

split_fastq

input

<input function>
<input function>

output →

workup/split_fastq/sample,{\..+}_R1.part_000.fastq
workup/split_fastq/sample,{\..+}_R1.part_001.fastq
workup/split_fastq/sample,{\..+}_R2.part_000.fastq
workup/split_fastq/sample,{\..+}_R2.part_001.fastq

compress_fastq

input

workup/split_fastq/sample_R1.part_{splitid}.fastq
workup/split_fastq/sample_R2.part_{splitid}.fastq

output →

workup/split_fastq/sample,{\..+}_R1.part_{splitid}.fastq.gz
workup/split_fastq/sample,{\..+}_R2.part_{splitid}.fastq.gz

adaptor trimming

input

workup/split_fastq/sample_R1.part_{splitid}.fastq.gz
workup/split_fastq/sample_R2.part_{splitid}.fastq.gz

output →

workup/trimmed/sample,{\..+}_R1.part_{splitid}.fastq.gz.trimming_report.txt
workup/trimmed/sample,{\..+}_R1.part_{splitid}_val_1.fq.gz
workup/trimmed/sample,{\..+}_R2.part_{splitid}.fastq.gz.trimming_report.txt
workup/trimmed/sample,{\..+}_R2.part_{splitid}_val_2.fq.gz

barcode_id

input

workup/trimmed/sample_R1.part_{splitid}_val_1.fq.gz
workup/trimmed/sample_R2.part_{splitid}_val_2.fq.gz

output →

workup/fastqs/sample,{\..+}_R1.part_{splitid}.barcoded.fastq.gz
workup/fastqs/sample,{\..+}_R2.part_{splitid}.barcoded.fastq.gz

split_bpm_dpm

input

workup/fastqs/sample_R1.part_{splitid}.barcoded.fastq.gz

output →

workup/fastqs/sample,{\..+}_R1.part_{splitid}.barcoded_bpm.fastq.gz
workup/fastqs/sample,{\..+}_R1.part_{splitid}.barcoded_dpm.fastq.gz
workup/fastqs/sample,{\..+}_R1.part_{splitid}.barcoded_other.fastq.gz
workup/fastqs/sample,{\..+}_R1.part_{splitid}.barcoded_short.fastq.gz

barcode_identification_efficiency

input

workup/fastqs/sample_R1.part_{splitid}.barcoded.fastq.gz

output →

workup/sample,{\..+}.part_{splitid}.bid_efficiency.txt

cutadapt_dpm

input

workup/fastqs/sample_R1.part_{splitid}.barcoded_dpm.fastq.gz

output →

workup/trimmed/sample,{\..+}_R1.part_{splitid}.barcoded_dpm.RDtrim.fastq.gz
workup/trimmed/sample,{\..+}_R1.part_{splitid}.barcoded_dpm.RDtrim.qc.txt

bowtie2_align

input

workup/trimmed/sample_R1.part_{splitid}.barcoded_dpm.RDtrim.fastq.gz

output →

workup/alignments_parts/sample,{\..+}.part_{splitid}.DNA.bowtie2.mapq20.bam

rename_and_filter_chr

input

workup/alignments_parts/sample.part_{splitid}.DNA.bowtie2.mapq20.bam

output →

workup/alignments_parts/sample,{\..+}.part_{splitid}.DNA.chr.bam

cutadapt_oligo

input

workup/fastqs/sample_R1.part_{splitid}.barcoded_bpm.fastq.gz

output →

workup/trimmed/sample,{\..+}_R1.part_{splitid}.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample,{\..+}_R1.part_{splitid}.barcoded_bpm.RDtrim.qc.txt

fastq_to_bam

input

workup/trimmed/sample_R1.part_{splitid}.barcoded_bpm.RDtrim.fastq.gz

output →

workup/alignments_parts/sample,{\..+}.part_{splitid}.BPM.bam
workup/alignments_parts/sample,{\..+}.part_{splitid}.BPM.unsorted.bam

make_clusters

input

workup/alignments_parts/sample.part_{splitid}.BPM.bam
workup/alignments_parts/sample.part_{splitid}.DNA.chr.masked.bam

output →

workup/clusters_parts/sample,{\..+}.part_{splitid}.clusters
workup/clusters_parts/sample,{\..+}.part_{splitid}.unsorted.clusters

