

ER-flow Application Description Template

Application Name: RNA-seq meta workflow
Application domain: RNA Sequencing
Brief description of application The analysis of RNA-seq data involves various steps that can be roughly divided into two parts: discovery of candidate genes and annotation of genes. The first part is provided by the Tuxedo workflow. The second part consists of enrichment of metadata about the list of genes generated by the first group using additional databases and ontologies. This second part is performed by a TAVERNA workflow. The meta-workflow combines both parts together.
input data: <ul style="list-style-type: none">List of files containing paired-end reads in FASTQ or FASTA format, both left and right reads.Reference annotation to guide Reference Annotation Based Transcript (RABT) assembly output data: <ul style="list-style-type: none">Enriched dataset by identifying: which pathways genes are involved in; the biological processes and molecular functions the genes are involved in; functional annotation clustering of the genes. sample data: (lfn: /grid/vlmed/AMC-e-BioScience/RNASeq/data) application (http://ccb.jhu.edu/software/tophat/ ; http://cufflinks.cbc.umd.edu/ ; http://www.myexperiment.org/workflows/4453.html) docs (http://ccb.jhu.edu/software/tophat/manual.shtml , http://cufflinks.cbc.umd.edu/manual.html) publication (http://www.nature.com/nprot/journal/v7/n3/full/nprot.2012.016.html)
Execution environment DCI: (EGI, SRM/LFC, vlmed VO) middleware: gLite workflow system: WS-PGRADE, Taverna
Execution characteristics input: 10MB-100GB output: 10MB-100GB processing time (per unit): 1 - 36 hours memory usage: n.a. disk usage: ~10MB-100GB
Target users Bioinformatitions (LUMC, AMC); number of users: 10+; user type: end-user
Usage scenario for workflow in the ER-FLOW A meta-workflow as a combination of WS-PGRADE and Taverna workflows has been implemented to port this application to EGI for the vlmed VO. The workflow is published on the SHIWA repository with appropriate documentation, metadata and sample data.
Contact information (workflow developers) name: Vladimir Korkhov e-mail: vkorkhov@gmail.com