	KNIME	Galaxy	CLC Genomics Workbench
Specific/General purpose	General (for many fields of work/research)	Specific (Biological/Bioinformatics research)	Specific (Biological/Bioinformatics research)
License	GNU GPL / Commercial	Academic Free License	Commercial
Open/Closed Source	Partially open/closed source	Open Source	Closed source
Available contact	• Contact via Knime website	• No specific contact point/address	• Contact via Qiagen website
Available help online	• KNIME website • KNIME forums • KNIME YouTube videos • Other forums: seqanswers, biostars, stackoverflow, etc • Example workflows available	• Galaxy website • Galaxy dev site • Galaxy videos on vimeo • other forums: seqanswers, biostars, stackoverflow, etc	• Detailed manuals (online and in Workbench) • Step by step tutorials • Qiagen website (non-contact) • Qiagen videos on Qiagen website
Google searching	• Often search results for KNIME forums • Often relevant hits to searches with general terms	• Often hits for galaxy website, galaxy biostars • Sometimes results for Samsung Galaxy phones (mainly with more technical searches)	• Often hits for Qiagen website • Also hits for Qiagen CLC tutorials
Cost/Pricing	Free (Knime analytics platform) and commercial (Knime server)	Free	Commercial (all versions)
Supported OS	Windows, OSX, Linux	OSX, Linux	Windows, OSX, Linux
Often used in Bioinformatics	Seemingly not so much	Seemingly quite often	Seemingly quite often
Add own programs/tools	Nodes to use the program/tool need to be programmed in Java	Is possible. An XML wrapper needs to be created in the same directory as the tool to use. An entry also needs to be added to an XML file of Galaxy.	Possible, requires users to create plugins with the CLC Developer Kit / SDK
Method of installation	• Windows: Installer • Linux: Extract archive	• Extract archive • Git versioning system	• Windows: Installer, • Linux: Extract archive
Ease of installation	Easy. After running the installer or extracting the archive, KNIME can be run as is.	Easy. After extraction of the archive Galaxy needs to be run once first to configure. Afterwards, it can be run as is.	• After installation/extraction CLC can be run as is
Requirements after installation	No other requirements after installation	• Run once to let Galaxy configure • Administrator account needs to be created and activated.	• Obtain/Activate a license
Other software required	• Java (is included) • R • R package Rserve • Tools required by knime4ngs already installed/available	• Python 2.7 • R • R packages readr & rhdf5 (required by DESeq2)	• None
Additional packages available	• Yes (nodes and additional	• Yes, many additional tools	• Plugins available to install

	• software) • Not many different bioinformatics packages/node collections	• available from several repositories.	
Installing additional packages	Users can install additional nodes and packages	• Tools and additional tools need to be installed by an administrator	Users can install plugins
Shared tools	Nodes can be shared with Knime Server	• Tools installed on a galaxy instance by an administrator can be used by all registered users.	All instances have the same tools
Updating	Yes, via KNIME GUI	Use git commands to update as administrator	Obtain installer
RNA-Seq packages support (how often updated)	Created nodes and packages do not seem to be updated often? (But maybe also not required as much?)	Popular tools seem to be updated every now and then by either the devteam or other teams. (Checking the version of tools seems to be a bit more complicated)[*1]	Tools for RNA-Seq are built in and updated with development.
Workflow creation and setting parameters	• Occurs simultaneously • Can be done separately via workflow variables	• Partially simultaneously, partially separate • Data input separate from workflow creation	• Can occur completely simultaneously as well as completely separate
Differentiation between users	Workspaces from Knime Analytics Platform	User accounts	• License, workbench instance
RNA-Seq workflows available to import and modify	• Barely any (or they are hard to find).	• Published workflows on usegalaxy.org • Workflows at MyExperiment	• Don't seem to be available (or hard to find).
Availability of usable nodes for RNA-Seq	Nodes to perform essential steps are available but only for a few programs.	Nodes available for most types and flavours of programs.	All necessary tools are available.
Ease of using programs that have no nodes	• External Tool node • Java/Python code snippets • External SSH Tool node	• No tool available	• Requires Workbench to be connected to CLC Server and have the 'External Applications Plugin' installed. • CLC Server needs to be configured via administrative web interface.
Workflow readability	• Easy to see the steps of the workflow	• Large workflows can become somewhat cluttered	• Large workflows can become somewhat cluttered
Workflow flexibility	• Two way branching with if switch • Two way branching with if switch controlled by java code • Three way branching with case switch	• Galaxy does not offer branching/decisions (is planned to be incorporated however)	• No options for branching or decision making
Workflow robustness	• Workflow nodes can be run and rerun individually.	• Workflow steps can be rerun if necessary	• Workflow steps can be run as individual programs

