**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.x  argparse 1.4.0  pandas 0.25.3 | bicf/python3 | File.csv  Experiment Settings.csv Experiment.csv | Replicate RID  Strandedness  Spike  Species |
| trimData | TrimGalore 0.6.4 | bicf/trimgalore | Strandedness  \*.fastq.gz | \*trimmed.fq.gz |
| alignData | hisat 2.1.0  samtools 1.9 | bicf/hisat2 | Strandedness  Spike  Species \*trimmed.fq.gz | \*.bam  \*.bai |
| dedupData | Picard |  | \*.bam | \*dedup.bam |
| countData | Subread  python OR R |  | \*dedup.bam | \*.count.csv  \*.tpm.csv  \*.fpkm.csv |
| makeBidWig | deeptools |  | \*.bam (dedup) | \*.bw |
| fastqc | fastqc |  | \*.fastq.gz | \*.fastqc.zip |
| qc | python or R |  | 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 |  |