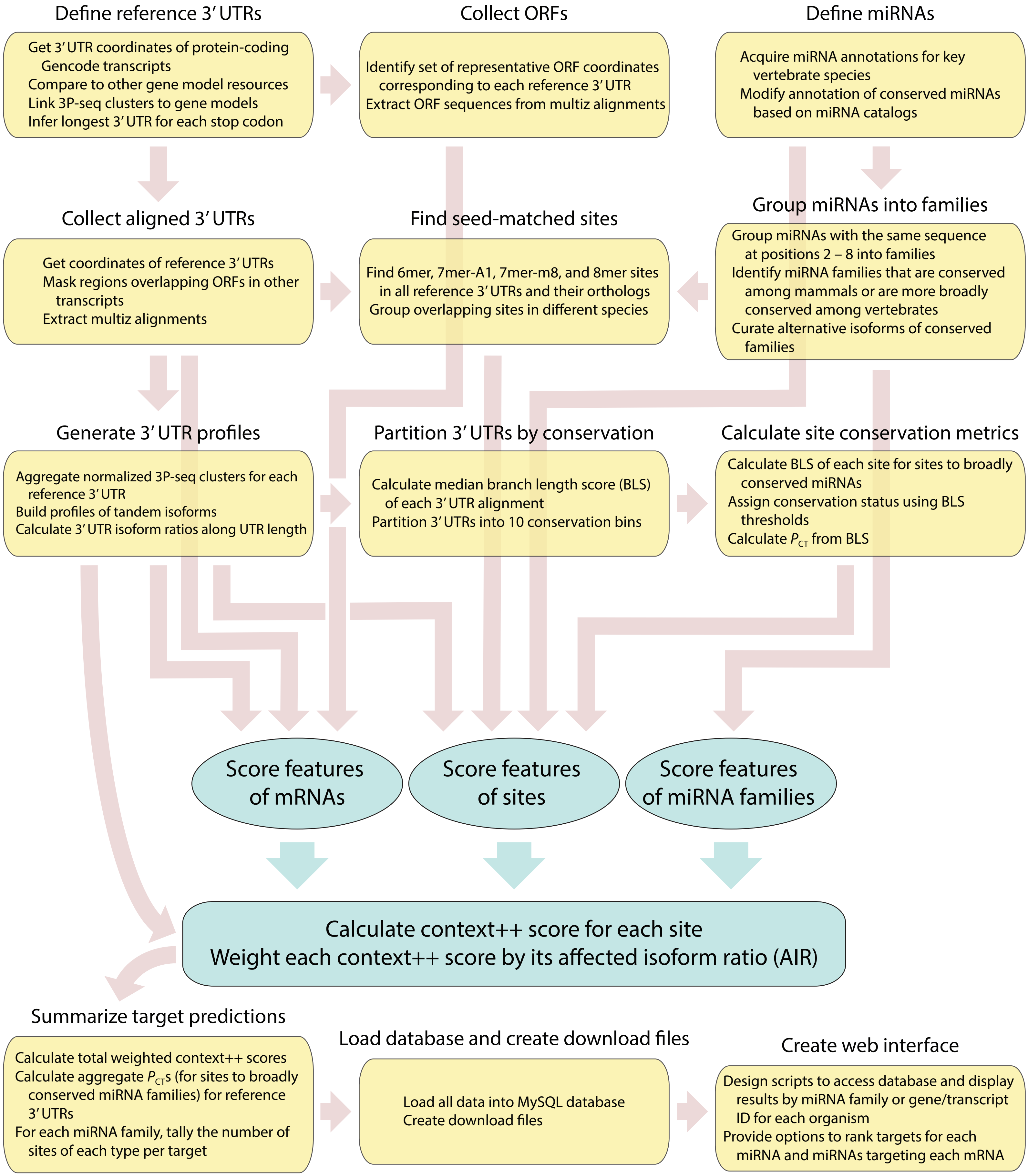

Define reference 3′ UTRs

- Get 3′ UTR coordinates of protein-coding Gencode transcripts
- Compare to other gene model resources
- Link 3P-seq clusters to gene models
- Infer longest 3′ UTR for each stop codon

Collect ORFs

- Identify set of representative ORF coordinates corresponding to each reference 3′ UTR
- Extract ORF sequences from multiz alignments

Define miRNAs

- Acquire miRNA annotations for key vertebrate species
- Modify annotation of conserved miRNAs based on miRNA catalogs

Collect aligned 3′ UTRs

- Get coordinates of reference 3′ UTRs
- Mask regions overlapping ORFs in other transcripts
- Extract multiz alignments

Find seed-matched sites

- Find 6mer, 7mer-A1, 7mer-m8, and 8mer sites in all reference 3′ UTRs and their orthologs
- Group overlapping sites in different species

Group miRNAs into families

- Group miRNAs with the same sequence at positions 2 – 8 into families
- Identify miRNA families that are conserved among mammals or are more broadly conserved among vertebrates
- Curate alternative isoforms of conserved families

Generate 3′ UTR profiles

- Aggregate normalized 3P-seq clusters for each reference 3′ UTR
- Build profiles of tandem isoforms
- Calculate 3′ UTR isoform ratios along UTR length

Partition 3′ UTRs by conservation

- Calculate median branch length score (BLS) of each 3′ UTR alignment
- Partition 3′ UTRs into 10 conservation bins

Calculate site conservation metrics

- Calculate BLS of each site for sites to broadly conserved miRNAs
- Assign conservation status using BLS thresholds
- Calculate P_{CT} from BLS

Score features of mRNAs

Score features of sites

Score features of miRNA families

Calculate context++ score for each site
Weight each context++ score by its affected isoform ratio (AIR)

Summarize target predictions

- Calculate total weighted context++ scores
- Calculate aggregate P_{CT}s (for sites to broadly conserved miRNA families) for reference 3′ UTRs
- For each miRNA family, tally the number of sites of each type per target

Load database and create download files

- Load all data into MySQL database
- Create download files

Create web interface

- Design scripts to access database and display results by miRNA family or gene/transcript ID for each organism
- Provide options to rank targets for each miRNA and miRNAs targeting each mRNA