**H1299 RNA sequencing data processing**

For the analysis of the known reference transcriptome annotation, RNAseq data were processed with the nf-core RNAseq pipeline 1 v1.2 with default parameters, unless mentioned otherwise (**SupplementaryData1\_known\_reference\_multiqc\_report**). In short, raw sequences were aligned to the human reference genome hg19 by the aligner HISAT2 23 v2.1.0 with the -*reverseStranded* option in *--pairedEnds* mode. Transcripts were assembled using StringTie 4 v1.3.4d and gene code annotation v29lift37 5. Gene counts were generated with the prepDE.py script, accompanying the StringTie 4 v1.3.4d software.

Coverage tracks were generated from bam files utilizing deeptools 6 v3.3.1 function bamCoverage with the option –*normalizeUsing RPKM*. Replicates of the different treatments were merged using the software kenUtils ([www.github.com/Arunken/kenutils](http://www.github.com/Arunken/kenutils)) v377.

For the deNovo assembly of the H1299 transcriptome, sorted bam files from the hg19 alignment were assembled using StringTie 4 v2.1.1, evoking *-m 200 -f 0.05 -c 1.5 -p 10* parameters. Individual assemblies were merged with *stringtie --merge* and default parameters. DeNovo assembled transcripts were compared and annotated with the gene code annotation v29lift375 utilizing GFF Utilities 7v0.11.2 function *gffcompare* with -*R* and *-r* parameters. Transcripts were classified according to their relationship with the closest reference transcript (class\_code). The class\_code was further simplified by merging transcripts classified as s (intron match on the opposite strand), x (exonic overlap on the opposite strand), i (fully contained within a reference intron), y (contains a reference within its intron(s)), p (possible polymerase run-on (no actual overlap)), or u (unknown, intergenic) as non-chimeric (novel) transcripts, transcripts classified as m, n (retained intron(s), not all introns matched/covered), j (multi-exon with at least one junction match), e single exon transfrag partially covering an intron, possible pre-mRNA fragment), or o (other same strand overlap with reference exons) as chimeric (novel) transcripts and transcripts classified as = (complete, exact match of intron chain), as known transcripts. Furthermore, the TSS of transcripts were annotated to the closest or overlapping TEs using the function *distanceToNearest* of the R package GenomicRanges 71 v1.38.0.

Quantification of the RNAseq data with the deNovo assembled H1299 transcriptome was performed with the nf-core RNAseq pipeline 1 v1.2 as described earlier (**SupplementaryData2\_deNovo\_reference\_multiqc\_report**). Transcripts were assembled using StringTie 4 v1.3.4d and the deNovo assembled H1299 transcriptome annotation. Transcript counts were generated with the prepDE.py script, accompanying the StringTie 4 v1.3.4d software.

**Prediction of open reading frames**

ORFs were predicted using the software TransDecoder ([www.github.com/TransDecoder/TransDecoder](http://www.github.com/TransDecoder/TransDecoder)) v2.0. Therefore, transcript sequences were extracted using the *gffread* function from the software GFF Utilities 7v0.11.2, and ORFs were predicted using the function *TransDecoder.LongOrfs* evoking the parameters *-m 8 -S*. Furthermore, ORFs and their resulting peptide sequences were selected for complete or 5’prime ORFs. For 5’prime ORFs, including multiple Methionines, the longest ORF was maintained.

**Differential expression analysis**

For the identification of differentially expressed genes in the known transcriptome annotation, the R package DESeq2 8 v1.26.0 was used. Gene counts were applied for a group-wise comparison of the different treatments. An adjusted *P* value cutoff < 0.05 and an absolute log2 fold change of >1 was applied to fulfill statistical significance.

Similar analyses were applied for the identification of differentially expressed transcripts in the deNovo assembled transcriptome annotation. Transcript counts were applied for group-wise comparison of the different treatments. An adjusted *P* value cutoff < 0.01 and an absolute log2 fold change of >2 was applied to fulfill statistical significance.

**Gene set enrichment analysis**

GSEA against AIM, CTA, and Interferon-alpha response gene sets was performed utilizing the R package clusterProfiler 9 v3.12.0 and the function *GSEA*. Therefore, genes were ordered according to their log2 fold change and the fgsea algorithm was applied. Gene sets were acquired from Li et al. 10 and the MSigDB 11.

**Analysis of transposable elements expression**

Overlap of aligned reads with TEs was performed using subread’s 12 v1.6.4 function *featureCounts* evoking *-p -f -F 'GTF'* parameters. Counts were summarized over TE subfamilies and normalized using DESeq2 8 v1.26.0. For the hierarchical cluster analysis, of LTR12 subfamily expression, by the R package pheatmap 13 v1.0.12, the *vst* function was applied to normalized counts using the R package DESeq2 8 v1.26.0.

**Enrichment of transposable elements**

The R package LOLA 14 v 1.16.0 was used to enrich TSSs with TE classes, families, and subfamilies. The TE LOLA database was manually generated using the genomic regions of TE classes, families, and subfamilies. The TSS of TINPAs was enriched against a background containing all TSSs from the deNovo assembled transcriptome annotation. ??In case we also perform a LOLA analysis with peptide transcripts, this has to be further elaborated here??

**References**

1. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* vol. 38 276–278 (2020).

2. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357 (2015).

3. Brind’Amour, J. *et al.* LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat. Commun.* **9**, 3331 (2018).

4. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

5. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

6. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

7. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, 304 (2020).

8. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

9. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

10. Li, H. *et al.* Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget* **5**, 587–598 (2014).

11. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

12. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

13. Raivo Kolde. Pheatmap. *cran* (2019).

14. Sheffield, N. C. & Bock, C. LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).

15. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, (2020).

16. Stark, R. & Brown, G. DiffBind : differential binding analysis of ChIP-Seq peak data. *bioconductor* 1–27 (2011).