RNA-Seq Analytic Pipeline for GUDMAP/RBK

Process Design Table

Process	Tool (version)	Docker Container (version)	Inputs	Outputs
getBag	deriva 1.0.0	bicf/gudmaprbkfilexfer	Replicate RID	bagit.zip
getData	deriva 1.0.0	bicf/gudmaprbkfilexfer	bagit.zip	*.fastq.gz File.csv Experiment Settings.csv Experiment.csv
parseMetadata	python 3.7.4 argparse 1.4.0 pandas 0.25.3	bicf/python3	File.csv Experiment Settings.csv Experiment.csv	Replicate RID Strandedness Spike Species
getRef	aws-cli 1.16.293 OR deriva 1.0.0	bicf/awscli OR bicf/gudmaprbkfilexfer	Spike Species	HiSat2 Reference Reference .bed
trimData	TrimGalore 0.6.4	bicf/trimgalore	Strandedness *.fastq.gz	*trimmed.fq.gz
alignData	hisat 2.1.0 samtools 1.9	bicf/gudmaprbkaligner	Strandedness Spike Species *trimmed.fq.gz HiSat2 Reference	*.bam *bam.bai
dedupData	Picard 2.21.7 samtools 1.9	bicf/gudmaprbkdedup	*.bam *.bai	*dedup.bam *dedup.bam.bai
makeBigWig	Deeptools 3.3	bicf/deeptools3.3	*dedup.bam	*.bw
countData	Subread OR RSEM python OR R		*dedup.bam	*.count.csv *.tpm.csv *.fpkm.csv
fastqc	Fastqc 0.11.8		*.fastq.gz	*.fastqc.zip
inferMetadata	RSeQC 3.0	bicf/rseqc3.0	*.dedup.bam *dedup.bam.bai Reference .bed	
qc	python or R 3.0		Replicate RID Strandedness Spike Species *.fastqc.zip additional qc files	*.qc.json multiqc_report.html
uploadData	deriva 1.0.0	bicf/gudmaprbkfilexfer	*dedup.bam *.count.csv *.tpm.csv *.fpkm.csv *.fastqc.zip *.qc.json	