merge_clusters

input

workup/clusters_parts/sample.part_000.clusters
workup/clusters_parts/sample.part_001.clusters

output →

workup/clusters/sample,{\..+}.clusters
workup/clusters/sample,{\..+}.duplicated.clusters

clusters_all

input

workup/clusters/sample1.clusters
workup/clusters/sample2.clusters

output →

workup/clusters/all.clusters

merge_beads

input

workup/alignments_parts/sample.part_000.BPM.bam
workup/alignments_parts/sample.part_001.BPM.bam

output →

workup/alignments/sample,{\..+}.merged.BPM.bam

merge_dna

input

workup/alignments_parts/sample.part_000.DNA.chr.masked.bam
workup/alignments_parts/sample.part_001.DNA.chr.masked.bam

output →

workup/alignments/sample,{\..+}.DNA.merged.bam

threshold_tag_and_split

input

workup/alignments/sample).DNA.merged.bam
workup/clusters/sample).clusters

output →

workup/splitbams/sample,{\..+}.DNA.merged.labeled_AB1-A1.bam
workup/splitbams/sample,{\..+}.DNA.merged.labeled_AB2-A2.bam

splitbams_merged

input

workup/splitbams/sample1.DNA.merged.labeled_{target}.bam
workup/splitbams/sample2.DNA.merged.labeled_{target}.bam

output →

workup/splitbams/{target_AB1-A1|AB2-A2}.bam

pipeline_counts

input

workup/alignments/sample1.DNA.merged.bam
workup/alignments/sample1.merged.BPM.bam
workup/alignments/sample2.DNA.merged.bam
workup/alignments/sample2.merged.BPM.bam
workup/alignments_parts/sample1.part_000.BPM.bam
workup/alignments_parts/sample1.part_000.DNA.bowtie2.mapq20.bam
workup/alignments_parts/sample1.part_000.DNA.chr.bam
workup/alignments_parts/sample1.part_000.DNA.chr.masked.bam
workup/alignments_parts/sample1.part_000.DNA.chr.BPM.bam
workup/alignments_parts/sample1.part_001.DNA.bowtie2.mapq20.bam
workup/alignments_parts/sample1.part_001.DNA.chr.bam
workup/alignments_parts/sample1.part_001.DNA.chr.masked.bam
workup/alignments_parts/sample2.part_000.BPM.bam
workup/alignments_parts/sample2.part_000.DNA.bowtie2.mapq20.bam
workup/alignments_parts/sample2.part_000.DNA.chr.bam
workup/alignments_parts/sample2.part_000.DNA.chr.masked.bam
workup/alignments_parts/sample2.part_001.BPM.bam
workup/alignments_parts/sample2.part_001.DNA.bowtie2.mapq20.bam
workup/alignments_parts/sample2.part_001.DNA.chr.bam
workup/alignments_parts/sample2.part_001.DNA.chr.masked.bam
workup/clusters/sample1.clusters
workup/clusters_parts/sample1.part_000.clusters
workup/clusters_parts/sample1.part_001.clusters
workup/clusters_parts/sample2.part_000.clusters
workup/clusters_parts/sample2.part_001.clusters
workup/fastqs/sample1_R1.part_000.barcoded.fastq.gz
workup/fastqs/sample1_R1.part_000.barcoded_bpm.fastq.gz
workup/fastqs/sample1_R1.part_000.barcoded_dpm.fastq.gz
workup/fastqs/sample1_R1.part_001.barcoded.fastq.gz
workup/fastqs/sample1_R1.part_001.barcoded_bpm.fastq.gz
workup/fastqs/sample1_R1.part_001.barcoded_dpm.fastq.gz
workup/fastqs/sample1_R2.part_000.barcoded.fastq.gz
workup/fastqs/sample1_R2.part_001.barcoded.fastq.gz
workup/fastqs/sample1_R2.part_001.barcoded_dpm.fastq.gz
workup/fastqs/sample2_R1.part_000.barcoded.fastq.gz
workup/fastqs/sample2_R1.part_001.barcoded.fastq.gz
workup/fastqs/sample2_R2.part_000.barcoded.fastq.gz
workup/fastqs/sample2_R2.part_001.barcoded.fastq.gz
workup/split_fastq/sample1_R1.part_000.fastq.gz
workup/split_fastq/sample1_R1.part_001.fastq.gz
workup/split_fastq/sample1_R2.part_000.fastq.gz
workup/split_fastq/sample1_R2.part_001.fastq.gz
workup/split_fastq/sample2_R1.part_000.fastq.gz
workup/split_fastq/sample2_R1.part_001.fastq.gz
workup/split_fastq/sample2_R2.part_000.fastq.gz
workup/split_fastq/sample2_R2.part_001.fastq.gz
workup/splitbams/sample1.DNA.merged.labeled_AB1-A1.bam
workup/splitbams/sample1.DNA.merged.labeled_AB2-A2.bam
workup/splitbams/sample2.DNA.merged.labeled_AB1-A1.bam
workup/splitbams/sample2.DNA.merged.labeled_AB2-A2.bam
workup/trimmed/sample1_R1.part_000.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample1_R1.part_000.barcoded_dpm.RDtrim.fastq.gz
workup/trimmed/sample1_R1.part_001.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample1_R1.part_001.barcoded_dpm.RDtrim.fastq.gz
workup/trimmed/sample1_R2.part_000.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample1_R2.part_001.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample2_R1.part_000.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample2_R1.part_001.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample2_R2.part_000.barcoded_bpm.RDtrim.fastq.gz
workup/trimmed/sample2_R2.part_001.barcoded_bpm.RDtrim.fastq.gz
workup/pipeline_counts.txt
workup/qc/pipeline_counts.csv

