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 Ensembl Rnor_6.0 genome by HiSAT2(v2.1.0). Transcripts were assembled by StringTie (v1.3.4d). The differential expression analysis was performed with R Ballgown (R package v2.10.0) and cuffdiff (v2.2.1). LncRNA prediction was performed by CPAT analysis of cuffdiff output.

Figure description

Fig.1 Transcript length distribution

The X-axis is the transcript length for each gene (Unit: bp). The Y-axis is the gene density for each gene at the corresponding transcript length. Density plot shows the distribution of transcript length across biological replicates.

Fig.2 FPKM distribution

Box plot shows the gene FPKM distribution for all biological replicates as a quality control measurement for RNA-Seq data.

Fig.3 Fold change distribution

Probability density histogram showing the distribution of log2 fold change (treatment / control) for each gene. The dash line indicates fold change = 2 (log2 fold change = 1) as a cut-off.

Fig.4 FPKM Scatterplot

Scatterplot matrix showing the pairwise scatter plots of the log2 FPKM value
Between treatment (Y-axis) and control (X-axis).

Fig.5 Transcript P value distribution

Probability density histogram showing the distribution of P value for each transcript between control and treatment comparison. Dash line indicates P value = 0.05 as a cut-off.

Fig.6 Gene P value distribution

Probability density histogram showing the distribution of P-value for each gene between control and treatment comparison. Dash line indicates P value = 0.05 as a cut-off.

File description

all_genes_and_transcripts/ con_treatment_all_gene.csv
all_genes_and_transcripts/ con_treatment_all_transcript.csv

Ballgown output of differential expression spreadsheets for all Ensembl Rnor_6.0 annotated genes and transcripts.

con_treatment_lncRNA.csv
con_treatment_mRNA.csv

lncRNA and mRNA genes separated from *all_genes_and_transcripts/ con_treatment_all_gene.csv* based on Ensembl Rnor_6.0 annotation.

The meaning of each column

geneIDs: Ensembl ID
geneNames: gene symbols
feature: genomic features
fc: fold change (treatment/control)
pval: p-value for testing whether this feature was differentially expressed
qval: estimated false discovery rate
log2fc: log2 fold change (treatment/control)
gene_biotype: Ensembl Rnor_6.0 annotation of the feature

diff_out/con_treatment_transcripts_coding.txt
diff_out/con_treatment_transcripts_non_coding.txt

Cuffdiff output of differential expression spreadsheets for all Ensembl Rnor_6.0 annotated and novel transcripts, followed by CPAT analysis to split the coding and non_coding transcripts.

Reference to the meaning of each column:
http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/index.html#differential-expression-tests
Differential expression tests section.