

Supplementary Method

Data Processing and Analysis

Paired-end RNA sequencing was performed on the Illumina NovaSeq 6000 platform. Raw reads were processed with Trimmomatic v0.38 to remove adapter sequences and trim low-quality bases. Cleaned reads were aligned to the *Homo sapiens* reference genome (GRCh38) using *HISAT2 v2.1.0* (Kim et al., 2015). The reference genome sequence and annotation files were obtained from the NCBI Genome Assembly and NCBI RefSeq databases.

Aligned reads in SAM format were sorted and indexed using *SAMtools v1.9*, and transcript assembly and quantification were performed using *StringTie v2.1.3b* (Pertea et al., 2015; 2016). Gene- and transcript-level expression was quantified as raw counts, FPKM (Fragments Per Kilobase of transcript per Million mapped reads), and TPM (Transcripts Per Million).

Differential Gene Expression Analysis

Differential expression analysis was conducted using *DESeq2 v1.38.3* (Love et al., 2014) with raw count data as input. During quality control, genes with non-zero counts across samples were retained. Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) were used to assess global expression patterns and sample clustering.

Normalization was performed using the Relative Log Expression (RLE) method to correct for library size variation. Statistical testing for differential expression was carried out using the DESeq2 Wald test, and resulting *p*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure. Genes were considered significantly differentially expressed at |fold change| ≥ 2 and adjusted *p* < 0.05. Hierarchical clustering of rlog-transformed expression values for significant genes was performed using Euclidean distance and complete linkage.

Functional enrichment and gene ontology analyses were performed with g:Profiler (Raudvere et al., 2019) against the Gene Ontology (GO) database. Adjusted *p*-values in enrichment results were computed using one-sided hypergeometric tests with Benjamini–Hochberg correction. Gene-wise boxplots for selected gene sets were generated from rlog-transformed expression values.

All statistical analyses and visualizations were conducted in *R v4.2.2* (www.r-project.org).

Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis was performed using the GSEAPy Python package (Fang et al., 2022). Differential expression results, including ranked gene lists based on \log_2 fold change, were used as input to identify significantly enriched pathways from curated databases such as KEGG. Enrichment scores and normalized enrichment scores (NES) were computed, and statistical significance was determined using phenotype-based permutation testing with false discovery rate (FDR) correction.

References

1. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–360.
2. Pertea M, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–295.
3. Pertea M, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–1667.
4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
5. Raudvere U, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191–W198.
6. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2022;39(1).