get_size_distribution

input

workup/clusters/all.clusters
workup/clusters/sample1.clusters
workup/clusters/sample2.clusters

output →

workup/clusters/BPM_cluster_distribution.pdf
workup/clusters/BPM_read_distribution.pdf
workup/clusters/OPM_cluster_distribution.pdf
workup/clusters/OPM_read_distribution.pdf

generate_cluster_ecdfs

input

workup/clusters/all.clusters
workup/clusters/sample1.clusters
workup/clusters/sample2.clusters

output →

workup/clusters/Max_representation_counts.pdf
workup/clusters/Max_representation_ecdf.pdf

generate_cluster_statistics

input

workup/clusters/all.clusters
workup/clusters/sample1.clusters
workup/clusters/sample2.clusters

output →

workup/clusters/cluster_statistics.txt

index_splitbams

input

workup/splitbams/{file}.bam

output →

workup/splitbams/{file}.bam.bai

generate_bigwigs

input

workup/effective_genome_size.txt
workup/splitbams/{file}.bam
workup/splitbams/{file}.bam.bai

output →

workup/bigwigs/{file}.bw

multiqc

input

workup/clusters/sample1.clusters
workup/clusters/sample2.clusters

output →

workup/qc/multiqc_report.html

log_config

output →

workup/logs/config_2025.02.07.json

all

input

workup/alignments/sample1.DNA.merged.bam
workup/alignments/sample1.merged.BPM.bam
workup/alignments/sample2.DNA.merged.bam
workup/alignments/sample2.merged.BPM.bam
workup/bigwigs/sample1.DNA.merged.labeled_AB1-A1.bw
workup/bigwigs/sample1.DNA.merged.labeled_AB2-A2.bw
workup/bigwigs/sample2.DNA.merged.labeled_AB1-A1.bw
workup/bigwigs/sample2.DNA.merged.labeled_AB2-A2.bw
workup/clusters/BPM_cluster_distribution.pdf
workup/clusters/BPM_read_distribution.pdf
workup/clusters/OPM_cluster_distribution.pdf
workup/clusters/OPM_read_distribution.pdf
workup/clusters/Max_representation_counts.pdf
workup/clusters/Max_representation_ecdf.pdf
workup/clusters/all.clusters
workup/clusters/cluster_statistics.txt
workup/logs/config_2025.02.07.json
workup/pipeline_counts.txt
workup/qc/multiqc_report.html
workup/splitbams/AB1-A1.bam
workup/splitbams/AB2-A2.bam
workup/splitbams/sample1.DNA.merged.labeled_AB1-A1.bam
workup/splitbams/sample1.DNA.merged.labeled_AB2-A2.bam
workup/splitbams/sample2.DNA.merged.labeled_AB1-A1.bam
workup/splitbams/sample2.DNA.merged.labeled_AB2-A2.bam