## **NCBI SRA** Filter output of Fetch data rRNA contamination initial search database search for analysis chimera analysis Necessary files: Search parameters/queries: Filtering conditions: Analysis steps: · Virome studies only • Number of reads SILVA SSU + LSU rRNA 1) BWA align each RNA-seg FASTQ on (>=100000 & <=15000000) the SILVA rRNA ref FASTA • Related to Homo sapiens or database (removed rRNAs human organism Average read size/length with >1% ambiguous N's) 2) BWA align each RNA-seq FASTQ on NO RNA-seg strategy only (>=100 & <=500) (SILVA rRNA ref FASTA) the corresponding virus ref FASTA Single-end layout only Context-based search of • RNA-seg data per sample 3) Detect rRNA-virus chimeric reads Illumina or Ion torrent virus-related samples (RNA-seg FASTQ) as found in common between the platform Exclusion of DNA and Studied viral genes / alignments of step 1 & 2 overrepresented viruses in genome per sample 4) Calculate rRNA contamination & the examined samples (virus ref FASTA) rRNA-virus chimera statistics rRNA-virus YES chimeras present? Legend: Implementation of the Virus mapping De novo assembly Individual file operation compared methods statistics / analysis statistics / analysis Per RNA-sea sample operation Compared methods: Analysis steps: Analysis steps: 1) MEGAHIT de novo assembly on RAW method i.e. the raw unprocessed 1) BWA align the treated RNA-seq Per compared reads FASTQ by each compared method the RNA-seg FASTQ files treated method operation • BWA method i.e. the unmapped reads on the virus ref FASTA by each compared method 2) Construction of a local BLASTn after BWA alignment on the SILVA 2) Calculate virus mapping / rRNA ref FASTA alignment statistics by samtools database based on the SORTMERNA method i.e. the corresponding virus ref FASTA unmapped reads after SortMeRNA 3) BLASTn of the generated alignment on the SILVA rRNA ref FASTA MEGAHIT contigs on the local · ZWA2 method i.e. the clean reads after viral database Decision 4) Calculate de novo assembly ZWA2 mapping and trimming based on node the SILVA rRNA ref FASTA statistics based on BLASTn output Analysis steps: report & virus-specific assembly 1) Implement the compared methods metrics **Process** separately on each RNA-seg FASTQ 2) Measure performance of each **Process** method notes