<https://github.com/B217754-2022/ICA1.git>

CCRYPT password for B217754-2022.ICA1.tar.gz.cpt

We\_love\_RNASeq

3 # This code should enable the user to take RNAi pair-ended data through a workflow including quality checks using fastqc, alignment using bowtie2, and gene expression quantification using bedtools/samtools, with a view to identify dif ferentially expressed genes in experimental setups with multiple organism lineages and testing conditions. The programme could accomodate data from additinoal cell lines in the future (ie. 'Clone 3').

7 #Copying the ICA data and Pair-ended RNAseq sequence data from the server to the present working directory

12 #we need to run the fastqc command on all of the files in the data directory, then outpout the resulting reports to a format we can analyse.

14 #-1

18 #Create a folder in the current directory for the fastqc output to be held

22 #Run fastqc on .fq.gz in the fastq folder of the downloaded directory, with -o specifying the newly created 'FASTQC\_OUTPUT' directory as the destination.

24 #--noextract is used to prevent output uncompressing, -t6 specifies that 6 threads should be used to process fastqc, -q suppresses all progress messag es to make the interface more appealing to the user.

25 #These are not required, deleting \*.html

28 #-2

31 #For each fastqc, use grep to identify the quality categories with WARN or FAIL (except per base sequence content, which is typically FAIL for RNAseq because the read generation method contains anomalous base content at the start of t he strands). The quality flag lines are then written to a new reads\_quality\_warnings\_list.txt and print a warning message.

53 #-3

57 #The reference genome is fasta.gz, so needs to be unzipped into .fasta, then used by bowtie2-build to make into an index for bowtie alignment