RNA Sequencing Analysis Pipeline v2.1.0 RNA-Seq Illlumina run FastQ files Genomes: Human: GRCh37 Rat: Rnor5.0 Mouse: GRCm38 Create directory structure Zebrafish: Zv9 Dog: CanFam3.1 + <run> + <sample> + fastac + jobs Align reads from FastQ files + logs Perform quality control on against reference genome + mapping FastO files STAR v2.4.2a + read counts FastQC v0.10.1 + read counts + jobs + logs For each sample: + bamMetrics -BAM file For each sample: -Chimeric junctions table FastQC report -Chimeric junctions SAM -Splice junctions table -Loafiles Splice junctions analysis Fusion genes analysis GTF: Add read groups to BAM (Picard v1.98) Human: GRCh37 v74 Index BAM (Sambamba v0.5.4) Rat: Rnor5.0 v71 Mouse: GRCm38 v70 Zebrafish: Zv9 v75 Dog: CanFam3.1 v75 Count reads in features (genes, For each sample: transcripts and/or exons) -Sorted BAM + BAI file HTSeq-count 0.6.1p1 Collect metrics about the alignment For each sample: Variant Calling, filtering and annotation of RNA to various functional classes Txt file with number of bamMetrics.pl (using Picard v1.135) bamMetrics.pl (using Picard v1.135) mapped reads per feature/ PDF report containing raw_variants.vcf Merge quality statistics of filtered variants.vcf Create table from all sample sorted BAM filtered variants snpEff.vcf count files filtered variants snpEff snpSift.vcf Example design file: Raw read counts table sample1 sample2 control sample3 test sample4 test sample5 Compute reads per kilobase per Normalize read counts by their Differential expression analysis million (RPKM) values size factor DESeq2 v1.6.3 EdgeR v3.8.6 DESeq v1.18.0 Result table of DE analysis RPKMs table Normalized read counts table - MA plot - PCA plot - Heatmap sample-to-sample correlation