Sample names

Control granulosa: AAcon-FFgn-TT-SS#, con_gn
DDT granulosa: AAddt-FFgn-TT-SS#, ddt_gn
Vinclozolin granulosa: AAvin-FFgn-TT-SS#, vin_gn

Control sertoli: AAcon-FFsc-TT-SS#, con_sc
DDT sertoli: AAddt-FFsc-TT-SS#, ddt_sc
Vinclozolin sertoli: AAvin-FFsc-TT-SS#, vin_sc

Control prostate epithelium: AAcon-FFpe-TT-SS#, con_pe
Vinclozolin prostate epithelium: AAvin-FFpe-TT-SS#, vin_pe

Control prostate stroma: AAcon-FFps-TT-SS#, con_ps
Vinclozolin prostate stroma: AAvin-FFps-TT-SS#, vin_ps

SS#: sample replicate number

Experiment design: con vs treatment (ddt/vin) in gn, sc, pe, or ps

Analysis workflow:

The raw data was processed by Trimmomatic (v0.33) to filter out the bad quality reads. The clean reads were then aligned to Rnor small RNA reference by AASRA pipeline with default parameters (https://www.biorxiv.org/content/early/2017/05/01/132928). The small RNA counts were then generated by featureCounts (v1.5.2). Counts normalization and statistically analysis was performed by DESeq2 (v1.18.1).

Rnor small RNA reference sources:
miRNA: Rattus norvegicus (miRBase release 21)
tRNA: Genomic tRNA Database http://gtrnadb.ucsc.edu/browse.html
piRNA: piRBase http://regulatoryrna.org/database/piRNA/download.html
mtRNA from ENSEMBL (release 76)

Figure description

Fig.1 Heat Map

The heat map shows the normalized count level and distribution for sncRNAs across biological replicates. The raw sncRNA reads were normalized by variance stabilizing transformation to eliminate the Poisson distribution effects from sncRNA-seq. Normalized counts less than one were then filtered out and 500 sncRNAs are randomly sampled to plot the heatmap for better visualization. Bottom-up hierarchical clustering was performed, and sncRNAs are reorganized to generate the dendrogram. Blue lines indicate the distribution of normalized counts.

Fig.2 Dispersion plot

Dispersion plot shows dispersion estimate at different normalized count levels as a quality control measurement for sncRNA-Seq data. Specifically, black dot represents the dispersion estimate for each sncRNA gene. Blue dot is the final dispersion estimates shrunk from the gene-wise estimates towards the fitted estimates. Red line is the fitted estimates showing the dispersions' dependence on the mean. Blue circles are the sncRNA genes which have high gene-wise dispersion estimates and are not shrunk toward the trend line, hence labeled and recognized as dispersion outliers.

Fig.3 Normalized count distribution

Box plot shows the normalized count distribution for all biological replicates as a quality control measurement for sncRNA-Seq data normalization.

File description

con_treatment_sncRNA.txt

DESeq2 output of differential expression spreadsheets for all Rnor sncRNAs.

The meaning of each column

id: sncRNA gene name and/or accession number, or Ensembl ID
baseMean: mean of normalized counts of all samples, normalized by sequencing depth
log2FoldChange: log2 fold change (treatment/control)
lfcSE: standard error of the log2FoldChange estimate
stat: Wald statistic
pvalue: Wald test p-value for testing whether the sncRNA was differentially expressed
padj: Benjamini-Hochberg adjusted p-value