

# ER-flow Application Description Template

<b>Application Name:</b> Tuxedo
<b>Application domain:</b> RNA Sequencing
<b>Brief description of application</b> Recent advances in high-throughput cDNA sequencing (RNA-seq) can reveal new genes and splice variants and quantify expression genome-wide in a single assay. The volume and complexity of data from RNA-seq experiments necessitate scalable, fast and mathematically principled analysis software. TopHat and Cufflinks are free, open-source software tools for gene discovery and comprehensive expression analysis of high-throughput mRNA sequencing (RNA-seq) data. Together, they allow biologists to identify new genes and new splice variants of known ones, as well as compare gene and transcript expression under two or more conditions. The Tuxedo protocol describes in detail how to use TopHat and Cufflinks to perform such analyses [Trapnell et al, 2012].  Trapnell et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7, 562–578 (2012)
<b>input data:</b> <ul style="list-style-type: none"><li>List of files containing paired-end reads in FASTQ or FASTA format, both left and right reads.</li><li>Reference annotation to guide Reference Annotation Based Transcript (RABT) assembly</li></ul> <b>output data:</b> <ul style="list-style-type: none"><li>Tar-gzipped cuffdiff output</li></ul> <b>sample data:</b> (lfn: /grid/vlmed/AMC-e-BioScience/RNASEq/data) <b>application</b> ( <a href="http://ccb.jhu.edu/software/tophat/">http://ccb.jhu.edu/software/tophat/</a> ; <a href="http://cufflinks.cbc.umd.edu/">http://cufflinks.cbc.umd.edu/</a> ) <b>docs</b> ( <a href="http://ccb.jhu.edu/software/tophat/manual.shtml">http://ccb.jhu.edu/software/tophat/manual.shtml</a> , <a href="http://cufflinks.cbc.umd.edu/manual.html">http://cufflinks.cbc.umd.edu/manual.html</a> ) <b>publication</b> ( <a href="http://www.nature.com/nprot/journal/v7/n3/full/nprot.2012.016.html">http://www.nature.com/nprot/journal/v7/n3/full/nprot.2012.016.html</a> )
<b>Execution environment</b> DCI: (EGI, SRM/LFC, vlmed VO) middleware: gLite workflow system: WS-PGRADE, Galaxy
<b>Execution characteristics</b> input: 10MB-100GB output: 10MB-100GB processing time (per unit): 1 - 36 hours memory usage: n.a. disk usage: ~10MB-100GB
<b>Target users</b> Bioinformatitions (LUMC, AMC); number of users: 10+; user type: end-user
<b>Usage scenario for workflow in the ER-FLOW</b> A WS-PGRADE workflow 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.
<b>Contact information</b> (workflow developers) name: Vladimir Korkhov e-mail: vkorkhov@gmail.com