CustomCDF); Normalization ExpressionFileCreator (Affy CDF); Normalization QuantileNormalization (singleProbe level, CustomCDF and Affy); Normalized Datasets; Validation; BatchCorrection ComBat; PCA / PCAViewer; Validated Datasets; Analysis; Significant regulation; Clustering; Correlations / Predictions; Candidate Selection; Modules altered or created upon request by MCIT

Permission workflow — CEL-Files; Pre-processing and Validation: - Keep preliminary files visible only to core - Create summary for customers - Share analysis details on demand; Validated Datasets / Preprocessing summary: - Publish data - Share with specific customers or - Continue analysis; Expression analysis: Possibility to - Keep preliminary data private - Share between experts