	• KNIME4NGS offers an extra layer of robustness through .klock files. • Successfully executed nodes remain in completed state	• Might be confusing as the step needs to be rerun from the History (not the workflow)	
Sharing of workflows between users	• Sharing between different users in KNIME Team Space, Server or Cloud Server (needs to be purchased). • Exporting and importing workflows	• Sharing with individual or multiple users via Galaxy • Share with everyone via Galaxy • Exporting and importing workflows	• Install workflow in CLC Genomics Workbench as individual user • Install workflow in CLC Genomics Server for multiple users
Sharing of used data for analyses between users	• Requires KNIME Team Space, Server or Cloud Server (commercial)	• Creation of data libraries (all users of galaxy instance) • Histories can be shared with individual or multiple users • Histories can be shared with everyone	• Through CLC Genomics Server (data in CLC Server) • Through CLC Bioinformatics Database (if data is in a databases)
Publishing workflows	N/A	• Workflows can be published to Galaxy's Published Workflows website • Workflows can be published on MyExperiment	N/A
Available QC programs	FastQC, FxFastQStats, FlexBar	FastQC, PRINSEQ, FlexBar	'Create Sequencing QC Report' tool
Available QC parameters	For FastQC not many if any at all.	For FastQC there do not seem to be many parameters other than the input files.	N/A
Available trimming/adapter removal programs	TrimGalore from KNIME4NGS, flexbar, KNIME4NGS RawReadManipulator	Cutadapt, TrimGalore, Trimmomatic, flexbar, FastqMcf, FastX, PRINSEQ, Sickle	'Trim Reads' tool
Available adapter removal parameters	Adapter removal, min/max length, min/max quality, 3/5' trimming available for	Quality limit, adapters to remove, remove short/long reads, remove leading/trailing bases	Quality limit, list of adapters to remove, remove leading/trailing nucleotides, removes short/long reads
Available mapping programs	Bowtie, Bowtie2, BWA, Masai, RazerS, YaraMapper, Segemehl, Star	Bowtie, Bowtie2, TopHat, TopHat2, BWA, STAR, HISAT2, Segemehl, Mosaik2, rqrnastar	'Map Reads to Reference' tool, 'RNA-Seq Analysis' tool
Available mapping parameters	• SeqAn and KNIME4NGS nodes offer many parameters to be set and changed.	• Reference genome/transcriptome, single/paired end reads, read group info, analysis mode (bwa,	• Only parameters involved with the alignment are configurable.

		bwa-mem), min/max intron length (tophat)	
Available read mapping QC programs	Picardtools from KNIME4NGS	RSeQC, Qualimap	No specific tool, tracks can be viewed however
Available transcriptome mapping quantification programs	N/A	Kallisto, RSEM, Salmon, Sailfish (for isoforms), eXpress	'Map reads to Reference' tool, 'RNA-Seq Analysis' tool
Available transcriptome mapping quantification parameters	N/A	Reference transcriptome, single/paired end reads, kmer/fragment length, bootstrap number and seed (Kallisto), type of indexing (Salmon)	Parameters involved in alignment (gap penalty, etc)
Availability of genome mapping quantification programs	FeatureCounts (KNIME4NGS), Insegt (SeqAn)	HTSeq-count, featureCounts, Cufflinks	'Map Reads to Reference' tool, 'RNA-Seq Analysis' tool
Available genome mapping quantification parameters	Feature type (exon, CS), ID attribute (gene_id), single/paired reads, dealing with nonunique/ambiguous mapped reads	Feature type (exon, CDS) to use, ID attribute (gene_id), minimum alignment quality, single/paired reads, dealing with nonunique/ambiguous mapped reads	Type of counts (TPM, RPKM, Unique counts, Total counts) Count
Availability of DE Analysis programs	DESeq, edgeR, Limma	DESeq2, edgeR, limma, Cuffdiff	'Differential Expression for RNA-Seq' tool
Available DE parameters	DESeq2: dispersion calculation, sharing mode Limma/EdgeR: normalization factor calculation, pvalue corre3ctgion method	• DESeq2: Factor levels, Type of input counts, Fit type to perform, Turn outlier filtering on/off • Limma/EdgeR: Filter low count, result filtering (log2fold change)	• Design matrix, what to test differential to and what kind of comparison to perform
Availability of transcriptome assembly programs	N/A	Trinity, Stringtie, velvet	'De Novo Assembly' tool
Available transcriptome assembly parameters	N/A	Minimum contig length, single/paired reads	• Parameters involved in alignment (gap penalty, etc)
Availability of variant discovery programs	SnpStore (SeqAn)	GATK, Freebayes, VarScan	Basic Variant Detection tool, Fixed Ploidy Variant Detection tool, Low Frequency Variant Detection tool
Availability of isoform discovery programs	N/A	Cufflinks	Not available in CLC Genomics Workbench
Possibility to view results from within workflow	Does not seem to be possible. This might only be possible for R snippets that display a figure.	Not during the workflow. Results might be viewed afterwards?	• Results can be viewed afterwards by opening files. • Good visualization abilities
Visualization options	Plotting nodes, IGV	Viewing plain text and html files, various plotting, genome browser	Viewing plain text and html files, various plotting, genome browser
Requirements for RNA-Seq nodes	• KNIME4NGS nodes require	• Required binaries and	• N/A (required software is

