

README: Integrate TE Expression with DMR Analysis

Overview

This R script integrates **Transposable Element (TE) expression data** from RNA-seq with **Differentially Methylated Regions (DMRs)** from EM-seq to identify repeat elements whose expression changes correlate with DNA methylation changes. The analysis examines radiation dose and time effects across multiple experimental comparisons.

Input Files

GRCh38_GENCODE_rmsk_TE.gtf GTF annotation file for transposable elements from RepeatMasker
GRCh38_centromere_UCSC_simplified.bed BED file with centromere coordinates
GRCh38_GENCODE_rmsk_TE_annotated_ChIPseeker.tsv Pre-computed genomic annotation for TEs (intron, exon, promoter, etc.)
*****_DESeq2_results_TE.csv**** DESeq2 differential expression results for TE transcripts
DMRs_annotated.xlsx Annotated DMR output from dmrseq/DSS analysis
sample_info.xlsx (implicit) Sample metadata

Analysis Procedure

1. TE Annotation Statistics

- Parses the TE GTF to count unique classes, families, genes, and transcripts
- Generates hierarchical counts at gene, family, and class levels

2. TE Distribution Visualization

- Creates bar plots showing transcript counts per TE family (top 30)
- Creates bar plots showing transcript counts per TE class
- Generates annotation-stratified plots (by genomic region: intron, exon, etc.)

3. Per-Comparison Analysis Loop

For each comparison (e.g., IR2Gy24h_vs_NIR, IR10Gy24h_vs_NIR):

- **DMR-TE Overlap:** Identifies DMRs overlapping TEs (± 1 kb maxgap)
- **Expression Integration:** Merges TE differential expression with DMR overlap status
- **Centromere Annotation:** Flags TEs located in centromeric regions
- **Regulation Counting:** Tallies up/down-regulated TEs by family
- **DMR-Expression Correlation:** Counts differentially expressed TEs within hyper/hypomethylated DMRs

4. Multi-Panel Summary Plots

- Combines individual comparison plots into faceted multi-panel figures
-

Output Files

Global Output (base_dir level)

TE_stats.tsv Summary counts: number of TE classes, families, genes, transcripts
TE_transcript_count_gene_level.tsv Transcript counts per gene with class/family info
TE_gene_count_family_level.tsv Gene counts per TE family
TE_family_count_class_level.tsv Family counts per TE class
TE_class_transcript_count.tsv Total transcript count per TE class
TE_transcript_count.pdf/png Bar plot of top 30 TE families by transcript count
TE_class_transcript_count.pdf/png Bar plot of transcript counts per TE class
TE_transcript_count_annotated.pdf/png Faceted plot of TE families by genomic annotation
TE_expression_counts_barplot_multiplot.pdf/png Multi-panel plot showing DE TE counts across all comparisons
TE_expression_in_DMR_barplot_faceted_multiplot.pdf/png Multi-panel plot showing TE expression in DMRs (including no-change)
regulated_TE_expression_in_DMR_barplot_faceted_multiplot.pdf/png Multi-panel plot showing only regulated TEs in DMRs
TE_DMR_overlap_counts_barplot_faceted_multiplot.pdf/png Multi-panel plot showing TE transcripts overlapping DMRs
DMR_TE_overlap_counts_barplot_faceted_multiplot.pdf/png Multi-panel plot showing DMR counts overlapping TEs
DMR_annotation_barplot_faceted_multiplot.pdf/png Multi-panel plot showing DMR distribution by genomic annotation

Per-Comparison Output (Integrate_TE_expression_with_DMR/ subfolder)

TE_annotated_with_DMR.xlsx Full table of all TEs with DMR overlap info, expression stats, centromere flag
TE_family_expression_counts.xlsx Up/down-regulated TE counts by family
TE_family_expression_counts_annotated.xlsx Same as above, stratified by genomic annotation
TE_family_expression_barplot.pdf/png Bar plot showing top 10 DE TE families
TE_family_expression_annotated_barplot.pdf/png Faceted bar plot showing DE TEs by genomic region
TE_family_expression_in_DMR_barplot_faceted.pdf/png Bar plot showing TE expression changes in hyper/hypomethylated DMRs (incl. no-change)
TE_family_regulated_in_DMR_barplot_faceted.pdf/png Bar plot showing only up/down-regulated TEs in DMRs
regulated_TE_family_expression_in_DMR.xlsx Filtered table of significantly regulated TEs overlapping DMRs
overlapping_dmr_count.pdf/png Bar plot showing DMR counts by TE family and methylation direction
overlapping_TE_transcripts_count.pdf/png Bar plot showing TE transcript counts overlapping DMRs

Key Parameters

- **DMR method:** dmrseq, DSS, or TE_targeted
- **Significance cutoffs:** $\text{padj} < 0.05$, $|\log_2\text{FoldChange}| > 1$ for regulated TEs
- **Overlap maxgap:** 1000 bp for DMR-TE and TE-centromere overlaps

Dependencies

```
library(dplyr)
library(tidyr)
```

```
library(purrr)
library(readr)
library(openxlsx)
library(ggplot2)
library(cowplot)
library(GenomicRanges)
library(plyranges)
library(GenomicFeatures)
library(ChIPseeker)
```

Comparisons Analyzed

1. IR2Gy24h_vs_NIR
2. IR10Gy24h_vs_NIR
3. IR10Gy24h_vs_IR2Gy24h
4. IR2Gy6d_vs_NIR
5. IR10Gy6d_vs_NIR
6. IR10Gy6d_vs_IR2Gy6d
7. IR2Gy6d_vs_IR2Gy24h
8. IR10Gy6d_vs_IR10Gy24h