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	replicateRID	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
				manualEnds
				metaEnds
				metaStrand
				metaSpike
				metaSpecies

getRef	aws-cli 1.16.293 OR deriva 1.0.0	bicf/awscli OR bicf/gudmaprbkfilexfer		*.ht2 genome.fna genome.gtf genome.bed
trimData	TrimGalore 0.6.4	bicf/trimgalore	manualEnds *.fastq.gz	*trimmed.fastq.gz *downsampled.fq.gz *.trimming_report.txt
alignData	hisat 2.1.0 samtools 1.9	bicf/gudmaprbkaligner	inferEnds inferSpike inferSpecies *trimmed.fq.gz ht2	*.bam *.bam.bai *.alignSummary.txt
dedupData	Picard 2.21.7 samtools 1.9	bicf/gudmaprbkdedup	*.bam *.bam.bai	*.dedup.bam *.dedup.bam.bai *.deduped.Metrics.txt
makeBigWig	Deeptools 3.3	bicf/deeptools3.3	*.dedup.bam	*.bw
countData	Subread 2.0.0 R 3.6.1 optparse 1.6.4	bicf/subread2:2.0.0	inferEnds *.dedup.bam *.dedup.bam.bai *.gtf	*.countTable.csv
alignSampleData	hisat 2.1.0 samtools 1.9	bicf/gudmaprbkaligner	*downsampled.fq.gz *.ht2	*downsampled.bam *downsampled.bai
inferMetadata	RSeQC 3.0	bicf/rseqc3.0	*.downsampled.bam *.downsampled.bai genome.bed	inferEnds inferStrand inferSpike inferSpecies
fastqc	Fastqc 0.11.8	bicf/fastqc:2.0.0	*.fastq.gz	*.fastqc.zip
tin	RSeQC 3.0	bicf/rseqc3.0	*.dedup.bam *.dedup.bam.bai genome.bed	*.tin.xls
qc	python or R		infer/meta metadata *.fastqc.zip *.trimming_report.txt *.alignSummary.txt *.deduped.Metrics.txt	*.qc.json multiqc_report.html
uploadData	deriva 1.0.0	bicf/gudmaprbkfilexfer	*.countTable.csv *.qc.json	