	binaries to be present, others not	dependencies are added when installing tools	included)
Automatisation possibilities	• Workflows can be run from command line • General workflow variables offers options to be set for the workflow similar to program parameters	• Workflows can be run from CLI and programs using the Galaxy API (requires a key to be generated for the account to access Galaxy)	• Workflows installed on CLC Server can be executed from the command line if CLC Server Command Line Tools is installed.
Reading multiple files	• FileLoader node from KNIME4NGS • List Files node	• Input dataset • Dataset collection	• Select folder or folders with the Batch option
Using/Processing multiple files	• Loop mechanisms such as chunk loop, parallel chunk loop, etc	• Dataset collection tool • Depends on tool if it can use dataset collections	• Handled via Batch processing • Some tools cannot operate in Batch mode
Change input files	• Reconfigure nodes • Change workflow variable controlling input data	• Change History • Select different files from History	• Reconfigure 'Workflow Input' tool • Select different input when running workflow
Use variables for flexibility	• Workflow variables can be used to make workflows more flexible.	N/A	N/A
Output production of files	• KNIME4NGS saves output in same folder as input • Other nodes offer different output location	• Output saved in History (can be saved in a new History) • History is located in subfolders of Galaxy instance	• Output is saved in user chosen workbench directory. • Option to save results in subfolder for each dataset in batch mode
Ease of managing output locations	• KNIME4NGS nodes do not offer to change output location. Other nodes may or may not. Changing output can be done with different input locations.	• Results for each run of a workflow can be send to a new History.	• Select output location prior to running a workflow
Ease of using same output in different steps	• Connect one output port to multiple input ports.	• Connect the output to multiple different inputs.	• Connect output to multiple different inputs.
Ease of tracking events for each step	• Nodes can be executed one at a time, all at once or selected nodes. • Nodes display their status.	• Viewable in History which output is being created. • Not viewable if a process is running when History is not updated	• Process window • Workflow execution log
Steps difficult to put into pipeline/workflow	• Transcriptome quantification (no specific node available) • Proper input format for DESeq analysis	• Transcriptome quantification with samtools not possible • Subworkflows not forwarding output to connected tool • DEA with DESeq and featurecounts (need to remove header from count files)	• Differential expression for many replicates
Transparency of issued	• KNIME4NGS offers .klock files	• Specific commands issued and	• Not possible to see what

commands/actions by programs	containing issued commands • Other nodes may not offer way to view issued commands	other details are available in the History for each output file	specifically happens
Exporting and reimporting own workflow	• Workflow remains the same • Provides errors if workflow required nodes are missing • Provides warnings if other versions of nodes are installed than used in workflow	• Workflow remains the same • Provides errors if workflow required tools are missing • Provides warnings if tool version in workflow differs from installed tool version	• Workflow remains the same
Workflow readability: General overview	• Easy to view which steps • Easy to view which output connects to which input	• Larger workflows more difficult to read • Input to output connections harder to distinguish in larger workflows	• Larger workflows can become more difficult to read • Tool specific inputs and outputs harder to distinguish in larger workflows
Workflow readability: Output creation	• Not shown without configuring or viewing output port specifically	• Easy to see which outputs are created as tools display which output they can/do create.	• Easy to see which outputs are created by which tool
Change settings per dataset when processing multiple datasets	• RawReadManipulator determines settings based on FastQC report • Might be possible by converting a table row with settings to flow variables.	N/A	N/A

To add as well

In Knime workflow, trimming/clipping settings can be determined per dataset by FastQC, this is not the case in Galaxy.

In Knime4ngs and Galaxy other parameters, such as mapping parameters, can't be changed per dataset in the workflow for multi mapping. This CAN be done when the workflow maps each F/R sets in separate flows and not in a loop! (Important)

Visualisation: The knime workflow hasn't offered many options to visualize the results. In Galaxy some results can be viewed as long as it's an image, plain text, html, etc. In CLCbio you can visualize most/all steps.

Accessing the platform: Knime Server allows several ways of accessing the Knime environment (knime analytics platform, your own application with rest api, with web browser through WebPortal)

Difficulties with workflows:

Knime

Galaxy

Sub workflows do not seem to forward output to a connected tool. Some tools such as Sickle actually can't use a dataset collection as input.

	KNIME	Galaxy	CLC Genomics Workbench
Workflow flexibility	• Two way branching with if switch • Two way branching with if switch controlled by java code • Three way branching with case switch	• Galaxy does not offer branching/decisions (is planned to be incorporated however)	• No options for branching or decision making
Available mapping programs	Bowtie, Bowtie2, BWA, Masai, RazerS, YaraMapper, Segemehl, Star	Bowtie, Bowtie2, TopHat, TopHat2, BWA, STAR, HISAT2, Segemehl, Mosaik2, rqrnastar	'Map Reads to Reference' tool, 'RNA-Seq Analysis' tool
Availability of DE Analysis programs	DESeq, edgeR, Limma	DESeq2, edgeR, limma, Cuffdiff	'Differential Expression for RNA-Seq' tool
Output production of files	• KNIME4NGS saves output in same folder as input • Other nodes offer different output location	• Output saved in History (can be saved in a new History) • History is located in subfolders of Galaxy instance	• Output is saved in user chosen workbench directory. • Option to save results in subfolder for each dataset in batch mode