GEP Annotation Workflow

Select gene prediction

Click on gene prediction and the "**Predicted Protein**" link
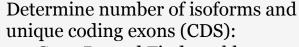
↓

Identify *D. melanogaster* ortholog (workflow)

FlyBase blastp search of predicted protein against the *D. melanogaster* "**Annotated Proteins**" database

↓

Review *D. melanogaster* gene structure

Determine number of isoforms and unique coding exons (CDS):
- Gene Record Finder tables
- Genome Browsers (*e.g.*, GBrowse)

↓

Select isoform to annotate

- Obtain *D. melanogaster* CDS sequence from Gene Record Finder
- Use NCBI **blastx → Align two or more sequences** to compare contig against CDS
  - ❖ Record **start** and **end** positions, **coverage**, and **reading frame**

↓

Map each coding exon (CDS) to contig

↓

Determine CDS coordinates (workflow)

- Identify the best supported donor (**GT**/GC) and acceptor sites (**AG**)
- Minimize changes compared to *D. melanogaster*

↓

Verify isoform model

Provide explanations for:
- Errors and warnings in checklist
- **Large gaps in dot plot**
- **Gaps near splice sites**

Repeat for each isoform

| UCSC Genome Browser | Gene Record Finder | Gene Model Checker |
| FlyBase | NCBI BLAST (bl2seq) | |