A Protocol for m6A-seq Data Analysis

This protocol serves as a step-by-step guide to m6A-seq data analysis. We include all the major steps involved in seven modules: **database checking, data preprocessing, differential expression analysis, site detection(host), site detection(virus),differential methylation analysis, and reference-based analysis**. We demonstrate the procedures by analyzing two published MeRIP-seq datasets from scratch based on UNIX shell and R system. The data information are summarized in the "Dataset" subsection. The software information and workflow are summarized in the following table.

Software	Function	MarkDown	M0. Checking Existing Databases	M1. Data Preprocessing	M2. Differential Expression	M3. Site Detection - Host	M4. Site Detection - Virus	M5. Differential Methylation	M6. Reference based analysis
		MarkDown		Yes	Yes	Yes	Yes	Yes	Yes
m6A-Atlas	m6A database		R			O		O	
RMDisease	SNP disease database		R			O		O	
Whistle Server	Site prediction	Less Reliable	O			O		O	
SRA toolkit	Obtaining GEO data			O					
FASTQC	Quality evaluation			O					
Trim Galore	Reads trimming			O					
Trumpet	Quality evaluation	Correct / Evalu		O					
Hisat2	Reads alignment			R					
Ballgown	Differential expression				R				
exomePeak2	m6A-seq data analysis					R	R	R	R
STREME	Motif finding	Positive strand				R		N	
MetaTX	Distribution of m6A sites	Isoform ambiguity				R		O	
IGV	Visualization of reads					R	R	N	
DAVID	GO enrichment analysis	hyper/hypo vs both				N		R	
RNAmod	Annotation web server					O		O	
		Key Points			Input control samples of m6A-seq data	Strand-specificity GC correction	Strand-specificity GC correction M level problem	GC correction Absolute vs relative DM	Differential analysis M level quantification (reasonable)