TSS Annotation Workflow

Identify ortholog

Characterize TSS in the *D. melanogaster* ortholog

- Identify isoforms with unique TSS
  - FlyBase JBrowse
  - Gene Record Finder

For each unique TSS

- Classify core promoter
  - peaked / intermediate / broad / insufficient evidence

GEP UCSC Genome Browser (*D. melanogaster*)
1. Transcription start sites (Celniker > Embryonic)
2. DNase I hypersensitive sites
3. 9-state model

Annotate TSS in the target species (e.g., *D. ananassae*)

If RAMPAGE data is available

- Narrow TSS search region: RAMPAGE peak from TSRchitect
- TSS position: Position with the highest RAMPAGE read density within the RAMPAGE peak

If RNA PolII ChIP-Seq or ATAC-Seq data are available

- Wide TSS search region: RNA PolII ChIP-Seq peak (preferred) or ATAC-Seq peak from MACS2

If RAMPAGE, RNA PolII, and ATAC-Seq data are unavailable, annotate TSS based on other lines of evidence

1. H3K4me2 ChIP-Seq
2. RNA-Seq (multiple stages, tissues)
3. Splice junction predictions
4. *blastn* (change search parameters)
5. Conservation track
6. Core promoter motifs (e.g., Inr)
TSS Annotation (Recommended Parameters)

Characterize TSS in the *D. melanogaster* ortholog

**GEP UCSC Genome Browser (D. melanogaster)**

**Transcription start sites**
- TSS (Celniker) (R5): pack

**DNase I hypersensitive sites**
- Detected DHS Positions (Cell Lines) (R5): pack
- DHS Read Density (Cell Lines) (R5): full

**9-state model**
- BG3 9-state (R5): dense
- S2 9-state (R5): dense

**Types of core promoters**

<table>
<thead>
<tr>
<th>Classification</th>
<th># TSS</th>
<th>#DHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peaked</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Intermediate</td>
<td>≤ 1</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>Broad</td>
<td>&gt; 1</td>
<td>&gt; 1</td>
</tr>
</tbody>
</table>

Annotate TSS in the target species (e.g., *D. ananassae*)

**Define TSS based on experimental data**

**Combined RAMPAGE TSS**
- Combined RAMPAGE Read Density: full
- RAMPAGE Peaks: dense

**RNA PolII ChIP-Seq or ATAC-Seq**
- MACS2 Peaks: dense
- Enrichment (logLR) Profile: full

**RNA-Seq tracks**
- RNA-Seq Coverage (Stages, Tissues): full
- Combined Splice Junctions: squish

**blastn search parameters**
- Word size: 7
- Match / Mismatch scores: 1 / -1
- Gap costs: Existence: 2, Extension: 1
- Low complexity filter: off

**Core promoter motifs search**
1. Click on the “Short Match” link
2. Enter the motif consensus sequence
3. Change the display mode to “pack”
4. Click “Submit”

**Histone modifications**
- H3K4me2 MACS: dense
- H3K4me2 logLR: full

**Additional resources**
- Genome alignments of 28 *Drosophila* species
- Augustus, RefSeq Genes, N-SCAN PASA-EST TSS predictions